

**ARE SIGNALING ARGUMENTS COSTING US TOO MUCH?
Reassurance, Rationalism and Costly Signaling
in International Relations**

Kyle M. Lascurettes and Brandon K. Yoder

Paper prepared for the
Politics Graduate Student Conference
Woodrow Wilson Department of Politics
University of Virginia
March 2009

**[VERY ROUGH, VERY EARLY DRAFT.
You can cite or quote if you want, but we would strongly advise against it.]**

INTRODUCTION

The concept of costly signaling is seemingly everywhere in the international relations literature. Originally applied only to situations of crisis bargaining, where states seek to credibly demonstrate their resolve, it has more recently been adapted to fit situations of possible reassurance, where states seek to credibly demonstrate their deeper, underlying motives and objectives.

In this paper, we argue that the rationalist costly signaling logic may not be a good fit for reassurance scenarios. Specifically, we identify some important differences between crisis bargaining (CB) and reassurance (RE) scenarios that ultimately suggest that costly signaling is less likely to effectively reveal relevant information in reassurance in the way it does in crisis bargaining. Our fundamental goal is to show that by adopting rationalist assumptions, successful costly signaling is likely to be only extremely rare in reassurance contexts. Yet because we nonetheless see significant cases of successful reassurance in the real world, we must ultimately look for other hypotheses and potential explanations that are not so firmly wedded to rationalist assumptions and the costly signaling logic.

We proceed as follows: after briefly introducing the concept of costly signaling in international relations, we identify three important differences between crisis bargaining and reassurance situations. We devote significant time to the third of these differences—the type of uncertainty endemic in CB versus RE scenarios—because we have found that previous scholarship has downplayed the differences here and fundamentally mischaracterized the nature of reassurance itself. We then build on these three observations to offer three corresponding arguments for why rationalist costly signaling is less likely to be used effectively in reassurance contexts. Next, we examine a weaker version of the rationalist logic that has more recently been applied to cases of reassurance in international relations—here called semi-costly signaling—and assess whether it succeeds in overcoming the problems of the stronger logic while remaining true to rationalism’s foundational assumptions. Finally, we conclude by briefly advocating some avenues for future research in the hopes of moving beyond rationalist signaling logic in order to more effectively understand how, when, and why reassurance between states succeeds in international relations.

I. THE RATIONALIST APPROACH

1.1 Costly Signaling in International Relations

The concept of ‘costly signaling’ has had a tremendous impact on the international relations field over the past 15 years. Originally developed in economics, a costly signal is conceptualized as an action taken by an actor to convey some attribute about itself that an actor lacking that attribute would be unable or unwilling to take. The main idea is that such gestures will only actually carry a prohibitively high ‘cost’ for a actor that attempts to falsely signal a bluff related to some aspect of the issue at hand.

The costly signaling logic was first brought the international relations literature by James Fearon.¹ Specifically, Fearon argued that much of the conflict we observe between states in international relations comes as the result of states’ lack of information about other states. States often do not simply tell each other what they want to know, however, because they frequently have incentives to misrepresent some privately held information about themselves that they can use to their advantage. Because of these incentives to misrepresent oneself as, for example, a) highly resolved to fight in order to convince another to back down, or b) highly trustworthy and non-aggressive in order to convince another to let its guard down, Fearon argued that honest states need a way to credibly separate themselves from bluffers. Thus, a highly resolved or genuinely benign and trustworthy state can take some action that a state of the opposite type hoping to bluff would find too costly to take. Rational states observing these actions should thus only ‘update’ their beliefs about the state if the action itself constitutes a costly signal. Any gesture less than a costly signal is ‘cheap talk’ that is no more difficult for a bluffer to engage in than an honest actor, and thus should be relatively worthless to rational observers.

In game theoretic terms, a costly action that leaves no doubt in the receiver of the signal about the sender’s type is known as a fully-separating equilibrium (or simply a separating equilibrium). If the action is equally costly/beneficial to honest and bluffing states of both types, it gives the receiver no reliable information about the sender of the signal (cheap talk), and thus constitutes a pooling equilibrium.

¹ See Fearon, “Threats to Use Force: Costly Signals and Bargaining in International Crises,” PhD Dissertation (1992).

1.2 Crisis Bargaining and Reassurance: What's the Difference?

The conventional understanding of the principal distinction between crisis bargaining and reassurance is relatively straightforward. In crisis bargaining, states have openly conflictual preferences about something, be it a territory, a weapons system, or some other valued good. The sender intentionally attempts to credibly convince the receiver of both its willingness and ability to go to war to achieve its preferred outcome. In other words, states aim to discover the other's immediate resolve. In reassurance, by contrast, states are unsure about what the other actually really seeks from the situation. The receiver's goal here is thus to determine—through signaling—whether the other's underlying objectives are compatible with and benign in relation to one's own. In other words, states aim to demonstrate that their underlying preferences or motivations are benign and non-aggressive, and thus that their general character is status quo-oriented and trustworthy. We take issue with this conventional conceptualization of reassurance below, but it serves our general purpose for now. This section identifies three observable distinctions between crisis bargaining and reassurance scenarios related to their duration and scope, their incentive structures for the actors involved, and the types of uncertainty about one another that each actor is attempting to learn. Because the first two points are intuitive, we devote the majority of our attention to the third distinction. In the following section, we use these observations to offer three corresponding arguments for why the conventional rationalist costly signaling logic is often inapplicable to reassurance scenarios.

1. First, at least in reassurance scenarios where the potential use of force is at issue, *the opportunities and dangers of revealing privately held information in RE are distinct from those inherent in CB*. In CB, the primary danger is inadvertent war. Highly resolved states may only be able to send costly signals of their willingness to fight by taking steps that increase the probability of war. War is an inefficient, risky outcome, suboptimal to reaching a settlement short of war. For a resolved state, brinkmanship in CB risks a conflict in which it likely holds an advantage, since the other may be bluffing and weaker than he says. In contrast, the greatest danger in RE when military expansion is at issue is not inadvertently provoking war, but instead learning too late that the other harbors aggressive intentions, and hence being unprepared for its eventual attack. The costs of being taken for a sucker are therefore often higher in RE than in CB.

2. Second, *the length and scope of the 'game' in CB is likely to be much shorter and more contained than in RE*. RE scenarios often occur continuously in a broad context and over a long and indefinite period of time. In contrast, CB typically takes place over a specific issue and a finite period of time. States within a general reassurance context are thus likely to develop longer-term time horizons than states engaged in crisis bargaining, and are thus likely to think and worry about the distant future more by comparison. [note: section needs to be developed more]

Redefining Reassurance

As noted above, expounding our third distinction between CB and RE requires some refining some of the key concepts of the existing reassurance literature. Because Andrew Kydd's *Trust and Mistrust in International Relations* (2005) is the most ambitious and prominent exploration of reassurance, and because he explicitly refers to his as "the costly signaling theory of reassurance," we focus primarily on his arguments and interpretations here.² Kydd conceives of states as being of two types: trustworthy or untrustworthy. 'Trustworthiness' here refers to the likelihood that an actor will reciprocate a cooperative gesture with one of its own rather than exploit it. In game theoretic terms, if a state prefers mutual cooperation (CC) to exploitation (DC) then, all else being equal, it is likely to be trustworthy. To a certain extent, Kydd is to be commended for invoking the concept of 'trustworthiness,' to define state type, since it can be generalized across security and non-security issue areas, rather than relying on the problematic dichotomy of 'status quo/revisionist' that is inapplicable to non-security issue areas and invites a variety of subjective and normative ambiguities. 'Trustworthiness' is also inherently relational, another conceptual advantage. Rather than restricting preferences to a reified conceptualization of states as inherently aggressive or benign (see, for instance, Randall Schweller's work),³ Kydd

² For an earlier, non-formal version of much of his same foundational argument, see Kydd, "Sheep in Sheep's Clothing: Why Security Seekers Do Not Fight Each Other," *Security Studies* (1997).

³ The nomenclature varies across the literature, with states termed "nice", "security-seeking", or "status quo" on the one hand, and "aggressive", "greedy", or "revisionist" on the other. The former type obtains a low payoff from rearranging the international system, while the latter type obtains a relatively high payoff. Rationalist signaling arguments were originally introduced as a solution to the debate between offensive and defensive realists about the implications of the security dilemma for state behavior. Offensive realists maintain that anarchy alone can foster violent conflict among security-seeking states due to incomplete information about other states' preferences, while defensive and neoclassical realists claim that greedy states are necessary for violence to occur. The disagreement between the two camps hinges on the degree to which security seekers can signal their benign preferences to each

rightly recognizes what is most important about ‘type’ is not what a state “is,” but what it is “seen to be” by others. We do not take issue with this aspect of his argument.

More importantly, however, Kydd’s concept of ‘trustworthiness’ is what we refer to here as a *general* type: it refers to the portfolio of a particular state’s preferences across *all* issue areas. For Kydd, as for any model that conceptualizes a given state as being one of two possible dichotomous types, a state that successfully signals its trustworthiness to another on a particular issue is assumed to be a ‘trustworthy type’ in general. In other words, when state preferences are dichotomized into two general types (be is trustworthy vs. not, or status quo vs. revisionist), states are assumed (by the analyst, but also presumably by other states) to hold these general types across different issue areas and contexts. Kydd’s adoption of this ‘general type’ framework directly informs his conceptualization of the differences between crisis bargaining (CB) and reassurance (RE) scenarios:

In the crisis bargaining context, the signal is designed to show that one is resolute, tough, even belligerent, so that the other side should back down. In the reassurance context, the signal must achieve a different goal, demonstrating that one is moderate, not out to get the other side, willing to live and let live, preferring to reciprocate cooperation (Trust and Mistrust, 187).

It is the literature’s propensity to conceptualize states in reassurance scenarios as possessing ‘general types’ that we take issue with. In contrast to the claims of Kydd and Charles Glaser⁴ (among others) we argue that one’s ‘general type’ is likely to be impossible to signal in reassurance, both because no rational actor is likely to automatically behave similarly across all issue areas, and, as a result, no rational actor is likely to believe that another actor carries a general type in this way. Kydd thus vastly downplays the differences between CB and RE.⁵ *Following from this, we posit that RE differs from CB not only in the type of information that*

other. On these terms and this debate, see especially Randall Schweller, “Bandwagoning for Profit: Bringing the Revisionist State Back In,” *International Security* (1994); “Neorealism’s Status-Quo Bias: What Security-Dilemma?” *Security Studies* (1996); and Jeffrey W Taliaferro, “Seeking Security Under Anarchy: Defensive Realism Revisited,” *International Security* (2000/2001).

⁴ See “Realists as Optimists: Cooperation as Self-Help,” *International Security* (1994-1995); “The Security Dilemma Revisited,” *World Politics* (1997).

⁵ While we take issue with Kydd’s conceptualization of uncertainty and preferences in reassurance, we recognize the immense contribution he has made to the intra-realist dialogue and cooperation literature by launching a research program that asks how states know each other’s goals. Indeed the larger goal of this project and Kydd’s are the same: if reassurance is possible, and anarchy does not necessitate an offensive realist-like world of constant fear and power maximization, there must be some mechanism (or mechanisms) by which states are able to send and receive credible signals about what they really want.

states wish to learn about one another, but also by the amount and complexity of information that they need to learn and deal with about each other's preferences. Specifically, we argue that in RE situations, a state is uncertain about two dimensions of the other's preferences: a) their preferences over outcomes on a particular issue area, and b) the importance the other assigns to that issue relative to others. We now discuss each dimension of preferences in turn⁶ (*see this important footnote for important definitions regarding preferences*).

First, reassurance scenarios are characterized by a particular state (A) wanting to learn another state's (B) underlying preferences on an issue of significant importance to A. Whereas previous scholars assume that states in RE attempt to assess the feasibility and likelihood of cooperation by attempting learn each other's general type, we posit that states want to assess the chances for cooperation by learning how compatible others' preferences are with theirs on a particular issue area. This is our redefinition of reassurance. The reasons for this shift are straightforward: Conceptually, there are a large number of discrete issue areas that states might concern themselves with. Different states throughout history have prioritized issue areas such as the distribution of territory, spread of cultural values, weapons proliferation, national honor, environmental issues, degree of systemic recognition of state sovereignty, openness of international markets, degree of trade protectionism, and the structure of international institutions to name but a few. States possess different goals on different issues, and there is no reason to think that a state that is 'nice' on one issue could not also be 'mean' on others.

We define what a single state wants on a particular issue as the *content* of their preferences in that issue area. Following from this, the relational aspect of the content of two states' preferences can be defined as the *compatibility* of their preferences on that issue. We find 'compatibility' to be a useful conceptualization because the term can be used across various issue areas to consistently denote a possibility for tacit or explicit cooperation on that particular issue.

⁶ NOTE ON DEFINITIONS: No different from Kydd and much of the rest of the literature, we use the terms 'goals,' 'preferences,' 'underlying preferences,' and 'preferences over outcomes' interchangeably to refer to what states ultimately want on a given issue regardless of systemic restraints or context specific contingencies. No different from Kydd or much of the rest of the literature, we define *preferences over strategies* (or just strategies) as the joint product of preferences over outcomes and structural constraints that dictate the means that are available to a state to maximize its probability of accomplishing these goals. The reason it is so valuable for states to know each other's preferences over outcomes is that these allow them to predict the other's behavior for any given configuration of the international system, i.e., the distribution of power and institutional structure, rather than just their behavior given the current systemic restraints. On some of these distinctions, see Robert Powell, "Anarchy in International Relations: The Neorealist-Neoliberal Debate," *International Organization* (1994).

Frequently used terms such ‘similar’ or ‘overlapping’ preferences are not as useful, because they do not necessarily tell us that cooperation is more likely than conflict. For instance, similarity or overlap of preferences on an issue like degree of international economic openness or the structure of an international organization would suggest significant room for cooperation, while holding similar/overlapping preferences for desiring control of the same territory or uninhibited access to the same scarce natural resource would suggest more room for conflict. The United States and the Soviet Union might have had similar preferences for the number allies or nuclear warheads they needed to feel secure, but this kind of overlap made them more likely to remain intensely suspicious of the other, not less. ‘Compatibility’ is a more universal and less ambiguous conceptualization.

The second dimension of uncertainty over another’s preferences that states attempt to learn in reassurance is their rank ordering of issue areas according to which are most and least important to them. For instance, state A desires to know not only whether state B shares their preference for global economic openness and low tariff barriers, but also how salient this issue is to B relative to other issues. Why would a receiver be concerned with a sender’s ranking over the salience of issues? First, while states no doubt have underlying preferences on any and all issue areas, the constraints of the real world dictate that no state is likely to be able to enact its preferences on many different issues at a given point in time, much less across most or all issue areas at once. States must therefore prioritize by devoting their efforts to achieving their preferences on some issues while sacrificing their preferred outcomes on others. Thus, the less importance a state places on a particular issue area, the more likely they are to sacrifice their underlying preferences on it to outside events and circumstances. Likewise, the less they value the issue area, the more likely their behavior is to diverge from their underlying preferences on the issue. Thus, even if the first state knows the other’s preferences within that issue area are compatible, it is unable to make predictions about the other’s likely future behavior as it affects the first state’s central goals. Second, a rational actor might fear that the other is willing to cooperate on the issue only in order to devote scarce resources toward achieving their preferences in other issue areas. They would thus want to know if cooperation on one issue might perversely lead to the other undermining their preferences on other issues. For the sake of clarity (as we assume we have already sacrificed simplicity) we refer to the importance a state attributes to a particular issue area as the *intensity* of its preference on that issue. Because so

many issues are potentially ‘competing’ for the attention of states in general reassurance scenarios, the *relative intensity* of preferences is a particularly onerous form of uncertainty in these contexts.

In crisis bargaining, by contrast, it is only the *absolute intensity* of the other’s goals on a particular issue that is the source of uncertainty. Both states in a CB scenario presumably already know that they are both focusing on the same issue and that the other holds underlying preferences on the issue as incompatible with their own. All that remains uncertain is the degree to which the issue passes some ‘threshold’ of importance for the other to be willing to fight for their underlying goals on it. It is the *absolute intensity* of preferences that matters: if a state is highly resolved on the issue at stake, it will go to war regardless of whether it holds another preference even more dearly.

States remain uncertain about the other’s *relative intensity* of preferences in reassurance scenarios, a more problematic unknown. For instance, even if an issue is very important to a state in absolute terms (they are willing to engage in at least some conflict on it, for instance), its preferences on that isolated issue give others very limited insight into that state’s likely future behavior on that state actually happens to value a number of other issues even more. And once we drop the ideal assumption that states in reassurance already know they can cooperate in a given issue area, we remember that states in reassurance are also uncertain about the compatibility of preferences on the issue, an additional type of uncertainty that, again, is not at issue in CB.

* * *

3. In sum, rather than states simply being unsure about a particular kind of information in the two scenarios (resolve versus general preferences such as ‘trustworthiness’) *the uncertainty in RE is qualitatively more vast, complex, and difficult to disentangle than the uncertainty in CB.* This constitutes our third observable difference between CB and RE scenarios.

1.3 So What? Three Arguments

The above distinctions have three important implications, each of which suggests that costly signaling will either be less effective or less frequently attempted in reassurance cases. Specifically, the nature of reassurance scenarios suggests that costly signals entail greater risks

for potential sender states, and, from the receiver's perspective, are less likely to be a) believed as sufficiently costly, and b) even correctly recognized/interpreted to begin with.

Risking Signals

Following from observation 1, reassurance scenarios involving security issues pose risks for sending costly signals beyond those present in situations of crisis bargaining, thus making potential senders less likely to risk costly signaling. In CB, states aim to convince each other that they have both high capabilities and resolve. Gestures that are likely to demonstrate these qualities, such as raising the domestic audience costs of backing down, sinking costs by moving troops to a critical border, or investing in theater-specific military technology, do not typically increase the vulnerability of the sender to attack.⁷ This is not the case in military reassurance scenarios. Since the goal is to credibly demonstrate benign or unobtrusive preferences on some security-related issue, costly signals necessarily involve some unilateral strategic concession or sign of restraint. The mechanisms proposed in the rationalist reassurance literature include troop or arms reductions, investments in suboptimal defensive military technology, or the imposition of some binding restraint on oneself that would decrease its ability or likelihood for taking military action in a particular region or issue area. The problem here, however, is that taking such an action increases one's vulnerability to a surprise attack from the other, whose preferences and motives by definition still remain unknown..⁸ As Evan Montgomery has shown, the fallacy of existing reassurance arguments is that they simply assume that the sender already believes the receiver to be benign (Montgomery, 160-161). Yet the dilemma of reassurance is what motivates states to send signals likely to increase their vulnerability when others' motivations are unknown. As Montgomery notes,

Not only must a signaling state endeavor to reveal its benign preferences; it must also attempt to discover whether its adversary is a security seeker. Although the first goal calls for significant gestures that will serve as adequate proof of the signaling state's motives, the second calls for smaller gestures as a test of the adversary's reaction. Yet smaller gestures will not be viewed as

⁷ The caveat here is, of course, are a) situations where signals of resolve or strength give away crucial military strategies or capabilities that could hurt the sender's performance in war, or b) situations that have escalated to the point where the only costly signal of credibility/capability left to statesmen is some action that starts or provokes military conflict. See James Fearon, "Rationalist Explanations for War," *International Organization* (1995), pp. 397-400.

⁸ This is the principal argument of Evan Braden Montgomery, "Breaking Out of the Security Dilemma: Realism, Reassurance, and the Problem of Uncertainty," *International Security* (2006).

credible signals of reassurance and are unlikely to be reciprocated...Credible gestures are therefore less likely to be made when are most needed—when uncertainty is a significant constraint (162-163).

Sending a costly signal in this context thus poses a larger and qualitatively more ominous security risk in reassurance than it does in crisis bargaining.

Ironically, this constraint on signaling corresponds with one of James Fearon's rationalist paths to war in crisis bargaining that he identifies as a 'commitment problem.' Specifically, Fearon argues that states may rationally accept the inefficient outcome of war to peaceful negotiation if the settlement would shift the balance of power in favor of the other state.⁹ Because the advantaged state would have the ability to revise the arrangement in the future due to the increased power afforded by the bargain, it would be unable to commit itself to not seeking more advantageous terms once this has occurred. By comparison, it is the signal itself in reassurance (rather than some settlement) that alters the balance of power in the other's favor. Even if potential sender state A suspects that B would not exploit its resulting vulnerability from sending the signal, state A still cannot be sure that its costly concession will not still enable B to capitalize on A's weakening concession in some other way in the unforeseen future. In short, a costly signal itself in almost always begets a commitment problem in scenarios of military reassurance.

Reciprocating Signals

Second, the less clear-cut and more hazily specified temporal boundaries for interaction in reassurance scenarios significantly raise the threshold for what actions constitute credible, costly signals in these situations. Signals that clearly appear costly to observers in shorter and finite durations of crisis bargaining scenarios are less likely to be believed in general reassurance scenarios, where a signaling 'game' could be played between actors for a long and often indefinite period of time. Take, for example, the two most prominent strategies of credible signaling in both CB and RE scenarios, tying hands and sinking costs. Where both types of actions are relatively easy to make credible and costly in CB, this is much more difficult to do in

⁹ As he puts it, "...if the territory in question is strategically vital or economically important, its transfer could radically increase one side's future bargaining leverage...In principle, one state might prefer war to the status quo but be unable to commit not to exploit the large increase in bargaining leverage it would gain from limited territorial concessions." See Fearon, "Rationalist Explanations for War," p. 408.

general reassurance contexts. Tying hands is more difficult because creating believable audience costs becomes harder over time. Foreign observers are much less likely to believe, for instance, that a leader is domestically bound to remain true to some public promise about another state if is made to apply to a long, indefinite and unknowable future, rather than to a specific issue applicable for 2 weeks. The longer the elapsed time period since the pronouncement, the more easily leader will be able to uses changes in leadership or circumstances in the interim to successfully evade the audience costs.¹⁰ Likewise, sinking costs into things such as higher military spending or troop deployment (or, conversely, showing restraint on these things to show benignity of preferences) is likely to look less meaningful, effective, and costly as time elapses, since an observer will be more prone to weighing such costs against the sender state's overall capabilities, commitments, and investments as years rather than simply days or months go by.

More specifically, longer and indefinite time horizons produce greater uncertainties about the unpredictable future, making costly signals sent in the present either a) less likely to be believed as intended, or b) unlikely to make much of an impact relative to other factors. On the former, longer durations of interaction and broader time horizons make states more apt to interpret otherwise costly actions to situational rather than dispositional factors.¹¹ For instance, a skeptical receiver with a longer time horizon might rationally conclude that a state that dramatically forgoes an immediate opportunity to militarily expand at the expense of weaker neighbors may be interpreted differently by states in different scenarios. In a scenario of short duration, a receiver would likely see the gesture as a credible signal of benign security preferences. In contrast, in a scenario with a long and indefinite duration, observer B is more likely than the observer A to see the gesture not as evidence of benignity of preferences/type, but as a sign that the sender state anticipates future trends in power, technology, etc. that will make a 'defection' on the issue more profitable and likely to succeed in the future. On the latter, scenarios of longer or indefinite duration are likely to condition states to adopt longer time

¹⁰ Fearon himself conceded this point, saying, "...it may be more possible to generate audience costs in crisis bargaining than regarding security guarantees in grand strategy because of the difficulties for a leader of projecting tied hands into an uncertain and distant future against unknown adversaries. Leaders do try to stake national honor, prestige, and reputation on the fulfillment of alliance or security guarantees, but..." such pledges are unlikely to generate believable costs...Also, because leaders and circumstances change over time...guarantees do not attach as directly to the person of the leader generating them as do efforts to tie hands in crisis bargaining." James Fearon, "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs," *Journal of Conflict Resolution* (1997), pp. 82-83.

¹¹ On the causes and consequences of attributing states' actions to situational versus dispositional factors, see Jonathan Mercer, *Reputation and International Politics* (1996).

horizons, thus increasing their sensitivities to unpredictable future shocks that alter the particular conditions of the present situation in unanticipated ways.¹² Where states are more apt to think and worry more about the distant future, the ‘prepare for the worst’ assumptions of pessimistic realists are more likely to take hold, thus dampening the possibility of any signal in the present being dramatic, credible, and costly enough to induce states to lower their guard and update their pessimistic beliefs.¹³

Recognizing Signals

Most importantly, signals sent in reassurance scenarios are less likely to be understood. Receivers often have difficulty judging the costliness of a signal in reassurance not only because of longer and more ambiguous time horizons, but also because they are often unsure about what the sender considers costs and gains in the first place. In CB situations, by contrast, this content is presumably already known, and signals can thus be evaluated more objectively. *In reassurance, what a sender constitutes as costs and benefits—what its utility formula is—is itself private information, unknown to the receiver.*¹⁴

This requires some elaboration. First, a state’s (A) principal goal in a reassurance scenario is to learn the content of another’s (B) preferences on an issue important to state A. Yet the very nature of uncertainty in reassurance makes even this task extraordinarily difficult. Why is this so difficult? It is hard because without already knowing which issues or issue areas B most values, it is virtually impossible for A to decipher the true costliness of any gesture that B might make. To put it in the language used earlier, the *relative intensity* of the sender’s preferences—or its issue area rankings by importance—remains unknown to the receiver. A receiver knows which issues are of greatest importance to itself, but it has no guarantee that a sender shares this ordering. The sender’s actions may be driven by a goal in a different issue area that it considers more important. It may take an action that is detrimental to its goals in the issue area of interest to the receiver, but only because it has subordinated that particular issue to another. These complexities hinder the ability of observers to gain information about preferences from overt

¹² These include future changes in relative power, unexpected international events, new strategic alignments, and changes in leadership, among numerous other factors.

¹³ On these kinds of ‘fear for the future’ pessimistic assumptions, see John J. Mearsheimer, *The Tragedy of Great Power Politics* (2001), but especially Dale C. Copeland, *The Origins of Major War* (2000), p. 246.

¹⁴ Harald Muller, “Arguing, Bargaining and All That: Communicative Action, Rationalist Theory and the Logic of Appropriateness in International Relations,” *European Journal of International Relations* (2004), pp. 398-401.

behavior, because they often render the actions and behaviors of senders with fundamentally different preferences on a particular issue in the long run observationally equivalent in behavior in the present. The receiver thus may come to doubt that *any* particular credible action accurately reflects the sender's underlying preferences on a particular issue, and at the very least raises the threshold of objective costliness necessarily for a receiver to update its beliefs about the sender on a particular issue.

For instance, imagine two states, A and B, deadlocked in a protracted global standoff over a range of issues. State A offers a concession on arms control to state B, announcing a plan to unilaterally phase out a class of battleships that had previously been seen by B's leaders as a sign of A's expansionist agenda in its area of the world. According to standard rationalist logic, B has every reason to update its beliefs about A's type. Yet B's leadership refuses to do this, believing instead that A's move signifies not an inherent preference for reduction of tensions, but instead merely a strategy to conserve its resources to facilitate economic autarky. This is not as far fetched as it may seem. From an objective rationalist perspective, for example, significant Soviet conciliatory gestures on nuclear weapons, conventional arms, and troop deployments in 1988 were clearly too costly to be taken by a revisionist power still dedicated to worldwide expansion. Yet at the time, key American officials refused to update their beliefs about the Soviets, and for reasons that cannot necessarily be dismissed as irrational. As then Deputy CIA Director Robert Gates observed, "[Soviet General Secretary Mikhail] Gorbachev intends improved Soviet economic performance, greater political vitality at home, and more dynamic diplomacy to make the U.S.S.R. a more competitive and stronger adversary in the years ahead... We must not mislead ourselves or allow ourselves to be misled into believing otherwise."¹⁵ These ambiguities about underlying preferences run rampant in reassurance, thus making the likelihood of credible, correctly interpreted signals much harder and more unlikely. Even in the face of incredibly costly signals, a failure of the other to update beliefs might remain a rational outcome.

Second, there are additional reasons why states want to learn about the other's relative intensity of preferences independent of their desire to learn their issue-specific goals. After all, the ultimate goal of discerning the other's true preferences on an issue is that doing so allows the

¹⁵ Quoted in Raymond L. Garthoff, *The Great Transition: American-Soviet Relations and the End of the Cold War* (1994), pp. 339-340.

receiver to better anticipate the future behavior of the sender independent of future structural changes and other unanticipated contingencies (since as discussed earlier, we are less likely to observe changes in strategies/behaviors on a particular issue as environmental changes occur the higher that state values that issue area relative to others). Even if the content of the other's preferences within an isolated issue area is known, it remains difficult for a receiver to anticipate a sender's likely future behavior if the relative salience of issues to the sender is unknown to the receiver. The issue area of importance to the receiver may be relatively unimportant to the sender, resulting in the sender simultaneously taking actions detrimental to the receiver's goals in other issue areas despite cooperating with it for the time being on that one particular isolated issue most valued by the receiver.

Overall, then, even if state A judges state B's actions in the present to be compatible with its underlying preferences on an issue of high importance to A, it is still rational for A to desire information about how much B values the issue relative to others for three reasons: If B doesn't value the issue much, a) its behavior/strategy on that issue will be determined more by external circumstances than internal preferences, thus making it highly susceptible to unpredictable future changes in systemic constraints; b) its current behavior/strategy might not reflect its underlying preference on the issue, and a change in future circumstances may permit B to more actively pursue its *actual* preferences on the issue that could be contrary to A's; and c) it may be simply appeasing A in order to devote scarce resources to other issue areas where it could severely conflict with and hurt A's preferences.

* * *

In sum, these dual sources of uncertainty about the costliness of another's actions from the receiver's point of view—caused by the longer and indefinite duration of reassurance scenarios and the complex uncertainty about the other's preferences in reassurance (arguments 2 and 3 above)—mean that the objective costliness of a signal must be greater, on average, for the receiver to update its beliefs about the sender's preferences at all. What quickly becomes clear when attempting to assess the meaning of signals in a reassurance scenario is that there is significantly greater ambiguity than in a crisis bargaining scenario. Even if the objective costliness of the signal isn't up for dispute, the receiving state often still has grounds for rationally remaining wary of the sender's underlying motives. To be clear, we are *not* arguing that costly signals never result in updating in reassurance situations. Instead, we posit that the

threshold for what actions are objectively costly enough to be recognized as sufficiently costly, ‘separating’ gestures is significantly higher in reassurance scenarios than in crisis bargaining. If we add to this the fact that signaling non-conflictual preferences in reassurance often involves increasing one’s vulnerability/insecurity—by itself an often prohibitively risky venture (argument 1 above)—we can surmise that a higher threshold for clearly establishing costliness only makes sufficiently costly signals even that much less likely to be sent.¹⁶

If these arguments about the extreme difficulty of effective costly signaling in reassurance are convincing, one of two possible conclusions must naturally follow: Either successful reassurance is rare in international politics, or rationalist signaling arguments are largely incapable of explaining many such cases. To the extent that we can observe numerous significant cases of successful reassurance occurring not infrequently throughout history (and we think we can), the rationalist signaling logic may not be a good fit for understanding such scenarios.

1.4 Semi-Costly Signals?

Thus far, we have argued that the logic of the rationalist signaling thesis appears to be largely unable to explain or account for cases of successful reassurance between rational states. Yet perhaps in anticipation of these critiques and concerns, numerous scholars have adopted a weaker version of the rationalist argument. This weaker version tacitly admits that even costly signals often contain some degree of uncertainty, and receivers often have to make updating assessments based on percentages and probabilities about the other’s type rather than on absolutes. In game theoretic terms, this description roughly fits with the concept of a ‘semi-separating equilibrium.’ Instead of allowing observers to conclude that only one type of actor *always/never* takes a certain action in the given situation, a semi-separating signal provides us only with a *probability* of what type of actor is *more likely* to take that action. This variant of the argument posits that rational actors update beliefs based on *probabilistic* calculations rather than adopting *possibilistic* worst-case assumptions about others that do not the most costly signals possible. The implication here is rather than the strict dichotomous classification of signals as

¹⁶ As one scholar has put it, because “rationalist theory [correctly] identifies the problem of the inability of either state to credibly communicate information about its preferences, its assumptions largely preclude a satisfying conclusion,” and it thus “strongly predisposes analysis towards expecting the failure of engagement.” Marc Lynch, “Why Engage? China and the Logic of Communicative Engagement,” *European Journal of International Relations* (2002), p. 408.

‘costly’ or ‘costless,’ the difference is actually continuous, moving from ‘least costly’ to ‘most costly’ with many possible additional types of signals in between. Many prominent rationalist scholars, such as James Fearon and Eric Gartzke, have implicitly or explicitly adopted this weaker costly signaling logic.¹⁷

Most notably, Kydd has embraced the probabilistic logic in the very type of situation—reassurance—that this paper focuses upon. Kydd argues that the costliness of signals in reassurance must be somewhere between too costless and too costly to be successful, but he acknowledges that there is a continuum between these extremes that he does not see as mutually exclusive.¹⁸ Specifically, he argues that reassurance—or what he calls trust¹⁹—can develop between two security-seeking states that are unsure about each other’s underlying preferences so long as they begin by making and reciprocating smaller but still costly signals. Sending and receiving these types of signals allows the states to gradually build up to making larger, more costly signals and concessions on the issues they care about the most. Kydd formally models this by depicting a game with a series of consecutive rounds where the early rounds mean less to the players than later rounds. Though the payoffs or losses in the early rounds matter less to states, they are still large enough to constitute partially separating (he says fully separating) equilibria that allow the other to probabilistically but conclusively update their beliefs about the other’s type and act in the next round, worth a bit more than the previous round, accordingly. In this way, his argument is a formalized, costly signaling version of GRIT—or the strategy of Graduated Reduction in International Tensions pioneered by Charles Osgood in the 1960s²⁰—that shows how non-revisionist, security-seeking states should often be able to overcome uncertainty and mistrust through starting small and gradually working up to larger and significant levels of cooperation.²¹

¹⁷ See for example, Fearon, “Signaling Foreign Policy Interests”; Gartzke, “War Is In the Error Term,” *International Organization* (1999).

¹⁸ As he at one point notes, for example, “The [trick] is finding a signal that is adequately costly to deter the mean types from sending it but not so costly that the nice type is afraid to send it.” Kydd, “Trust, Reassurance, and Cooperation,” *International Organization* (2000), p. 338.

¹⁹ The concept of trust for Kydd does not seem to carry any more meaning for Kydd than the belief that the other prefers to reciprocate cooperation with you rather than take advantage of it. As he says, “To trust someone, then...is to believe it relatively likely that they would prefer to reciprocate cooperation. To mistrust someone is to think it is relatively likely that they prefer to defect even if they think one will cooperate” (*Trust and Mistrust*, 9).

²⁰ On GRIT, see Charles Osgood, *An Alternative to War or Surrender* (1962).

²¹ In a nutshell, “[t]he mechanism that enables states to learn about each other’s motivations is cooperation...states are often willing to take a chance by cooperating with another country, in hopes of establishing trust” (*Trust and Mistrust*, 19).

This weaker version of the signaling logic potentially allows signaling models to account for much of the interaction that takes place between states in the real world. Rather than strictly separating actions into categories of costly or costless, gestures could be identified as ‘semi-costly,’ ‘not very costly,’ ‘sufficiently costly,’ or ‘highly costly,’ for example.²² More importantly for our purposes here, adopting this weaker logic suggests that if states can partially and probabilistically update beliefs about the other, they are thus able to learn about the other amidst greater uncertainty. If this weaker approach is logically sound, it thus deals a potentially fatal blow to our arguments about costly signaling being less applicable in reassurance scenarios. In effect, it solves the problem of our first argument that costly signals are too costly to send to military reassurance scenarios. For instance, Kydd posits that “...in devising these signals, care must be taken that they are adequately costly, so that untrustworthy types would not send them too in an effort to trick the other side. At the same time they cannot be made too costly, or the trustworthy types may be unwilling to send them because they expose the trustworthy type to too great a level of risk in case the other side turns out to be untrustworthy...” (*Trust and Mistrust*, 170). The implicitly argument here is that there are in fact signals that can avoid the difficulties posed at either too costly or costless extreme of the continuum of signals. This weaker logic also potentially mitigates the problems posed in our second and third arguments, since it allows states to partially but not fully update beliefs about the other amidst greater amounts of uncertainty (about costliness over time, about shifting exogenous systemic factors, about the issues areas that the other state cares about). If signals can be both small and costly, states in any type of reassurance scenario should be able to gradually learn the other’s underlying preferences on the issue areas they care about without having to make ‘shot in the dark’ and potentially debilitating guesses about the other’s gestures no matter how much external ‘noise’ may be getting in the way. In short, if states probabilistically update within the foundational framework adopted by rationalists, the logic of their arguments is significantly more likely to apply to cases of reassurance than we have argued here.

Alas, for a number of reasons, we have reason to remain skeptical about this weaker, probabilistic signaling logic. Most broadly, this weaker version of the argument may actually violate or even undermine the larger logic of the rationalist signaling paradigm. If signals can be

²² From this perspective, presumably any signal that is not exactly 50% likely to be taken by opposite types of actors would provide the receiver with some information about the receiver, since they could conclude that one type is more likely to act in this way than another type.

partially costly and/or partially costless, can we (the analyst or the statesman) ever even draw a definitive line between what signals are costly for bluffers and which are costless? If the line can be drawn, wouldn't the point on the continuum where signals become definitively costly just create the same sort of dichotomy between costly and costless that characterizes the strong logic? If the line cannot be drawn, how are the signals between completely costly and costless on the continuum to be read? If statesmen must make probabilistic judgment calls on the costliness of these signals, wouldn't this violate the foundational premise that only signals that would be too costly for a bluffer to make will lead rational actors to update?²³

Ultimately, however, we do not aim to conclusively answer whether or not this weaker version fits within the broader signaling logic, a topic that can surely be better addressed by scholars more well versed in the rationalist and modeling literature than us. In fact, for our purposes here we are happy to simply assume that the weaker version is theoretically acceptable. Nonetheless, three notable problems with the weaker rationalist logic still remain. First, in practice rationalist scholars problematically shirk explicitly explaining what makes a signal too costly or too costless. Instead, they simply assume that senders are capable of making semi-costly signals. This is a problem for Kydd specifically, who never establishes *a priori* criteria for what types of actions constitute these 'middle-of-the-road' signals. His argument among others thus falls victim to the Goldilocks Problem: the signal must be "just right" in terms of costliness (not too cheap or too costly), but analysts and statesmen are given no idea of how to distinguish "just right" signals from those that are too costly or costless.²⁴

²³ These puzzles can be illustrated by observing the tensions between the weak and strong logics present in one of the most prominent of costly signaling articles in the international relations field, James Fearon's "Signaling Foreign Policy Interests" (1997). Surprisingly, Fearon occasionally embraces the language of the probabilistic logic, saying for instance that "it must be *more likely* that a resolved state would make the threat than an unresolved state" (69) and discussing smaller costly signals that "*relatively less resolved* types are inclined to use" (81, emphasis added). Yet more often, he emphatically embraces the strong logic that would seem to preclude the effectiveness of semi-costly signals noting that "the possibility of sending convincing costly signals will make it impossible for a state to 'partially convince' by sending less costly signals" while uncertainty "forc[es] the defender to signal 'all or nothing,' because signaling anything less...leads to the inference that the defender" is bluffing (71, 75. See also 77, 80, and 81). Ultimately, Fearon acknowledges the frequent discrepancy between these implications of the strong argument and empirical reality, where smaller and less costly signals are not only sent but also often successful. Rather than fully embracing the weaker logic, however, Fearon ultimately kicks the can down the road, concluding that "[b]ecause we do observe efforts to partially commit in international relations, even when it seems that stronger signals are possible, this observation poses a puzzle that future work, both empirical and theoretical, might usefully address" (87).

²⁴ Kydd is paradoxically overly pessimistic in his assumptions and overly optimistic in his conclusions. On the one hand, he assumes that states are relatively untrustworthy and will neither believe cheap talk nor risk sending an overt costly signal, both of which might expose them to attacks by the other to their detriment. On the other hand, he argues that there are "just right" costly signals that can escape the reassurance dilemma identified by Montgomery,

Second, rationalists have neglected to offer logical or empirical support for why a series of less costly signals is likely to create an opportunity for more costly signals, gestures, and concessions in the future. Kydd cites the earlier literature on GRIT as the theoretical foundation for the argument that small cooperative gestures can lead to larger ones. Yet what he downplays is that the (vague) mechanism for connecting small to larger signals in GRIT was rooted in human psychological tendencies involving trust, subjective reference points and identity, not rationalism. More recently, political scientist Aaron Hoffman has convincingly argued that “unless the process of completing a series of small transactions...transforms the relations among suspicious parties,” a logic that is notably outside a rationalist framework, “decomposing big issues into small issues will not induce suspicious parties to entrust their interests to one another”⁴⁰. In short, even if we can identify what semi-costly signals are, we still lack a mechanism linking less costly to more costly gestures that is rooted in rationalism.²⁵

Finally and most importantly, adopting this probabilistic weaker signaling logic greatly increases the likelihood that analysts will have to regularly incorporate variables far outside of the rationalist model in order to offer complete and accurate empirical accounts.²⁶ This is true because if signals are not definitively either costly or costless and instead fall somewhere between these extremes on a ‘costliness continuum,’ assessments of costliness are almost always bound to be subjective. In other words, different statesmen (not to mention analysts) are likely to evaluate differently whether a particular semi-costly signal is sufficiently costly enough to update beliefs (whether it is, for example, likely to be made honestly 60% of the time and dishonestly 30%, or instead 95% / 25%), based as much on idiosyncratic expectations, prior beliefs, and worldviews as on ‘objective’ evaluations of costliness. For one thing, an actor’s prior experiences and expectations play a huge role in how signals and conciliatory or confrontational

and that even the most untrusting, suspicious states can identify and successfully use these symbols to set off on the path toward possible reconciliation.

²⁵ Aaron M. Hoffman, *Building Trust: Overcoming Suspicion in International Conflict* (2006). As Ken Booth and Nicholas J. Wheeler put it in a slightly different way, “The weakness of the rationalist reassurance arguments is that they “provid[e] little or no account of how decision-makers might come to appreciate that fear and uncertainty rather than ambition might motivate others. Moreover, even if actors have exercised security dilemma sensibility, and are ready to send a ‘costly signal’—the mechanism by which states communicate peaceful intent for [rationalism]—the latter is silent as to where the trust comes from that leads states to make the first move...” Booth and Wheeler, *The Security Dilemma: Fear, Cooperation and Trust in World Politics* (2008), pp. 104-105.

²⁶ As Marc Lynch argues, “The reliance of rationalist models upon the assumption of common knowledge suggests the importance of the background knowledge constructed through [previous interactions]. For rationalists, common knowledge means that the actors are aware of the nature of the game being played, its stakes and possible moves. This building of common knowledge shifts the strategic structure, even if capabilities and preferences remain constant,” (“Why Engage?” 193).

gestures are interpreted. Kydd himself admits that non-rational factors involving prior beliefs and experience condition how signals are interpreted.²⁷ Moreover, signals that are not either hugely costly or costly are likely to be assessed subjectively and differently by different actors regardless of prior history. As Kydd again admits about his own models, they “assume that state actions are automatically recognizable. Observers are not uncertain about who does what and what it means. This assumption is a simplification...” (*Trust and Mistrust*, 255).

The larger point here is while adopting a weaker rationalist model that allows probabilistic updating of beliefs by receivers and semi-costly signaling by senders might make the basic rationalist signaling logic more applicable in reassurance scenarios, it introduces so much perceptual uncertainty and potential subjectivity that remains only a shell of the foundational rationalist logic that it is meant to represent.

Probabilistic Signaling and the Cold War's End

These weaknesses are on display in Kydd's most prominent empirical test case of his argument, successful reassurance between the United States and the Soviet Union and the end of the Cold War. Kydd's story depicts Mikhail Gorbachev's Soviet Union gradually offering larger and larger unilateral concessions that were increasingly seen by key officials in the Ronald Reagan administration as too costly for a bluffing state not truly interested in reducing superpower security tensions to make. Specifically, Kydd treats Gorbachev's decision to accept huge Soviet absolute and relative losses in the INF Treaty of 1987 as passing some 'costliness threshold' that made American leaders begin to take notice. The more and most costly actions that followed, including militarily withdrawing from Afghanistan, unilaterally withdrawing substantial conventional forces from Europe, and, most notably, renouncing the use of force to prevent the losses of the Eastern European satellites in the revolutions of 1989, succeeding in more fully 'updating' the beliefs of American leaders about the Soviet Union, even leading them to see Gorbachev as a crucial ally rather than a suspicious adversary.

²⁷ In a rather frank and surprising admission, Kydd concedes, “[t]he common prior assumption behind most game theoretic models is extremely unrealistic... Individuals interact with others who have radically different worldviews. Beliefs are influenced by many factors besides a rational assessment of observed facts” (*Trust and Mistrust*, 256). In the context of his specific account, he notes that prior levels of trust are crucial for understanding how a reassurance scenario between two actors might work. This prior level of trust, he says, is itself conditioned on a) prior interactions with this particular state, b) prior generalized experience with other states, or c) particular theories or beliefs about interactions in certain situations! (331-332).

What Kydd's account fails to explain, however, is why Gorbachev began making costly gestures that could have potentially increased Soviet vulnerability in the first place, nor why he continued making such gestures without tangible costly reciprocations from the United States. Thus, what is most interesting and in need of an explanation *isn't* why the United States came to update their beliefs about the Soviet Union in response to enormous unilateral concessions—this would be the optimal choice for any rational actor—but why the USSR came to send such signals in the first place. “What Kydd's analysis overlooked,” say Ken Booth and Nicholas Wheeler, was that “[t]rust began to grow before Gorbachev's costly signaling, and while the latter reinforced the former, what has to be explained is how the trust became possible that led Gorbachev in the first place to make his dramatic moves.”²⁸ (*The Security Dilemma*, 155). Most importantly, Kydd himself even admits that the Soviet shift to ‘new thinking’ preferences was the crucial first step to understanding why Gorbachev was willing to begin offering unilateral concessions, an argument clearly outside of his logic.²⁹ The logic of rationalist signaling thus could be said to begin operating only after the Soviet Union took crucial actions outside of the model's logic. Furthermore, Kydd commits at least two additional fallacies here: First, he often relies on the subjective conclusions of what gestures Soviet and American officials *think* are significant to assign degree of costliness to particular signals. Second he paradoxically does not conduct *enough* process tracing to determine that that the relevant actors interpreted the signals he sees as the most costly as the most influential to their changing beliefs. Contrary to his assumptions, many historians and scholars of the Cold War have found that “what is striking...is how little significance the key policy-makers in Washington attached to Moscow's” costly unilateral gestures, and instead more often “attached most importance to the personal relations that developed between the key players” (157). In other words, while leaders often did find the signals Kydd thinks are most important to be significant and revealing, it was often actually non-

²⁸ *The Security Dilemma*, p. 155. For their excellent summary of the rationalist and constructivist interpretations of the end of the Cold War in both political science and historical literatures, see pp. 145-158. See also Mark L. Haas, “The United States and the End of the Cold War: Reactions to Soviet Power, Policies, or Domestic Politics?” *International Organization* (2007).

²⁹ As Kydd notes about the shift to ‘new thinking,’ “the rewards for mutual cooperation...were perceived to be much more desirable...[t]hus, I would argue that the new Soviet leadership viewed the gains of mutual cooperation as greater and the cost of exploitation as lesser than previous Soviet leaderships...” (*Trust and Mistrust*, 343). For similar non-rationalist views of ‘new thinking's’ origins, see Robert English, *Russia and the Idea of the West* (2006); Jeffrey T. Chechel, *Ideas and International Political Change: Soviet/Russian Behavior and the End of the Cold War* (1997); Sarah E. Mendelson, *Changing Course: Ideas, Politics, and the Soviet Withdrawal from Afghanistan* (1998); Jeffrey W. Legro, *Rethinking the World: Great Power Strategies and International Order* (2005).

costly, ‘cheap talk’ gestures, such as trusting bonds developed between particular personalities, that might have played an even more significant role for the successful revealing of compatible security interests that completed Soviet-American reassurance and ended the Cold War.

* * *

Summing up, if the arguments put forward both in this section and overall are convincing, we must come to one of two logical conclusions regarding strong and weak rationalist signaling logic and reassurance: Either we accept that the strong version of the logic (exclusively costly or costless signals) fails to explain almost any case of reassurance we observe empirically, or we adopt the weak form of the argument that can technically explain more cases of reassurance, but at the same time opens interpretations of ‘costliness’ up to subjective, non-rational explanations. Either line of defense admits the inability of rationalist costly signal arguments to completely explain cases of successful reassurance frequently observed in international relations. In either case, recognizing a need for exploring non-rationalist hypotheses that incorporate psychological or social constructivist mechanisms for updating beliefs—well off of the Bayesian path—is thus a reasonable conclusion.

II. BEYOND THE RATIONALIST APPROACH (woefully incomplete)

2.1 Reassurance Really Happens

[We have cases! really! Honestly, this project started with our empirical observations that there have been important cases of successful reassurance that do not fit easily within the rationalist logic. Specifically, we observe states coming to trust each other enough to cooperate through signals that either are not costly or are ambiguously and subjectively costly. In some of these cases we do see costly signaling, yet process tracing reveals that it was often actually ‘cheap talk’ that led receivers to update. In other cases, less-than-costly signaling leads to reassurance and successful cooperation where rationalist arguments would not expect it. CASES IN THIS SECTION TO DEMONSTRATE THE PLAUSIBILITY OF OUR CLAIMS TO COME!!!]

2.2 Exploring New Hypotheses

[In this section, we survey some arguments for non-rational reassurance and updating beliefs from social constructivism in the international relations literature as well as arguments from social psychology, sociology, and evolutionary biology. (Some of these are foreshadowed in the

conclusion). Overall, the goal is to formulate new hypotheses for why states might update their beliefs about each other that are not dependent on the restrictive rationalist costly signaling assumptions. Specifically, we hope to be able to derive hypotheses that are clear and simple enough to encourage future testing in IR cases.]

CONCLUSION [section still rough]

In this paper, we have argued that the mechanisms for signaling in reassurance scenarios are fundamentally different from those in crisis bargaining. Signaling often poses greater risks in RE than in CB because of the nature of the actions necessary to send credible signals. The indefinite duration of reassurance scenarios results in states possessing longer time-horizons, a factor that makes the threshold at which a signal is identified as definitively costly significantly higher. Most importantly, rationalist scholars have mischaracterized the complexity of uncertainty over preferences that states must overcome for reassurance to succeed. Whereas crisis bargaining is limited to a single known issue area over which the content of both actors' preferences is also already known, both of these dimensions—the underlying preferences of both actors on a particular issue and the importance each assigns to that issue relative to others—remain unknown and difficult to learn in reassurance. While a weaker version of the rationalist signaling logic potentially averts some of these difficulties in reassurance, employing this weaker argument ultimately raises more questions than answers about whether reassurance can actually *ever* succeed if states act according to the assumptions of rationalism.

Our arguments potentially call into question the practical applicability of costly signaling as a means of updating beliefs about others' preferences, and render rationalist theories of reassurance potentially unable to explain many cases of successful reassurance. Yet the fact that costly signaling is unlikely to be effectively used in reassurance does not, as some rationalists have argued (see Montgomery), preclude reassurance from succeeding altogether. Rather than simply giving up on reassurance, states often turn to other means for updating their beliefs that often involve ideational and social cues. These alternative types of signals—including arguments over norms, communicative action, or gestures that trigger an emotional response or that defy a receiver's previous expectations, biases or prejudice—are often are not objectively costly, yet produce the same effects—updating of beliefs—that rationalists would expect only costly signaling could achieve. Such alternative signaling mechanisms deserve further study.

In conclusion, we encourage scholars to depart from the costly signaling dogma of information transfer, and to open themselves up to the possibility that some alternative perspective might be able to systematically account for when, why and how leaders update their beliefs about others. Previous constructivist and psychological arguments in the international relations literature (such as those of Wendt, Risse, Mercer, Jervis, Larson, Lebow, Stein) have touched on these concepts, yet for the most part have failed to do so in a focused, methodologically rigorous, and falsifiable way. We therefore advocate the resuscitation of these research programs, in the context of reassurance scenarios, with a goal of constructing a comprehensive framework for better understanding—or perhaps ultimately refuting—the logic of Bayesian updating.