

THE PRICE OF PEACE: MOTIVATED REASONING AND COSTLY
SIGNALING IN INTERNATIONAL RELATIONS
Supplementary Appendix

Joshua D. Kertzer, Brian Rathbun, and Nina Srinivasan Rathbun

May 3, 2019

Contents

1	Differentiating motivated skepticism and selective attention	1
2	Supplementary analysis: survey experiment I	3
	Table 1: The effects of costly signals on DVs apart from trustworthiness	5
	Table 2: Results from fully-saturated three-way interaction models	6
	Figure 2: Results from fully-saturated three-way interactions	7
	Table 3: Results robust to controlling for heterogeneous information	8
	Table 5: Results using political ideology as moderator	11
	Figure 3: Conditional effects of signals by political ideology	12
	Table 6: Conditional effects of signals by militant internationalism	13
	Figure 4: Conditional effects of signals by militant internationalism	14
	Table 7: Trustworthiness of Iran	15
	Table 8: Conditional effects of signals by cooperative internationalism	16
	Table 9: Conditional effects of signals by partisanship	17
	Figure 5: Conditional effects of signals by partisanship	18
	Figure 6: Foreign policy orientations outperform partisan identity	19
3	Supplementary information: terms of the Iran Deal	20
4	Supplementary analysis: survey experiment II	21
4.1	Experimental design	21
	Figure 7: Experimental design: 2018 survey	22
4.2	Results	24
	Table 10: Change in trustworthiness following costly signals	24
	Table 11: Change in trustworthiness following each type of costly signal	25
4.2.1	Testing for variance of priors	27
	Table 12: Variance of priors (I)	29
	Table 13: Variance of priors (II)	30
5	Motivated reasoning and costless signals	31

1 Differentiating motivated skepticism and selective attention

In the paper we suggest that many discussions of motivated reasoning tend to conflate two mechanisms that should be treated as conceptually distinct for analytic purposes, even if they often coexist in practice. The first is the notion of *motivated skepticism*, which we use here to refer to asymmetries in how individuals *evaluate* information. The second is the notion of *selective attention*, which we use here to refer to asymmetries in how individuals *seek out* information. The tendency to uncritically accept attitudinally congruent arguments and dismiss contrary ones (the notion of disconfirmation bias) implicates the former.¹ The tendency to seek out confirmatory evidence in support of your beliefs (the notion of confirmation bias) implicates the latter (Taber and Lodge, 2006). While individuals who engage in one tendency often also engage in the other, we can treat them as analytically distinct.

Thus, for example, members of the George W. Bush administration in the lead-up to the Iraq War were both looking for signs that Iraq had weapons of mass destruction — paying selective attention — and dismissing evidence to the contrary — engaging in motivated skepticism (Jervis, 2006). When we combine these two dynamics, we end up with a model of individuals not as “intuitive scientists” motivated by accuracy goals, but as “intuitive lawyers” motivated by directional goals (Lodge and Taber, 2000). Whereas the former consider information in an evenhanded and dispassionate fashion, the latter both process and consider it selectively, seizing on information consistent with their pre-existing beliefs, and ignoring information inconsistent with it.

Two points are relevant here, one methodological, and one substantive. First, in a natural setting, differentiating between these two mechanisms can be challenging, because individuals have the option of self-selecting into the kinds of information they receive; Democrats may choose to watch MSNBC, and Republicans Fox News (Iyengar and Hahn, 2009), each of which may choose to cover very different sets of stories, or cover similar stories in different ways. In contrast, experiments are better suited to differentiating between the two mechanisms. Even though respondents may come to the experiment having been exposed to different prior information (Druckman and Leeper, 2012), the experimenter can control what information respondents receive in the experiment itself. In the analysis in the main text, we are therefore able to disentangle the two mechanisms from one another.

Second, there are a variety of debates in the study of political behavior about how to reconcile motivated skepticism with Bayesian models of learning (e.g. Gerber and Green, 1999; Bartels, 2002; Bullock, 2009). We are deliberately agnostic on this point. As Bullock (2007, 90) notes, Bayes’ theorem “is merely an accounting identity”, capable of being configured in an infinite number of ways, such that almost any behavior, regardless how puzzling, can be retrofit into a Bayesian framework. “The problem isn’t that the Theorem is not restrictive enough”, but that it is so flexible that “it permits what we should reject as irrational” (Bullock, 2007, 121).

Indeed, in Figure 1 below, we show how the type of motivated skepticism we study here can be incor-

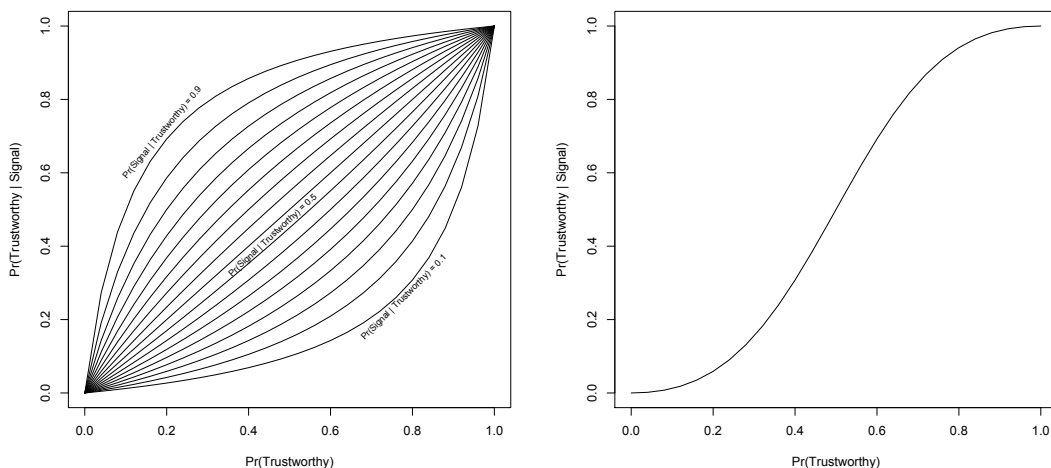
¹See also Bisgaard (2019) on motivated attribution.

porated into a Bayesian framework. Suppose an observer with prior beliefs ($Pr(T)$) about Iran’s trustworthiness; Iran sends a signal (S), the observer assesses the signal’s informativeness ($Pr(S|T)$), and updates accordingly, producing posterior belief $Pr(T|S)$. For a motivated skeptic, the degree to which a new piece of information is seen as adjudicating between alternative states of the world depends whether it forces them to change their beliefs or not; in other words, the likelihood function is partially a function of one’s prior beliefs. Although there is nothing in Bayes’ rule precludes this possibility (Kahan, 2016, 4), it violates the classic image of rational information processing as objective, deliberative and dispassionate; it means that observers who begin with divergent priors may not necessarily move towards similar conclusions even as they receive identical information. In this sense, we can be concerned about motivated reasoning even in the absence of “backfire” effects (Wood and Porter, 2019; Guess and Coppock, Forthcoming).

Figure 1: Modeling the relationship between prior beliefs and interpretation of evidence

(a) Likelihood independent of priors

(b) Likelihood correlated with priors



In a classic Bayesian setup, where S is the signal, and T is trustworthiness, $Pr(T|S) = \frac{Pr(S|T) \times Pr(T)}{Pr(S|T) \times Pr(T) + Pr(S|\sim T) \times Pr(\sim T)}$. Figure 1(a) presents simulations where priors about Iran’s trustworthiness ($Pr(T)$) are on the x axis, and posteriors about Iran’s trustworthiness given the signal ($Pr(T|S)$) are on the y axis. Each line on the plot itself is a simulation for a different likelihood ($Pr(S|T) = \{0.1 \dots 0.9\}$). The plot shows that as long as costly signals are informative (that is, as long as the likelihood is 0.5), the slope of the curve is steeper for individuals with lower priors, such that the individuals who see Iran as the least trustworthy should respond the *most* to the signal: e.g. if $Pr(S|T = 0.7)$, $Pr(T|S) - Pr(T) = 0.2$ for a skeptical individual with a prior of 0.3, but 0.1 for a trusting individual with a prior of 0.8. Figure 1(b) presents similar simulations, but this time the likelihood is a function of the prior (that is, $Pr(S|T) = Pr(T)$). Now, the curve is concave for individuals with high priors, and convex for individuals with low ones: signals work as expected for individuals who are motivated to find evidence of trustworthiness, but either fail to work or backfire among individuals who are motivated skeptics. As we note below, our argument is not that motivated skepticism is inconsistent with a Bayesian framework, but rather, that the lack of truth convergence it displays is inconsistent with normatively desirable theories of information processing, and has important implications for how costly signals are interpreted.

The interesting question for us, then — and, we hope, for others — is not the narrower challenge that motivated skepticism may or may not pose to Bayesian models, but the broader challenge it poses for standards of normatively desirable decision-making. In American politics, the questions at stake concern the quality of democratic deliberation: how can democracy flourish in an era when politics has become our

identity (Mason, 2018), and efforts to correct misperceptions go so awry? How can we bring together those who are of different minds about a problem if they cannot agree about what is fact and what is fiction? These insights, however, are also important to IR scholars who ask a similar question: how do individuals respond to credible new information about a political question when it contradicts their previous beliefs? If individuals see what they want to see, and view information that challenges their preconceptions as “fake news”, is there such a thing as a credible signal of reassurance that all respond to similarly?

2 Supplementary analysis: survey experiment I

To build a greater understanding of the microfoundations of costly signals of reassurance, we fielded our primary survey experiment on 1815 American adults recruited through Amazon Mechanical Turk (MTurk). MTurk is increasingly used in experimental research in the social sciences to obtain more larger and more diverse samples than the convenience samples of college students typically used in this type of work (Berinsky, Huber, and Lenz, 2012), and experimental studies conducted on MTurk have been published in a variety of prominent outlets, including the *American Political Science Review* (Aarøe, Petersen, and Arceneaux, 2017; McEntire, Leiby, and Krain, 2015), the *American Journal of Political Science* (Bishin, Hayes, Incantalupo et al., 2016; Huff and Kertzer, 2018), *International Organization* (Wallace, 2013; Chaudoin, 2014; Brutger and Kertzer, 2018) and others. Respondents were paid \$1.50 for completing a 15-20 minute survey. The survey was fielded following the conclusion of a general framework of an agreement with Iran but before final details of the pact were concluded in June 2015. The study was fielded from April 18-21, 2015. This was an optimal time to collect data because the issue was in the public eye but not being widely discussed in the media.

Table 1 presents a series of regression models similar to Table 1 in the main text, but with danger, support for the deal, and the likelihood of Iran making the deal as dependent variables instead. The first model in each table panel estimates the effects of the treatments (the signals sent by Iran in the negotiations, either regarding the inspections regime or the status of Iran’s fuel cycle capabilities). The second model in each table panel adds foreign policy orientations (militant internationalism, isolationism, and most importantly for our purposes, cooperative internationalism) as covariates, as well as a series of standard demographic characteristics. The third adds a feeling thermometer variable capturing feelings towards Iran; in this manner, the tables control for heterogeneous affect towards Iran itself. Finally, the fourth model in each table panel adds need for cognition (Cacioppo and Petty, 1982), to control for the possibility that some individuals are less dispositionally inclined to change their beliefs by virtue of their cognitive style. This is also relevant given work by Pennycook and Rand (2018), who argue that individuals are “lazy, not biased”, and that some of what we think of as normatively undesirable forms of information processing are driven less by motivated reasoning, and more by a failure to think analytically. The results for the first two panels strongly parallel those from Table 1 in the main text; the results for the third panel differs, showing the extent to which

assessments of the likelihood of Iran making the deal are fundamentally different than questions about how trustworthy or dangerous Iran is, and participants' own level of support for the deal.

Table 2 replicates Table 2 from the main text, but estimating a three-way interaction model, in which the effects of each type of costly signal (regarding inspections, or the fuel cycle) are conditional on both cooperative internationalism *and* the other type of costly signal. As Braumoeller (2004) shows, if we believe that the effect of the proposed inspections regime is contingent on cooperative internationalism, and that the effect of the proposed fuel cycle status is contingent on cooperative internationalism, it is necessary to estimate the fully-saturated three-way interaction, with all of the lower-order terms, rather than making implicit and untested restriction assumptions. Since three-way interactions are difficult to substantively interpret, the two panels of Figure 2 presents the conditional effect of each set of treatments, paralleling the presentation of the results in Figure 2 of the main text. The same substantive points remain: in Figure 2(a), individuals high in CI respond to the inspections treatment by being significantly more likely to see Iran as trustworthy, significantly less likely to see Iran as dangerous, and significantly more likely to support the deal. The same pattern is also evident with respect to the fuel cycle treatments in Figure 2(b). In both panels, individuals who are low in CI update much less; given the width of the 95% confidence intervals, in many of the plots we fail to find evidence they significantly update at all. The figures thus reconfirm our conclusions from the main text about how costly signals of reassurance are most effective for those who already want to be reassured.

Table 3 replicates Table 2 from the main text, but also controlling for heterogeneous information, modeled by a series of domain-specific political knowledge questions (McGraw and Pinney, 1990) measuring participants' knowledge about Iranian domestic politics. It is possible that our results reflect unobserved heterogeneous beliefs that might be correlated with our predispositions of interest and which might affect the weight attached to the information conveyed in the costly signal. For example, if individuals high in CI systematically differ from their low-CI counterparts in pretreatment exposure to information, this would induce a confound in our estimate of heterogeneous treatment effects. To eliminate this possibility, as part of the block of demographic and individual difference measures administered either before or after the survey, participants completed a questionnaire measuring their levels of knowledge about Iranian politics (one item tested respondents' ability to correctly identify the dominant branch of Islam in Iran; another involved correctly identifying the Iranian President, Hassan Rouhani, and so on). Not only is knowledge about Iran uncorrelated with CI, but our results in Table 3 remain unchanged compared to Table 2, such that we cannot interpret the patterns we report here to be an artifact of heterogeneous pretreatment exposure to information about Iran. Thus, we find no evidence that those who react less strongly to costly signaling treatments are doing so based on preexisting knowledge.

Table 4 replicates Table 2 from the main text, but this time testing the effects of feelings towards Iran as a moderator of the experimental treatments. The results show that preexisting attitudes towards Iran are far weaker moderators of the treatments than CI is.

Table 1: The effects of costly signals on DVs apart from trustworthiness

	Danger posed by Iran			Support for deal with Iran			Likelihood of Iran making deal					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Inspections: None	0.816*** (0.144)	0.947*** (0.130)	0.973*** (0.127)	0.970*** (0.127)	-1.166*** (0.080)	-1.241*** (0.076)	-1.255*** (0.074)	-1.256*** (0.074)	0.153* (0.091)	-0.021 (0.051)	-0.021 (0.051)	-0.018 (0.051)
Inspections: All	-0.501*** (0.144)	-0.363*** (0.130)	-0.389*** (0.127)	-0.393*** (0.128)	0.316*** (0.080)	0.279*** (0.076)	0.293*** (0.075)	0.292*** (0.075)	0.050 (0.091)	-0.032 (0.051)	-0.032 (0.051)	-0.029 (0.051)
Fuel Cycle: Continue	0.814*** (0.144)	0.724*** (0.130)	0.719*** (0.127)	0.718*** (0.127)	-0.508*** (0.080)	-0.462*** (0.076)	-0.459*** (0.074)	-0.460*** (0.075)	0.004 (0.091)	0.036 (0.051)	0.036 (0.051)	0.037 (0.051)
Fuel Cycle: Dismantle	-0.280* (0.144)	-0.318** (0.129)	-0.289** (0.127)	-0.289** (0.127)	0.124 (0.080)	0.152** (0.075)	0.137* (0.074)	0.137* (0.074)	-0.046 (0.091)	-0.029 (0.051)	-0.029 (0.051)	-0.029 (0.051)
Militant Intl.		4.203*** (0.325)	3.394*** (0.331)	3.381*** (0.331)		-1.163*** (0.190)	-0.726*** (0.193)	-0.730*** (0.194)		-1.234*** (0.127)	-1.232*** (0.132)	-1.220*** (0.132)
Cooperative Intl.		-0.636** (0.303)	-0.289 (0.299)	-0.246 (0.303)		0.991*** (0.177)	0.803*** (0.175)	0.818*** (0.177)		1.327*** (0.119)	1.326*** (0.120)	1.288*** (0.121)
Isolationism		-0.888*** (0.308)	-0.959*** (0.302)	-0.961*** (0.302)		0.700*** (0.180)	0.738*** (0.177)	0.737*** (0.177)		-0.062 (0.121)	-0.062 (0.121)	-0.060 (0.121)
Feelings toward Iran			-0.023*** (0.003)	-0.023*** (0.003)			0.013*** (0.002)	0.013*** (0.002)			0.00005 (0.001)	-0.0001 (0.001)
Need for cognition												
Male		-0.581*** (0.107)	-0.579*** (0.105)	-0.567*** (0.106)		-0.048 (0.063)	-0.050 (0.061)	-0.045 (0.062)		-0.041 (0.042)	-0.041 (0.042)	-0.053 (0.042)
log(Age)		0.656*** (0.184)	0.547*** (0.180)	0.552*** (0.180)		-0.422*** (0.107)	-0.363*** (0.105)	-0.361*** (0.105)		-0.138* (0.072)	-0.138* (0.072)	-0.143** (0.072)
White		-0.191 (0.133)	-0.247* (0.130)	-0.238* (0.131)		-0.029 (0.078)	0.002 (0.076)	0.005 (0.076)		0.036 (0.052)	0.036 (0.052)	0.028 (0.052)
Education		-0.058 (0.043)	-0.069* (0.042)	-0.062 (0.042)		0.034 (0.025)	0.040* (0.024)	0.043* (0.025)		0.005 (0.017)	0.005 (0.017)	-0.001 (0.017)
Party ID		0.617*** (0.238)	0.583** (0.233)	0.587** (0.233)		-0.282** (0.139)	-0.264* (0.136)	-0.263** (0.136)		-4.144*** (0.093)	-4.144*** (0.093)	-4.147*** (0.093)
Constant	5.161*** (0.131)	2.048** (0.834)	3.428*** (0.830)	3.480*** (0.833)	4.014*** (0.073)	5.051*** (0.486)	4.304*** (0.486)	4.324*** (0.487)	4.630*** (0.083)	6.566*** (0.327)	6.563*** (0.332)	6.517*** (0.333)
N	1,815	1,815	1,815	1,815	1,815	1,815	1,815	1,815	1,816	1,816	1,816	1,816
Adjusted R ²	0.074	0.251	0.283	0.283	0.199	0.282	0.308	0.308	-0.0004	0.690	0.690	0.691

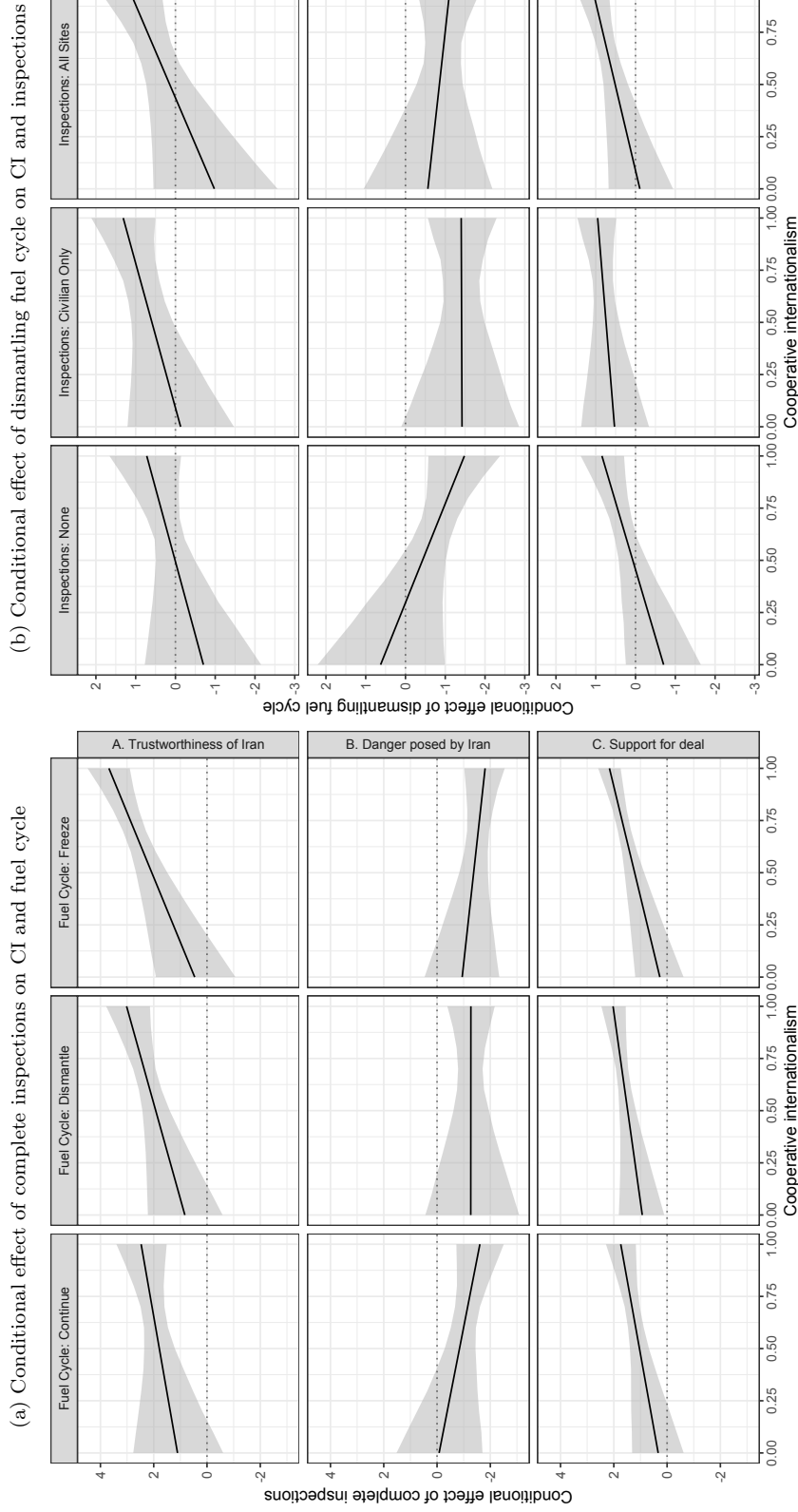
*p < .1; **p < .05; ***p < .01

Table 2: Results from fully-saturated three-way interaction models

	Trustworthy	Dangerous	Approval of Deal
	(1)	(2)	(3)
Inspections: None	-1.054 (0.751)	-0.250 (0.747)	-0.090 (0.433)
Inspections: All	-0.575 (0.734)	-1.173 (0.730)	0.163 (0.423)
Fuel Cycle: Continue	0.036 (0.708)	0.385 (0.704)	-0.120 (0.408)
Fuel Cycle: Dismantle	-0.103 (0.692)	-0.926 (0.687)	0.365 (0.398)
Cooperative internationalism	3.049*** (0.747)	-1.621** (0.742)	1.598*** (0.430)
Militant internationalism	-2.774*** (0.328)	4.222*** (0.326)	-1.189*** (0.189)
Isolationism	0.913*** (0.311)	-0.882*** (0.309)	0.693*** (0.179)
Male	-0.014 (0.108)	-0.593*** (0.107)	-0.042 (0.062)
log(Age)	-0.891*** (0.186)	0.656*** (0.185)	-0.423*** (0.107)
White	-0.251* (0.134)	-0.175 (0.134)	-0.037 (0.077)
Education	0.010 (0.043)	-0.057 (0.043)	0.033 (0.025)
Party ID	-0.349 (0.239)	0.607** (0.238)	-0.270* (0.138)
Inspections: None x Fuel Cycle: Continue	0.501 (1.067)	-0.366 (1.061)	0.209 (0.614)
Inspections: None: Fuel Cycle: Dismantle	-0.057 (1.059)	1.590 (1.052)	-0.992 (0.609)
Inspections: All x Fuel Cycle: Continue	1.137 (1.033)	0.510 (1.027)	0.276 (0.595)
Inspections: All x Fuel Cycle: Dismantle	0.299 (1.015)	1.212 (1.009)	-0.315 (0.584)
Inspections: None x CI	-1.850* (1.048)	1.968* (1.042)	-1.715*** (0.603)
Inspections: All x CI	1.369 (1.048)	1.049 (1.041)	0.214 (0.603)
Fuel Cycle: Continue x CI	-1.147 (1.028)	0.803 (1.022)	-0.650 (0.592)
Fuel Cycle: Dismantle x CI	0.335 (1.011)	0.653 (1.005)	-0.158 (0.582)
Inspections: None x Fuel Cycle: Continue x CI	0.352 (1.510)	-0.200 (1.502)	0.065 (0.870)
Inspections: None x Fuel Cycle: Dismantle x CI	0.276 (1.501)	-2.175 (1.492)	1.133 (0.864)
Inspections: All x Fuel Cycle: Continue x CI	-1.526 (1.495)	-0.844 (1.486)	-0.441 (0.860)
Inspections: All x Fuel Cycle: Dismantle x CI	-0.770 (1.454)	-1.183 (1.446)	0.293 (0.837)
Constant	6.991*** (0.956)	2.667*** (0.950)	4.673*** (0.550)
N	1,816	1,815	1,815
Adjusted R ²	0.281	0.252	0.293

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

Figure 2: Results from fully-saturated three-way interactions



The figure depicts the predicted values from a set of three-way interaction models in Table 2; in panel a, the effects of the inspection regime on each dependent variable (trustworthiness of Iran, danger posed by Iran, and support for the deal) is conditional on both CI (depicted on the x axis) and the fuel cycle regime, while in panel b, the effects of the fuel cycle regime on each dependent variable is conditional on both CI and the inspections regime. In panel a, the left-hand column shows the conditional effects of complete inspections when Iran continues its current fuel cycle; the middle column shows the conditional effects of complete inspections when Iran freezes its fuel cycle; the right-hand column shows the conditional effects of complete inspections when Iran dismantles its fuel cycle. In panel b, the left-hand column shows the conditional effects of dismantling its fuel cycle when Iran allows no inspections; the middle column shows the conditional effects of dismantling its fuel cycle when Iran allows civilian inspections only; the right-hand column shows the conditional effects of dismantling its fuel cycle when Iran allows complete inspections. Both panels present 95% confidence intervals and point estimates derived from 1500 bootstraps.

Table 3: Results robust to controlling for heterogenous information

	Trustworthy	Dangerous	Approval of Deal
	(1)	(2)	(3)
Inspections: None	-0.869** (0.433)	0.169 (0.429)	-0.357 (0.249)
Inspections: All	-0.109 (0.415)	-0.570 (0.411)	0.126 (0.239)
Cooperative internationalism	3.234*** (0.580)	-0.836 (0.574)	1.365*** (0.333)
Fuel Cycle: Continue	0.522 (0.433)	0.464 (0.429)	0.022 (0.249)
Fuel Cycle: Dismantle	-0.033 (0.428)	-0.124 (0.424)	0.001 (0.246)
Militant internationalism	-2.808*** (0.336)	3.967*** (0.333)	-1.097*** (0.193)
Isolationism	0.912*** (0.311)	-0.910*** (0.308)	0.705*** (0.179)
Male	-0.008 (0.109)	-0.581*** (0.108)	-0.047 (0.062)
log(Age)	-0.850*** (0.189)	0.669*** (0.187)	-0.427*** (0.108)
White	-0.235* (0.134)	-0.180 (0.132)	-0.039 (0.077)
Education	0.015 (0.043)	-0.054 (0.042)	0.031 (0.025)
Knowledge about Iran	-0.165 (0.190)	-0.238 (0.188)	0.123 (0.109)
Ideology	-0.234 (0.251)	1.033*** (0.249)	-0.428*** (0.144)
Inspections: None x CI	-1.703*** (0.617)	1.147* (0.611)	-1.304*** (0.355)
Inspections: All X CI	0.635 (0.600)	0.323 (0.594)	0.210 (0.345)
Fuel Cycle: Continue x CI	-1.439** (0.617)	0.415 (0.611)	-0.741** (0.355)
Fuel Cycle: Dismantle x CI	0.181 (0.607)	-0.301 (0.601)	0.228 (0.349)
Constant	6.663*** (0.898)	2.207** (0.889)	4.800*** (0.516)
N	1,816	1,815	1,815
Adjusted R ²	0.279	0.257	0.294

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

Table 4: Conditional effects of signals by feelings towards Iran

	Trustworthy	Dangerous	Support for deal
	(1)	(2)	(3)
Inspections: None	-1.932*** (0.206)	1.125*** (0.213)	-1.408*** (0.125)
Inspections: All	0.097 (0.201)	-0.266 (0.208)	0.190 (0.122)
Fuel Cycle: Continue	-0.351* (0.203)	0.716*** (0.210)	-0.591*** (0.123)
Fuel Cycle: Dismantle	0.080 (0.205)	-0.499** (0.212)	0.181 (0.124)
Inspections: None x Iran	-0.004 (0.006)	-0.005 (0.006)	0.005 (0.003)
Inspections: All x Iran	0.010* (0.006)	-0.005 (0.006)	0.004 (0.003)
Fuel Cycle: Continue x Iran	-0.002 (0.006)	0.0003 (0.006)	0.005 (0.003)
Fuel Cycle: Dismantle x Iran	-0.001 (0.006)	0.007 (0.006)	-0.001 (0.003)
Cooperative internationalism	1.890*** (0.298)	-0.011 (0.308)	0.699*** (0.180)
Militant internationalism	-1.335*** (0.326)	3.128*** (0.338)	-0.613*** (0.198)
Isolationism	1.056*** (0.291)	-0.964*** (0.301)	0.734*** (0.176)
Feelings towards Iran	0.040*** (0.005)	-0.023*** (0.005)	0.009*** (0.003)
Male	-0.022 (0.101)	-0.591*** (0.105)	-0.043 (0.061)
log(Age)	-0.698*** (0.174)	0.523*** (0.180)	-0.349*** (0.105)
White	-0.135 (0.126)	-0.252* (0.130)	0.005 (0.076)
Education	0.037 (0.040)	-0.067 (0.042)	0.039 (0.024)
Ideology	-0.207 (0.235)	1.033*** (0.243)	-0.448*** (0.142)
Constant	4.753*** (0.809)	3.281*** (0.837)	4.444*** (0.490)
N	1,816	1,815	1,815
Adjusted R ²	0.366	0.287	0.311

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

Table 5 replicates Table 2 from the main text, but this time using political ideology as a moderator instead of CI. To facilitate a more straightforward interpretation, the results are visually represented in Figure 3. The results for the fuel cycle treatment are not as strong with ideology as they are for CI, but the results for the inspections treatment is the same, in which more conservative respondents generally respond less to inspections than their liberal counterparts. Table 6 does the same with respect to militant internationalism; the conditional effects are plotted in Figure 4. As was the case with Table 5, the results for the fuel cycle treatment are not as strong with MI as they are for CI, but the results for the inspections treatment is strikingly similar, in which hawks respond less to inspections than doves do. Liberals and doves update the most, even though they need the signals the least.

We also rule out a number of alternative interpretations for our results hinging on a lack of receptivity to the treatment. First, it might be the case that those who do not demonstrate a sensitivity to Iranian signals do so out of the belief that the signals are uninformative, as talk is cheap in diplomatic negotiations. In this line of argument, Iran’s bargaining positions do not discriminate between a dangerous, untrustworthy country and a peaceful, trustworthy one. If this were the case, however, they would change their tune if Iran actually agreed to these terms, that is when what could be cheap talk becomes actual bargaining concessions. This seems not to be the case. Our results in Table 2 in Appendix §2 show that those high in CI are both more likely to support a deal with Iran, and as model 3 of Table 8 show, become more so the more that Iran makes costly concessions. In contrast, those low in CI are no more likely to support the deal even as Iran gives away more.

This is particularly noteworthy in light of another set of results. We also asked our survey subjects how likely they thought it that Iran would be willing to make a deal on the terms described in their particular treatment. Model 3 in Table 1 shows that cooperative internationalism strongly predicts optimism irrespective of the experimental conditions, which themselves do not have any effect. This indicates a belief on the part of those low in CI not that Iran is toying with the United States but rather that Iran is not seriously committed to international cooperation. The lack of an interaction between CI and the treatments indicates that those low in CI do not expect Iran to make even the slightest of concessions.

Finally, given prominent discussions about partisan polarization in American politics (e.g. [Iyengar and Westwood, 2015](#); [Mason, 2018](#)), and the undeniable partisan dynamics in elite debates about the Iran deal, it is also worth examining the role of partisanship in some detail. Three points are worth noting. First, the results reported in the main text are robust to controlling for partisanship, as Table 7 and Table 1 illustrate. Second, as Table 9 and Figure 5 show, we recover similar, though slightly weaker, patterns with partisanship as a moderator of our treatment effects as we do with the main individual difference variables in the main text. In particular, democrats respond more to intrusive inspections regimes than Republicans do when it comes both to assessments of Iran’s trustworthiness, and respondents’ approval of the proposed deal; the interaction effects with the fuel cycle treatments are relatively weak.² Third, and most importantly,

²Importantly, the results for the moderating effects of partisanship in Table 9 are strikingly similar when the foreign policy orientations and political ideology are dropped from the models, so it is not the case that the effect of partisanship is being

Table 5: Results using political ideology as a moderator

	Trustworthy (1)	Dangerous (2)	Approval of Deal (3)
Inspections: None	-2.448*** (0.232)	1.164*** (0.230)	-1.658*** (0.134)
Inspections: All	0.545** (0.233)	-0.167 (0.230)	0.347*** (0.134)
Inspections: None x Ideology	1.156** (0.502)	-0.583 (0.496)	1.112*** (0.288)
Inspections: All X Ideology	-0.539 (0.490)	-0.516 (0.484)	-0.171 (0.281)
Fuel Cycle: Continue	-0.722*** (0.232)	1.223*** (0.229)	-0.682*** (0.133)
Fuel Cycle: Dismantle	-0.011 (0.231)	0.095 (0.228)	0.016 (0.133)
Fuel Cycle: Continue x Ideology	0.766 (0.499)	-1.277*** (0.493)	0.566** (0.287)
Fuel Cycle: Dismantle x Ideology	0.283 (0.492)	-1.092** (0.486)	0.372 (0.283)
Ideology	-0.567 (0.552)	2.411*** (0.545)	-1.156*** (0.317)
Cooperative internationalism	2.476*** (0.317)	-0.336 (0.313)	0.860*** (0.182)
Militant internationalism	-2.714*** (0.336)	3.928*** (0.332)	-1.042*** (0.193)
Isolationism	0.946*** (0.311)	-0.898*** (0.307)	0.724*** (0.179)
Male	-0.020 (0.108)	-0.596*** (0.107)	-0.042 (0.062)
log(Age)	-0.894*** (0.185)	0.616*** (0.183)	-0.405*** (0.106)
White	-0.207 (0.134)	-0.181 (0.133)	-0.020 (0.077)
Education	0.015 (0.043)	-0.060 (0.042)	0.034 (0.025)
Party ID	-0.321 (0.348)	-0.241 (0.343)	0.105 (0.200)
Constant	7.383*** (0.854)	1.552* (0.843)	5.285*** (0.490)
N	1,816	1,815	1,815
Adjusted R ²	0.276	0.258	0.293

*p < .1; **p < .05; ***p < .01. Reference categories for treatments are the middle categories in each factor (Inspections: Civilian Only, and Fuel Cycle: Freeze, respectively).

Figure 3: Conditional effects of signals by political ideology

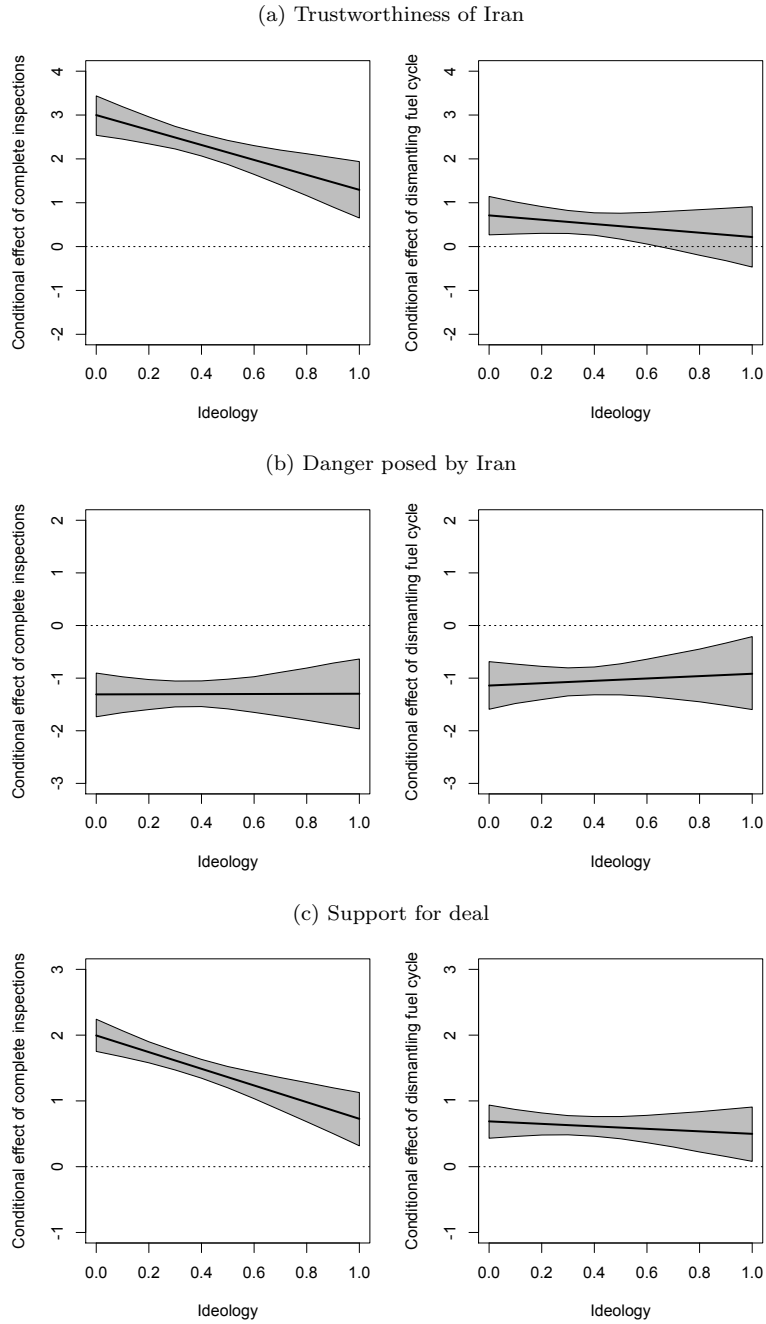


Figure 3 plots the conditional effect of each type of treatment (either complete inspections or dismantling fuel cycle) by political ideology (such that higher values correspond to more conservative ideologies), on each of the three main dependent variables (perceived trustworthiness of Iran, danger posed by Iran, and support for a deal). Liberals respond the most to the signals regarding inspections, while the interaction effects in regards to the fuel cycle treatments are weaker. For results in regression table form, see Table 5.

Table 6: Conditional effects of signals by militant internationalism

	Trustworthy (1)	Dangerous (2)	Approval of Deal (3)
Inspections: None	-2.816*** (0.368)	1.299*** (0.364)	-1.641*** (0.212)
Inspections: All	0.359 (0.363)	-0.017 (0.358)	0.107 (0.209)
Inspections: None x MI	1.528** (0.654)	-0.670 (0.646)	0.762** (0.377)
Inspections: All X MI	-0.083 (0.656)	-0.664 (0.648)	0.319 (0.378)
Fuel Cycle: Continue	0.162 (0.368)	0.675* (0.364)	-0.229 (0.212)
Fuel Cycle: Dismantle	0.202 (0.353)	0.142 (0.349)	-0.032 (0.203)
Fuel Cycle: Continue x MI	-1.138* (0.661)	0.096 (0.653)	-0.447 (0.381)
Fuel Cycle: Dismantle x MI	-0.218 (0.641)	-0.906 (0.633)	0.360 (0.369)
Cooperative internationalism	2.490*** (0.317)	-0.342 (0.313)	0.867*** (0.183)
Militant internationalism	-2.777*** (0.618)	4.655*** (0.611)	-1.392*** (0.356)
Ideology	-0.021 (0.365)	1.277*** (0.360)	-0.532** (0.210)
Isolationism	0.958*** (0.311)	-0.911*** (0.308)	0.719*** (0.179)
Male	-0.030 (0.108)	-0.591*** (0.107)	-0.046 (0.062)
log(Age)	-0.912*** (0.186)	0.633*** (0.183)	-0.417*** (0.107)
White	-0.228* (0.134)	-0.187 (0.133)	-0.032 (0.077)
Education	0.013 (0.043)	-0.057 (0.042)	0.033 (0.025)
Party ID	-0.335 (0.348)	-0.266 (0.344)	0.086 (0.201)
Constant	7.305*** (0.885)	1.560* (0.874)	5.298*** (0.510)
N	1,816	1,815	1,815
Adjusted R ²	0.275	0.256	0.286

*p < .1; **p < .05; ***p < .01. Reference categories for treatments are the middle categories in each factor (Inspections: Civilian Only, and Fuel Cycle: Freeze, respectively).

Figure 4: Conditional effects of signals by militant internationalism

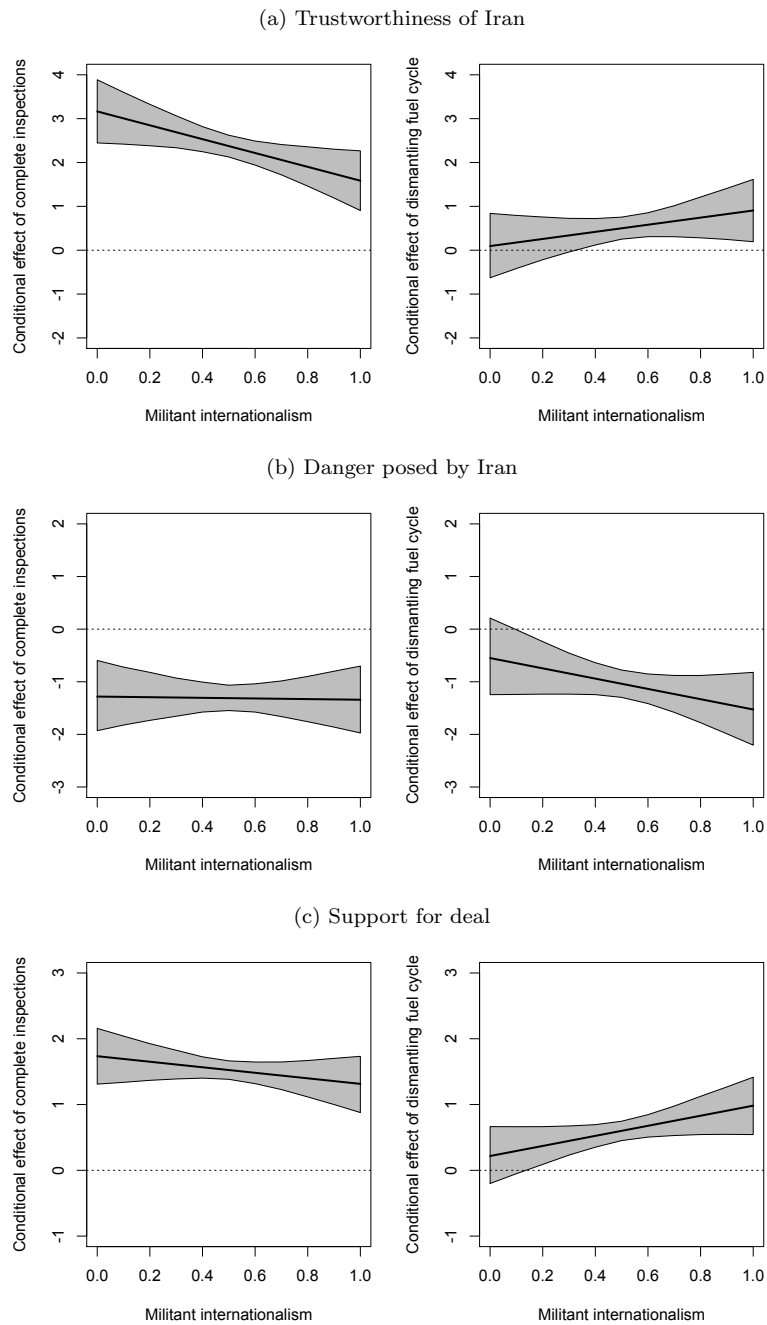


Figure 4 plots the conditional effect of each type of treatment (either complete inspections or dismantling fuel cycle) by militant internationalism, on each of the three main dependent variables (perceived trustworthiness of Iran, danger posed by Iran, and support for a deal). Doves, who are low in MI, respond the most to the signals regarding inspections, while the interaction effects in regards to the fuel cycle treatments are weaker. For results in regression table form, see Table 6.

Table 7: Trustworthiness of Iran

	(1)	(2)	(3)	(4)
Inspections: None	-1.839*** (0.142)	-2.015*** (0.131)	-2.063*** (0.123)	-2.072*** (0.123)
Inspections: All	0.433*** (0.142)	0.337** (0.132)	0.384*** (0.123)	0.375*** (0.123)
Fuel Cycle: Continue	-0.557*** (0.142)	-0.429*** (0.131)	-0.418*** (0.123)	-0.422*** (0.123)
Fuel Cycle: Dismantle	0.014 (0.141)	0.093 (0.131)	0.045 (0.122)	0.047 (0.122)
Militant internationalism		-2.733*** (0.329)	-1.320*** (0.319)	-1.354*** (0.319)
Cooperative internationalism		2.492*** (0.307)	1.885*** (0.289)	1.999*** (0.292)
Isolationism		0.928*** (0.312)	1.052*** (0.291)	1.046*** (0.291)
Feelings towards Iran			0.041*** (0.003)	0.041*** (0.003)
Need for cognition				-0.557** (0.230)
Male		-0.025 (0.108)	-0.027 (0.101)	0.007 (0.102)
log(Age)		-0.899*** (0.186)	-0.709*** (0.174)	-0.695*** (0.174)
White		-0.221 (0.134)	-0.124 (0.126)	-0.101 (0.126)
Education		0.017 (0.043)	0.036 (0.040)	0.054 (0.041)
Party ID		-0.355 (0.240)	-0.291 (0.224)	-0.282 (0.224)
Constant	4.583*** (0.129)	7.217*** (0.843)	4.804*** (0.801)	4.943*** (0.802)
N	1,816	1,816	1,816	1,816
Adjusted R ²	0.145	0.272	0.366	0.367

*p < .1; **p < .05; ***p < .01. Reference categories for treatments are the middle categories in each factor (Inspections: Civilian Only, and Fuel Cycle: Freeze, respectively). See Table 1 in the main text for equivalent results from model controlling for ideology instead of partisanship.

Table 8: Conditional effects of signals by cooperative internationalism

	Trustworthy	Dangerous	Approval of Deal
	(1)	(2)	(3)
Inspections: None	-1.206*** (0.405)	0.315 (0.422)	-0.448* (0.246)
Inspections: All	-0.083 (0.388)	-0.599 (0.404)	0.138 (0.235)
Inspections: None x CI	-1.278** (0.577)	0.976 (0.601)	-1.197*** (0.349)
Inspections: All x CI	0.666 (0.560)	0.326 (0.583)	0.212 (0.339)
Fuel Cycle: Continue	0.291 (0.405)	0.638 (0.422)	-0.057 (0.245)
Fuel Cycle: Dismantle	0.007 (0.400)	-0.048 (0.417)	-0.015 (0.243)
Fuel Cycle: Continue x CI	-1.077* (0.578)	0.134 (0.602)	-0.615* (0.350)
Fuel Cycle: Dismantle x CI	0.056 (0.567)	-0.355 (0.591)	0.221 (0.344)
Cooperative internationalism	2.492*** (0.544)	-0.581 (0.567)	1.238*** (0.330)
Militant internationalism	-1.421*** (0.319)	3.402*** (0.332)	-0.776*** (0.193)
Isolationism	1.031*** (0.290)	-0.969*** (0.302)	0.734*** (0.176)
Feelings towards Iran	0.041*** (0.003)	-0.023*** (0.003)	0.012*** (0.002)
Need for cognition	-0.547** (0.230)	-0.223 (0.239)	-0.063 (0.139)
Male	0.012 (0.102)	-0.569*** (0.106)	-0.042 (0.062)
log(Age)	-0.693*** (0.173)	0.544*** (0.180)	-0.355*** (0.105)
White	-0.111 (0.126)	-0.233* (0.131)	-0.003 (0.076)
Education	0.051 (0.041)	-0.060 (0.042)	0.040 (0.025)
Party ID	-0.273 (0.224)	0.575** (0.233)	-0.252* (0.136)
Constant	4.684*** (0.855)	3.710*** (0.891)	4.077*** (0.518)
N	1,816	1,815	1,815
Adjusted R ²	0.371	0.283	0.316

*p < .1; **p < .05; ***p < .01. Reference categories for treatments are the middle categories in each factor (Inspections: Civilian Only, and Fuel Cycle: Freeze, respectively). See Table 2 in the main text for equivalent results from model controlling for ideology instead of partisanship.

Table 9: Conditional effects of signals by partisanship

	Trustworthy	Dangerous	Approval of Deal
	(1)	(2)	(3)
Inspections: None	-2.464*** (0.238)	1.069*** (0.236)	-1.580*** (0.137)
Inspections: All	0.524** (0.236)	-0.297 (0.234)	0.440*** (0.136)
Inspections: None x Party ID	1.174** (0.513)	-0.327 (0.508)	0.899*** (0.295)
Inspections: All x Party ID	-0.471 (0.494)	-0.197 (0.488)	-0.395 (0.284)
Fuel Cycle: Continue	-0.816*** (0.237)	1.034*** (0.235)	-0.559*** (0.136)
Fuel Cycle: Dismantle	-0.137 (0.237)	0.065 (0.234)	0.018 (0.136)
Fuel Cycle: Continue x Party ID	0.962* (0.504)	-0.763 (0.499)	0.219 (0.290)
Fuel Cycle: Dismantle x Party ID	0.566 (0.503)	-0.988** (0.498)	0.333 (0.289)
Cooperative internationalism	2.499*** (0.317)	-0.348 (0.313)	0.898*** (0.182)
Militant internationalism	-2.689*** (0.337)	3.911*** (0.333)	-1.028*** (0.193)
Isolationism	0.961*** (0.311)	-0.910*** (0.308)	0.743*** (0.179)
Male	-0.014 (0.108)	-0.588*** (0.107)	-0.040 (0.062)
log(Age)	-0.895*** (0.186)	0.623*** (0.183)	-0.412*** (0.107)
White	-0.192 (0.134)	-0.189 (0.133)	-0.010 (0.077)
Education	0.017 (0.043)	-0.059 (0.042)	0.035 (0.025)
Ideology	-0.020 (0.364)	1.278*** (0.360)	-0.512** (0.209)
Party ID	-1.048* (0.535)	0.492 (0.529)	-0.247 (0.308)
Constant	7.413*** (0.854)	1.701** (0.844)	5.147*** (0.491)
N	1,816	1,815	1,815
Adjusted R ²	0.276	0.256	0.291

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

Figure 5: Conditional effects of signals by partisanship

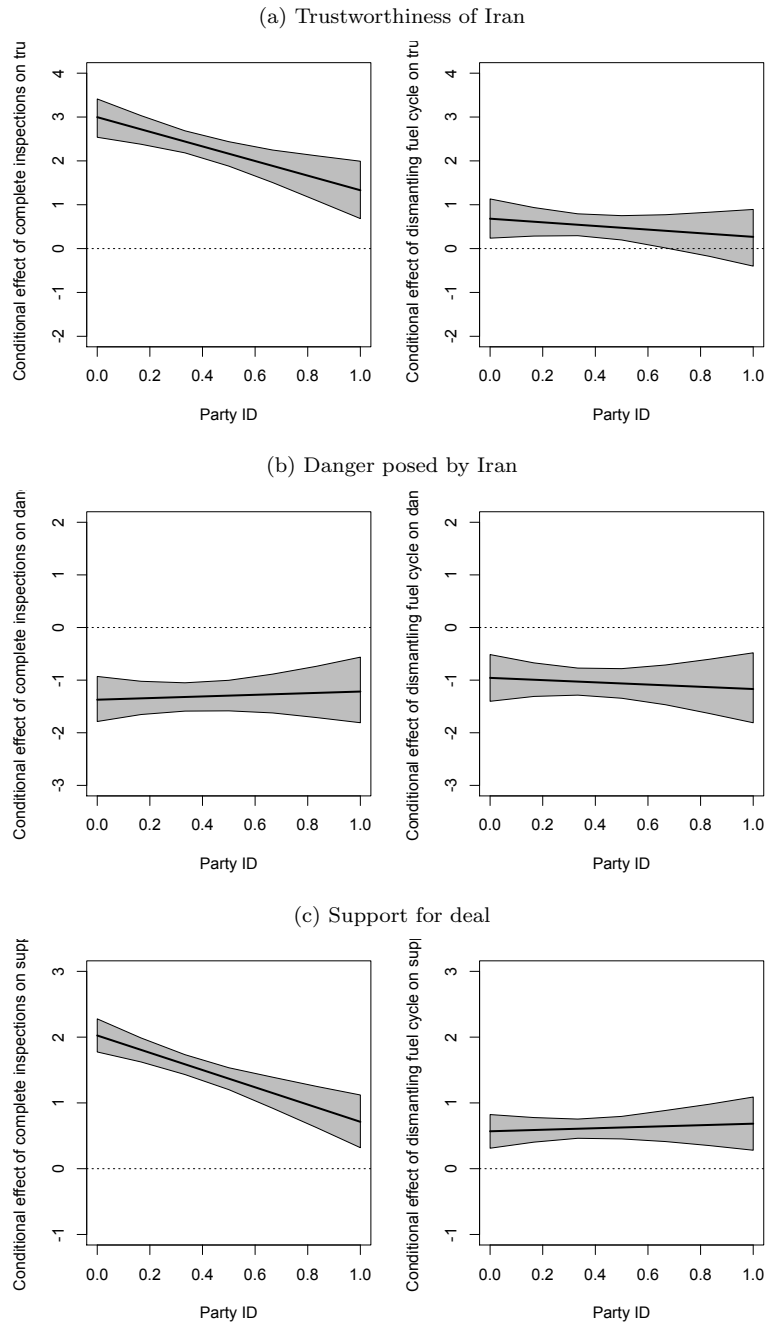


Figure 5 plots the conditional effect of each type of treatment (either complete inspections or dismantling fuel cycle) by party ID, on each of the three main dependent variables (perceived trustworthiness of Iran, danger posed by Iran, and support for a deal). The x-axis is scaled such that higher values = stronger Republicans. Democrats respond the most to the signals regarding inspections, while the interaction effects in regards to the fuel cycle treatments are weaker. For results in regression table form, see Table 9.

Figure 6: Foreign policy orientations outperform partisan identity

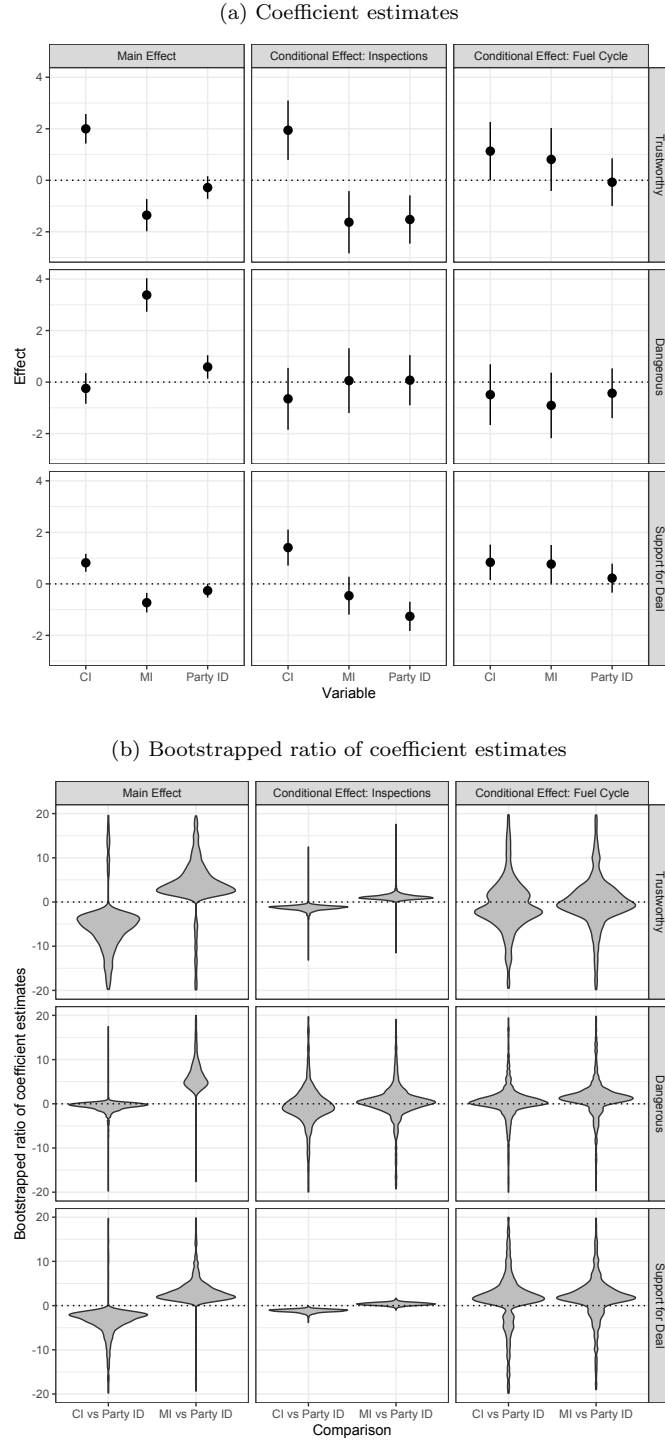


Figure 6(a) presents a coefficient plot (with 95% SEs) comparing the effects of our two central foreign policy orientations (CI, and MI) with that of party ID, for each of our three main dependent variables (trustworthiness of Iran, dangerousness of Iran, and approval of the proposed deal), both in terms of main effects (the left-most panels), and as moderators of the inspection and fuel cycle treatments (the middle and right-hand panels); for the main effects, the estimates are derived from model of Table 7 and models 4 and 8 of Table 1; for the conditional effects, the estimates are derived from the interaction terms from Tables 8, 6 and 9. The violin plots in Figure 6(b) make similar comparisons, but expressed as the bootstrapped ratio of coefficient estimates instead. Across both figures, however, it is clear that foreign policy orientations loom larger than partisanship does, both in terms of main effects, and as moderators of the inspections treatments with regard to trustworthiness and support for the deal.

if compare the effects of foreign policy orientations and partisanship across the models we estimate here, foreign policy orientations clearly trump partisanship. Figure 6(a) presents a coefficient plot comparing the effects of our two central foreign policy orientations (CI, and MI) with that of party ID, for each of our three main dependent variables (trustworthiness of Iran, dangerousness of Iran, and approval of the proposed deal), both in terms of main effects (the left-most panels), and as moderators of the inspection and fuel cycle treatments (the middle and right-hand panels); for the main effects, the coefficient estimates and accompanying 95% standard errors are derived from model of Table 7 and models 4 and 8 of Table 1; for the conditional effects, the coefficient estimates and accompanying 95% standard errors are derived from the interaction terms from Tables 8, 6 and 9. The violin plots in Figure 6(b) make similar comparisons, but expressed as the bootstrapped ratio of coefficient estimates instead. Across both figures, however, it is clear that foreign policy orientations loom larger than partisanship does in regard to our dependent variables here. The differences are the starkest for the main effects: when it comes to perceptions of trustworthiness, for example, the effect of CI is nearly 6 times larger than that of partisanship; for assessments of the danger that Iran poses, the effect of MI is nearly 12 times larger than that of partisanship; for support for the deal, the effect of CI is over 5 times larger than that of partisanship. Among the conditional effects, the preeminence of foreign policy orientations over partisanship is also evident in regards the conditional effect of inspections with respect to both trustworthiness and support for the deal, although less so with the uranium enrichment treatments, where the interaction effects are generally weaker regardless. Like other recent work critical of partisan and elite-driven theories of public opinion in foreign policy, then (e.g. [Kertzer and Zeitzoff, 2017](#); [Rathbun, Kertzer, Reifler et al., 2016](#)), we find that policy orientations, not partisan identities, are doing more work here. When we discuss polarization in the main text, we are therefore careful to emphasize ideological polarization, rather than partisan polarization.

3 Supplementary information: terms of the Iran Deal

Iran agreed to freeze and roll back enrichment capability by reducing its installed centrifuges by two-thirds, restricting its uranium enrichment to 3.67 percent (far below the 20 percent threshold for low enriched uranium (LEU)), reducing its current LEU stockpile by 98 percent to a cap of 300 kg of 3.67 percent LEU, restricting its' enrichment technology to first generation centrifuges in only one facility (Natanz) and dismantling/removing more advanced centrifuges and converting its research facility at Fordow that had been producing 20 percent LEU to no longer enrich uranium. Iran is also to ensure IAEA regular access to all of Iran's nuclear facilities (including Natanz and Fordow) using the most up-to-date monitoring technologies, continuous IAEA surveillance at uranium mills, continuous IAEA surveillance and freezing of the centrifuge rotors and bellows production and storage facilities, continuous IAEA monitoring and approval for all supply/sale/transfer to Iran of nuclear-related and dual use materials and technology, implementation suppressed by controlling for individual differences that are potentially post-treatment.

of the IAEA Additional Protocol to allow IAEA access to both declared and undeclared facilities and to investigate suspicious sites anywhere in the country, and comply with measures to allow IAEA oversight of the possible military dimensions of Iran’s nuclear program.

The JCPOA also went beyond the framework plan in a number of ways relevant to enrichment and inspection. First, it is the first time that a state has authorized the IAEA to verify nuclear weaponization work that does not involve nuclear materials. Iran commits to never engage in neutron initiator development and multi-point explosive detonation as well as to allow the IAEA to verify this (Fitzpatrick, 2015). Second, the snap-back procedure in the event of a U.S. charge that Iran is in significant non-performance of its commitments automatically restores the UN Security Council sanctions unless a new Security Council resolution is adopted that maintains the suspension of sanctions (which the U.S. could, of course, veto).

While the limits on civilian nuclear enrichment and heavy water reactors expire in 10 to 15 years, the IAEA monitoring of the production of key centrifuge parts will last for 20 years and the additional level of safeguards and inspections from the Additional Protocol and the Subsidiary Arrangements to its Safeguards Agreement (i.e. to provide designs for new nuclear facilities for IAEA approval before beginning construction) are permanent. Most experts thereby expect that breakout times will remain significantly longer than pre-JCPOA (estimated at about 2-3 months) before year 14, and that even then the move towards breakout would be significantly harder due to IAEA inspections and accumulated knowledge (Einhorn, 2015). While Iran would normally have 24 hours to respond to an IAEA inspection request, the JCPOA places a maximum for how long Iran can delay an IAEA inspection of a suspicious site to 24 days, which the U.S. Department of Energy National Laboratories proved was too short of a time to clean up even very limited quantities of uranium (Fitzpatrick, 2015, 49). While Iran can contest an IAEA request for access to undeclared sites, the JCPOA procedure for resolving disputes allows for the U.S. together with its European allies to overrule Iran, regardless of possible support from Russia and China, thereby allowing IAEA access to any location, including military ones within 24 days. In other words, objectively this is a highly costly signal of assurance by Iran.

4 Supplementary analysis: survey experiment II

4.1 Experimental design

The experimental results in the main text come from a survey experiment fielded in the spring of 2015, right before the terms of Iranian nuclear deal (JCPOA) were announced. As noted in the main text, this study employs a factorial experimental design, in which respondents were randomly assigned different potential versions of a prospective Iranian nuclear proposal, varying the extent to which Iran would be willing to limit its production and enrichment of uranium (the *fuel cycle* manipulation), and the degree to which Iran would be willing to allow monitoring of its nuclear activities (the *inspection* manipulation). Through random assignment at the between-subjects level, the experiment enables us to estimate the average effects

of different types of signals of reassurance on beliefs about Iran, while also testing which types of respondents demonstrate greater responsiveness to the signals. We find that, on average, respondents change their beliefs in response to costly signals of reassurance, but we also find crucial evidence of asymmetric updating, in that respondents predisposed to find evidence of costly signals of reassurance (respondents high in cooperative internationalism, for example) embrace corrective information that bolsters their beliefs, while others (such as respondents low in cooperative internationalism) downplay the signals altogether.

Because the experiment controls the information presented to the respondents, we interpret these findings as being consistent with motivated skepticism rather than selective attention, since respondents are in an environment where information search is minimized. However, another potential mechanism at work apart from selective attention involves the variance of respondents' priors: if some respondents have low variance priors about Iran, and others have high variance ones, we might expect the two groups to update at different rates. It therefore would be beneficial to have an alternate research design capable of adjudicating between these two claims. Additionally, we chose a between-subjects design for our 2015 experiment. As in all questions of research design, there are tradeoffs here. While a between-subjects setup has advantages in terms of leveraging the power of random assignment, and avoiding potential concerns about anchoring effects, demand effects, maturation effects, and pre-treatment effects that can occur with within-subject setups where participants' attitudes are measured at time $t = 0$ and again at time $t = 1$ after a signal has been sent, it also means that the design can only study how subgroups of individuals respond to costly signals in the aggregate, rather than measuring how much each *individual* respondent updates in response to the signal.

We thus also designed a supplementary follow-up experiment, which we fielded on $N = 1001$ respondents on Amazon Mechanical Turk in April 2018, right before President Trump announced his decision to pull out of the agreement.³ Here, we adopted an experimental design that also has a within-subject component, illustrated in Figure 7. We first assess respondents' prior beliefs about Iran's trustworthiness and dangerousness. We then present respondents with one of three different information treatments, each of which sends a different signal about Iranian trustworthiness. As we show below, we can also use this design to approximate the variance of respondents' priors, and thus adjudicate an alternative mechanism.

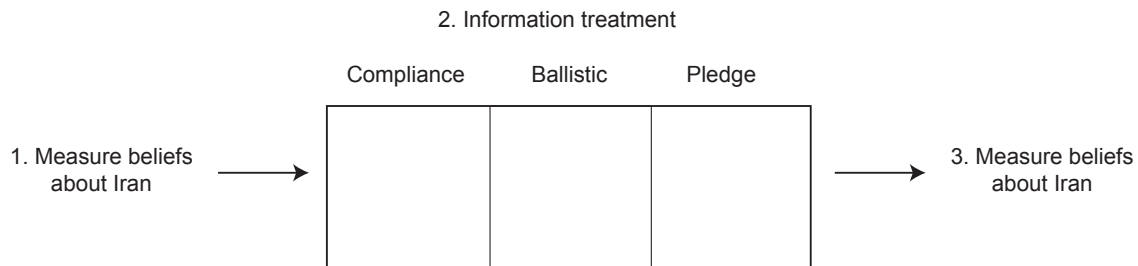
All respondents were presented with a short synopsis about the Iranian deal, focusing in particular on the two dimensions manipulated in the 2015 experiment:

In 2015 the Iran and the United States agreed on a deal in which Iran committed never to pursue nuclear weapons, promised to freeze their production of the highly enriched uranium necessary for nuclear weapons, and agreed to allow international inspections to verify that they are living up to the deal. In exchange, economic sanctions against Iran were lifted.

Respondents were then randomly received one of three potential sets of signals about Iranian trustwor-

³Following [Burleigh, Kennedy, and Clifford \(2018\)](#), we drop respondents using Virtual Private Servers (VPS) to mask their location, or whose locations we could not validate as being in the United States, leaving an effective sample size of $N = 912$.

Figure 7: Experimental design: 2018 survey



thiness. In the Compliance condition, respondents were told that the Iranian government had complied with the terms of the 2015 deal:

Iran is generally regarded to have complied with the agreement by the Department of Defense.

In the Ballistic condition, respondents were told that the Iranian government had abided by the terms of the deal, but was also behaving in an untrustworthy manner in terms not covered by the original agreement – the type of critique of the deal often raised by some of its critics in the current administration:

The Department of Defense judges that Iran has complied with the deal up to this point. However, the United States has concerns about Iran’s behavior in areas that were not part of the deal. Iran has been testing ballistic missiles. Although this is not illegal under the agreement, their testing raises questions about Iran’s long-term goals. Some are concerned that these missiles could even be used to deliver nuclear warheads in the future.

In this sense, unlike the other treatment conditions, which should reassure respondents, the ballistic condition sends a mixed signal about Iranian trustworthiness, indicating that Iran has complied with the letter of the agreement, but is behaving in threatening ways in other domains.

In the Pledge condition, respondents were presented with the same information as above about Iranian ballistic missile testing, but then were also told that Iran had pledged to halt its testing:

Imagine if Iran agreed to a ban on testing these missiles for as long as the nuclear deal lasts.

Testing is necessary to develop these missile and is detectable by the United States.

Following the information treatment, respondents were asked to provide their assessments of Iranian trustworthiness and danger in light of this information. In this manner, we obtain a measure both of respondents’ prior beliefs about Iran, and their beliefs about Iran following the presentation of information that provides signals about Iranian trustworthiness. As noted above, there are methodological challenges with this type of research design: how informative the information treatments are, for example, will depend on how attentive respondents have been to discussions about the deal in the news. Nonetheless, triangulating

the findings from this supplementary experiment with those from the experiment reported in the main text offers more complete insights into how costly signals are interpreted.

4.2 Results

Table 10: Change in perceived trustworthiness following costly signals

	(1)	(2)	(3)	(4)
Complied condition	1.576*** (0.147)		1.584*** (0.145)	1.585*** (0.145)
Pledge condition	0.920*** (0.147)		0.910*** (0.145)	0.917*** (0.146)
CI		0.928*** (0.292)	0.936*** (0.275)	0.820** (0.373)
MI		-0.255 (0.328)	-0.156 (0.309)	-0.262 (0.361)
Iran feeling		-0.834*** (0.293)	-0.828*** (0.276)	-0.835*** (0.280)
Follows news about Iran deal		-0.521*** (0.193)	-0.557*** (0.182)	-0.453** (0.194)
Isolationism				-0.313 (0.321)
Knowledge				-0.358 (0.228)
Need for closure				0.108 (0.370)
Ideology				-0.010 (0.263)
Education				-0.198 (0.231)
Male				-0.075 (0.122)
Intercept	0.153 (0.104)	0.924*** (0.291)	0.049 (0.287)	0.630 (0.591)
N	912	912	912	912
Adjusted R ²	0.112	0.018	0.131	0.130

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

We begin by simply looking at the within-subject effects. Since our quantity of interest is the extent to which respondents update in response to costly signals, we present a series of regression models in Table 10 where the dependent variable is the difference in respondents' perceptions of Iranian trustworthiness before and after receiving a positive signal about Iranian compliance. Here, we begin with a pooled analysis, where we estimate how much respondents update as a function of each treatment, as well as foreign policy orientations like cooperative internationalism, and a series of demographic controls. To facilitate easier interpretation of the coefficients, all dispositional and demographic variables are rescaled from 0 to 1. Model 1 depicts the treatment effects for the compliance and pledge conditions, using the ballistic condition as a

Table 11: Change in trustworthiness following each type of costly signal

	Compliance			Ballistic			Pledge		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	1.729*** (0.111)	2.250*** (0.471)	1.121 (1.052)	0.153* (0.088)	-0.552 (0.409)	2.638*** (0.888)	1.073*** (0.110)	0.863 (0.542)	0.777 (1.083)
CI		0.852* (0.478)	1.351** (0.667)		0.742* (0.405)	-0.574 (0.534)		1.193** (0.543)	1.585** (0.752)
MI		-0.469 (0.538)	-0.603 (0.638)		0.490 (0.457)	0.196 (0.515)		-0.263 (0.609)	-0.067 (0.726)
Iran feeling		-1.233** (0.482)	-1.068** (0.489)		-0.090 (0.399)	-0.349 (0.401)		-1.251** (0.556)	-1.295** (0.564)
Follows news about Iran deal		-1.314*** (0.326)	-1.134*** (0.347)		-0.071 (0.295)	-0.089 (0.310)		-0.234 (0.319)	-0.011 (0.346)
Isolationism			0.105 (0.597)			-1.439*** (0.465)			0.426 (0.596)
Knowledge			0.013 (0.408)			-0.581 (0.360)			-0.474 (0.407)
Need for closure			1.449** (0.679)			-1.289** (0.562)			-0.017 (0.673)
Ideology			0.396 (0.492)			-0.502 (0.371)			0.064 (0.500)
Education			-0.069 (0.418)			-0.395 (0.354)			-0.349 (0.422)
Male			-0.124 (0.223)			0.165 (0.178)			-0.196 (0.231)
N	303	303	303	305	305	305	304	304	304
Adjusted R ²	0.000	0.063	0.064	0.000	0.001	0.040	0.000	0.013	0.006

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

baseline. They show that both treatments significantly increase perceptions of Iranian trustworthiness, with the complied condition displaying a stronger effect than the pledge condition.

More important for our purposes, though, are the results of models 3 and 4. Model 3 adds a set of foreign policy orientations (CI, MI, and feelings about Iran) as well as a variable indicating the extent to which respondents reported following the news about the Iran deal. This news measure is important to include because whereas the 2015 experiment was fielded before the actual terms of the Iranian deal had been announced (minimizing the likelihood of pre-treatment effects), the 2018 experiment was fielded after the deal had been brokered, and a variety of debates had broken out in the news media about Iranian compliance. Controlling for attention to the deal in the news thus helps address concerns about pre-treatment. Importantly, the coefficient for CI is positive and significant, and equivalent in magnitude to the pledge condition: individuals who are higher in CI update more, on average, than individuals who are lower in CI. This is consistent with the motivated skepticism results from the previous experiment: it is those respondents who are predisposed to find evidence of reassurance who are ultimately more reassured. A similar pattern holds in model 3, which adds a variety of demographic control variables, as individuals who are high in CI update more than individuals who are lower in CI. Interestingly, the news attentiveness variable has significant negative effects across both models: the more respondents indicate they follow news about the Iran deal, the less they update. We interpret this pattern as evidence of pre-treatment effects, in which pre-treated respondents are less likely to change their beliefs in response to the treatments because it accords with information they already know. It is also possible to interpret the result as due to motivated skepticism of a different form (if respondents who report following news about the Iran deal more closely are more concerned about the deal, and thus more skeptical of signals of reassurance), but we consider this less likely, in that the news attentiveness measure is only weakly correlated with partisanship, and in the opposite direction ($r = -0.03$).

Table 11 presents additional analyses, where we disaggregate each of the three treatments to estimate the effect of CI on receptivity to each. Here, the dependent variable now represents the within-subject treatment effect for a given condition. Model 1 depicts the treatment effect for the compliance condition, model 4 for the ballistic condition, and model 7 for the pledge condition; because the dependent variable in each model is the first-difference within a given treatment, the average within-subject treatment effect in each of the experimental conditions is depicted by the intercept. Thus, model 1 shows that telling respondents that Iran has complied with the terms of the JCPOA significantly increases respondents' assessments of Iranian trustworthiness, with respondents updating an average of 1.73 points on a 10 point scale. In contrast, model 4 shows that the effect of the ballistic condition is relatively weak, with respondents updating an average of only 0.15 points. There are two interpretations for this modest effect. The first is that the treatment doesn't provide as much information that respondents weren't already aware of; the second is that it provides countervailing information, both telling respondents that Iran is in compliance with the deal (bolstering trustworthiness), and that Iran is testing ballistic missiles (undermining trustworthiness), such that the two

largely cancel one another out. Model 7 shows that the pledge condition has a stronger treatment effect: on average, respondents told of an Iranian pledge to halt testing updated assessments of Iranian trustworthiness by an average of 1.07 points on a 10 point scale.

In the aggregate, then, we find that the signals conveyed in the treatments are effective, just as they were in the previous experiment, and in a manner consistent with what our theoretical models of signals in IR would expect: respondents perceive Iran as more trustworthy when told of its compliance with the JCPOA, or of a pledge to halt missile testing. For our purposes, however, the important question concerns variation in how much each respondent updates in response to the signals. Models 2-3 regress the within-subject treatment effect in the compliance condition on our core dispositional variables from the 2015 study, along with a set of demographic controls; models 5-6 does the same in the ballistic condition, and models 8-9 in the pledge condition. As before, because of concerns about pre-treatment effects, our list of covariates includes a control variable for the extent to which respondents report closely following news about the Iran deal.

Across all three treatment conditions, we find evidence that the amount respondents update in response to costly signals of reassurance is partially a function of their levels of Cooperative Internationalism (CI), consistent with the findings from the previous experiment. Models 2 and 3 suggest that high CI respondents update from the compliance condition to a greater extent than their low-CI counterparts. Model 5 suggests that Cooperative Internationalism is associated with changes in the magnitude at which respondents update to the ballistic condition, although the evidence here is weaker, perhaps reflecting the small size of the ballistic treatment effect, and is the effect of CI is no longer significant once a broader set of demographic controls are included in Model 6. Models 8 and 9 suggest that high CI respondents update from the pledge condition at a larger rate as well.

In general, then, despite its divergent experimental design, we find evidence in the 2018 experiment consistent with the main findings of its predecessor. First, in the aggregate, the signals conveyed in the treatments are effective in bolstering perceptions of Iranian trustworthiness, operating in the manner that our theories of signaling in IR would expect. At the same time, however, there is considerable heterogeneity among our respondents, and some evidence that high CI respondents are more likely to update in response to the signals, particularly the clearest signal of reassurance offered by the compliance condition.

4.2.1 Testing for variance of priors

As noted above, because of the setup of the research design, our interpretation of the above results emphasizes motivated skepticism (how individuals evaluate information), rather than selective attention (how individuals seek out information). However, an alternate mechanism somewhat distinct from our motivated skepticism account involves the variance of respondents' priors, in which respondents with higher variance priors respond more to the treatments. The nature of the 2018 study allows us to test this alternative mechanism.

Prior to participating in the experiment, respondents completed a thought listing soliciting what thoughts come to mind when they think about the Iran deal. Based on a free recall task used by [Berinsky](#)

and Kinder (2006), each respondent could provide up to seven thoughts each. However, whereas Berinsky and Kinder use a post-treatment thought listing task to study what respondents had learned from treatments provided by the experiment, our pre-treatment thought listing task was included to lend the study some *verstehen* and understand what people think of when they think of the Iranian nuclear deal — which was particularly pertinent in our case because of concerns about pretreatment effects, given that we were presenting respondents with information about Iranian behavior in a period in which many respondents had already been exposed to such information.

A rich literature in American political behavior uses similar thought listing techniques to study the attitudinal consequences of ambivalence on voting behavior (Alvarez and Brehm, 1995; Lavine, 2001; McGraw, Hasecke, and Conger, 2003; Keele and Wolak, 2008). Individuals who, when they think of political issues like abortion, tend to think of both positive and negative considerations, tend to display very different political behaviors than those who don't. Our interest here is in using automated content analysis of our thought listing about the Iran deal to similar effect, as a rough test of how the variance about individuals' priors about the Iran deal affect how they respond to costly signals.

To that end, our empirical strategy has six parts. First, we tokenize the thought listings using the `tidytext` package in R. We then run a sentiment analysis — a supervised learning method that codes the tone or valence of each response as positive or negative using a reference dictionary (Hopkins and King, 2010; Young and Soroka, 2012); to avoid “placing our thumbs on the scale” we use pre-defined sentiment dictionaries rather than generating our own.⁴ To the use of these sentiment scores as measures of respondents' prior attitudes, we then calculate each respondents' total sentiment score (the sum of the sentiment scores of each thought listing), and examine its correlation with a variety of dispositional measures. The correlations are modest, but in the right direction: CI, for example, is correlated with the summed sentiment score at $r = 0.19$ or 0.22 , depending on the dictionary, meaning that individuals high in CI express more positive sentiment about the Iranian nuclear deal than individuals low in CI. We then calculate the variance of the sentiment scores across each respondent's thought listings, to treat it as a proxy of the variance of respondents' priors.

Importantly, there is no significant correlation between CI and the variance of respondents' priors ($r = 0.05$); thus, we should not attribute the increased updating of our high-CI respondents to them possessing more diffuse priors. Nonetheless, it is possible that the variance of respondents' priors affects how much respondents update more generally. Thus, Table 12 replicates Table 10, but including the variance of respondents' priors as an additional covariate in each model; Table 13 does the same for Table 11. A direct comparison between the two sets of results is complicated by missing data (since variance estimates cannot be calculated for respondents who provided only a single free response item, or who provided multiple free responses but did not include words included in the sentiment dictionaries), but across every model in Tables 12-13, the variance of priors covariate lacks statistical significance. In so far as our study is equipped to test

⁴For the analyses reported below, we use Bing as our sentiment dictionary; we obtain similar results when we use AFINN.

Table 12: Variance of priors (I)

	(1)	(2)	(3)
Complied condition	1.502*** (0.175)	1.510*** (0.173)	1.503*** (0.174)
Pledge condition	1.022*** (0.175)	1.005*** (0.172)	1.011*** (0.174)
CI		1.199*** (0.330)	1.097** (0.460)
MI		0.204 (0.363)	0.295 (0.427)
Iran feeling		-0.611* (0.333)	-0.634* (0.338)
Follows news about Iran deal		-0.686*** (0.215)	-0.600*** (0.230)
Variance of priors	0.014 (0.053)	0.011 (0.052)	0.017 (0.053)
Isolationism			-0.020 (0.383)
Knowledge			-0.314 (0.269)
Need for closure			-0.131 (0.447)
Ideology			-0.160 (0.326)
Education			-0.269 (0.284)
Male			-0.059 (0.145)
Intercept	0.168 (0.139)	-0.314 (0.356)	0.191 (0.723)
N	652	652	652
R ²	0.106	0.138	0.141

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

Table 13: Variance of priors (II)

	Compliance			Ballistic			Pledge		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	1.846*** (0.186)	1.892*** (0.600)	0.553 (1.253)	0.126 (0.134)	-1.246** (0.497)	0.727 (1.024)	1.099*** (0.181)	0.767 (0.676)	1.107 (1.427)
CI		0.916 (0.579)	1.371* (0.819)		1.082** (0.456)	0.287 (0.616)		1.761*** (0.674)	2.264** (0.999)
MI		0.049 (0.642)	0.327 (0.772)		1.011** (0.511)	0.954 (0.590)		-0.350 (0.728)	0.069 (0.857)
Iran feeling		-0.735 (0.585)	-0.590 (0.594)		0.340 (0.453)	0.058 (0.463)		-1.715** (0.693)	-1.687** (0.696)
News about Iran		-1.262*** (0.381)	-1.216*** (0.409)		-0.074 (0.340)	-0.180 (0.362)		-0.537 (0.393)	-0.229 (0.423)
Variance of priors	-0.127 (0.106)	-0.106 (0.105)	-0.107 (0.106)	0.048 (0.073)	0.065 (0.073)	0.059 (0.074)	0.087 (0.095)	0.095 (0.093)	0.121 (0.094)
Isolationism			0.288 (0.692)			-0.829 (0.526)			0.471 (0.762)
Knowledge			0.181 (0.474)			0.013 (0.407)			-1.034** (0.499)
Need for closure			1.825** (0.835)			-1.708*** (0.644)			-0.298 (0.832)
Ideology			-0.044 (0.616)			-0.277 (0.442)			-0.027 (0.627)
Education			-0.325 (0.508)			0.007 (0.432)			-0.717 (0.538)
Male			0.090 (0.268)			0.048 (0.199)			-0.195 (0.286)
N	216	216	216	219	219	219	217	217	217
R ²	0.007	0.069	0.094	0.002	0.046	0.088	0.004	0.055	0.090

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

the variance of priors mechanism, we therefore fail to find evidence supportive of it.

5 Motivated reasoning and costless signals

In our research design itself, we steered clear of the question of costless signals, since, following [Jervis \(1976, 2002\)](#), we didn't want to be in a position of needing to adjudicate the objective costliness of signals when setting up the experiments (it is for this same reason that our analysis in the results focuses on exploring the variation in how much respondents updated, rather than determining how much respondents *should* have updated). Nonetheless, if our chief finding is that motivated skepticism means that signals work the most among the recipients that need them the least, this suggests another potential explanation for the surprising efficacy of cheap talk or costless signals in recent work ([Trager, 2010](#); [Tingley and Walter, 2011](#); [Yarhi-Milo, 2014](#); [Holmes, 2018](#)). If individuals see what they want to see when interpreting signals, then signals need not necessarily be costly in order to exert effects.⁵ In this sense, it is not just that motivated reasoning makes signaling harder in some circumstances than canonical models presume, but that it also makes signaling easier in other circumstances, depending on the recipient. This of course makes it all the more important for signalers to understand the psychological dynamics of their target. Future work should consider investigating these questions further.

⁵Note that motivated reasoning therefore offers a very different way of understanding costless signals than that usually offered by this literature, which tends to explain cheap talk's efficacy by finding ways to render it costly (e.g. [Guisinger and Smith, 2002](#); [Sartori, 2005](#)).

References

- Aarøe, Lene, Michael Bang Petersen, and Kevin Arceneaux. 2017. The Behavioral Immune System Shapes Political Intuitions: Why and How Individual Differences in Disgust Sensitivity Underlie Opposition to Immigration. *American Political Science Review* 111 (2):277–294.
- Alvarez, R. Michael, and John Brehm. 1995. American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values. *American Journal of Political Science* 39 (4):1055–1082.
- Bartels, Larry M. 2002. Beyond the Running Tally: Partisan Bias in Political Perceptions. *Political Behavior* 24 (2):117–150.
- Berinsky, Adam J, Gregory A Huber, and Gabriel S Lenz. 2012. Evaluating online labor markets for experimental research. *Political Analysis* 20 (3):351–368.
- Berinsky, Adam J., and Donald R. Kinder. 2006. Making Sense of Issues Through Media Frames: Understanding the Kosovo Crisis. *Journal of Politics* 68 (3):640–656.
- Bisgaard, Martin. 2019. How Getting the Facts Right Can Fuel Partisan-Motivated Reasoning. *American Journal of Political Science* Forthcoming.
- Bishin, Benjamin G., Thomas J. Hayes, Matthew B. Incantalupo, and Charles Anthony Smith. 2016. Opinion Backlash and Public Attitudes: Are Political Advances in Gay Rights Counterproductive? *American Journal of Political Science* 60 (3):625–648.
- Braumoeller, Bear F. 2004. Hypothesis Testing and Multiplicative Interaction Terms. *International Organization* 58 (4):807–820.
- Brutger, Ryan, and Joshua D. Kertzer. 2018. A Dispositional Theory of Reputation Costs. *International Organization* 72 (3):693–724.
- Bullock, John G. 2007. Experiments on partisanship and public opinion: Party cues, false beliefs, and Bayesian updating. Ph.D. thesis, Stanford University, Palo Alto, CA.
- . 2009. Partisan Bias and the Bayesian Ideal in the Study of Public Opinion. *Journal of Politics* 71 (3):1109–1124.
- Burleigh, Tyler, Ryan Kennedy, and Scott Clifford. 2018. How to Screen out VPS and International Respondents Using Qualtrics: A Protocol. Working paper.
- Cacioppo, John T., and Richard E. Petty. 1982. The Need for Cognition. *Journal of Personality and Social Psychology* 42 (1):116.
- Chaudoin, Stephen. 2014. Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions. *International Organization* 68 (1):235–256.
- Druckman, James N., and Thomas J. Leeper. 2012. Learning More from Political Communication Experiments: Pretreatment and Its Effects. *American Journal of Political Science* 56 (4):875–896.
- Einhorn, Robert. 2015. Debating the Iran nuclear deal: A former American negotiator outlines the battleground issues. Brookings Report.
- Fitzpatrick, Mark. 2015. Iran: A good deal. *Survival* 57 (5):47–52.
- Gerber, Alan, and Donald P. Green. 1999. Misperceptions about Perceptual Bias. *Annual Review of Political Science* 2:189–210.
- Guess, Andrew, and Alexander Coppock. Forthcoming. Does Counter-Attitudinal Information Cause Backlash? Results from Three Large Survey Experiments. *British Journal of Political Science* .
- Guisinger, Alexandra, and Alastair Smith. 2002. Honest Threats: The Interaction of Reputation and Political Institutions in International Crises. *Journal of Conflict Resolution* 46 (2):175–200.
- Holmes, Marcus. 2018. *Face-to-Face Diplomacy: Social Neuroscience and International Relations*. Cambridge, UK: Cambridge University Press.
- Hopkins, Daniel J., and Gary King. 2010. A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science* 54 (1):229–247.
- Huff, Connor, and Joshua D. Kertzer. 2018. How The Public Defines Terrorism. *American Journal of Political Science* 62 (1):55–71.
- Iyengar, Shanto, and Kyu S Hahn. 2009. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication* 59 (1):19–39.
- Iyengar, Shanto, and Sean J Westwood. 2015. Fear and Loathing Across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science* 59 (3):690–707.

- Jervis, Robert. 1976. *Perception and misperception in international politics*. Princeton University Press.
- . 2002. Signaling and Perception: Drawing Inferences and Projecting Images. In *Political Psychology*, edited by K.R. Monroe, 304–405. Hillsdale, NJ: Erlbaum.
- . 2006. Reports, politics, and intelligence failures: The case of Iraq. *Journal of Strategic Studies* 29 (1):3 – 52.
- Kahan, Dan M. 2016. The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It. In *Emerging Trends in the Social and Behavioral Sciences*. Wiley Online Library.
- Keele, Luke, and Jennifer Wolak. 2008. Contextual Sources of Ambivalence. *Political Psychology* 29 (5):653–673.
- Kertzer, Joshua D., and Thomas Zeitzoff. 2017. A Bottom-Up Theory of Public Opinion about Foreign Policy. *American Journal of Political Science* 61 (3):543–558.
- Lavine, Howard. 2001. The Electoral Consequences of Ambivalence Toward Presidential Candidates. *American Journal of Political Science* 45 (4):915–929.
- Lodge, Milton, and Charles Taber. 2000. Three Steps toward a Theory of Motivated Political Reasoning. In *Elements of Reason: Cognition, Choice, and the Bounds of Rationality*, edited by Arthur Lupia, Mathew D. McCubbins, and Samuel L. Popkin, 183–213. Cambridge: Cambridge University Press.
- Mason, Lilliana. 2018. *Uncivil Agreement: How Politics Became Our Identity*. Chicago, IL: University of Chicago Press.
- McEntire, Kyla Jo, Michele Leiby, and Matthew Krain. 2015. Human Rights Organizations as Agents of Change: An Experimental Examination of Framing and Micromobilization. *American Political Science Review* 109 (3):407–426.
- McGraw, Kathleen M., Edward Hasecke, and Kimberley Conger. 2003. Ambivalence, Uncertainty, and Processes of Candidate Evaluation. *Political Psychology* 24 (3):421–448.
- McGraw, Kathleen M., and Neil Pinney. 1990. The effects of general and domain-specific expertise on political memory and judgment. *Social Cognition* 8 (1):9–30.
- Pennycook, Gordon, and David G. Rand. 2018. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* Forthcoming.
- Rathbun, Brian C, Joshua D. Kertzer, Jason Reifler, Paul Goren, and Thomas J Scotto. 2016. Taking foreign policy personally: Personal values and foreign policy attitudes. *International Studies Quarterly* 60 (1):124–137.
- Sartori, Anne E. 2005. *Deterrence by Diplomacy*. Princeton, NJ: Princeton University Press.
- Taber, Charles S, and Milton Lodge. 2006. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* 50 (3):755–769.
- Tingley, Dustin H., and Barbara F. Walter. 2011. Can Cheap Talk Deter? An Experimental Analysis. *Journal of Conflict Resolution* 55 (6):996–1020.
- Trager, Robert F. 2010. Diplomatic Calculus in Anarchy: How Communications Matter. *American Political Science Review* 104 (2):347–368.
- Wallace, Geoffrey PR. 2013. International law and public attitudes toward torture: An experimental study. *International Organization* 67 (01):105–140.
- Wood, Thomas, and Ethan Porter. 2019. The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence. *Political Behavior* 41 (1):135–163.
- Yarhi-Milo, Keren. 2014. *Knowing the adversary: leaders, intelligence, and assessment of intentions in international relations*. Princeton University Press.
- Young, Lori, and Stuart Soroka. 2012. Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication* 29 (2):205–231.