Shooting From the Hip: The Counterintuitively Intuitive Nature of Deterrence

Brian C Rathbun
University of Toronto
brian.rathbun@utoronto.ca

Caleb Pomeroy
University of Toronto
caleb.pomeroy@utoronto.ca

Joshua D Kertzer
Harvard University
jkertzer@gov.harvard.edu

August 21, 2025

Abstract

Research in international relations typically associates deterrence with rational choice and deliberative thought. Drawing on insights from cognitive and evolutionary psychology, we expect the opposite: intuition and gut feelings underpin deterrence. By thinking from the gut and shooting from the hip, humans solved deterrence challenges in our evolutionary past. By contrast, the careful, analytical thought implied by rationalist theory begets support for diplomacy and cooperation. An innovative time pressure experiment, an observational survey of thinking styles, and a large-scale text analysis of Cold War US foreign policy elites lend support to these expectations. The findings help to explain the basic human tendency to err on the side of threat overreaction, as well as the impulse to solve challenges through deterrence rather than the empathy required to solve security dilemmas. Beyond theory, our experimental and text analytic methods pave new avenues for the study of intuition in international relations.

1   Introduction

Most of human judgment, it is now generally understood, is not deliberative, self-conscious, and reflective but rather intuitive, automatic, and reflexive (Kahneman, 2011; Gigerenzer, 2007; Damasio, 1999).[1] However, knowing that intuition plays a large role in behavior, even in foreign affairs, is different from knowing which choices we tend to make when we act quickly, intuitively, and spontaneously as opposed to deliberately and reflectively, what we could call rationally. Decisions may be made according to gut feelings or hunches, but is there something generalizable about those judgments?

Drawing on insights from cognitive and evolutionary psychology, we argue that intuition generates deterrence tendencies in international politics, namely retaliatory aggression (McDermott, Lopez and Hatemi, 2017) and a belief in the importance of standing firm to project resolve (Kertzer et al., 2014). By contrast, deliberative thinking makes individuals more likely to consider cooperation, de-escalation, and negotiation to achieve state aims. In international relations research, deterrence is traditionally the remit of rational choice theories, in which strategic calculation is necessary to effectively signal resolve. We argue instead that deterrence is typically the product of gut and feel. Deterrence is counterintuitively intuitive.

Deterrence tendencies are a manifestation of hawkishness, the focus of a burgeoning research tradition in decision-making and international relations (Kahneman and Renshon, 2007; Kertzer et al., 2022). This research finds that systematic cognitive biases tend to activate hawkish tendencies in foreign policy, namely "a propensity for suspicion, hostility, and aggression in

---

[1]For applications in international relations, see Holmes and Traven (2015); Yarhi-Milo (2014); Rathbun, Kertzer and Paradis (2017); Porter (2018); Lee (2020).

the conduct of conflict" (Renshon and Kahneman, 2017, 51). One of the most basic features of hawkishness is the endorsement of what Jervis (1976) calls the "deterrence model": one must stand firm lest would-be aggressors infer a lack of resolve. Studies indicate that hawkishness pervades group decision-making at both the mass and elite levels (Kertzer et al., 2022; Jost et al., 2024).

Intuition drives hawkishness for good evolutionary reasons. Hawkishness is *ecologically* rational (Todd and Gigerenzer, 2012; Gigerenzer, 2007), a powerful, evolutionarily-informed version of rationality typically overlooked in IR theory. What makes a behavior ecologically rational is its ability to generate better outcomes than other strategies in any particular environmental niche. Many claim that automatic and escalatory responses to threats—what is known as retaliatory aggression (McDermott, Lopez and Hatemi, 2017)—served as credible signals of resolve in our ancestral past. The more that individuals seem to be thinking about their response, the more adversaries will believe they are contemplating the costs and benefits of escalation. It is precisely by signaling that we are *not* deliberating—that we are hot-headed and trigger-happy—that we are more likely to force others to back down (Frank, 1988; McDermott, Lopez and Hatemi, 2017). For deterrence, intuition rather than rational deliberation might have been more ecologically rational over the course of human evolution. Rather than deviations from rationality, our hawkish biases aid survival (Johnson, 2020). Retaliatory aggression serves a deterrent purpose.

The use of our reasoning faculties, by contrast, likely evolved to resolve conflicts through dialogue and the exchange of arguments and justifications (Mercier and Sperber, 2017). Humans are a social species, and the ability to coordinate and cooperate contributed to human success. An individual or group can deter without any exchange of reasons at all; a smack across the

face in response to a perceived slight might serve this purpose. Many species do this. Reaching mutually beneficial agreements, by contrast, requires the ability to communicate and persuade others rather than simply coerce them. In short, reasoning is the precondition for diplomacy.

To assess the expectation that intuition leads to hawkishness in general, and deterrence tendencies in particular, we present three forms of evidence. Our first study overcomes a key methodological challenge: by its very nature, intuition is difficult to identify and observe. We use an innovative time pressure method, pioneered in influential studies by Rand and colleagues (Rand, Greene and Nowak, 2012; Rand et al., 2014), to experimentally induce intuitive versus deliberative responses. Simply put, people reveal their gut feelings when they are not given time to contemplate. In a survey experiment of US-based adults, we find that forcing a quick decision leads to a more forceful and aggressive response. By contrast, when we compel subjects to think carefully and deliberately—more aligned with the necessary cognitive preconditions for rational deterrence—subjects actually become more cooperative and diplomatic.

Second, we find the same pattern of results in a correlational survey fielded on a second sample in the US. We use well-established scales to measure reliance on intuition as well as beliefs in the deterrence model, finding that the former associates with agreement that demonstrating resolve is important in foreign affairs and that foreign policy losses tend to fall like dominoes. By contrast, individual-level measures of rational thinking display no association with deterrence beliefs. In addition to showing that our experimental results are robust to alternative measures, this survey allows us to connect intuition with explicit beliefs about deterrence. It is not merely that provocation leads to retaliation. Rather, those who rely on intuition consider retaliation necessary to demonstrate resolve.

3

Finally, we conduct a word embedding analysis of the universe of Cold War documents collected in the *Foreign Relations of the United States* (FRUS). Recent work shows that word embeddings, which quantify associations between words, provide the ability to capture relatively intuitive thinking patterns in textual data (Caliskan, Bryson and Narayanan, 2017). We find that terms indicative of reliance on intuition associate more positively with terms indicative of deterrence thinking than deliberative terms. By contrast, deliberative terms associate more positively with cooperative thinking than do intuition terms. This foreign policy decision-making takes place behind closed doors and in an environment that should promote careful calculation. The overlap in findings across these public- and elite-level studies suggests that this phenomenon is a relatively widespread human tendency, as evolutionary psychologists would expect.

Together, these studies offer consistent support for the expectation that our intuitive reaction to challenges and threats is typically escalatory and aggressive.[2] After all, Wild West tales reveal that gunslingers are also deeply invested in their reputation. The fact that concerns about credibility and reputation are intuitive explains their ubiquity and perhaps even unthinking institutionalization in policy, something that IR scholars who doubt the importance of credibility for deterrence have long criticized (Mercer, 1997; Press, 2005). Further, the fact that humans are psychologically primed to intuitively deter likely helps to explain why individuals are more prone to overestimate, rather than underestimate, threats (Stein, 1988, 2013; Kahneman and Renshon, 2007), as well as why deterrence is far easier than the cognitively taxing empathy required to solve security dilemmas (Kertzer, Brutger and Quek, 2024).

---

[2]See also Powers and Altman (2023) for a similar argument in the context of coercion.

2   Rational deterrence theory: I think, therefore I deter?

The IR literature on rational deterrence theory is enormous. However, as many scholars note, a common core unites this tradition. We follow Jervis's (1976, 58) line: "For our purposes we need not be concerned with the many subtleties and complexities of deterrence theory, but only with the central argument that great dangers arise if an aggressor believes that the status quo powers are weak in capability or resolve. This belief will lead the former to test its opponents, usually starting with a small and apparently unimportant issue. If the status quo powers retreat, they will not only lose the specific value at stake but, more important in the long run, will encourage the aggressor to press harder." Other scholars offer nearly identical conceptions of deterrence (e.g., Tetlock, 1983; Weisiger and Yarhi-Milo, 2015, 475).

This concept places a premium on the credible conveyance of one's willingness to pay the costs of conflict to potential adversaries. States must invest in reputations for resolve to deflect adversaries' challenges on a given issue, as well as third parties drawing inferences from those interactions. What Schelling (1966) famously calls the "interdependence of commitments" is now "orthodox reputation theory" (Weisiger and Yarhi-Milo, 2015, 475). This escalates the stakes for even minor conflicts. As Jervis (1976, 56) explains, "To avoid this disastrous situation, the state must display the ability and willingness to wage war. It may not be able to ignore minor conflicts or to judge disputes on their merits. Issues of little intrinsic value become highly significant as indices of resolve." Since conflict is costly, even for powerful states, mere declarations of resolve are cheap talk (Jervis, 1989; Fearon, 1994). To credibly convey resolve, signals must be costly. Following through on such threats is one way to invest in reputation.

In the discussion below, we utilize this well-established understanding of rational deterrence rather than narrower variants, since there is still considerable debate within the rationalist literature about what constitutes "perfect deterrence" (Zagare and Kilgour, 2000).

Although rational choice models often make no explicit assumptions about cognitive processes (e.g. Lake and Powell, 1999; Zagare, 1990, 7), there are strong reasons to believe that instrumentally rational behavior in the real world requires deliberative thinking. Decision-makers need to consider the motivations of others and how they will assess their actions, including the long-term consequences of their actions. Attention to a long shadow of the future (Keohane, 1984; Oye, 1985) and consideration of second-order beliefs (known in the formal theoretic literature as "k-level reasoning") are both hallmarks of strategic decision-making (Arad and Rubinstein, 2012). Bueno de Mesquita (2013, 28), a prominent rational choice scholar, describes the thought process necessary for such encounters: "[E]ach decision maker and each individual or group looks ahead, contemplating what the likely responses are if they choose this action or that action. Then they choose the action that they believe, based on looking ahead and working back to the current situation, will give them the best result." Given the Cartesian nature of deterrence, it seems that deterrence thinking must emerge out of a deliberative thought process as decision-makers contemplate how to confront threats and aggression. In fact, empirical evidence suggests that a commitment to deliberative thought underpins behavior consistent with rational modeling (Rathbun, Kertzer and Paradis, 2017). As much as rationalist work deliberately tries to avoid the phenomenology of decision-making, it often relies on a set of cognitive preconditions that are aspirational and normative rather than empirically common.

Deterrence thinking is an expression of hawkishness more generally. As Rathbun et al.

(2016, 125-126) explain, "Hawks embrace the 'deterrence model,' in which strength and the demonstration of resolve best achieve peace…Doves, in contrast, point out the often self-defeating nature of such displays, which risk inciting fear on the other side and escalating hostilities in such a way that leaves both sides worse off." Hawkishness is strongly ingrained in human nature, even if perhaps not to the degree that we find in humankind's closest biological cousins (Wrangham, 2018). Notably, Kahneman and Renshon (2007) construct a list of biases uncovered across four decades of psychological research—such as loss aversion, the endowment effect, reactive devaluation, and overconfidence—finding that each bias pushes individuals towards hawkishness when applied to conflict situations. Moreover, hawkish biases do not disappear in group settings and might even worsen (Kertzer et al., 2022). And, they are evident at the highest levels of government. In an analysis of all available meeting records of the US National Security Council, as well as 2,088 informal meetings in which presidents discussed foreign policy issues with advisers, Jost et al. (2024) show that hawks tend to prevail in American national security decisions.

Systematic biases tend to exhibit a lack of self-reflective and cognitively costly deliberation. They are the product of intuition. However, the connection between hawkish bias and intuition has not been made in the international relations literature even though Kahneman himself popularized both the notion that "hawks win" (Kahneman and Renshon, 2007) and also that much of human cognition is automatic and intuitive (Kahneman, 2011). We draw this missing link. Deterrence "thinking" is more often a function of the gut than head.

3 "We do not think, therefore we deter": Retaliatory aggression as an adaptive intuition

Psychologists now frequently distinguish between two different types of thinking: System 1 and System 2 (Stanovich, 2011; Evans, 2008; Kahneman, 2011; Epstein et al., 1996), a distinction that has received growing attention in behavioral IR research (Hafner-Burton et al., 2017; Yarhi-Milo, 2014; Holmes, 2015; Rathbun, 2019; Rathbun, Kertzer and Paradis, 2017; Lee, 2020). This "dual-process" account of judgment "conceptualizes decisions as resulting from the interaction between more intuitive versus reflective processes. Intuitive processes are typically characterized as being fast, automatic, effortless, and emotional. Reflective processes, by contrast, are typically slower and more controlled, effortful and deliberative" (Rand et al., 2014, 2). System 1 is a "hot" system, often emotional in character, which induces individuals to act quickly based on their "gut feelings," without explicit thinking. By contrast, System 2 is a "cold" system that proceeds slowly as individuals carefully consider their beliefs and choices. System 2 indicates "procedural rationality," the process that individuals go through when thinking rationally. These systems are differentiated by thinking *process*, not the final outcome of the decision per se.

With some caveats, the distinction between System 1 and System 2 is useful for theorizing about the psychological bases of deterrence. Particularly in Kahneman's (2011) framework, System 1 sometimes serves as a residual category encompassing everything except reasoning, a category that is far too broad (Evans, 2008). Mercier and Sperber (2017) usefully distinguish entirely unconscious processes and behaviors from those that rely on intuition. "The content of

an intuition is conscious. It would be paradoxical to say, 'I have an intuition, but I am unaware of what it is about.' *There is no awareness, on the other hand, of the inferential processes that deliver an intuition*" (Mercier and Sperber, 2017, 75). We follow these authors' "simple first-pass way" in defining intuition as "judgments (or decisions, which can also be quite intuitive) that we make and take to be justified without knowledge of the reasons that justifies them" (Mercier and Sperber, 2017, 73). This stands in contrast to reasoning. "When we reason, conclusions do not just pop up in our mind as self-evident; we arrive at them by considering reasons to accept them" (Mercier and Sperber, 2017, 61). The content of an intuition is conscious but not reflective. Intuitions present themselves to us as needing no evidentiary basis.[3]

Whereas much psychological IR research conceptualizes biased and systematic deviations from rational thinking as *irrational*, as noted (but not endorsed) by Mercer (2005), evolutionary psychology offers a very different perspective: our intuitions, dispositions, and impulses are *ecologically rational*, that is, optimal given the environment in which humans evolved (Todd and Gigerenzer, 2012; Johnson, 2020). Ecological rationality refers to the efficiency of some trait or tendency of the organism in a recurrent situation, such as the need to deter aggression. It has nothing to do with reasoning per se and everything to do with behavior and outcomes. As Frank (1988, 6) writes, "Innate behavior patterns... are adaptive in a broad sense. Yet the behaviors they motivate should not be viewed as having sprung from a rational, case-by-case

---

[3]We note that our argument is also consistent with a second leading paradigm in psychological IR research. As Pauly and McDermott (2022, 42) describe, this paradigm emphasizes the emotional basis of reason, perhaps most associated with Antonio Damasio's somatic marker hypothesis. As Bechara and Damasio (2005, 352) explain, "The somatic marker hypothesis provides neurobiological evidence in support of the notion that people often make judgments based on 'hunches,' 'gut feelings,' and subjective evaluation of the consequences." Somatic predictions overlap with what Verweij and Damasio (2019, 5) refer to as Gigerenzer's (2007) "gut feeling" model. In IR, influential work by scholars like McDermott (2004) and Mercer (2013) emphasize the primacy of emotion in downstream cognition, questioning the conceptual distinction between emotion and reason (see also Grayot, 2020). For our purposes, both paradigms point in the same direction: most of human thought is intuitive, and we expect these intuitions to generate deterrence thinking.

assessment of self-interest." Something is rational if it enhances genetic fitness and success on average over tens of thousands or even millions of years, thereby promoting its selection. Applied to international relations, the idea is that threats to collective interests activate ingrained psychological mechanisms for dealing with challenges. "Models drawn from evolutionary psychology emphasize the centrality of environmental triggers and explain how contextual cues, such as those embedded in the social or institutional environment, activate different psychological strategies according to an ancestrally adaptive logic. When modern situations mirror these cues, the relevant psychological mechanisms become activated and shape our perceptions of threat and opportunity, and they instigate a repertoire of behavioral responses" (McDermott, Lopez and Hatemi, 2017, 77).

Deterrence is one of the most recurrent of human challenges. As McDermott, Lopez and Hatemi (2017, 70) explain, "As long as humans have lived and competed in groups, the question of deterring threats from one's adversaries has been of central importance...For all of human history, people have had to deal with challenges to their physical security and that of their family and friends." Instinctive reactions to aggression and notably the felt impulse for revenge, they argue, serve the evolutionary purpose of deterring threat. To deter, humans and other species engage in retaliatory aggression, meeting blows with blows (McDermott, Lopez and Hatemi, 2017). "Aggression as an adaption for conflict resolution has been extensively studied in primates and in humans. What has broadly been labeled 'retaliatory aggression'...is one of the most zoologically common, well-recognized and well-studied behavioral responses for dealing with threats and challenges." Retaliatory aggression is "designed by natural selection to deal with the challenges that threaten inherently social species. Aggression as an adaptation for conflict resolution...can be at least partly explained as a common component

of an evolved psychology, refined over millions of years of human evolution" (McDermott, Lopez and Hatemi, 2017, 76).

Retaliatory aggression is thought to be so quick and unreflective—that is, intuitive—precisely because deliberative and rational thought might have undermined our interests in certain types of situations that we commonly confronted in our evolutionary past. Retaliatory aggression is costly, but rationally calculating its advantages and disadvantages indicates less resolve to others, thereby inviting challenges. Those who retaliate impulsively, without thought of the cost, are more credible adversaries. This tendency to act without thought unwittingly serves self-interest. In other words, making more use of intuition than reason generates the ecologically rational outcome of deterrence. This process often works through feeling and emotions, such as the anger that accompanies retaliatory aggression. As Frank (1988, 4) argues, in discussing the chicken dynamics inherent to deterrence situations, "Passions often serve our interests. . . because we face important problems that simply cannot be solved by rational action. The common feature of these problems is that to solve them we must commit ourselves to behave in ways that may later prove contrary to our interests." Not swerving might lead to a head-on collision, so threats to not swerve lack credibility for rational people.

This seeming paradox—that we benefit ourselves by not thinking deliberatively—mirrors Thomas Schelling's concept of the "rationality of irrationality." Schelling (1960, 18) writes that "an explicit theory of 'rational' decision, and of the strategic consequences of such decisions, makes perfectly clear that it is not a universal advantage in situations of conflict to be inalienably and manifestly rational in decision and motivations." Here, he uses rationality in two senses. It is not always ecologically rational to be procedurally rational in thought. The fact that Schelling, the spiritual father of so much of formal modeling in international rela-

tions, distinguishes between non-reflective gut reactions and deliberation and calls the former "irrational" brings home that the default assumption and cognitive precondition in rationalist work is indeed rational thought.

None of this is to say that we cannot deter based on careful deliberation concerning the situation at hand. Some might come to the conclusion, through a careful rational calculation, of the importance of responding to provocations to preserve reputation and prevent future challenges, just as game theorists do. Rather, our claim is that IR research over-rationalizes, so to speak, what is typically a non-reflective impulse. Indeed human beings frequently rationalize their intuitions as well thought-out beliefs with an evidentiary basis (Haidt, 2001; Liu and Ditto, 2013). Precisely because thinking about deterrence indicates the rationality that makes it less credible, deterrence is counterintuitively intuitive.

Indeed, we have reason to believe that the use of rationality and reason is more likely to lead to efforts to conciliate and negotiate than to demonstrate resolve and credibility. Mercier and Sperber (2017, ch. 10) maintain that reason evolved to help humans settle conflicts and reap the gains of coordination and cooperation. Mutually beneficial coordination and cooperation promotes evolutionary success, a longstanding theme in the literature (Boehm, 2012; Bowles, 2011; Alexander, 2017). Doing so requires extensive communication and the giving of reasons to justify behavior and positions. By one argument, our reasoning faculties evolved originally as one of several "special-purpose mechanisms, designed to solve an array of recurrent, highly specialized adaptive problems" (Cosmides, 1989, 190).

Whereas intuition might have facilitated deterrence, reason promotes what we could call diplomacy. Therefore, deliberation is actually more likely to lead to an effort to conciliate and engage rather than to deter. Indeed, diplomacy, to the extent that it means something

more than the simple delivery of threats or "coercive diplomacy," is associated with reasoned dialogue, in which opposing sides attempt to work through their differences without rush to action (Rathbun, 2014; Risse, 2000). Reason allows cooler heads to prevail.

In light of the above, we examine the following overarching expectation:

> *H1:* Intuitive thinking is more predictive than deliberative thinking of retaliatory aggression and deterrence beliefs.

Related to this hypothesis, but of secondary concern, we consider the following expectation:

> *H2:* Deliberative thinking is more predictive than intuitive thinking of conciliation and diplomacy.

Again, we do not claim that individuals cannot rationally invest in reputation, or that individuals cannot intuitively cooperate. Rather, our expectation is that intuitive and deliberative thinking disproportionately associate with deterrence and cooperation, respectively.

Furthermore, we note that intuition and deliberation are not entirely antithetical to one another, nor are they wholly discrete systems. Our emotions and intuitions can shape higher order thinking and deliberation, as Damasio and colleagues maintain (Bechara and Damasio, 2005; Damasio, 1999). Individuals also rely on these systems to varying degrees, a finding we exploit later. Still, research shows that intuition and deliberation as cognitive styles are sufficiently orthogonal, such that it is reasonable to think of System 1 and System 2 as embodying differences in kind rather than degree (Pacini and Epstein, 1999).

4 Deterring, fast or slow: Survey evidence

To examine whether intuition generates deterrence tendencies, we first present evidence from two surveys. Given the challenges associated with studying intuition in international politics, surveys afford useful measurement precision. An innovative time pressure experiment first shows that intuition activates aggressive tendencies in a deterrence scenario, whereas deliberation begets diplomacy. A second survey shows strong correlations between well-known measures of reliance on intuition and alternative measures of deterrence tendencies, finding the same relationship reported in the experiment.

4.1 Under the gun: Time pressure experiment

Our experiment draws on an influential behavioral paradigm that uses variation in time pressure to experimentally activate intuition versus reasoning. As Rand et al (2012, 427) explain, "one of the psychological features most widely used to distinguish intuition from reflection is processing speed: intuitive responses are relatively fast, whereas reflective responses require additional time for deliberation." In an experimental setting, then, random variation in time pressure can activate intuitive versus deliberative thinking. Tasking individuals to make decisions quickly "reduces their ability to reflect and gives them less opportunity to override their intuitive responses. Conversely, forcing people to stop and think has the opposite effect, allowing for more deliberation" (Rand et al., 2014, 2). To be clear, time pressure here is a methodological device that elicits intuition, not a theoretical claim that decisions in the real

world are necessarily made quickly.[4] We believe that time pressure experiments greatly expand

our scholarly ability to prime, and distinguish between, intuitive versus deliberative thinking

processes in political contexts.

We fielded the experiment on a sample of US adults recruited from Prolific in June 2023

($N = 1,253$). The survey was housed on Qualtrics' online platform. Subjects first completed

standard demographic and foreign policy orientation measures. We then asked subjects to

imagine that they are the leader of the United States and presented all subjects with the fol-

lowing hypothetical scenario:

> The United States faces a potential challenger to its influence in an important region
>
> of the world. This potential challenger has mobilized its forces along the border
>
> with a smaller neighbor, threatening invasion.

We explained that the smaller country is important to US economic and security interests, the

challenger is not a democracy, the challenger lacks nuclear weapons, and its military is half as

powerful as US forces in the region.[5] Further, we explained that "You and your administration

have demanded that the challenger withdraw its troops from the border," but the challenger

has failed to comply. The scenario was designed to place credibility on the line, with the

expectation that the intuitive response is to protect it.

Following this scenario, participants were randomized into "fast" or "slow" conditions, as

well as conditions that manipulated the existence or absence of US reputational stakes, form-

ing a fully-crossed 2×2 design. Subjects in the fast condition were told "A rapid response is

necessary. Please try to choose one of the following within 15 seconds." We programmed the

---

[4]For more on the separate question of time constraints in international crises, see Holsti (1965); Carnevale and Lawler (1986); Brecher and Wilkenfeld (1997).

[5]Appendix A1.1 provides further instrumentation.

survey so that respondents could see a timer counting down the seconds. By contrast, subjects in the slow condition were asked to "Please think carefully for at least 30 seconds about the advantages and disadvantages of the different options. It is important to weigh the pros and cons." A timer was displayed and respondents were not allowed to make a decision until 30 seconds passed. We note that the treatments effectively induced fast versus slow decision-making, with the former subjects submitting responses in an average of 7.02 seconds and the latter submitting responses in an average of 43.4 seconds ($p < .001$, $t = -31.8$).

We also included an explicit manipulation of reputational stakes. A central premise of deterrence theory is that individuals can arrive at decisions to escalate a conflict due to concerns about reputational damage when threats, such as ultimatums, are not carried through. Half of the subjects were told: "You have threatened that there will be significant military consequences if the challenger does not yield," whereas we omitted this information for subjects in the no explicit reputational stakes condition. The reputation treatment allows us to examine whether the fast treatment pushes subjects in the same direction as a more common and well-established indicator of deterrence thinking. It further allows us to assess the substantive effect size of fast thinking relative to reputational concerns.

Following this experimental intervention, subjects responded to our dependent variable of interest. We told subjects that "Your advisors have given you the following options:"

- Continue to try to find a diplomatic solution to the conflict.

- Order U.S. fighter planes from an aircraft carrier nearby to strike challenger ground and air forces.

We expect that the fast treatment, as well as the reputational stakes treatment, will make

respondents more likely to opt for airstrikes against the challenger. By contrast, we expect that the slow treatment will make respondents more likely to pursue a diplomatic solution.

Figure 1 displays logistic regression estimates for the treatment effect of fast thinking (relative to slow thinking) on support for airstrikes (relative to diplomacy).[6] The fast treatment increases the odds of airstrikes by 1.48 (coef = 0.39, $p < .01$), suggesting that intuition generates an impulse towards escalatory aggression. Said differently, slow and deliberative thinking induces diplomatic tendencies. In the terminology of predicted probabilities, intuition (relative to deliberation) increases the probability of selecting airstrikes from 0.21 to 0.28, an increase of 33.3%. Reputational stakes also increase the odds of airstrikes (coef = 0.50, $p < .001$), suggesting that intuition leads subjects in the same direction as concern for reputation, a hallmark feature of deterrence thinking. Finally, the treatment effect of fast thinking is substantively noteworthy. The effect size is statistically indistinguishable from the effect size of reputational stakes and self-reported isolationist beliefs, a well-established predictor of foreign policy attitudes. All of this indicates that intuitive thinking generates hawkishness in this deterrence scenario.

Given that behavioral IR research often points to individual-level variation in deterrence tendencies (Mercer, 1997; Yarhi-Milo, 2018; Casler, 2025), Appendix A1.3 examines whether the treatment effect of intuition on escalation varies by trait-level hawkishness, college education, and conservative political ideology. We find a lack of interaction between the intuition treatment and these self-reported measures. That is, consistent with arguments from evolutionary theory that emphasize similarity across individual psychology, the treatment effect of intuition on escalation is not driven by these specific subgroups.

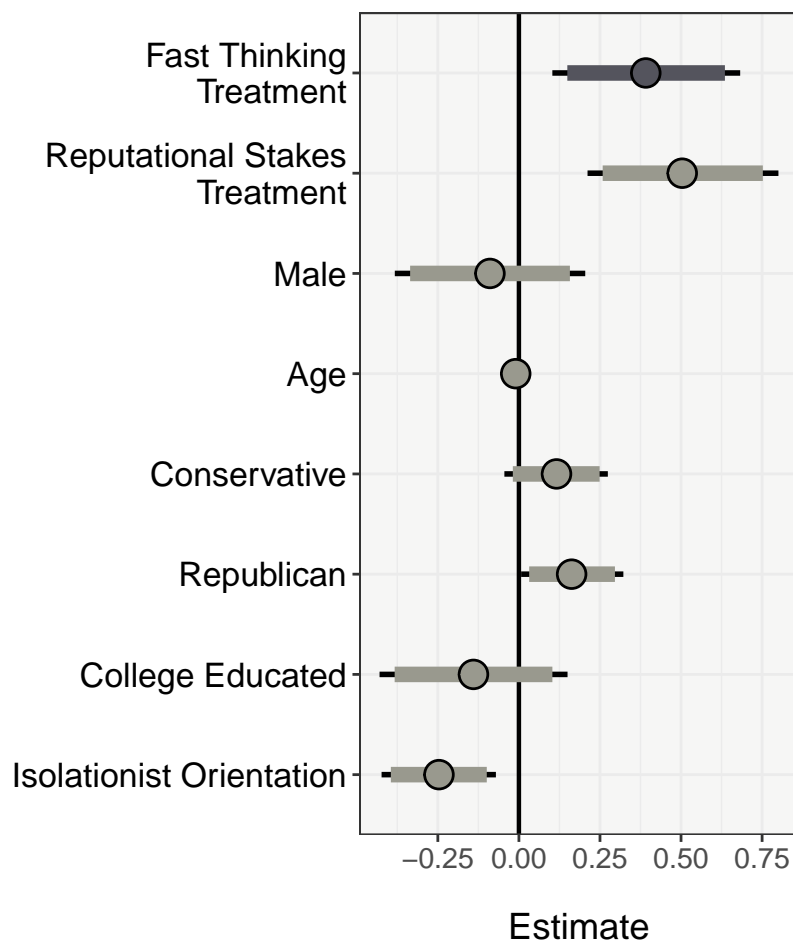[6]Appendix A1.2 presents the full regression results.

Figure 1: *Time Pressure Experiment Results.* Logistic regression estimates of the fast thinking treatment effect (relative to slow treatment) on support for airstrikes (relative to diplomacy), with 90% and 95% CIs. Sample consists of $N = 1,253$ US adults.

4.2 Faith in deterrence: Correlational evidence

Our experimental results suggest that intuition begets conflict escalation and retaliatory aggression, whereas deliberation induces diplomatic responses. We report evidence from a second survey for two reasons. First, given that we draw upon a well-established time pressure paradigm, we have strong reason to believe that the fast treatment above induces intuitive responses. Here, we further validate this expectation. We exploit the fact that although all individuals rely on intuition and reason to some degree, there is individual-level variation

in these tendencies. This allows us to make use of a previously validated measure of these thinking styles, the rational-experiential inventory (REI) (Epstein et al., 1996). This scale is grounded in a dual-process model of judgment, consistent with the dual-process paradigm that we describe above.

Second, we more firmly examine whether intuition leads to something more than just automatic, retaliatory aggression. Specifically, does intuition associate with explicit *beliefs* about the use of force to demonstrate credibility and resolve? If this is not the case, retaliatory aggression might simply be revenge. As described above, retribution might serve a deterrent function, even if that is not the explicit intent of the aggressor. However, we expect that, beyond simply automatic and unconcious responses, our intuitions associate with explicit beliefs about the efficacy of deterrence. Intuitions are conscious but unreflective in character, as discussed above.

In May 2020, we fielded a survey on 1,615 US-based adults recruited from the MTurk platform.[7] To measure reliance on intuition, we use the following items from the REI's "Faith in Intuition" subscale:

- I trust my initial feelings about people.

- I believe in trusting my hunches.

- My initial impressions of people are almost always right.

- When it comes to trusting people, I can usually rely on my "gut feelings."

- I can usually feel when a person is right or wrong even if I can't explain how I know.

---

[7]Subjects were directed to the study housed on the Qualtrics online platform.

By contrast, to measure relatively deliberative thinking styles, we use items from the REI's "Need for Cognition" subscale. Individuals vary in the degree to which they engage in effortful cognitive activity. As Cacioppo et al. (1996, 197) explain, "some individuals tend to act as cognitive misers in circumstances that call forth effortful problem solving in most individuals, whereas others tend to be concentrated cognizers even in situations that lull most individuals into a cognitive repose." Need for Cognition captures a greater commitment to the System 2 processing of which all are capable but to which not all are equally inclined. Rathbun, Kertzer and Paradis (2017) demonstrate that the Need for Cognition leads egoists to respond in ways that rationalist models predict in repeated bargaining games, both for undergraduates in incentivized bargaining experiments and elite decision-makers. To measure Need for Cognition, we use the following items:

- I don't like to do a lot of thinking (reverse coded).

- I prefer to do something that challenges my thinking abilities rather than something that requires little thought.

- I prefer complex to simple problems.

- I try to avoid situations that require thinking in depth about something (reverse coded).

- Thinking hard and for a long time about something gives me a lot of satisfaction.

We use factor analysis to reduce responses to unidimensional factor scores, where higher values indicate more intuitive and deliberate thinking styles, respectively.[8]

---

[8]Cronbach's $\alpha = 0.894$ and 0.864 for the Faith in Intuition and Need for Cognition items, respectively. SS loadings = 3.16 and 2.83, respectively. Appendix A2 presents the distribution of scores for each subscale.

To measure deterrence beliefs, we use items from the well-established militant internation-alism (MI) scale in foreign policy research (Wittkopf, 1990; Herrmann, Tetlock and Visser, 1999; Kertzer et al., 2014; Rathbun et al., 2016). Two items most directly measure beliefs in the efficacy of deterrence, gathered on five-point (dis)agree scales:

- The United States must demonstrate its resolve so that others do not take advantage of it.

- There is considerable validity in the theory that when one nation falls to aggressor na-tions, others nearby will soon follow a similar path.

While MI is understood as a hawkish posture in general, these two items provide the most direct test of our expectations (see also Jervis, 1976, 56-58). If the deterrence model is intuitive, then intuitive thinking will show a stronger association with these deterrence beliefs than deliberate thinking shows with these beliefs.

Figure 2 presents coefficients from linear regressions that model the relationship between intuition and the deterrence items from the MI scale.[9] Faith in Intuition strongly and posi-tively associates with deterrence beliefs, namely the need to demonstrate resolve and belief in the validity of the domino theory.[10] By contrast, deliberate thinking (measured with "Need for Cognition") displays a null association with deterrence beliefs. Said differently, intuition corre-lates much more strongly with deterrence beliefs than deliberate thinking. These results help to validate our experimental results above with alternative measures of intuition and deter-rence. Furthermore, these results extend our findings beyond the escalatory focus of the above

---

[9]Appendix A2.1 presents the full regression results.
[10]Appendix A2.2 shows that our results are robust to the exclusion of low-quality survey responses.

experiment, showing that intuition associates with specific beliefs about resolve and dominoes that are central to the deterrence model.
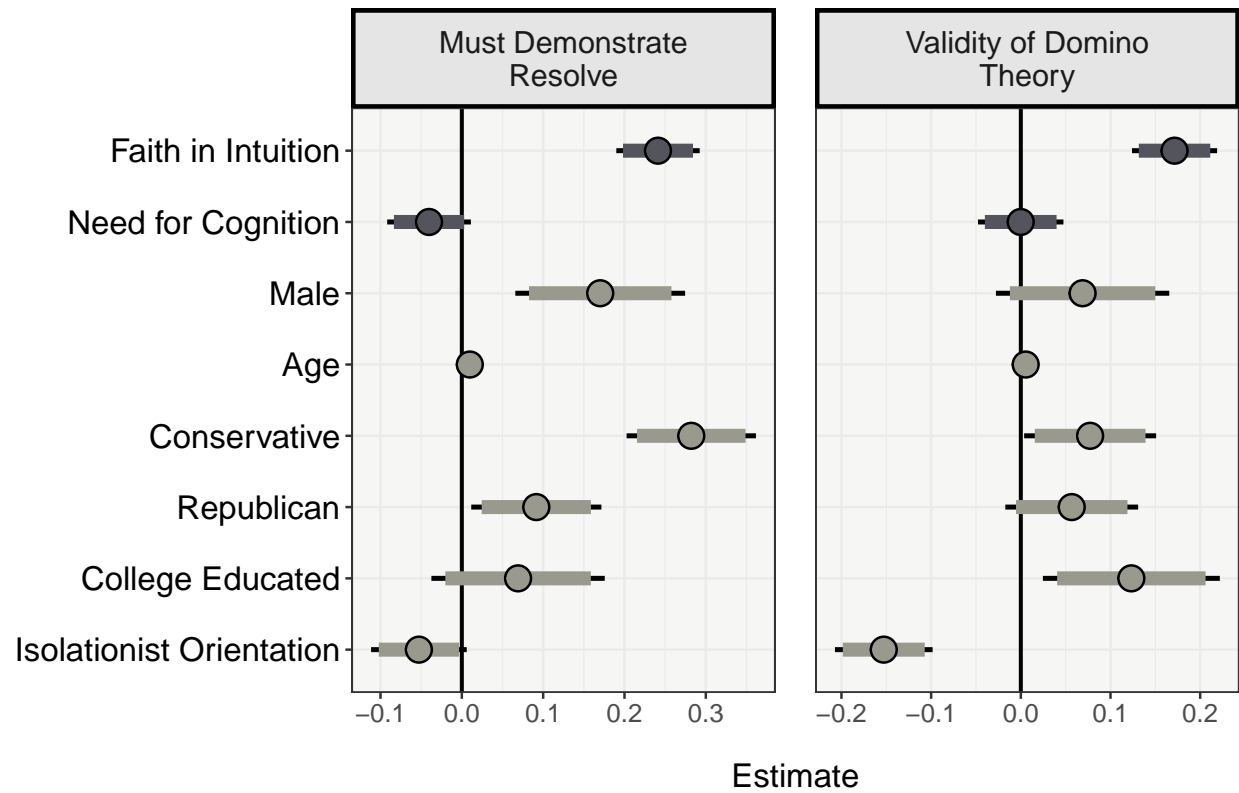

Figure 2: *Correlational Survey Results.* Linear regression estimates of the relationship between Faith in Intuition, Need for Cognition, and deterrence beliefs, with 90% and 95% CIs. The "resolve" and "domino" items are on five-point numeric scales. Sample consists of $N = 1,615$ US adults.

Aware that the demographic profile of MTurk respondents might differ from the general US population (Berinsky, Huber and Lenz, 2012), Appendix A2.3 assesses potential heterogeneity in the relationship between intuitive thinking and deterrence beliefs. We find a lack of interaction between intuitive thinking and education level, and a slight negative interaction between intuitive thinking and self-reported conservatism. Consistent with evolutionary arguments, the latter suggests that the correlation between intuitive thinking and deterrence beliefs is not uniquely driven by conservative respondents, who tend to be more dispositionally hawkish

(Kertzer et al., 2014). Furthermore, as with education, both liberal and conservative respondents show a positive association between intuition and deterrence beliefs, which suggests that our results are not driven by the demographics of MTurk respondents.

5   US elites during the Cold War: Text analytic evidence

Our experimental and observational survey evidence examines ordinary individuals, finding that intuition generates hawkishness in general and deterrence thinking in particular. We expect that these are basic human tendencies and, therefore, expect to see this same phenomenon among elite foreign policy decision-makers. This final empirical section assesses our expectations across all available Cold War documents in the *Foreign Relations of the United States*. Recent research suggests that textual data provide rich insights into psychological mechanisms (Boyd and Schwartz, 2021). We extend this text analytic research by assessing the relationship between intuition and deterrence thinking.

To obtain the documents, we manually downloaded every available Cold War era volume in the *FRUS* from the website of the US State Department's Office of the Historian. The *FRUS* represents the "official documentary historical record of major US foreign policy decisions and diplomatic activity."[11] Entries in the *FRUS* include memoranda, correspondence, and diplomatic cables, among others. Indeed, to the extent that some documents in the *FRUS* were written in a more deliberative fashion, this corpus provides a harder test for intuitive thinking. We focus on the Cold War, because perhaps no other era is more associated with the concept of deterrence in US foreign policy. The volumes under consideration begin in 1945 with the

---

[11]Office of the Historian, available at <https://history.state.gov/historicaldocuments/about-frus>, accessed 9 December 2024.

Truman Administration and thin out by 1988 with the Reagan Administration, where many volumes are still under declassification review. The result is 300 total volumes, which we convert into plain text files for analysis. The corpus contains over 142 million words.

To analyze the texts, we use a method well-suited to capture intuitions: word embeddings. Traditional models of political text rely on word *frequencies*, namely counts of terms in a given document (Grimmer and Stewart, 2013). By contrast, embedding models utilize word *co-occurrences* to operationalize the notion that we can know a word by the company it keeps (Rodriguez and Spirling, 2022). Prominent social science research finds that embeddings more powerfully capture language meaning than count-based models of text. For example, embeddings have been used to quantify psychological biases and implicit associations in human language (Caliskan, Bryson and Narayanan, 2017) as well as racial and gender stereotypes at societal scales (Garg et al., 2018; Kozlowski, Taddy and Evans, 2019).

To estimate the embeddings, we use the global vectors for word representation (GloVe) model (Pennington, Socher and Manning, 2014), perhaps the most common embedding model in the social sciences (Rodriguez and Spirling, 2022).[12] Before fitting the model, we stem the terms (e.g., `deterrence` and `deterrent` both become `deter*`), lowercase the terms, remove punctuation, and remove terms that appear fewer than 10 times across the corpus. To construct the co-occurrence matrix, we use a fixed window of six words before and after a target feature, a standard choice in embeddings research (Rodriguez and Spirling, 2022). For example, we would expect the term `feel*` (relative to `investig*`) to co-occur more often with `deter*`. Finally, the GloVe model uses an ordinary least squares estimator to reduce this matrix to a lower dimensional representation. The output of this reduction procedure

---

[12]Appendix A3 describes model hyperparameter choices and embedding quality checks.

is an "every unique term" by "arbitrary coordinate" dimension matrix. We fit the model in 300 arbitrary dimensions, perhaps the most-standard dimensionality choice for these models (Rodriguez and Spirling, 2022). The output of the model is a set of 300 coordinates (i.e., the beta coefficients fitted by the OLS estimator) for each unique term in the corpus. These coordinates quantify the location of every unique term in the estimated vector space.

With these word coordinates in hand, the final step before analysis is the definition of dictionary terms that capture our theoretical constructs. Although psychologists often use textual data to capture system 1 processes, like emotion or implicit bias (see, e.g., Boyd and Schwartz, 2021), there is no work we know of that operationalizes the presence of intuition. We therefore take a theory-driven approach to define our dictionary terms, conduct close readings of *FRUS* documents to confirm the face validity of these terms, and conduct robustness checks to ensure that our results are not sensitive to choices of specific dictionary terms. To measure observed thinking styles, we use the following terms:

- <u>Intuition</u>: `intuit*, instinct*, hunch*, sens*, feel*`.

- <u>Deliberation</u>: `examin*, investig*, calcul*, study*, diagnos*`.

The former terms serve as proxies for intuitive thinking, and the latter terms proxy for deliberative thinking that might appear in these texts. As described above, intuition is not unconscious and, therefore, surfaces in writing. For example, our intuition terms would capture sentences like the following, written by Special Assistant for Vietnamese Affairs George Carver in 1968: "I have a strong visceral **<u>hunch</u>**, unsupported by evidence, that our friends in Hanoi are contemplating some major move—either military, political or both—within the next few days."[13]

---

[13]*Foreign Relations of the United States*, 1964–1968, Vol. VI, Vietnam, January-August 1968, Document 342.

By contrast, our deliberation terms would capture statements like Dean Acheson's suggestion in 1968 that the US should conduct "An **examination** of what has happened over the last year in Vietnam, including, of course, the Tet offensive and its aftermath. This should be a careful, factual effort."[14]

The empirical question, of course, is whether intuitive or deliberative thinking is more likely to co-appear with evidence of deterrence thinking. We use the following terms to measure deterrence versus cooperative thinking:

- <u>Deterrence</u>: **deter*, resolut*, credibl*, reput*, strength*.**

- <u>Cooperation</u>: **cooper*, negot*, coordin*, collabor*, discuss*.**

We selected the deterrence terms in line with standard theoretical understandings of deterrence, which emphasize concerns for reputation and credibility, as well as the projection of strength and resolve. We measured cooperative tendencies to align with our experiment presented above, which offered subjects an escalation versus a diplomatic option.

With these constructs in hand, we expect that intuition (relative to deliberation) correlates positively with deterrence thinking. We use cosine similarity as the metric of interest, essentially a correlation coefficient well-suited for vector spaces like embeddings. We use a similar procedure to Caliskan, Bryson and Narayanan (2017), which estimates the average cosine similarities between the intuition terms and deterrence, as well as the similarities between the deliberation terms and deterrence, and then takes the difference between these two quantities. Akin to a difference-in-differences estimate, larger values indicate that intuition correlates more strongly with deterrence than deliberation correlates with deterrence. We also

---

[14]Ibid, Document 125.

estimate the same relationship between intuition and cooperation. In line with the experiment above, we expect a null or negative association between intuition terms and cooperation. Said differently, we expect deliberation terms to correlate positively with cooperation.

Finally, we use the resampling procedure described by Kozlowski, Taddy and Evans (2019) to construct nonparametric confidence intervals. Specifically, we resample the underlying corpus with replacement to construct twenty resampled corpora, refit an embedding model to each resampled corpus, and then re-calculate the estimate of interest. The result is twenty cosine estimates, where the second smallest and nineteenth largest estimates span a 90% nonparametric confidence interval.

Figure 3 displays the correlation between intuition terms (relative to deliberation terms) and deterrence terms, as well as the correlation between intuition terms and cooperation terms. Intuition terms associate positively and significantly with deterrence terms (cosine sim = 0.10, 90% CI[0.04, 0.17]). The interpretation is that when US decision-makers use terms that indicate intuitive judgment (e.g., `feel*`), they are more likely to use terms that indicate deterrence thinking (e.g., `deter*`). By contrast, the use of intuitive terms associates negatively and significantly with cooperation terms (cosine sim = -0.17, 90% CI[-0.09, -0.22]). In other words, deliberation terms associate positively with cooperation terms. These results strongly align with our survey findings above.

Appendix A3 reports a series of robustness checks. We validate our deterrence terms with a ground-truth example, finding that our deterrence terms correlate positively with terms indicative of Soviet threat. Given that much of Cold War US deterrence strategy was indeed targeted at the Soviet Union, this increases our confidence in the validity of our deterrence terms. Furthermore, we use the nearest neighbors functionality in the `conText` package to
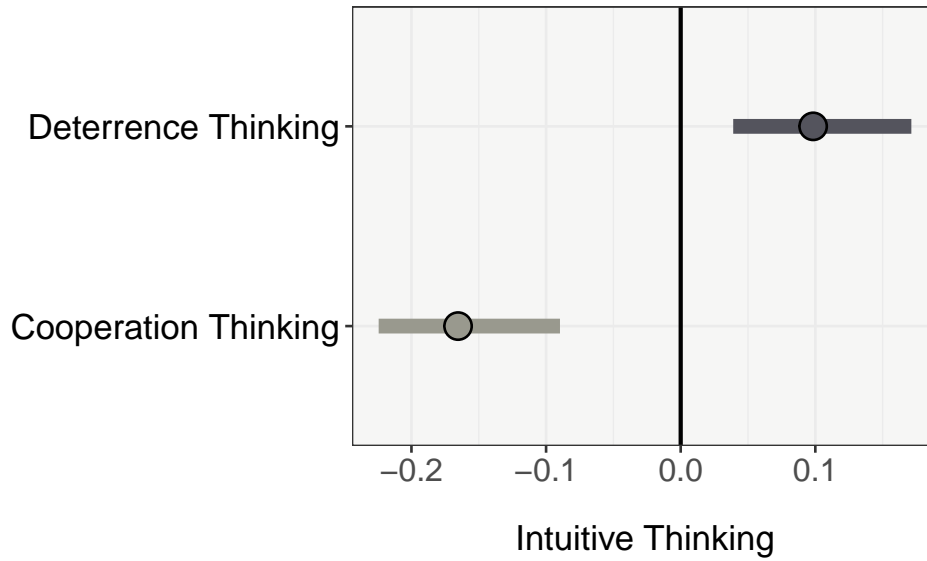
Figure 3: *Text Analysis Results.* Intuition terms (relative to deliberation terms) correlate positively with deterrence terms and negatively with cooperation terms. Nonparametric 90% confidence intervals derive from twenty resamples of the underlying corpus.

assess the quality of the fitted embeddings (Rodriguez, Spirling and Stewart, 2023), and we extend our intuition and deliberation dictionaries. We find that our dictionary terms are close to other reasonable synonyms in the fitted vector space, and the results reported above are robust to choices of different dictionary terms. These analyses bolster our results. IR scholarship typically portrays the Cold War as a strategic, calculating contest of deterrence. These results instead suggest that intuition associates with deterrence thinking among Cold War US decision-makers. Deliberative, calculating terms associate with the opposite: cooperation.

6   Conclusion

Research in international relations typically associates deterrence with a deliberative and calculating mindset. By contrast, our findings suggest that intuition underpins deterrence thinking. What is typically explained as the outcome of a rational process is instinctive rather than reflec-

tive. Rational deliberation likely generates very different behavioral inclinations: the pursuit of diplomacy, cooperation, and conflict de-escalation. More often than not, deterrence is counterintuitively intuitive.

The impulse to match blows with blows helps to explain the basic human tendency to err on the side of threat overreaction rather than underreaction (Stein, 2013, 1988; Pomeroy, 2025). Some earlier IR scholarship portrays threat inflation as the product of thoughtful and strategic manipulation by elites (Mueller, 2006; Kaufmann, 2004). Evolutionary psychology instead suggests that threat inflation is deeply ingrained in human nature, encouraging us to "err on the side of caution" and therefore survive (Johnson, 2020, 9). Our results help to explain why members of the public might be intuitively receptive to elite narratives of threat inflation, and why elites tell such narratives in the first place. Rational models of deterrence could be updated to incorporate a bias in favor of threat overreaction.

Relatedly, our findings help to explain why the deterrence model comes more naturally than the empathy required to solve security dilemmas. Security dilemmas require deliberative thought, a conscious attempt to understand the appearance of one's actions to the adversary, what Wheeler and Booth (2008) call "security dilemma sensibility." Before we can end spirals, we have to realize that we are in one. To place oneself in another's shoes is cognitively taxing (Kertzer, Brutger and Quek, 2024). In an early account of the security dilemma, Butterfield (1951, 154) describes this challenge: we feel the "terrible fear" of the adversary yet fail to appreciate the adversary's "counter-fear." This is why Kydd (2005) argues that rational signaling can reduce fears and convey status quo intentions. This is also why Holmes (2015) suggests that decision-makers are pulled towards responding to hostility with hostility, the result of an unthinking, "affective intuition." All of this points to rich avenues for future work on deliber-

ation and de-escalation. If the lack of reflection begets deterrence, is it also the case that "I think, therefore I cooperate?"

References

Alexander, Richard. 2017. *The biology of moral systems*. Routledge.

Arad, Ayala and Ariel Rubinstein. 2012. "The 11-20 money request game: A level-k reasoning study." *American Economic Review* 102(7):3561–73.

Bechara, Antoine and Antonio R Damasio. 2005. "The somatic marker hypothesis: A neural theory of economic decision." *Games and economic behavior* 52(2):336–372.

Berinsky, Adam J, Gregory A Huber and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political analysis* 20(3):351–368.

Boehm, Christopher. 2012. *Moral origins: The evolution of virtue, altruism, and shame*. Soft Skull Press.

Bowles, S. 2011. "A cooperative species: Human reciprocity and its evolution.".

Boyd, Ryan L and H Andrew Schwartz. 2021. "Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field." *Journal of Language and Social Psychology* 40(1):21–41.

Brecher, Michael and Jonathan Wilkenfeld. 1997. *A study of crisis*. University of Michigan Press.

Bueno de Mesquita, Bruce. 2013. *Principles of international politics*. Sage.

Butterfield, Herbert. 1951. "The Tragic Element in Modern International Conflict." *The Review of Politics* 12(2):147–164.

Cacioppo, John T, Richard E Petty, Jeffrey A Feinstein and W Blair G Jarvis. 1996. "Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition." *Psychological bulletin* 119(2):197.

Caliskan, Aylin, Joanna J Bryson and Arvind Narayanan. 2017. "Semantics derived automatically from language corpora contain human-like biases." *Science* 356(6334):183–186.

Carnevale, Peter J. D. and Edward J. Lawler. 1986. "Time Pressure and the Development of Integrative Agreements in Bilateral Negotiations." *Journal of Conflict Resolution* 30(4):636–659.

Casler, Don. 2025. "Credibility, Organizational Politics, and Crisis Decision Making." *Journal of Conflict Resolution* 69(6):1059–1088.

Cosmides, Leda. 1989. "The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task." *Cognition* 31(3):187–276.

Damasio, Antonio R. 1999. *The feeling of what happens: Body and emotion in the making of consciousness*. Houghton Mifflin Harcourt.

Epstein, Seymour, Rosemary Pacini, Veronika Denes-Raj and Harriet Heier. 1996. "Individual differences in intuitive–experiential and analytical–rational thinking styles." *Journal of personality and social psychology* 71(2):390–405.

Evans, Jonathan St BT. 2008. "Dual-processing accounts of reasoning, judgment, and social cognition." *Annu. Rev. Psychol.* 59:255–278.

Fearon, James D. 1994. "Signaling versus the balance of power and interests: An empirical test of a crisis bargaining model." *Journal of Conflict Resolution* 38(2):236–269.

Frank, Robert H. 1988. *Passions within reason: The strategic role of the emotions.* WW Norton & Co.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky and James Zou. 2018. "Word embeddings quantify 100 years of gender and ethnic stereotypes." *Proceedings of the National Academy of Sciences* 115(16):E3635–E3644.

Gigerenzer, Gerd. 2007. *Gut feelings: The intelligence of the unconscious.* Penguin.

Grayot, James D. 2020. "Dual process theories in behavioral economics and neuroeconomics: A critical review." *Review of Philosophy and Psychology* 11(1):105–136.

Grimmer, Justin and Brandon M Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis* 21(3):267–297.

Hafner-Burton, Emilie M, Stephan Haggard, David A Lake and David G Victor. 2017. "The behavioral revolution and international relations." *International Organization* 71(S1):S1–S31.

Haidt, Jonathan. 2001. "The emotional dog and its rational tail: a social intuitionist approach to moral judgment." *Psychological review* 108(4):814.

Herrmann, Richard K, Philip E Tetlock and Penny S Visser. 1999. "Mass public decisions on go to war: A cognitive-interactionist framework." *American Political Science Review* 93(3):553–573.

Holmes, Marcus. 2015. "Believing this and alieving that: theorizing affect and intuitions in international politics." *International Studies Quarterly* 59(4):706–720.

Holmes, Marcus and David Traven. 2015. "Acting rationally without really thinking: The logic of rational intuitionism for International Relations Theory." *International Studies Review* 17(3):414–440.

Holsti, Ole R. 1965. "The 1914 case." *American Political Science Review* 59(2):365–378.

Jervis, Robert. 1976. "Perception and misperception in international politics." *Princeton University Press* 13.

Jervis, Robert. 1989. *The logic of images in international relations.* Columbia University Press.

Johnson, Dominic D. P. 2020. *Strategic Instincts: The Adaptive Advantages of Cognitive Biases in International Politics*. Princeton, NJ: Princeton University Press.

Jost, Tyler, Joshua D Kertzer, Eric Min and Robert Schub. 2024. "Advisers and Aggregation in Foreign Policy Decision Making." *International Organization* 78(1):1–37.

Kahneman, Daniel. 2011. *Thinking, fast and slow*. Macmillan.

Kahneman, Daniel and Jonathan Renshon. 2007. "Why hawks win." *Foreign policy* pp. 34–38.

Kaufmann, Chaim. 2004. "Threat Inflation and the Failure of the Marketplace of Ideas: The Selling of the Iraq War." *International Security* 29(1):5–48.

Keohane, Robert. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton University Press.

Kertzer, Joshua D, Kathleen E Powers, Brian C Rathbun and Ravi Iyer. 2014. "Moral support: How moral values shape foreign policy attitudes." *The Journal of Politics* 76(3):825–840.

Kertzer, Joshua D., Marcus Holmes, Brad L. LeVeck and Carly Wayne. 2022. "Hawkish Biases and Group Decision Making." *International Organization* 76(3):513–548.

Kertzer, Joshua D., Ryan Brutger and Kai Quek. 2024. "Perspective Taking and the Security Dilemma: Cross-National Experimental Evidence from China and the United States." *World Politics* 76(2):374–78.

Kozlowski, Austin C, Matt Taddy and James A Evans. 2019. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84(5):905–949.

Kydd, Andrew H. 2005. *Trust and mistrust in international relations*. Princeton University Press.

Lake, David A and Robert Powell. 1999. *Strategic choice and international relations*. Princeton University Press.

Lee, Yoon Jin. 2020. Thinking Fast and Slow in Alliance Politics PhD thesis Harvard University.

Liu, Brittany S and Peter H Ditto. 2013. "What dilemma? Moral evaluation shapes factual belief." *Social Psychological and Personality Science* 4(3):316–323.

McDermott, Rose. 2004. "The feeling of rationality: The meaning of neuroscientific advances for political science." *Perspectives on politics* 2(4):691–706.

McDermott, Rose, Anthony C Lopez and Peter K Hatemi. 2017. "'Blunt Not the Heart, Enrage It': The Psychology of Revenge and Deterrence." *Texas National Security Review* 1(1):68–88.

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

Mercer, Jonathan. 2005. "Rationality and psychology in international politics." *International organization* 59(1):77–106.

Mercer, Jonathan. 2013. "Emotion and strategy in the Korean War." *International Organization* 67(2):221–252.

Mercier, Hugo and Dan Sperber. 2017. *The enigma of reason*. Harvard University Press.

Mueller, John E. 2006. *Overblown: How politicians and the terrorism industry inflate national security threats, and why we believe them*. Simon and Schuster.

Oye, Kenneth A. 1985. "Explaining cooperation under anarchy: Hypotheses and strategies." *World politics* 38(1):1–24.

Pacini, Rosemary and Seymour Epstein. 1999. "The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon." *Journal of personality and social psychology* 76(6):972.

Pauly, Reid BC and Rose McDermott. 2022. "The psychology of nuclear Brinkmanship." *International security* 47(3):9–51.

Pennington, Jeffrey, Richard Socher and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pp. 1532–1543.

Pomeroy, Caleb. 2025. "The Damocles Delusion: The Sense of Power Inflates Threat Perception in World Politics." *International Organization* 79(1):1–35.

Porter, Patrick. 2018. "Why America's grand strategy has not changed: Power, habit, and the US Foreign policy establishment." *International Security* 42(4):9–46.

Powers, Kathleen E and Dan Altman. 2023. "The psychology of coercion failure: How reactance explains resistance to threats." *American Journal of Political Science* 67(1):221–238.

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

Rand, David G, Alexander Peysakhovich, Gordon T Kraft-Todd, George E Newman, Owen Wurzbacher, Martin A Nowak and Joshua D Greene. 2014. "Social heuristics shape intuitive cooperation." *Nature communications* 5(1):1–12.

Rand, David G, Joshua D Greene and Martin A Nowak. 2012. "Spontaneous giving and calculated greed." *Nature* 489(7416):427–430.

Rathbun, Brian C. 2014. *Diplomacy's value: Creating security in 1920s Europe and the contemporary Middle East*. Cornell University Press.

Rathbun, Brian C. 2019. *Reasoning of state: Realists, romantics and rationality in international relations*. Vol. 149 Cambridge University Press.

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.

Rathbun, Brian C, Joshua D Kertzer and Mark Paradis. 2017. "Homo diplomaticus: Mixed-method evidence of variation in strategic rationality." *International Organization* 71(S1):S33–S60.

Renshon, Jonathan and Daniel Kahneman. 2017. "Hawkish biases and the interdisciplinary study of conflict decision-making." *Advancing Interdisciplinary Approaches to international relations* pp. 51–81.

Risse, Thomas. 2000. ""Let's argue!": communicative action in world politics." *International organization* 54(1):1–39.

Rodriguez, Pedro L and Arthur Spirling. 2022. "Word embeddings: What works, what doesn't, and how to tell the difference for applied research." *The Journal of Politics* 84(1):101–115.

Rodriguez, Pedro L, Arthur Spirling and Brandon M Stewart. 2023. "Embedding regression: Models for context-specific description and inference." *American Political Science Review* 117(4):1255–1274.

Schelling, Thomas C. 1960. *The strategy of conflict*. Harvard university press.

Schelling, Thomas C. 1966. "Arms and influence Yale University press." *New Haven* .

Stanovich, Keith. 2011. *Rationality and the reflective mind*. Oxford University Press.

Stein, Janice Gross. 1988. "Building politics into psychology: The misperception of threat." *Political psychology* pp. 245–271.

Stein, Janice Gross. 2013. "Threat perception in international relations." *The Oxford Handbook of Political Psychology* 2.

Tetlock, Philip E. 1983. "Policy-makers' images of international conflict." *Journal of Social issues* 39(1):67–86.

Todd, Peter M and Gerd Gigerenzer. 2012. *Ecological rationality: Intelligence in the world*. OUP USA.

Verweij, Marco and Antonio Damasio. 2019. The somatic marker hypothesis and political life. In *Oxford research encyclopedia of politics*.

Weisiger, Alex and Keren Yarhi-Milo. 2015. "Revisiting reputation: How past actions matter in international politics." *International Organization* 69(2):473–495.

Wheeler, Nicholas and Ken Booth. 2008. *The security dilemma: Fear, cooperation, and trust in world politics*. Palgrave Macmillan.

Wittkopf, Eugene R. 1990. *Faces of internationalism: Public opinion and American foreign policy*. Duke University Press.

Wrangham, Richard W. 2018. "Two types of aggression in human evolution." *Proceedings of the National Academy of Sciences* 115(2):245–253.

Yarhi-Milo, Keren. 2014. *Knowing the Adversary: Leaders, Intelligence, and Assessment of Intentions in International Relations*. Princeton University Press.

Yarhi-Milo, Keren. 2018. *Who Fights for Reputation: The Psychology of Leaders in International Conflict*. Princeton University Press.

Zagare, Frank C. 1990. "Rationality and deterrence." *World Politics* 42(2):238–260.

Zagare, Frank C and D Marc Kilgour. 2000. *Perfect deterrence*. Vol. 72 Cambridge University Press.

Supplementary Appendix for:
Shooting From the Hip: The Counterintuitively Intuitive Nature of Deterrence

August 21, 2025

A1 Time Pressure Survey Experiment

The main text presents the results of a time pressure experiment. Here, we provide additional sample details, instrumentation, and full regression tables. The sample consisted of 1,253 US-based adults, recruited from Prolific and directed to the survey housed on the Qualtrics online platform (54.2% male, 45.8% female or non-binary; 14.0% Republican, 44.0% Democrat, 42.0% Independent or Independent but leaning; 30.1% age 18-30, 32.0% age 31-40, 17.2% age 41-50, 20.7% age 51 or older). The survey was fielded in June 2023.

A1.1 Additional Instrumentation

Subjects first completed standard demographic questions and were then presented with the following instrumentation. The experiment began with the following:

> We are going to describe a situation the United States has faced in the past and could likely face again in the future. For scientific validity, the situation is general, and is not about any specific country in the news today. After describing the situation, we will ask a few questions. We would like you to imagine that you are the President of the United States and to tell us how you would respond to the following hypothetical scenario.

Then, we presented subjects with the following scenario:

> The United States faces a potential challenger to its influence in an important region of the world. This potential challenger has mobilized its forces along the border with a smaller neighbor, threatening invasion.
>
> - That smaller country is important to U.S. economic and security interests.
> - The challenger's military is about half as powerful as American forces in the region but the challenger does NOT have nuclear weapons.
> - The challenger is not a democracy and shows no signs of becoming a democracy.
>
> As the leader of our country, you must consider the best way to defend American interests in the region at a reasonable cost.

The next screen presented the reputation treatment (namely reputation information or no reputation information).

- You and your administration have demanded that the challenger withdraw its troops from the border.

- [You have threatened that there will be significant military consequences if the challenger does not yield. / No statement.]

- However, the challenger has not removed its forces and shows no sign it will do so.

The final screen presented the time pressure treatment and measured the study's dependent variables.

Your advisors have given you the following options.

- Continue to try to find a diplomatic solution to the conflict.

- Order U.S. fighter planes from an aircraft nearby to strike challenger ground and air forces.

[A rapid response is necessary. Please try to choose one of the following within 15 seconds. / Please think carefully for at least 30 seconds about the advantages and disadvantages of the different options. It is important to weigh the pros and cons.]
(Forced choice selection)

- Continue to try to find a diplomatic solution to the conflict.

- Order U.S. fighter planes from an aircraft nearby to strike challenger ground and air forces.

A1.2 Full Regression Results

Table A1 presents the results of the time pressure experiment, estimated using logistic regression. The outcome measure is a binary variable capturing support for airstrikes, relative to continuing to try to find a diplomatic solution to the conflict. The treatment effect of fast

3

thinking (relative to slow thinking) is substantively noteworthy: the effect is statistically indistinguishable from the reputation treatment (a known predictor of deterrent tendencies) and self-reported isolationist beliefs (a well-known foreign policy orientation).[1]

Table A1: Time Pressure Experiment: Logistic Regression Estimates

	Support Airstrikes
Intercept	−2.34***
	(0.32)
Fast Treatment	0.39**
	(0.15)
Reputation Treatment	0.50***
	(0.15)
Male	−0.09
	(0.15)
Age	−0.01
	(0.01)
Conservative	0.12
	(0.08)
Republican	0.16*
	(0.08)
College Educated	−0.14
	(0.15)
Isolationist Orientation	−0.26**
	(0.09)
AIC	1197.60
BIC	1243.79
Log Likelihood	−589.80
Deviance	1179.60
Num. obs.	1252

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

---

[1]By statistically indistinguishable, we mean the fast treatment estimate of 0.39 falls within the the 95% CI of the reputation treatment effect [0.21, 0.80] and isolationist beliefs coefficient [−0.44, −0.08]. We measured isolationism with three commonly-used items: (1) Our allies are perfectly capable of defending themselves and they can afford it, thus allowing the United States to focus on internal rather than external threats to its well-being; (2) The U.S. needs to play an active role in solving conflicts around the world (reverse-coded); (3) We should not think so much in international terms but concentrate more on our own national problems. See Kertzer et al. (2014) for more on this scale.

A1.3   Assessing Potential Treatment Heterogeneity

Although we primarily focus on the average treatment effects of intuition on deterrence tendencies, we also assess whether subjects who score higher in trait hawkishness disproportionately drive our treatment effect. This moderation analysis helps to adjudicate between an individual-differences story, in which hawks become more hawkish when compelled to think intuitively, versus a more general story in which all subjects become more escalatory regardless of trait hawkishness or dovishness. We measure hawkishness using common items from the militant internationalism scale.[2] Table A2 presents the results of this moderation analysis, estimated with logistic regression. We find a lack of significant interaction between the fast treatment and militant orientation (coef $= 0.19$, $p = 0.36$). This result is aligns with evolutionary arguments that emphasize similarity across individuals and suggests that hawkish subjects do not uniquely drive our treatment effect.

As a further check on potential treatment heterogeneity beyond trait hawkishness, we assess interactions between the intuition treatment and measures of education level and political ideology. Given our paper's focus on differences in more intuitive versus deliberative and analytical thinking, we check whether the treatment effect varies by education level. Furthermore, research shows that individuals with more conservative beliefs tend to display more hawkish foreign policy attitudes (Kertzer et al., 2014). Table A3 presents these moderation analyses, estimated with logistic regressions. We find a lack of significant interaction between the fast treatment and college education (coef $= -0.30$, $p = 0.31$), as well as the fast treatment and conservative ideology (coef $= -0.01$, $p = 0.94$). As with trait hawkishness above, these null effects further align with evolutionary arguments that emphasize similarity, rather than differences, across human psychology.

---

[2]The wording of the three MI items is as follows: (1) There is considerable validity in the theory that when one nation falls to aggressor nations, others nearby will soon follow a similar path; (2) The United States should take all steps including the use of force to prevent aggression by any expansionist power; (3) Rather than simply countering our opponents' thrusts, it is necessary to strike at the heart of an opponent's power.

Table A2: Time Pressure Experiment: Lack of Interaction Between Hawkishness and the Fast Treatment

	Support Airstrikes
Intercept	−2.16***
	(0.33)
Fast Treatment	0.29
	(0.16)
Reputation Treatment	0.53***
	(0.15)
Male	−0.20
	(0.15)
Age	−0.01
	(0.01)
Conservative	0.05
	(0.08)
Republican	0.17*
	(0.08)
College Educated	−0.05
	(0.15)
Isolationist Orientation	−0.01
	(0.10)
Militant Orientation	0.70***
	(0.15)
Fast Treatment × Militant Orientation	0.19
	(0.20)
AIC	1139.32
BIC	1195.78
Log Likelihood	−558.66
Deviance	1117.32
Num. obs.	1252

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table A3: Time Pressure Experiment: Lack of Interaction Between College Education, Conservative Ideology, and the Fast Treatment

	Support Airstrikes (College Education Model)	Support Airstrikes (Ideology Model)
Intercept	−2.44***	−2.36***
	(0.33)	(0.36)
Fast Treatment	0.55*	0.41
	(0.21)	(0.33)
Reputation Treatment	0.51***	0.50***
	(0.15)	(0.15)
Male	−0.10	−0.09
	(0.15)	(0.15)
Age	−0.01	−0.01
	(0.01)	(0.01)
Conservative	0.12	0.12
	(0.08)	(0.09)
Republican	0.16*	0.16*
	(0.08)	(0.08)
College Educated	0.03	−0.14
	(0.22)	(0.15)
Isolationist Orientation	−0.25**	−0.26**
	(0.09)	(0.09)
Fast Treatment × College Educated	−0.30	
	(0.30)	
Fast Treatment × Conservative		−0.01
		(0.08)
AIC	1198.55	1199.60
BIC	1249.88	1250.92
Log Likelihood	−589.28	−589.80
Deviance	1178.55	1179.60
Num. obs.	1252	1252

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

A2 Intuition and Deterrence Correlational Survey

The main text presents the analysis of an original survey that uses alternative measures of intuitive thinking and deterrence beliefs. This survey helps to serve as a robustness check on our experimental results while also extending the experiment's results with explicit deterrence *beliefs*, as opposed to simply retaliatory aggression in a deterrence context. Here, we provide additional sample details and full regression tables.

The survey was fielded in May 2020 on 1,615 US-based adults recruited from the MTurk platform (59% male, 41% female or non-binary; 23.4% Republican, 38.8% Democrat, 37.8% Independent or Independent but leaning; 26.7% age 18-30, 36.6% age 31-40, 16.7% age 41-50, 19.8% age 51 or older). The survey was housed on the Qualtrics online platform. Subjects first completed standard demographic questions. As described in the main text, we then measured subjects self-reported thinking styles using items from the Rational Experiential Inventory, as well as beliefs about deterrence using the resolve and domino items from the militant internationalism scale. The main text presents the relevant item wording. Figure A1 displays the distributions of the Faith in Intuition and Need for Cognition factor scores. Previous research indicates that the distribution of need for cognition in MTurk samples is similar to high quality online samples, notably the American National Election Studies (Berinsky, Huber and Lenz, 2012, 359-360).

A2.1 Regression Results

Table A4 presents the results of this correlational analysis, estimated using linear regression. As described in the main text, the outcome measures capture five-point agreement with the resolve and domino items from the militant internationalism scale. We represent Faith in Intuition and Need for Cognition using unidimensional factor scores. As an example of interpretation, a one standard deviation increase in intuitive thinking corresponds to a 0.24 increase in the belief that the US must demonstrate its resolve so that others do not take advantage of it.
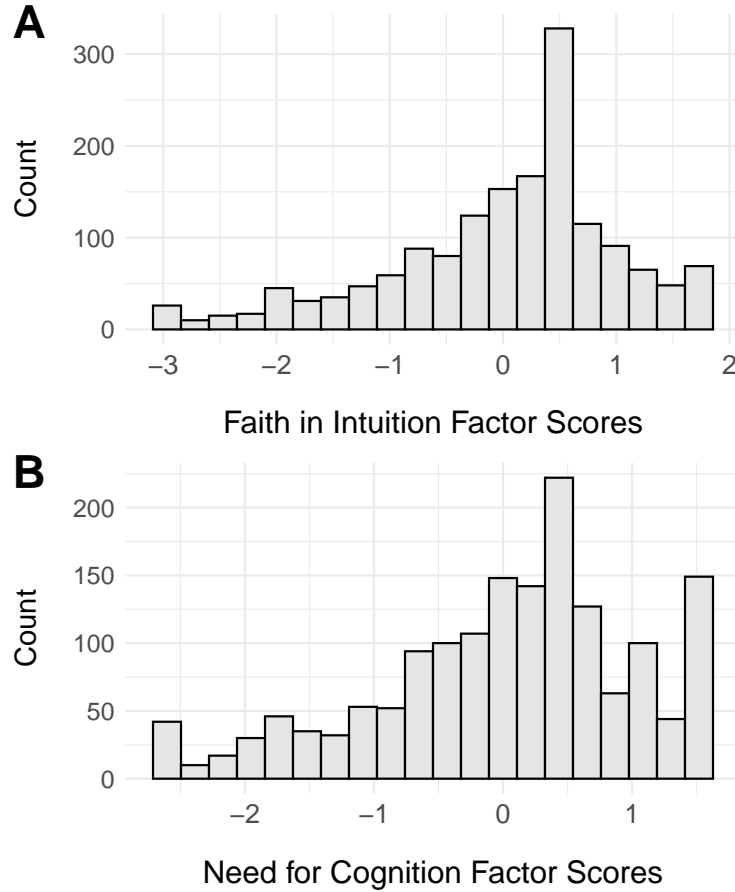
Figure A1: *Rational Experiential Inventory Factor Scores.* Unidimensional factor scores for the Faith in Intuition and Need for Cognition scales in our MTurk sample. Sample consists of $N = 1,615$ US adults.

A2.2 Low-Quality Survey Responses Removed

In line with best practices for online survey sampling, particularly on the MTurk platform, here we verify the robustness of our results with potentially low quality responses removed. Recent work suggests that potentially low quality responses on MTurk can be traced to respondents outside of the United States who use virtual private servers to mask their locations (Kennedy et al., 2020). Such subjects have been shown to provide substantially lower quality responses, often aided through the use of semi- or fully-automated code to complete surveys (also known as "bots").

To check the quality of our sample, we use the IPHub online platform — a commonly used IP verification database — to screen subjects' internet protocol (IP) addresses and flag respon-

Table A4: Faith in Intuition Associates Positively with Deterrence Beliefs

	Must Demonstrate Resolve	Validity of Domino Theory
Intercept	3.13***	3.20***
	(0.11)	(0.10)
Faith in Intuition	0.24***	0.17***
	(0.03)	(0.02)
Need for Cognition	−0.04	−0.00
	(0.03)	(0.02)
Male	0.17**	0.07
	(0.05)	(0.05)
Age	0.01***	0.01**
	(0.00)	(0.00)
White	−0.22***	−0.15**
	(0.06)	(0.06)
College Educated	0.07	0.12*
	(0.05)	(0.05)
Republican	0.09*	0.06
	(0.04)	(0.04)
Conservative	0.28***	0.08*
	(0.04)	(0.04)
Isolationist Orientation	−0.05	−0.15***
	(0.03)	(0.03)
R^2	0.18	0.07
Adj. R^2	0.17	0.06
Num. obs.	1613	1613

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

dents who are likely using virtual private servers and/or are located outside of the United States.[3] Of the 1,615 total respondents, this procedure flagged 116 respondents (approximately 7.2% of our MTurk sample) as potential bots. Table A5 re-estimates our correlational results with these respondents removed, showing that our key results for Faith in Intuition and Need for Cognition remain statistically and substantively unchanged. Thus, low quality responses do not appear to affect our conclusions.

[3]See www.iphub.info

Table A5: Faith in Intuition Associates Positively with Deterrence Beliefs (Potential Bots Removed)

	Must Demonstrate Resolve	Validity of Domino Theory
Intercept	3.14***	3.25***
	(0.11)	(0.10)
Faith in Intuition	0.23***	0.17***
	(0.03)	(0.03)
Need for Cognition	−0.04	−0.00
	(0.03)	(0.03)
Male	0.16**	0.08
	(0.06)	(0.05)
Age	0.01***	0.00
	(0.00)	(0.00)
White	−0.23***	−0.13*
	(0.06)	(0.06)
College Educated	0.04	0.10
	(0.06)	(0.05)
Republican	0.11*	0.06
	(0.04)	(0.04)
Conservative	0.30***	0.10*
	(0.04)	(0.04)
Isolationist Orientation	−0.07*	−0.18***
	(0.03)	(0.03)
R^2	0.19	0.07
Adj. R^2	0.18	0.07
Num. obs.	1499	1499

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

A2.3 Assessing Potential Heterogeneity in Intuition

As with our experiment above, our observational survey primarily focuses on the average correlation between intuitive thinking and deterrence beliefs. However, here we also assess whether this correlation varies by individual differences, namely education level and political ideology. Again, because our paper focuses on thinking styles, we check for possible interactions between education and self-reported intuitive thinking. Further, to assess whether conservative individuals disproportionately drive the correlation between intuition and deterrence beliefs, we model an interaction between political ideology and self-reported intuitive thinking. Table A6 reports a lack of interaction between intuitive thinking and education level, suggesting the relationship between intuition and deterrence (including both the resolve and domino items from the MI scale) does not vary by education level. Table A7 reports negative and statistically significant interactions between conservatism and deterrence beliefs. Intuitive thinking positively associates with deterrence beliefs. However, interestingly, the negative interaction coefficient suggests that conservative subjects show a slightly less pronounced relationship between intuition and deterrence, whereas liberals show a slightly *more* pronounced relationship between intuition and deterrence. In short, conservatives tend to show higher levels of deterrence beliefs regardless of intuition level, while intuition tends to reveal stronger beliefs about deterrence among liberals. Either way, it is clear that conservative subjects do not uniquely drive the relationship between intuition and deterrence beliefs.

Table A6: Lack of Interaction Between Intuition and College Education

	Must Demonstrate Resolve	Validity of Domino Theory
Intercept	3.13***	3.20***
	(0.11)	(0.10)
Faith in Intuition	0.28***	0.15***
	(0.04)	(0.04)
Need for Cognition	−0.04	0.00
	(0.03)	(0.02)
Male	0.17**	0.07
	(0.05)	(0.05)
Age	0.01***	0.01**
	(0.00)	(0.00)
White	−0.22***	−0.15**
	(0.06)	(0.06)
College Educated	0.07	0.12*
	(0.05)	(0.05)
Republican	0.09*	0.06
	(0.04)	(0.04)
Conservative	0.28***	0.08*
	(0.04)	(0.04)
Isolationist Orientation	−0.05	−0.15***
	(0.03)	(0.03)
Faith in Intuition × College Educated	−0.06	0.04
	(0.05)	(0.05)
R^2	0.18	0.07
Adj. R^2	0.17	0.06
Num. obs.	1613	1613

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table A7: Interactions Between Intuition and Political Ideology

	Must Demonstrate Resolve	Validity of Domino Theory
Intercept	3.16***	3.23***
	(0.11)	(0.10)
Faith in Intuition	0.23***	0.16***
	(0.03)	(0.02)
Need for Cognition	−0.04	0.00
	(0.03)	(0.02)
Male	0.17**	0.07
	(0.05)	(0.05)
Age	0.01***	0.01*
	(0.00)	(0.00)
White	−0.22***	−0.15**
	(0.06)	(0.06)
College Educated	0.06	0.12*
	(0.05)	(0.05)
Republican	0.09*	0.06
	(0.04)	(0.04)
Conservative	0.29***	0.08*
	(0.04)	(0.04)
Isolationist Orientation	−0.06	−0.16***
	(0.03)	(0.03)
Faith in Intuition x Conservative	−0.06*	−0.05*
	(0.02)	(0.02)
R^2	0.18	0.07
Adj. R^2	0.18	0.06
Num. obs.	1613	1613

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

A3 Word Embedding Analysis

The embedding analysis uses the global vectors for word representation (GloVe) model to "locally" estimate the embeddings. That is, rather than use a model pre-trained on quotidian texts, like Wikipedia entries and newswire texts, we fit the model to the actual empirical materials of interest.[4] The main text reports nearly all of the relevant details of this analysis. Here, we note further model hyperparameter choices and robustness checks.

The model was fitted using the `text2vec` package in the R statistical programming environment (Selivanov, Bickel and Wang, 2022). We train the embeddings in 300 dimensions with an x_{max} (i.e., maximum number of term co-occurrences used in the weighting function) of 15. We use an algorithm convergence tolerance of 0.05, a learning rate of 0.001, and fit the model over 50 iterations. All of these choices are quite standard in embeddings research (Rodriguez and Spirling, 2022).

Furthermore, we conduct three robustness checks to assess the quality of the fitted embedding space, as well as our dictionary choices. First, we use the `find_nns()` function in the `conText` (Rodriguez, Spirling and Stewart, 2023) package to assess the quality of the estimated embeddings and find that our key terms are close to other reasonable synonyms in the fitted vector space. For example, the term `cooper*` is close to terms like `assist*`, `mutual*`, `consult*`, and `understand*`. Similarly, the term `deter*` is close to terms like `counter*`, `aggress*`, `threat*`, and `prevent*`.

Second, as a check on the validity of our deterrence terms as a measure of deterrence thinking—that is, to assess whether we are measuring what we think we are measuring—we assess the relationship between deterrence terms and a real world adversary. Given that much of Cold War US foreign policy was concerned with the Soviet Union, we assess the relationship between our deterrence (relative to cooperation) terms and the terms `soviet, russian, aggress*` and `hostil*`. The deterrence terms associate more positively (and significantly) with this Soviet threat construct than cooperation terms (cosine sim = 0.09, 90% CI[0.04, 0.13]). This increases our confidence that our deterrence dictionary actually captures deterrent tendencies expressed in these documents.

---

[4]For more on this distinction, see Rodriguez and Spirling (2022) and Pennington, Socher and Manning (2014).

Finally, to ensure that our results are robust to different dictionary term choices, we expand our dictionary with synonymous terms for intuitive versus deliberative thinking. In the main text, we report results for the following terms:

- Intuition: **intuit***, **instinct***, **hunch***, **sens***, **feel***.

- Deliberation: **examin***, **investig***, **calcul***, **study***, **diagnos***.

Here, we expand these dictionaries to include the following synonyms:

- Intuition: **gut***, **inclin***, **sentiment***, **urg***, **impuls***.

- Deliberation: **analyz***, **ponder***, **inspect***, **survey***, **apprais***.

We re-run the analyses presented in the main text with these expanded dictionaries and find substantively identical results. Intuition terms associate positively and significantly with deterrence terms (cosine sim = 0.15, 90% CI[0.11, 0.21]). That is, when US decision-makers use terms that indicate intuitive judgment (e.g., **impuls***), they are more likely to use terms indicative of deterrence thinking (e.g., **deter***). By contrast, again, the use of intuitive terms associates negatively and significantly with cooperation terms (cosine sim = −0.13, 90% CI[−0.08, −0.18]). Said differently, deliberation terms associate positively with cooperation terms. These expanded dictionary results align with the text analytic results presented in the main text, as well as our experimental and observational survey evidence.

A4   Ethical Considerations

The original surveys reported in this paper and appendix were deemed exempt from review by the Institutional Review Board at [university name anonymized]. All subjects completed a consent form with study details, were given the option to withdraw from the study with no penalty, and were presented with a debrief form that described the study's purpose and any experimental deception involved. The debrief form also provided the chance for subjects to withdraw their data from analysis. Subjects opt-in to receive survey invitations from MTurk and Prolific, the two subject recruitment platforms used in the paper's original surveys. All subjects completed the studies anonymously.

All samples were convenience samples. The models reported in the paper and appendix do not use weighting. In terms of question order and survey flow, the paper's original surveys consisted of (1) a consent form, (2) demographic and individual-level measures, (3) random assignment to experimental conditions if relevant, (4) post-treatment measures of dependent variables as relevant, and (5) a debrief form.

References

Berinsky, Adam J, Gregory A Huber and Gabriel S Lenz. 2012. "Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk." *Political analysis* 20(3):351–368.

Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Philip D Waggoner, Ryan Jewell and Nicholas JG Winter. 2020. "The shape of and solutions to the MTurk quality crisis." *Political Science Research and Methods* 8(4):614–629.

Kertzer, Joshua D, Kathleen E Powers, Brian C Rathbun and Ravi Iyer. 2014. "Moral support: How moral values shape foreign policy attitudes." *The Journal of Politics* 76(3):825–840.

Pennington, Jeffrey, Richard Socher and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pp. 1532–1543.

Rodriguez, Pedro L and Arthur Spirling. 2022. "Word embeddings: What works, what doesn't, and how to tell the difference for applied research." *The Journal of Politics* 84(1):101–115.

Rodriguez, Pedro L, Arthur Spirling and Brandon M Stewart. 2023. "Embedding regression: Models for context-specific description and inference." *American Political Science Review* 117(4):1255–1274.

Selivanov, Dmitriy, Manuel Bickel and Qing Wang. 2022. *text2vec: Modern Text Mining Framework for R*. R package version 0.6.3.
  URL: *https://CRAN.R-project.org/package=text2vec*