THE UNIVERSITY OF CHICAGO

FANATICS, FOOLS, AND MADMEN:
HOW PERCEPTIONS OF BIAS, INCOMPETENCE, AND IRRATIONALITY IMPACT INTERNATIONAL POLITICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF POLITICAL SCIENCE

BY
NICHOLAS CAMPBELL-SEREMETIS

CHICAGO, ILLINOIS
AUGUST 2020

For my father, Bruce:
He thought I was crazy, we fought constantly, and I loved every minute with him
He never stopped believing

Fanatics, Fools and Madmen
How Perceptions of Bias, Incompetence and Irrationality Impact International Politics

Nicholas Campbell-Seremetis

Table of Contents

List of Figures……………………………………………………………………………..vi

List of Tables………………………………………………………………………………vii

Chapter 1: An Introduction to Perceived Irrationality in International Politics 1

Extant Work: Perceived Irrationality in Political Science:...3

Is Madness in the Eye of the Beholder, or, what is Perceived Irrationality Anyway?:.............11

Beliefs About Judgment: The Conceptual Spectrum 14

Argument, Dependent Variable, and Scope Conditions:..23

Plan of the Dissertation: 30

Chapter 2: Theory: Unpacking Perceived Irrationality and its Impact on Conflict Diplomacy................. 35

Beliefs About Judgment: A Spectrum for Perceived Irrationality:...39

Diplomatic Bargaining Preferences: Definition and Scope:..61

Beliefs about Judgment and Diplomacy: A Theory of Systematic Preference Shifts: 65

Observable Implications, Caveats and Conditions of the Theory:..83

Diplomatic Bargaining Preferences: Alternative Explanations in Existing Work:.................103

Conclusion: 111

Chapter 3: An Experimental Test of the Consequences of Perceived Irrationality 116

Existing Experimental Evidence: 120

Why Survey Experiments and why Domestic Public Preferences?: 127

Hypotheses: 129

Alternative Hypotheses: 133

Results 1: Beliefs About Judgment and Diplomatic Bargaining Preferences in a Hypothetical Scenario: 142

Experimental Design: Study 2: 150

Results 2: Beliefs About Judgment in the US/North Korea Nuclear Crisis: 152

Discussion: 154

Conclusion: Implications and Next Steps: 158

Chapter 4: The Korean War: 167

The Korean War: A Brief Timeline of The Forgotten Quagmire: 170

US Interests and Capabilities in Korea: 171

Same War, Divergent Diplomacy: The Principle Puzzles in the US Approach: 175

Alternative Theories and Arguments: 178

The Explanatory Utility of Beliefs About Judgment: 185

Communist China: The Irrational Foe: 187

The USSR: The Reasonable Adversary: 199

A Note on North Korea: The Neglected Adversary: 207

Conclusion: 210

Chapter 5: Madness in the Eye of the Beholder: How Individuals Form Beliefs about the Rationality of Foreign Leaders 214

How Beliefs About Judgment Form: Explanations from Extant Literature: 217

Methods Outline: 229

Survey 1: Choice-Based Conjoint Design: 234

Survey 2: Multivariate Factorial Experiment: 242

Survey 2 Results: 249

Discussion: 254

Chapter Six: Conclusion: Implications, and Directions for Future Work 258

Theoretical Implications: 263

Policy Implications: 267

Avenues for Future Work: 268

BIBLIOGRAPHY 272

Appendices: Appendix A: 'Consequences' Survey Design: North Korea Version: 278

Appendices: Appendix B: Survey 1 (Chapter 5) 320

Appendices: Appendix C: Survey 2 (Chapter 5): 331

List of Figures

Figure 1.1 IDEAL TYPE BELIEFS ABOUT JUDGMENT………………………………..17

Figure 1.2 IDEAL TYPES, TRAITS, AND LEADERS…………………………………..21-22

Figure 2.1 IDEAL TYPE BELIEFS ABOUT JUDGMENT………………………………..54

Figure 2.2 PUBLIC VS ELITE SHIFTS………………...……………………………. ..82

Figure 3.1 SURVEY FLOW…………………………….…………………………..141

Figure 3.2 MEANS AND CONFIDENCE INTERVALS………………………………….144-5

Figure 3.3 HISTOGRAMS OF EFFICACY MEASURES…………………………………147

Figure 3.4 THE ELITE/PUBLIC CYCLE……………….……………………………………163

Figure 4.1 SUMMARY OF BELIEFS, PREFERENCES, AND BEHAVIORS…………...187

Figure 5.1 SAMPLE PAIR OF LEADER PROFILES…………………………………...240

Figure 5.2 TREATMENT EFFECTS ON RATIONALITY RATING……………………. 252

Figure AB.1 AVERAGE MARGINAL UTILITY GRAPHS………………………………321-3

Figure AB.2 RELATIVE UTILITY GRAPHS ……………………………………………324

Figure AC.1 REGRESSION TABLES…………………………………………………363-8

List of Tables

Table 3.1 DESCRIPTIVE RESULTS OF STUDY 1, MAIN……………………………………143

Table 3.2 DESCRIPTIVE RESULTS OF STUDY 1, INTERACTION………………………143

Table 3.3 DESCRIPTIVE RESULTS OF STUDY 1, SSI……………………………………..150

Table 5.1 CONJOINT STUDY TREATMENTS…………………………………………237-9

Table 5.2 OLS ON RATINGS OF LEDER RATIONALITY……………………………………250

Table 5.3 OLS ON RATINGS OF LEDER RATIONALITY (LIBERALS)…..………………251

Chapter 1: An Introduction to Perceived Irrationality in International Politics

That saying, 'it's the quiet ones you gotta watch:' I don't know about that. If you're sitting in a café with two guys, and one if them is sitting quietly and minding his own business while the other one is banging on a table with a knife and screaming "I'll kill the next person who comes in here!' then I ask you: who you gonna watch? -George Carlin

Speaking of Adolph Hitler on September, 1938, Neville Chamberlain lamented that "the fate of millions depends on one man and he is half mad,"[1] and that same month British ambassador warned his American counterpart that Hitler may "have crossed the borderline into insanity." At the height of the Cold War, US officials worried that while former Soviet leaders had been as chess players, Nikita Khruschev was "drunk constantly,"[2] "emotionally unstable," and "volatile and unpredictable:" less a chess player than an irrational tantrum thrower. Privately and publicly, many in the George W. Bush administration called Saddam Hussein a dangerous madman,[3] a perception not unique to Saddam among the club of so-called 'Rogue Leaders;'[4] Kim Jong-Un has been labeled everything from "crazy fat kid" to "incompetent clown," and unnerved much of the world by trading insults with President Donald Trump over the size of their respective nuclear buttons. Meanwhile, while Trump infamously brags that he is a "very stable genius," his critics at home and many counterparts abroad seem convinced that he is anything but.[5] These leaders all have at least one important thing in common: they are, or were, seen by some as biased, delusional, erratic, incompetent, or not strictly speaking rational.

---

[1] David Faber, *Munich, 1938: Appeasement and World War II* (New York: Simon & Schuster, 2008), 251

[2] Fursenko and Naftali,,*Khrushchev's Cold War*, 25

[3] Chaim Kaufmann,. "Threat inflation and the failure of the marketplace of ideas: The selling of the Iraq war." *International Security* 29, no. 1 (2004): 5-48.

[4] Kenneth Pollock. *The Threatening Storm: What Every American Needs to Know Before an Invasion in Iraq*. Random House, 2003.

[5] Examples from US newspapers are too numerous to meaningfully cite. The following two were the first resilts of a google search of the term "trump is crazy newspaper articles:" Jennifer Rubin. "Stop the Crazy." *The Washington Post.* August 26, 2019. Greg Sagent. "Trump's Unhinged Rant About a New Attack ad Shows his Weakness." *The Washington Post*. May 5, 2020.

According to Schelling's famous argument about the 'rationality of irrationality,' the apparent irrationality ascribed to these leaders should be a great asset to them in international politics, particularly in violent as well as non-violent interstate negotiations.[6] According to the 'rationality of irrationality,' a madman image enhances the credibility of threats thereby making it easier to coerce or deter counterparts without resorting to actual violence, disincentivizes adversaries from making threats or using force against the apparent madman, and provides many other boons besides. This is a widely accepted argument not just in academic, journalistic, or policy-wonk circles, but among some of the very state leaders to which it refers: Richard Nixon is infamous for his subscription to what he told his secretary of state he called "the Madman Theory, Bob. I want the North Vietnamese to believe I've reached the point where I might do anything to stop the war... and Ho Chi Min himself will be in Paris in two days begging for peace." Trump openly subscribes to the idea of acting erratic, unpredictable, or downright crazy to keep negotiating counterparts on their toes, and it is not implausible to infer at least an element of intentional performativity in the personas of the Kims. When the leaders of nuclear-capable states are modeling their behavior based on a theory, it seems prudent to be certain that the theory in question is accurate.

The issue is that as it stands such certainty is beyond our grasp. The rationality of irrationality is theoretically underdeveloped in the context of international politics: despite its import and ubiquity, Schelling devoted little space to elucidating the logic of his argument and none to testing its veracity with anything but anecdotal evidence, if that. Scholars of International Relations (IR) require a more detailed theory of the international political consequences of perceived irrationality: one that is not only logically consistent but supported by empirical evidence. This dissertation meets that requirement.

---

[6] Thomas C. Schelling. *Arms and influence*. Yale University Press, 2020.

I argue that perceived irrationality functions precisely counter to the expectations of the 'rationality of irrationality' theory: apparently mad leaders are more likely to be subject to threats or force and to experience unfavorable outcomes in interstate negotiations. Furthermore, I argue that a dichotomous conception of perceived irrationality is unhelpful and not reflective of reality, and that to gain useful insight scholars must break perceived irrationality down into its constituent dimensions; after all, even if Hitler, Khruschev, Saddam, Kim Jong-Un, and Trump are or were in some way all seen as irrational, the precise contours of this perception differed between these leaders.

Before outlining the theoretical and empirical contributions of this dissertations, I first discuss the existing treatment (or lack thereof) of perceived irrationality and its consequences in IR.

Extant Work: Perceived Irrationality in Political Science:

Apparently insane leaders feature frequently in both contemporary politics and historical accounts of interstate relations: from the Mad King George III against whom the American colonies rebelled, to Kaiser Wilhelm and Hitler, all the way up to modern speculation about the mental faculties of President Donald Trump and Supreme Leader Kim Jong-Un. Yet despite the ubiquity of this phenomenon, scholars of political science and international relations have devoted minimal time, effort, and ink to developing detailed theories on how beliefs about the madness or sanity of foreign leaders impacts international politics. However, theorization is at least somewhat present, most famously in Thomas Schelling's musings about the rationality of irrationality and the writings of scholars, policy-makers, and journalists who reference Schelling or take inspiration from his ideas.

The same cannot be said of published empirical research on the interstate implications of perceived irrationality, which was completely non-existent until 2019, when a small group of scholars (with whom I was in consultative contact) began attempting to fill this void. Below, I outline the three principle theoretical schools that either speak to, or notably ignore, the subject of perceived irrationality and its consequences in international politics: Schelling's 'Rationality of Irrationality,' the 'misperception' literature pioneered by Robert Jervis, and the slate of contemporary political psychology research collectively known as the behavioral revolution in international politics.

Madman Theory and the 'Rationality of Irrationality:'

Thomas Schelling's "rationality of irrationality" is one of his better-known contributions to international relations, despite occupying a scant eight pages of his book *Arms and Influence*. Schelling posited that a 'madman' image helps state leaders to accomplish several desirable ends. First, it enhances the credibility of their own compellent or deterrent coercive threats by convincing observers that their behavior was apt to be self-destructive, unpredictable, or at best not strictly speaking rational.[7] Second, it inures them to compellent or deterrent coercive threats by making these seem more likely to fail or be counterproductive; apparently irrational leaders might have an insurmountable cost-tolerance or be prone to respond to threats with self-destructive escalation. Therefore, adversaries would be less likely to threaten or use force against, and more likely to concede to the demands of, apparently irrational leaders in the context of international crises. In Schelling's thinking, frequently cited in modern journalistic and policy circles, an actor that convinces others it is irrational cannot be expected to accurately appraise the constraints of its situation or the likely costs of its action, nor to recognize the best strategy to attain its goals. Therefore, its behavior cannot be predicted by observing elements of the external situation,[8] nor its actions influenced in reliable directions by manipulating the situation, rendering coercion (deterrent or compellent) ineffective.

---

[7] Thomas C. Shelling. *Arms and Influence: With a New Preface and Afterword*. Yale University Press, 2008, ch. 2 esp pg. 37-49

Despite its fame, Schelling's work provides little beyond anecdotes and logic without evidence to support his theoretical claims. Neither the empirical record[9] nor relevant studies in psychology[10] corroborate this direct relationship between assessments of adversary rationality and proclivity towards coercive threats and compromise in interstate conflict. In sum, Schelling's treatment of the strategic consequences of perceived irrationality is cursory and empirically unsupported. Moreover, the utility he ascribes to a madman image is limited to bargaining under a very particular distribution of the “power to hurt,” its veracity remains untested, and the potential implications of madman images beyond this manifestation of coercive strategic interaction go unexplored.

---

[8] Schelling introduced the idea of the rationality of irrationality in the narrow context of nuclear compellence, but did not address it in other situations nor elaborate upon the mechanisms by which one might credibly signal irrationality, here understood as a complete disregard for or insensitivity to costs. Since Schelling, the concept has occasionally been referenced but never further explored.

[9] Nixon famously tried and failed to convey the impression that he was an irrational 'madman.'

[10] For example racial-stereotypes and facial features have been shown to impact assessments of competence in both individual and inter-group contexts. For how facial features impact assessments of competence or similar characteristics see for example Chappell Lawson, Gabriel S. Lenz, Andy Baker, and Michael Myers. "Looking like a winner: Candidate appearance and electoral success in new democracies." *World Politics* 62, no. 4 (2010): 561-593, or Panu Poutvaara, Henrik Jordahl, and Niclas Berggren. "Faces of politicians: Babyfacedness predicts inferred competence but not electoral success." *Journal of Experimental Social Psychology* 45, no. 5 (2009): 1132-1135. For the link between racial stereotypes and competence assessments see a wide range of work: Susan T. Fiske,"Stereotyping, prejudice, and discrimination at the seam between the centuries: Evolution, culture, mind, and brain." *European Journal of Social Psychology* 30, no. 3 (2000): 299-322, Lisa D. Cothran, "Facial affect and race influence threat perception." *Imagination, Cognition and Personality* 30, no. 3 (2011): 341-354, or for a more general look such work as Gordon W. Allport's *The nature of prejudice*. Basic books, 1979.

Moreover, Schelling's definition of perceived irrationality is fuzzy at best and non-existent at worst: he frequently moves between metaphors and anecdotes that respectively differ in their suggestion of what it mean for a leader to appear irrational. For example at one point in his discussion of Khruschev he suggests Khrusvhev seemed irrational because he appeared to base decisions on emotion rather than reason, while a different anecdote implies it was Khruschev's erratic behavior rather than his emotional decision-making that made him an apparently irrational head of state.[11] At still other points Schelling discusses apparent irrationality as: insurmountable cost-tolerance, insatiable aggression or desire for conquest, or a puppy's inability to understand human speech or behavior. The specific contours of each of these conceptions, the lines between them, and the mechanisms through which they are supposed to shape international politics are all unclear.

Subsequent scholars have modeled the consequences of including an irrational actor in formal theoretic strategic interaction games, but these scholars do not agree on what a narrowed definition irrationality should be in the context of international politics. For example, they respectively define irrational actors as those that are boundedly rational, always choose the aggressive option, disbelieve unfavorable information, overinflate their expectations of success in a conflict, cease gathering information after a minimal amount of time or effort, and any number of other psychologically non-rational behavioral tendencies.[12] Yet while this definitional divergence is already problematic for the purposes of interrogating or expanding upon the rationality of irrationality, the more pressing issue from the perspective of this project is that research in this vein is all dealing with real rather than perceived irrationality.

---

[11] Schelling 2008

[12] See Jervis 1976, Avidit Acharya and Edoardo Grillo. "War with crazy types." *Political Science Research and Methods* 3, no. 2 (2015): 281-307, Brendan Nyhan and Jason Reifler. "When corrections fail: The persistence of political misperceptions." *Political Behavior* 32, no. 2 (2010): 303-330, and a wide range

A few scholars have recently attempted to unpack or test the rationality of perceived irrationality, but there are limitations to their approaches, which my conceptual and empirical work seek to address. Acharia and Grillo incorporate the possibility that one player is 'crazy,' defined as always making unreasonable offers and taking the aggressive option when presented with a choice, into a conflict bargaining model that includes the possibility of war.[13] Yet this framing of 'crazy' is unrealistically dichotomous, and conflates madness with aggression. McManus takes an approach that is conceptually closer to this project, conducting an empirical examination of how perceived madness affects leaders' international crises outcomes.[14] However she operationalizes perceived madness according to descriptions of leaders as 'crazy' by media, and therefore cannot distinguish between media rhetoric and the beliefs of key decisionmakers. Nor can she establish a nuanced conception of perceived madness, a shortcoming she acknowledges and attempts to address in a related paper where she proposes a typology of perceived madness that, while an important step, is conceptually constrained. Her typology concerns itself only with perceived madness, so it neglects the vital task of theorizing perceived madness in relation to perceived rationality: I address this issue by conceptualizing perceived madness as part of a larger typology of beliefs about judgment that includes various manifestations of perceived rationality. Finally, this work does not establish empirical microfoundations for the interstate impacts of perceived irrationality, an issue for a thoroughly psychological phenomenon.

---

[13] Avidit Grillo 2015.

[14] McManus, Roseanne W. "Crazy Like a Fox? Are Leaders with Reputations for Madness More Successful at International Coercion?." *British Journal of Political Science* (2019): 1-19.

‘Misperception’ and the Behavioral Revolution:

There is a robust scholarship critical of the 'rational actor assumption'[15] in IR, the modern manifestation of which can be traced to psychological work on biases and misperceptions in human judgment which gained traction in IR through the pioneering work of Robert Jervis. Jervis’ work focuses on actors' misunderstandings of the actions and intentions of their counterparts,[16] and subsequent work examines the roots and effects of a range of biases that invite actors to embrace erroneous beliefs about the world.[17] Yet the integration of this work into broader strategic frameworks was initially limited by a “failure to identify scope conditions” that made it “difficult to establish which bias would matter when and to specify its impact on choice.”[18] As a result, this foundational literature provides little theoretical or empirical bases for thinking about the impacts of perceived irrationality in international politics.

In fact, this foundational theoretical literature on misperception holds that a major source of error in interstate relations is that states always expect counterparts to perceive the world as they do: in other words, many poor or inexplicable decisions in international politics are the product of a mirror imaging fallacy. Jervis laments that decisionmakers see international politics as chess when in fact it is rashomon, a point he makes to assert that many problems could be solved if decisionmakers believed their counterparts' judgment differed from their own.[19] Yet Jervis' prescription explicitly roots such higher order awareness of heterogeneous perceptions in empathic perspective taking. This fails to capture that such awareness can carry connotations of perceived irrationality, and so introduce rather than eliminate problems in interstate relations. Romanticizing greater awareness of heterogeneous judgment as a way to overcome the *rashomon effect* illustrates just one of the issues with extant IR scholarship's cursory treatment of perceived irrationality and its consequences.

---

[15] In its strongest form this assumption treats states are risk-averse or neutral actors that seek out all available information on their situation, use it to update their beliefs in an unbiased way, and strategically choose the course of action that will realize the best possible outcome with respect to their subjective goals.

[16] This is the literature on “Misperception” most frequently associated with the work of Robert Jervis, *Perception and Misperception in International Politics.*

[17] See for example Jervis, *Perception and Misperception in International Politics*, Jack Snyder, *Myths of Empire*, cannonical works on the “cult of the offensive,” and work on the role of heuristics, metaphors, and analogies in shaping beliefs about the world.

[18] Janice Gross Stein, “The Micro-Foundations of International Relations Theory: Psychology and Behavioral Economics.” *International Organization* 71, Supplement 2017, p. S250.

[19] Jervis 2017

However, a recent 'behavioral revolution' in IR scholarship has taken steps to correct for the empirical as well as theoretical shortcomings of the work, by Jervis and others, that originally attempted to integrate psychology and international relations. This behavioral revolution traces its roots to pioneers of experimental IR like Rose McDermott[20] and Rick Hermann, and has since expanded to encompass a wide range of research covering myriad issue areas, including by scholars including Joshua Kertzer, Jonthan Renshon, and Mike Tomz. What unites scholarship that falls under the heading of the behavioral revolution is general approach and end goals. It is comprised of research that uses experimental methods to establish robust scope conditions, either in the environment or the heterogenous traits of actors within it, for various biases and so determine when decision makers should be expected to behave in line with behavioral, rather than rationalist, theories.

This research covers numerous manifestations of psychological heterogeneity, such as how actors perceive situations,[21] form beliefs,[22] or make choices,[23][24] across a massive range of international issue areas: the use of force, crises escalation, diplomatic style, cooperation, trust, reciprocity, economic policy, and more. Crucially for this project, "one of the key insights of the behavioral revolution has been to emphasize how actors perceive the situations they face."[25] In contrast to approaches that reluctantly acknowledge heterogenous actors only to explain what environmental factors cannot, "behavioral approaches emphasize the extent to which situational factors are indeterminate; dispositional differences matter not just because they reduce unexplained variance, but because *they affect how actors perceive and define the situations they face."*[26] In sum, the behavioral revolution stresses the importance of heterogeneity in the judgment of international actors; many of its key contributions emphasize that judgment under conditions of uncertainty is an inherently subjective process, one that is susceptible to heuristic biases[27] and shaped not only by current information but also past experience and an actor's body of knowledge. It is entirely possible, and indeed a frequent occurrence, for different actors to reach different conclusions about the likelihood of the same outcome on the basis of the same information because they perceive the information in fundamentally different ways, making the expectation of convergent judgments implicit in rationalist theories problematic.[28]

---

[20] Rose McDermott's "Experimental methods in political science." *Annual Review of Political Science* 5, no. 1 (2002): 31-61, remains a gold standard piece on the promises and pitfalls of experiments in political science and IR. For more work in this vein see e.g. McDermott, Rose. "The feeling of rationality: The meaning of neuroscientific advances for political science." *Perspectives on politics* 2, no. 4 (2004): 691-706. Abdelal, Rawi, Yoshiko M. Herrera, Alastair Iain Johnston, and Rose McDermott. "Identity as a Variable." *Perspectives on politics* 4, no. 4 (2006): 695-711.

[21] Herrmann 2017

[22] Yarhi-Milo 2018; Ryan and Kertzer 2018

[23] Kertzer 2017; Renshon, Lee and Tingley 2017

[24] Rathbun 2014

However, like the misperception literature that preceded them, scholars in the behavioral revolution have not examined the idea that actors are aware of psychological heterogeneity or addressed questions about the effects of this belief, accurate or not, that other actors or perceive situations or construct action plans differently. In other words, misperception is sometimes acknowledged but *perceived misperception* has been ignored.[29] Theories of strategic interaction implicitly assume the players have judged each other 'gameworthy;'[30] though some scholars have addressed why 'common knowledge' may not in reality hold, the assumption remains that interacting states have a mutual expectation of common knowledge. In sum, existing work in international relations neither acknowledges attributions of bias or incompetence nor examines their consequences in international relations.

---

[25] Joshua Kertzer, "Resolve, Time, and Risk." *International Organization* 71, Supplement 2017, p. S118

[26] Kertzer p. S131, emphasis mine

[27] Anchoring, Representativeness, and availability being the mot common

[28] Though this is *exactly* what the bargaining model of war expects, as it explicitly does not allow for uncertainty

[29] As extrapolated upon later, this issue *has* received attention in cognitive and social psychology. Some of the labels assosciated with perceived misperception are: bias attribution, the bias blind spot, naïve realism, bias-perception. See the work of Emily Pronin and Lee Ross.

Is Madness in the Eye of the Beholder, or, what is Perceived Irrationality Anyway?:

Before theorizing and testing the international consequences of perceived irrationality, we must first grapple with two of the aforementioned shortcomings in extant literature's definition of 'perceived irrationality.' First, the lack of clarity on what 'perceived irrationality' is, and by extension what it is not. Second, the treatment of perceived rationality and irrationality as a mutually exclusive dichotomy: actors are either perceived to be rational or irrational. These are problematic shortcomings of a theory dealing with a concept like irrationality; the concept lacks a self-evident 'common-language' definition and is inherently subjective, making it apparently ill-suited to a simplified binary treatment.[31] Moreover, 'irrationality' has been incorporated into a wide array of academic disciplines and subfields, each of which takes a particular view on the meaning of the concept that coheres with the needs of the field in question but often contradicts how other disciplines understand the term. In sum, a failure to explicate and justify what is meant by irrationality in a given theoretical context is a recipe for confusion. Therefore, for this project and the broader field of International Relations,

---

30 Goffman, *Strategic Interaction* pg. 97-98

[31] Indeed, the current reality is that most social sciences deal in some way with human irrationality but there is considerable divergence on how irrationality is in defined not only across but also within disciplines: often it is left as a fuzzy conceptual heading that subsumes whatever connotations are useful to a particular author's theory. Because irrationality is rarely the focus, the drawbacks of this definitional divergence are less readily apparent.

it is important to nail down what it means for one actor to perceive a counterpart to be irrational *before* we consider the consequences of that perception in any area of international politics.

For theoretical consistency, utility, and realism, it is important that we do not think about only about a single ideal type of perceived madness, or even multiple types of perceived madness. First, this would be such an oversimplification as to be empirically useless. People do not think of counterparts as either mad or sane with nothing in between. Second, restricting the scope to perceived irrationality alone is lazy and theoretically suspect. It obviates the need to posit a complete, consistent theory that encompass not only the impacts of perceived irrationality but also those of its absence, namely perceived rationality; the mechanisms used should to explain the consequences of the former should be transferable to the latter. Moreover, theorizing only one end of a spectrum is not a shortcut without drawbacks: it almost inevitably makes for a less complete or convincing argument, since it increases the potential for leaps, inconsistencies, and gaps in logic. Simply put, perceived irrationality is not an independent phenomenon but one extreme end of a spectrum: to build a theory that incorporates perceived irrationality we need to identify and address the spectrum on which it sits.

However, there is not a single spectrum into which perceived irrationality obviously fits, so defining that spectrum is not a given but a choice, and like any choice it has consequences for the scope, implications, and methods of testing a theory. The following are some examples of spectra according to which perceived irrationality could be, and has been, justifiably conceptualized:

- **Cost-Tolerance:** If irrationality is conceptualized on the cost-tolerance spectrum, then an apparently irrational actor is one with an insurmountable tolerance for costs, such that they will be destroyed before they experience costs they consider too painful to bear.

- **Preferences and Goals:** If irrationality is conceptualized on the spectrum of preferences and goals, then an apparently irrational actor is one with preferences that run counter to its 'objective' best interests: actors that are actively self-destructive or suicidal are the most extreme manifestation of this. However, this conceptualization of irrationality can and has also, for example, been used to characterize states with limitless appetites for territory (which therefore will not cease in attempts at conquest until they are either destroyed or rule the world) or voters who support trade policies that work against their economic self-interests. Put another way, conceptualizing irrationality on the spectrum of preferences and goals makes perceived irrationality equivalent to the belief that a counterpart has goals that, 'objectively,' it should not have. Alternatively, irrationality might be attributed to actors that have goals which seem so morally reprehensible as to be apparently insane.
- **Aggressiveness:** If irrationality is conceptualized as an extreme end on the spectrum of aggressiveness, then an apparently irrational actor is one that is expected to always act aggressively regardless of circumstance. Depending on the nature of the interaction in question, this could mean always resorting to violence (e.g. interstate or intergroup armed conflicts, violent one-on-one fights, or any other interaction that involves physical violence), always escalating interpersonal arguments, or otherwise choosing to escalate rather than deescalate when the opportunity presents itself.
- **Predictability:** If irrationality is conceptualized as an extreme end on the spectrum of predictability, then an apparently irrational actor is one that is expected to act as-if randomly.

The above are illustrative but not exhaustive of the various spectra on which perceived irrationality has been conceptualized in academic, journalistic, and literary work. It is, of course, not feasible to explicitly include all of these as main dimensions of an even somewhat parsimonious theoretical concept. However, we should seek to place perceived irrationality on a spectrum that has feasible observable implications for as many of the aforementioned variables (Cost-Tolerance, Preferences/Goals, Aggressiveness, Predictability) as possible.

Beliefs About Judgment: The Conceptual Spectrum

I conceptualize perceived irrationality on the spectrum of what I term '*Beliefs about Judgment*.' Below, I provide a basic outline of this conceptual spectrum, including the academic work on which it is based, the components of its constituent dimensions and the four ideal-type actors formed by their combinations, and how variation on these dimensions impacts perceiving actors' diplomatic preferences. In Chapter 2, I detail in-depth my conceptual typology and my theory of how it shapes choices and preferences regarding diplomacy, coercion, and violence in interstate conflict.

Judgment and choice "are related but not interchangeable"[32] components of decision-making: when I refer to beliefs about an actor's judgment, I refer to beliefs about how the actor thinks. Observers may (accurately or not) perceive variation in judgment between themselves and others as the reason a counterpart's behavior diverges from their own. Judgment consists of the interpretation of information into beliefs, the construction of a set of possible action plans in light of said beliefs, and the assignment of relative likelihoods to the possible outcomes of each action plan. Put another way, judgment processes are those by which actors answer the following questions:

---

[32] McDermott, Rose. *Risk-taking in international politics: Prospect theory in American foreign policy*. University of Michigan Press, 2001.

1) What is the world or situation I am in?
2) What are my possible plans of action, or strategies, in this situation?
3) What are the possible outcomes of each strategy or action plan, and what is the likelihood that each outcome manifests?

Notably, judgment does not include the assignment of value to possible outcomes (e.g. "how much do I want each possible outcome?") nor the choice of action on the basis of this valuation ("what will I do given the cost/benefit trade-off between the risk and reward of each plan of action?"). In other words, judgment comprises the processes by which actors perceive the situations they face and construct menus of options within those situations, but not the processes by which actors decide what they want.

My typology of *beliefs about judgment* has two dimensions: *perceived objectivity* and *perceived competence*, each of which denotes a spectrum along which observers' beliefs about the judgment of their counterparts can fall. The first is *perceived objectivity* (objective/biased): whether the other's perceptions and interpretations of information are believed to be objective or distorted by bias. The second is *competence* (competent/incompetent): whether the other is believed to have the experience or skill to translate preferences into policy.[33] These traits are conceptually linked because each 1) influences how an actor is expected to move from information about the external situation to choice but 2) is independent of values, goals, or desires. In sum, they indicate beliefs about how others think but not what others want. Actors believed to be sufficiently biased *and* incompetent could be characterized as expected to behave in ways that fulfill the criteria typically associated with irrationality: insensitive to costs, pursuing seemingly senseless goals, escalating even conflicts they cannot win, and otherwise behaving in ways that observers would characterize as erratic. This conceptualization draws from psychological work on 'mind reading', or the way people intuit the mental states of others, and naïve realism, a phenomenological stance that posits that people assume their judgment constitutes the standard of objectivity.[34]

---

[33] Elizabeth N. Saunders 2017.

As discussed above, there are theoretical advantages to placing perceived madness on this more complete spectrum of perceived judgment. Because this spectrum includes the 'rational' value of both its constituent dimensions, when theorizing the effects of variation within the spectrum we cannot consider *only* the effects of madness, as Schelling and other previous scholars have done.[35] A theory that posits international political consequences of perceived bias must also cogently do so for perceived objectivity, and an argument about how apparent incompetence impacts a given facet of interstate relations also needs to discuss how perceived competence impacts that same facet. Put another way, by necessitating the consideration of effects and mechanisms at both ends of the rationality spectrum, this approach facilitates the construction of theories that are more complete, internally consistent, and reflective of empirical reality. Furthermore, it helps us to disaggregate which aspects of rationality or irrationality drive effects or are central mechanisms in a given theory, and which aspects are tangential, which we cannot do if thinking about perceived irrationality as a binary variable.

For the purposes of interacting *perceived objectivity* and *perceived competence*, I treat these dimensions as binary: the combination of these dimensions yields four ideal-types (Figure 1.1).

---

[34] See Gilovich, Thomas, and Lee Ross 2015.

[35] Schelling 1968, McMannus 2019 and 2020,

These are objective/competent (*The Rational),* objective/incompetent (*The Fool),* biased/competent (*The Fanatic),* and biased/incompetent (*The Madman)*.

Figure 1.1: Ideal Type Beliefs about Judgment:

		Perceived Competence	
Perceived Objectivity		**Competent**	**Incompetent**
	Objective	The Rational	The Fool
	Biased	The Fanatic	The Madman

The first ideal type is the *Rational*, who the observer expects to accurately interpret information and ably construct optimal strategies to pursue their interests within situational constraints. This is the leader reflected in the rational-actor assumptions of most strategic interaction theories: a perfect player of any game in which they are engaged. The *Rational* accurately interprets information available to them, has beliefs that change in response to new information, and can correctly anticipate the actions of counterparts within the limits of the information available to them: they understand the world they are in and the rules of the game they are playing. Moreover, the *Rational* is able to construct and execute its optimum set of strategic options. President FDR, Chancellor Angela Merkel, Gorbachev, and the common trope of Vladamir Putin the mastermind[36] are all examples of the *Rational.*

---

[36] Those who subscribe to the image of Putin as a mastermind playing three-dimensional chess with his international counterparts

The second ideal type is the *Fool*, who observers expect to accurately interpret information but be incapable of constructing or implementing optimum action plans: this is the bungling leader stereotype. Like the *Rational*, the *Fool* interprets information accurately and updates its beliefs accordingly, yet where the *Rational* is a master the *Fool* is a novice that regularly makes mistakes or fails to recognize certain strategic options, often cannot translate desired outcomes into the best actions to achieve them, and has trouble anticipating the actions of more skilled counterparts. Put another way, the *Fool* understands rules of the game but not its mechanics or strategic nuances.

Following the style of Schelling, we can unpack the *Fool* ideal-type using two popular metaphors for interstate conflict and crises: chess and chicken. In chess[37] the *Fool* sees the board clearly as its opponent, knows the objective (checkmate) and understands *how* the pieces move. Yet it does not grasp the strategic *reasons* to move pieces to particular spots at particular times, so the *Fool's* moves tend to lack rhyme or reason: sacrificing pieces for no strategic gain, leaving valuable pieces open to capture, and generally making 'rookie mistakes.' Likewise, in the game of chicken, the *Fool* is the equivalent of a student driver: they can see the road and accurately perceive the distance between the cars, but that is not to say the will not mistake the gas pedal for the brake or twist the steering wheel the wrong way. This is not, it should be noted, to say that the *Fool* cannot learn or improve, move from incompetent to competent and so become a *Rational,* but this transition requires time, effort, and even if it occurs in reality may not be immediately registered by observers. Images of Boris Yeltsin as a boorish drunk, George W. Bush as a well-meaning but bumbling man who lacked the expertise to be an effective President, and Signman Rhee as a puppet leader incapable of managing his government alone[38], are all examples of the *Fool*.

---

[37] Usually used as a metaphor for international politics writ large

The third type is the *Fanatic,* who observers expect to hold a warped view of reality that means they misperceive situations, information and probabilities, but nonetheless adopt and capably implement ideal strategies within its warped reality; this is the brilliant ideologue. The *Fanatic* is strategically competent and capable of constructing or identifying optimal paths of action to advance its goals within the constraints of the situation it perceives. The *Fanatic's* 'biased' worldview[39] deterministically shapes how it perceives discrete information as well as general situations; therefore, it interprets information in ways outside observers consider warped or simply wrong, seems unaware of situational risks, may not update its beliefs in response to new information, and cannot reliably be made to 'see reason' through information or argument.

Let us again unpack the *Fanatic* ideal-type using chess and chicken, respectively. In chess, the *Fanatic* understands the objective, how the pieces move, and why to move them for strategic gain, but it is looking at a completely different board than its opponent: there is no common knowledge of the board and therefore the state of the game. Even a pair of grandmasters will find it difficult or impossible to play a coherent match without a shared view of the board, of where all the pieces are positioned at any given time. Indeed, this problem of two boards is why Jervis suggests that we should think of international politics not as chess but as rashomon, a game for which divergent perceptions of the playing board is a central feature. In the game of chicken, playing against the *Fanatic* is tantamount to playing against a licensed driver wearing a pair of badly cracked glasses; the *Fanatic* knows how to drive and has every desire to avoid crashing. The issue is that it is uncertain whether the *Fanatic* will accurately perceive the distance remaining between the cars, and so it may not know when to swerve to avoid collision. Images of Chairman Mao as brilliant but blinded by communist dogma, of Adolph Hitler as a volatile man so fervent in his nationalism as to render him incapable of comprehending risks, and of Khruschev as emotionally unstable and prone to erratic outbursts and decisions, are all examples of the *Fanatic.*

---

[38] Indeed, the paternalistic, dismissive image the USA and similar Western nations have frequently assigned to the leaders of underdeveloped or former colonial territories generally fit the *Fool* type.
[39] This warping prism could be a political ideology, a fervent nationalism, or a narcissistic self-assurance to name just a few.

The final type is the *Madman*, who the observer expects both to see the world through a cracked glass and be incapable of formulating intelligent strategies even within their own conception of reality. Neither the *Madman's* beliefs nor the actions it will consider can be ascertained by examination of the situation, since the *Madman* diverges from the observer on every factor relevant to the judgment phase of decision-making. The *Madman's* interpretations of discrete information and perceptions of the general situation seem hopelessly distorted: it is not looking at the same board (chess) or seeing the same road (chicken). The *Madman* has no understanding of strategy, comprehension of how to construct plans of action to realize desired outcomes, or organizational ability to turn intentions into actual actions: it does not know why to move particular pieces to particular places (chess) or which pedal is the gas and which is the brake (chicken). The *Madman* may, in short, not even know whether they are playing chicken, chess, or some other game all together; it may try to drive a car over the chess board, fling chess pieces at its adversary's car, or spray sprinkles at everything in the vicinity while yelling incoherently in a made-up language. This is, as Erving Goffman describes, a "game against nature," or in Schelling's terms a game against a puppy. It is pointless to try and predict a puppy's strategy, to attempt to see the world through its eyes, or to provide it new information or make rationally grounded arguments to get it to alter its behavior: a puppy, like a *Madman*, is going to do whatever it is going to do. Images of Saddam Hussein as the crazed dictator willing to commit suicide to destroy the US, of Kim Jong-Il and his son Kim Jong-Un as opaque eccentrics weirdly obsessed with Dennis Rodman and incapable of anything resembling calculated decision-making, and of Donald Trump as a fragile narcissist with no intellectual capability or curiosity and a penchant for petulant outbursts and short-termism, are all examples of the *Madman*.

Figure 1.2, below, summarizes the traits associated with each ideal type, the effect each ideal type should have on the diplomatic preferences of an adversarial actor, and the names of some real-world leaders who arguably fit the ideal-type in question.

Figure 1.2: Ideal Types, Traits, Predictions, and Leaders:

Ideal Type	**Traits (Non-Exhaustive Examples)**	**Diplomatic Preferences**	**Example Leaders**
The Rational	-*Objective*: • Interprets information 'accurately' (the same as observer) • Aware of risks when decision-making • Updates beliefs to incorporate new information -*Competent:* • High expertise/skill • Organized • Able to translate intention into action	-*De-escalatory Negotiation:* • Informational persuasion, dialogue • High cut-off for compromise in settlements (**less inclined to stand firm absent ideal outcome)**	Vladamir Putin, Mikahel Gorbachev, FDR

The Fool	*Objective*: -Interprets information ‘accurately’ (the same as observer) -Aware of risks when decision-making -Updates beliefs to incorporate new information *Incompetent:* -Low expertise/skill -Disorganized -Unable to translate intentions to action	*-Escalatory Negotiation:* • Informational persuasion, dialogue • Low cut-off for compromise in settlements (**more inclined to stand firm absent ideal outcome)**	Boris Yeltsin, Singman Rhee, George W. Bush
The Fanatic	*Biased*: -Interprets information ‘inaccurately’ (according to observer) -Ignores risk trade-offs in decision-making -Does not update beliefs to incorporate new information *Competent:* -High expertise/skill -Organized -Able to translate intention into action	*De-escalatory Coercion:* -Coercive persuasion, force -High cut-off for compromise in settlements (**less inclined to stand firm absent ideal outcome)**	Chairman Mao, Adolph Hitler, Nikita Khruschev
The Madman	*Biased*: -Interprets information ‘inaccurately’ (according to observer) -Ignores risk trade-offs in decision-making -Does not update beliefs to incorporate new information *-Incompetent:* -Low expertise/skill -Disorganized -Unable to translate intentions to action	*Escalatory Coercion:* -Coercive persuasion, force, ultimatums -Low cut-off for compromise in settlements (**more inclined to stand firm absent ideal outcome)**	Saddam Hussein, Kim Jong-Il and Un, Donald Trump

Treating perceived irrationality as a multi-dimensional variable moves beyond the binary distinction between ‘rational’ and ‘irrational,’ which makes perceived irrationality more theoretically useful by breaking it down into components. Some combination of *perceived* *objectivity* and *perceived competence* captures most common connotations of perceived irrationality. For example, unpredictability results from an observer's uncertainty about one or more components of a counterpart's decision-making process; it is difficult if not impossible to predict an actor's behavior if one has no sense of its strategic knowledge (*perceived incompetence*) or what situation it thinks it is in (*beliefs*). Thus, independently conceptualizing predictability would be tautological. Furthermore, this project is not concerned with another common connotation of perceived irrationality: beliefs that others are evil or intentionally suicidal. These subjective or normative beliefs about the rationality of what others *want* rather than how others *think* are outside the conceptual scope of beliefs about judgment as defined by this project.

Argument, Dependent Variable, and Scope Conditions:

Chapter 2 details in-depth my theory of how actors' beliefs about the judgment of adversary leaders shape their approaches to interstate bargaining in conflict, particularly their preferred approaches in diplomacy. I outline this theory below with a brief discussion of the argument, mechanisms of action, scope and limitations.

Contrary to the 'rationality of irrationality' thesis, I argue that leaders perceived as irrational on either dimension of the judgment spectrum are all-else-equal *more* likely to face coercion or violence, *less* likely to have adversaries concede to their full demands, and likely to get *fewer* of their interests fulfilled in a negotiated settlement. In brief, apparent irrationality is not rational for leaders seeking to avoid violence or enhance their chances of success in the context of interstate conflict or crisis diplomacy, which is the focus of both this project and (at least implicitly) Schelling's original thesis. Diplomacy is an aspect of bargaining wherein actors try to use information and/or argument to persuade a counterpart to behave in desired ways without exerting the necessary power to *force* the counterpart to alter its behavior.[40] Actors who perceive an adversary leader as irrational generally (and biased in particular) are more likely to conduct diplomacy through coercion, threats, and violence than through rational dialogue, information sharing, and argument. Actors who perceive an adversary leader as incompetent are less likely to make concessions or take conciliatory actions,[41] more likely to 'stand firm,' and more likely to exit negotiations and escalate violence if their desired settlement is not achieved.

The logic of this theory is straightforward and driven by two expectation-oriented mechanisms of action. First is the expected persuasive efficacy of information (driven primarily by the perceived objectivity/bias dimension), as information is the primary currency of diplomacy whether it takes the form of rational discussion and argument or coercion. Second is the expected costs of conflict (driven primarily by the perceived competence/incompetence dimension).

Impacts and Mechanisms of Perceived Objectivity and Bias:

Variation on the perceived objectivity dimension impacts actors' diplomatic bargaining preferences through their expectations regarding what we can think of as the *complexity-cut-off* for comprehensible information. Actors with dissimilar ways of thinking are less likely to have convergent interpretations of information that is complex,[42] plausibly ambiguous,[43] or inherently subjective,[44] yet likely to converge in their understanding of information that is simple or part of the 'universal language' of force. Chapter 2 expands upon what makes information complex and subjective vs. simple and universal in an interstate context, but at its most basic 'pure information' is more complex while both costly material actions (e.g. use of force or rewards) or statements of material intent (e.g. threats to use force, promises of reward) are simpler. Simply put, constructive dialogue is complicated, while coercive carrots and sticks are simple. This logic does not imply that simpler information is more effective at persuading adversaries to behave in desired ways: that is a question of their value trade-offs, not the complexity-cut-off for shared information comprehension. Instead it is intended to suggest that when the complexity-cut-off is low an actor can trust that simple but not complex information will convey to the counterpart the consequences of shifting their behavior: "if you do not give me what I want I will hurt you" is a pretty easy message to understand.

---

[40] General definition taken from Rathbun 2014. There are obviously other ways to define diplomacy but for the purposes of this project this is the fundamental conception I have chosen to work with.

[41] Conciliation is here distinct from concessions insofar as the latter implies at least an implicit demand by the adversary

[42] For example, information that requires specific expertise to fully comprehend, a.k.a. information that can only be fully or accurately comprehended by actors with high competence in the associated subject area.

[43] This can be hard to pin down because what one actor considers plausibly ambiguous It also, however, includes 'hard' data the veracity of which can be disputed either on its own evidentiary foundation or via the presentation of counter-evidence

Perceived bias creates expectations of divergent ways of thinking and interpreting information;[45] therefore the higher the level of bias observers ascribe to an adversary leader, the lower the complexity-cut-off for comprehensible information and the more likely said observer is to rely on simple information to affect desired changes in the leader's behavior. By contrast, perceived objectivity creates expectations of convergent ways of thinking and interpreting information; therefore the higher the level of objectivity observers ascribe to an adversary leader, the higher the complexity-cut-off for comprehensible information and the more likely said observer is to first attempt to use complex information and rational argument to affect desired changes in the leader's behavior.[46] Note that conducting diplomacy through simple information is almost always more costly than doing so through complex information. These costs include for example direct material costs (using force, providing incentives etc…), potential reputational costs, and escalation risks. By contrast complex information is effectively costless,[47] so it should at least be the preferred starting approach to diplomacy.

---

[44] For example, purely-opinion based information or presentations of information that frame or re-frame it through argumentation

[45] This conceptualization draws most immediately from psychology research on perceived bias in interpersonal and inter-group relationships. See Pronin 2007 and 2009, Pronin and Kennedy 2009, Kennedy, Kathleen A. *Conflict spirals, bias perceptions, and recommended interventions*. Princeton University, 2010.

Impacts and Mechanisms of Perceived Competence and Incompetence:

Variation on the perceived competence dimension impacts actors' diplomatic bargaining preferences through their expectations regarding the costs and risks of resuming or escalating conflict. A fundamental purpose of diplomacy is to obviate the need to fight a conflict to its conclusion, instead reaching a negotiated settlement that both sides prefer to accept instead of attempting to force a resolution they think is better through continued violence. The incentive to reach diplomatically negotiated settlements comes from the expected costs of continued conflict; fighting is inherently harmful to all participants, so state leaders should prefer to achieve aims nonviolently and weigh any potential benefits they could attain through violence against the expected costs of the violence itself.[48] There are several factors that affect how severe or negligible these costs are expected to be. Relative material endowments in military and economic resources, stakes in the issues at despite,[49] and relative public support for the conflict[50] are all variables impacting anticipated conflict cost that have been thoroughly covered in extant literature. However, material endowments alone are not determinative: the costs of conflict are also significantly impacted by how competently or incompetently each side utilizes the raw resources at its disposal. From the US in Vietnam to the USSR in Afghanistan, history is rife with examples of superpowers suffering immense costs at the hands of materially inferior but strategically sophisticated and opponents.[51]

---

[46] While niche or unknown in American international relations scholarship, the ideal of rational argument as a meaningful input into bargaining games that can shift actors' expectations about either the actions that will advance their interests or what their interests has been addressed by non-American political scientists. See for example Grobe, Christian. "The power of words: Argumentative persuasion in international negotiations." *European Journal of International Relations* 16, no. 1 (2010): 5-29. Risse, Thomas, and Mareike Kleine. "Deliberation in negotiations." *Journal of European public policy* 17, no. 5 (2010): 708-726.

[47] It all falls under the heading of what traditional rationalist and in particular bargaining models of interstate conflict and crises would call 'cheap talk;' statements, proclamations, and arguments that are meaningless because they are not backed by the material foundations that are necessary for states to credibly communicate.

[48] The basic principle behind the bargaining model of war as pioneered by James Fearon in his now canonical 1995 article.

In sum, the perception that an adversary leader is incompetent reduces the expected of costs of conflict, while the perception that an adversary leader increases the expected costs of conflict. The higher the level of incompetence observers ascribe to an adversary leader, the lower the degree of compromise said observer is willing to make to end a conflict through negotiated settlement: fighting for everything they want is less costly than sacrificing some of what they want to avoid the costs of fighting. Conversely, the higher the level of competence observers ascribe to an adversary leader, the higher the degree of compromise is willing to make to end a conflict through negotiated settlement: fighting for everything they want, even if successful, is likely to be more costly than accepting only a piece of their desired 'pie' to end violent conflict.

---

[49] This is the basis of the 'balance of resolve' explanation for triumph in crisis bargaining used for example in Richard K. Betts, *Nuclear blackmail and nuclear balance*. Brookings Institution Press, 2010.

[50] Variance in public support for war features prominently as an independent variable, explanatory mechanism, and dependent variable in an immense range of scholarship on the origins and effects of international conflict. For example the democratic peace and audience costs literature, two major if less contemporarily interesting areas of IR research, incorporate the role of public opinion as a crucial mechanism behind their arguments. An extensive literature with a political psychology bent concerns itself with the factors that shift public support for international conflict in a variety of manifestations, a research agenda that justifies its existence through the quite convincing contention that public opinion impacts state behavior in an interstate context.

[51] As is fiction: the rebel star fleet destroyed they Galactic Empire by besting their star destroyers while on the moon of Endor well armed storm troopers were bested by rock-throwing Ewoks with a sophisticated understanding of terrain driven combat and trap making.

Scope Conditions: Observers, Adversaries, and the Dependent Variable:

The theory presented in this dissertation applies whether observers are elite state decision-makers or members of the mass public, though for purposes of robustness the empirical portions of this dissertation distinguish between elites and members of the public. Apart from the obvious advantage of parsimony, there are several reasons this definition of observers is appropriate in the context of this theory.

First, there is ample evidence that patterns of human behavior identified by psychological research do not meaningfully differ between what political science would term 'elites' and 'average' members of the public.[52] Except for specific issue areas and mechanisms, causal evidence about the behavior of one group can therefore be used to make inferences about the other. Second, this is a theory of how variation in beliefs about judgment *move* diplomatic preferences: it does not claim to determine exactly where on the spectrum of diplomacy versus force an individual's preferences fall. Whatever an individual's baseline diplomatic preferences, and whatever their ultimate attitude, there is no reason to expect the belief that an adversary's judgment is a particular type to shift the preferences of elites in a *different direction* than the preferences the public. In other words, the theory's mechanism of action (beliefs) and dependent variable (attitude shifts) are not part of the class of theories where we would expect elites to differ from members of the public. Finally, political decision-makers monitor and are responsive to public policy preferences, particularly in democracies.[53] Therefore, we can expect shifts in public attitudes to be at least somewhat reflected by equivalent shifts in the expressed preferences of elites.

---

[52] Kertzer, Joshua D. *Resolve in international politics*. Vol. 2. Princeton University Press, 2016. Kertzer, Joshua D. "Microfoundations in international relations." *Conflict Management and Peace Science* 34, no. 1 (2017): 81-97. Hafner-Burton, Emilie M., Stephan Haggard, David A. Lake, and David G. Victor. "The behavioral revolution and international relations." *International Organization* 71, no. S1 (2017): S1-S31.

[53] Kertzer 2016, Foyle, Douglas C. *Counting the public in: Presidents, public opinion, and foreign policy*. Columbia University Press, 1999. Kertzer, Joshua D., and Thomas Zeitzoff. "A bottom-up theory of public opinion about

Furthermore, this theory is focused on how beliefs about the judgment of international counterparts impact preferences only in the area of conflict and crisis diplomacy. This scope condition reflects not only the importance of the issue area, but also that most extant scholarship on the rationality of irrationality has looked at crisis or conflict bargaining in some form. To draw the starkest possible contrast between my theory and the prior work it both expands and corrects, it is necessary to keep my theory's scope as similar to this prior work as possible: that way we can compare the predictions and mechanisms of the theories side-by-side. An important caveat is that the theory is designed to explain *preferences* in diplomatic bargaining, chiefly regarding what persuasive methods to use and what a state's maximum concession threshold should be. It is not, however, always the case that these preferences alone will be reflected in actual policies or behavior. For example, an actor may prefer to bargain through coercive threats and to exit negotiations rather than make even minor concessions, yet lack the material capability to make credible threats. Alternatively, the same actor could be pushed by allied third parties to open negotiations even though their overall preferences dictate ultimatum-based bargaining, to first attempt good faith dialogue before resorting to coercion, or to stay involved in negotiations even if this necessitates some compromise.

Finally, the conceptual contribution made by this dissertation in unpacking and typologizing the distinction between perceived rationality and irrationality has plausible

---

foreign policy." *American Journal of Political Science* 61, no. 3 (2017): 543-558. Baum, Matthew A., and Philip BK Potter. *War and democratic constraint: How the public influences foreign policy*. Princeton University Press, 2015.

implications for a myriad of questions in the study of international politics. For example, trade deals, alliance formation, burden sharing or shirking, and design or participation in international institutions are all areas that could conceivably be impacted by actors' beliefs about the rationality or irrationality of other participants. It is my hope that future work by myself and others will build upon the conceptual and empirical work in this project to explore these and other areas of interest to IR scholars.

Plan of the Dissertation:

The rest of the dissertation proceeds as follows.

Chapter two details my theory, which predicts that actors will be more likely to immediately resort to coercion against adversaries they perceive as biased, and less inclined to compromise for the sake of de-escalation during negotiations with adversaries they perceive as incompetent. Drawing on concepts from psychology, I break down the meaning of the binary dimensions of *beliefs about judgment:* objectivity and bias, competence and incompetence. I introduce the theoretical mechanisms by which variation on these dimensions shapes preferences in conflict diplomacy with reference extant psychology research, which suggests that ascriptions of bias undermine the expectation that information sharing will lead to convergent beliefs, and ascriptions of incompetence reduce perceived threat. I detail the logical mechanisms through which these foundational effects on perception, belief, and expectations shape the diplomatic bargaining preferences of actors in an interstate context. I then unpack the observable implications of variation in these basic preferences: what does diplomatic bargaining look like when conducted by an actor who prefers coercion without compromise versus dialogue with compromise? I outline how this theoretical framework improves upon our understanding of the role played by perception, leaders, and the mass public in interstate conflict. The chapter then examines three alternative arguments and outlines the methodology for testing the theory through a mix of experimental and case-study approaches.

Chapter three details my experimental tests of the theory's veracity at the level of public opinion, which consist of three novel survey experiments fielded to adult members of the US public. I discuss the unique advantages of survey experiments as a method to test the veracity of theories that center on the causal link between beliefs and preferences or behavior. I then address the 'levels-of-analysis' concern, and explain the two primary pathways through which I argue experimental evidence gathered at the level of the mass public can inform our expectations about the behavior of states and the elite decision-makers that directly execute their policies: a direct-inference pathway based on the similarity of human behavioral patterns, and a public-opinion pathway. I then detail the design of the survey experiments, which all shared a basic template utilizing a factorial treatment design but differed in their subject population or specific scenarios. The first survey was fielded through Amazon's Mechanical Turk (MTurk) and centered on a hypothetical nuclear crisis between the US and the imaginary state of Mykonara, the second survey was fielded through MTurk but centered on the real-world crisis between the US and North Korea, and the third survey was fielded on a nationally representative sample and again used Mykonara. Subjects were told about the crisis then randomly presented with between zero (control) and two opinion articles about the adversary leader which described the leader as biased or objective, and competent or incompetent. Following the treatments, subjects were asked a series of questions to ascertain their preferences and expectations for how the US should approach diplomatic bargaining with the adversary. The results from all three surveys provide strong support my theory, particularly the hypothesized effects and mechanisms of perceived bias; for example, in all three surveys subjects assigned the 'biased' article were significantly less likely to advocate diplomacy, or expect it to be effective, compared to subjects in the 'objective' or 'control' group. This chapter ends by addressing the 'levels of analysis; problem by discussing the pros and cons of these experimental findings as a means of making inferences about the behavior of political elites.

In Chapter four, I test my theory at the elite level through a case study of US decision-making during the Korean War, which permits me to trace variation in elites' diplomatic bargaining preferences towards three adversaries and across time in the context of a single conflict in which a ceasefire was the immediate aim of US policy. The Kaesong negotiations conducted under the Truman administration serve as a focal point for this case-study, though I draw evidence from 1946 through the end of the Truman administration and examine in some detail US discussions surrounding three additional noteworthy periods: the US/USSR Joint Commission on Korea, the North Korean invasion that began the Korean War, and the entrance of Chinese forces into the war. The case supports the veracity of my theory at the elite level. Most US decision-makers saw the Soviets as responsible statesman who could be reasoned with; officials were willing, even eager, to engage in iterative diplomatic dialogue with the Soviets in pursuit of a ceasefire, and evinced readiness to make some concessions to Soviet interests. However, these same officials thought diplomatic dialogue with the 'irrational' Chinese would be fruitless and concessions to their incompetent leadership unnecessary. It took immense pressure from a range of UN coalition members to bring the US to the negotiating table at Kaesong, and once at that table US officials mainly issued threats and ultimatums to Chinese representatives as opposed to engaging in an iterative dialogue. The case also shows that US policymakers were responsive to public beliefs about adversary rationality; Republicans seeking to brand Truman as soft on China, helped move the public towards the belief that the Chinese were irrational and neither could nor should be negotiated with. Growing subscription to this narrative put constraints on elites in power; many who believed the Chinese were rational were reluctant to negotiate for fear of political consequences. In short, both through the internalized views of decision-makers and the pressures of public opinion, US beliefs about the rationality of the Soviets, Chinese, and North Koreans shaped their diplomatic bargaining preferences in ways prominent alternative arguments fail to predict.

In Chapter five, I use experimental methods to probe the sources of beliefs about leaders' judgment so as to provide an empirical response to two major critiques of my theory's use of these beliefs as an independent variable; that the beliefs follow policy positions rather than the other way around, or that they are epiphenomenal to other variables that shape diplomatic bargaining preferences. I first detail some plausible hypotheses of belief formation derived from extant work in multiple fields; a rationalist pathway based on costly behavior, a motivated reasoning pathway shaped in which perceptions are driven by the intentions of the observer, and a range of hypothesis about how perceived rationality is influenced by a leader's identity and costless behavior. I describe a choice-based conjoint survey experiment that probes the plausibility of the wide range of potential explanations suggested by existing work, though it involves too many variables for causal inferences from the results to be valid. The conjoint results suggest that while material variables are important, leaders' ideology and personal behavior also carry significant weight in subjects' inferences about their rationality. I then describe a second survey experiment, which utilizes a multivariate factorial design to creates a more controlled setting in which to test the causal impact of a leader's race, ideology, personal behavior, and state economy on subjects' perceptions of their rationality and competence. The results provide strong support for the hypothesis that leaders with speaking styles that seem 'angry' or 'eccentric,' and leaders that are ideologically distant, are far more likely to be perceived as irrational than other leaders who possess equivalent material capability and make identical policy decisions. The second survey also contains questions about diplomatic bargaining preferences, and the results from these provide further support for the primary theory of the dissertation. The centrality of identity and costless behavior in inferences about foreign leaders' rationality suggests such inferences are exogenous to other important factors in IR, and so must be taken seriously as an explanatory variable.

Chapter 2: Theory: Unpacking Perceived Irrationality and its Impact on Conflict Diplomacy

I call it the Madman Theory, Bob. I want the North Vietnamese to believe I've reached the point where I might do anything to stop the war. We'll just slip the word to them that, "for God's sake, you know Nixon is obsessed about communism. We can't restrain him when he's angry—and he has his hand on the nuclear button" and Ho Chi Minh himself will be in Paris in two days begging for peace – Richard Nixon

Have you ever noticed that anyone driving slower than you is an idiot, and anyone driving faster than you is a maniac? - George Carlin

How do people approach international conflicts with adversaries they believe are foolish, delusional, or downright crazy? Do they advocate diplomacy out of an abundance of caution, or assume that military force is the only language such enemies understand? Theory and practice provide conflicting answers to these questions. Thomas Schelling, in positing the 'rationality of irrationality,' argued that leaders whose adversaries believed they were irrational would be free from coercion and deterrent or compellent violence.[54] This was the foundational logic of Nixon's 'madman theory:' people would quickly give in to the demands of, and refrain from escalatory violence against, apparently 'mad' leaders. On the one hand, leaders frequently brand adversaries as unpredictable madmen in an apparent attempt to underscore the futility of negotiation and garner support for military action from domestic or international audiences. For example, the Bush administration's propagation of the "Saddam the Madman" narrative was a major part of its efforts to sell the Iraq War to both the American public and prospective international allies.[55] In short, theoretically as well as empirically it remains an open question how interstate conflict diplomacy is affected by peoples' beliefs about the rationality or irrationality of foreign leaders, which is to say how foreign leaders *think*. This project provides both the theory and the evidence needed to fill this gap.

---

[54] Schelling 2008

[55] Lake, David A. "Two cheers for bargaining theory: Assessing rationalist explanations of the Iraq War." *International Security* 35, no. 3 (2010): 7-52.

I treat beliefs about the rationality of others' judgment as a spectrum with two dimensions, perceived objectivity and perceived competence, and theorize how variation in these beliefs shapes actors' *diplomatic bargaining preferences*. An actor's *diplomatic bargaining preferences* are constituted by its inclination toward conducting diplomacy via argument versus coercion and the degree of compromise or concession it is willing to make as part of a negotiated settlement: the combination of these components has a range of observable implications for attitude toward and conduct in interstate diplomacy. For example, a preference for coercion/concessions implies tit-for-tat coercive diplomacy wherein participants trade threats and concessions while trying hard not to cross the threshold of uncontrollable escalation, while a preference for coercion/no concessions implies the delivery of ultimatums that almost inevitably escalate conflict if they do not immediately end it.

I argue that perceived irrationality, and conversely perceived rationality, impacts diplomatic bargaining preferences in the polar opposite direction of that predicted by the 'rationality of irrationality' theory. Apparently mad leaders are more likely to be subject to threats or force, less likely to get the concessions they demand, and more likely to have counterparts demand more extensive concessions from them if a negotiated settlement is to be reached. Specifically, the more biased observers believe a leader to be, the greater their preference for coercion over dialogue: when at the bargaining table with apparently biased adversaries, actors are less likely to attempt rational persuasion and more likely to rely on the 'universal language' of threats or rewards. Meanwhile, actors have less incentive to initiate negotiations with adversaries they believe are incompetent, and if they do engage in negotiations are less inclined to make concessions or offer material incentives in order to avoid escalation. Finally, actors may abstain from all diplomacy and negotiation except to issue ultimatums with adversaries they believe are not objective or competent enough to be capable of sound judgment.

This theoretical contribution speaks to a range of topics in IR, and three recent research trends in particular: the ‘behavioral revolution’ in political psychology, the study of leaders in international politics, and the study of diplomacy. One of the key contributions of work in the behavioral revolution has been to stress differences in how actors perceive situations or construct policy, while using experimental methods to establish when and how this heterogeneity in judgment matters.[56] I expand this by stressing that actors may believe others′ judgment differs from their own,[57] then investigating how variation in these beliefs about judgment impact preferences and decisions in international politics and conflict. I contribute to leadership studies by focusing on the impacts beliefs about the judgment of foreign leaders, which highlights an understudied mechanism through which individual leaders can shape international politics: not through their own traits or decisions but through others’ perceptions of them, and in particular of a leader’s psychology or thought processes. Finally, I contribute to the expanding literature on international diplomacy by examining a neglected source of variation in actors' diplomatic preferences, styles, and choices: their beliefs about the judgment of others. What, in short, are the consequences for interstate diplomacy when leaders are believed to be not just unfriendly but fundamentally *unreasonable?* Answering this question promises to expand our understanding not only of non-violent diplomacy, crisis de-escalation or conflict termination but, by extension, their violent and traditionally more studied counterparts: crisis or conflict escalation and why what start as talks break down into violence.

---

[56] Joshua D. Kertzer "Resolve, time, and risk." *International Organization* 71, no. S1 (2017): S109-S136, Emilie M. Hafner-Burton, Stephan Haggard, David A. Lake, and David G. Victor. "The Behavioral Revolution and International Relations." *International Organization* 71, no. S1 (2017): S1-S31.

[57] Pronin 2009, Pronin 2007, Ross 2015

In this chapter, I detail my theory of consequences of variation in beliefs about the rationality of international adversaries. I first define key concepts: what do I mean by 'beliefs about judgment,' why does it make sense to think about perceived madness as part of this spectrum, and what are the constituent components of perceived objectivity, bias, competence, and incompetence? Second, I explicate my typology and the constituent components of its ideal types, the *Rational, Fool, Fanatic, and Madman*. Third, I define and scope the contours of the dependent variable, *diplomatic bargaining preferences*. Fourth, utilizing extant research in psychology, I detail the theoretical mechanisms by which variation on the dimensions of my typology shape diplomatic bargaining preferences in conflict diplomacy: evidence suggests that ascriptions of bias undermine the expectation that information sharing will lead to convergent beliefs, and ascriptions of incompetence reduce perceived threat. I detail the logical mechanisms through which these foundational effects shape interstate diplomatic bargaining preferences, and unpack the observable implications of variation in these preferences with reference to real-world examples: what does diplomatic bargaining look like when conducted by an actor who prefers coercion without compromise versus dialogue with compromise? Fifth, I describe the two primary reasons why this theory applies to individuals in both the mass public and elite decision-making circles: the broad applicability of the proposed psychological mechanisms, and the role of public opinion. Sixth, I discuss how this theoretical framework improves upon our understanding of diplomacy, and the role played by perception, leaders, and the mass public in interstate conflict. The chapter concludes with an examination of three alternative arguments.

Beliefs About Judgment: A Spectrum for Perceived Irrationality:

For theoretical consistency, utility, and realism, it is important that we think about perceived irrationality as not an independent phenomenon but one extreme end of a spectrum: to build a theory that incorporates perceived irrationality we need to identify and address the spectrum on which it sits. However, there is not a single spectrum into which perceived irrationality obviously fits, so defining that spectrum is not a given but a choice, and like any choice it has consequences for the scope, implications, and methods of testing a theory. As I outlined in Chapter One, I conceptualize madness as the 'extreme irrational' end on the spectrum of judgment, which includes the 'rational' value of its constituent dimensions. Therefore when theorizing the effects of perceived variation within the spectrum we cannot consider *only* the effects of the perceived madness extreme as Schelling and other previous scholars have done: we also have to consider perceived rationality.[58] This permits us to separate which aspects of rationality or irrationality drive effects or are central mechanisms in our theory and which are tangential, which we cannot do if thinking in binary terms.

I therefore conceptualize perceived irrationality as one extreme of the spectrum of what I term '*Beliefs about Judgment*;' judgment and choice "are related but not interchangeable"[59] components of decision-making. Judgment processes precede choice, even in cases where deliberation is impossible, and consists of the interpretation of information into beliefs, followed by the formulation of possible action plans considering said beliefs. The interpretation phase involves assessing evidence or other pieces of the external world to form beliefs about current and probable future states of the world; because the degree to which actors objectively interpret information can vary, actors in the same situation may form different beliefs.

---

[58] Schelling 2008, McMannus 2019 and 2020,

[59] McDermott, Rose. *Risk-taking in international politics: Prospect theory in American foreign policy*. University of Michigan Press, 2001.

However, beliefs about the world do not automatically generate some pre-existing set of possible actions that is constant across actors. Especially in complex arenas like interstate relations that involve myriad disparate factors, actors must construct a menu of options by drawing on their experience, skills, and knowledge:[60] these traits can be jointly defined as ‘competence.’ Since not all actors are equally competent, actors may construct different option sets even if they share identical beliefs and goals. Put another way, judgment processes are those by which actors answer the following questions: What is the external situation? What are my possible plans of action in this situation? And what is the likelihood of each possible outcome from each plausible plan of action? This does not include valuation of these possible outcomes (e.g. "how much do I want each possible outcome?").

In short, judgment processes include those by which actors perceive the situations they face and consider how they *might* act, but not those by which actors decide what they want and therefore how they *will* act. Observers may (accurately or not) perceive variation in judgment between themselves and others as the reason a counterpart's behavior diverges from their own.[61]

My typology of *beliefs about judgment* has two dimensions: *perceived objectivity* and *perceived competence*, each of which denotes a spectrum along which observers' beliefs about the judgment of their counterparts can fall. The first is *perceived objectivity* (objective/biased): whether the other's perceptions and interpretations of information are believed to be objective or distorted by bias. The second is *competence* (competent/incompetent): whether the other is believed to have the experience or skill to translate preferences into policy.[62] These traits are conceptually linked because each 1) influences how an actor is expected to move from information about the external situation to choice but 2) is independent of values, goals, or desires. In sum, they indicate beliefs about how others think but not what others want. Actors believed to be sufficiently biased *and* incompetent could be characterized as expected to behave in ways that fulfill the criteria typically associated with irrationality: insensitive to costs, pursuing seemingly senseless goals, escalating even conflicts they cannot win, and otherwise behaving in ways that observers would characterize as erratic.

---

[60] Horowitz, Allen and Stam 2015
[61] Promin 2007, 2009

This conceptualization draws from psychological work on ‘mind reading’, or the way people intuit the mental states of others, and naïve realism, a phenomenological stance that posits that people assume their judgment constitutes the standard of objectivity.[63] In an uncertain world where facts rarely 'speak for themselves,' people nonetheless tend to believe judgments which diverge from their own are not just different but 'wrong.’ They expect others' judgments to mirror their own, and when presented with evidence that this is not the case, they must explain the divergence. One particularly charitable explanation, which coheres with rationalist frameworks, is that the other is merely misinformed.[64] In theory observers could explain divergent judgments by re-examining their own beliefs, but in practice people face considerable cognitive and motivational barriers to questioning the validity of their own judgment. Less generous interpretations are that the other is incompetent, or that its judgment is distorted by endogenous biases that warp its view of reality and prevent it from reaching 'accurate' conclusions even when fully informed.

---

[62] Elizabeth N. Saunders 2017.

[63] See Gilovich, Thomas, and Lee Ross 2015.

[64] This is the default assumption of rationalist bargaining models, which assume a mutual expectation that information sharing will lead to belief convergence.

In brief, *perceived objectivity* ranges from entirely objective at one end to entirely biased at the other, and *perceived competence* ranges from entirely competent to entirely incompetent. Below, I detail these spectra in turn. In practice, observers are unlikely to categorize counterparts as the ideal-typical extreme on either dimension, so it useful to conceptualize these dimensions of my typology as continuous variables, though for the sake of parsimony I usually treat these dimensions as binary.

Perceived Objectivity:

Perceived Objectivity denotes an actor's beliefs about the information processing component of a counterpart's judgment process. This includes how the counterpart processes or interprets information, the reliability with which the counterpart receives information, whether and to what extent the counterpart's beliefs are resistant to updating in response to new information, and whether any ascribed biases operate in a predictable direction. The above is illustrative but not exhaustive; *perceived objectivity* refers broadly speaking to beliefs about how a counterpart interacts with new information, or novel framings of existing information, when constructing its beliefs about the current and likely future state of the world. Because perception is always at least somewhat subjective, an actor's reference point for perfect 'objectivity' is necessarily the way it itself interprets reality and updates beliefs in response to new information. 'Objective,' in other words, is in this formulation an alternative label for 'thinks like me.'[65] Conversely, 'Biased' is an alternative label for 'does not think like me.'

To attribute bias to an actor is to conclude that inherent flaws in its perception explain or predict disconnects between its beliefs and the observer's view of 'objective' reality. This divergence creates uncertainty about how apparently biased actors will translate information into beliefs about probabilities and the state of the world, and therefore uncertainty about what their beliefs will be in any particular situation or point in time.

---

[65] Again, this is the foundational concept of what psychology calls the naïve realism phenomenon.

The more *objective* an actor is believed to be, the less distance there is expected to be between the ‘reality’ implied by the information available to them and their beliefs about the world at any given time. An entirely objective actor would be expected to have an unfettered capacity to receive information, interpret it accurately, and instantaneously update its beliefs to reflect the sum total of information to which it has access. In short, an entirely objective actor processes information like an ideal-type rational actor: its beliefs are equivalent to the information to which it has access. This definition is complicated by the subjective nature of reality long since acknowledged in psychological scholarship: when moving outside of a formal theoretic framework, we must acknowledge that there are few if any instances in which information has a single ‘objective’ meaning. The same piece of information can mean very different things for different actors. However, because *perceived objectivity* is a relational variable, the objective reality to which it refers is dependent upon the observer. In other words, in the context of my conceptualization of *perceived objectivity*, to say an observer expects a counterpart to interpret information accurately is to say that the observer expects the counterpart to interpret information exactly as they themselves would.

By contrast, the more *biased* an actor is believed to be, the greater, and more persistent, the distance between the ‘reality’ of the information available to it and its beliefs about the world at any given time. An ideal-typically biased actor would be one that lacks any meaningful awareness of the external situation, that does not know the game it is playing and is inherently incapable of learning the game’s rules because it is inherently incapable of learning anything.

Such perfectly biased actors do not shape their actions in accord with the true state of their strategic environment; they consistently behave in ways that seem random or nonsensical because they are effectively responding to a world that exists only in their own mind and is independent of the reality perceived by others. Schelling posited that the behavior of such players in strategic interaction games would resemble that of a puppy, a metaphor he used to illustrate that the behavior of what he called irrational actors could not be 1) predicted based on the situation or 2) influenced in reliable directions through informational means. Put another way, there is no knowing what an irrational actor will or will not do, and no way to reliably influence the actions of an irrational actor without physically controlling them. The uncertainty implied in this framing is not intrinsic to an ideal-typically biased actor, as there may be cases when the direction in which an actor's biases impact its behavior is known with certainty; for example an actor may always be expected to fight if observers think said actor always believes it will win.[66] However, this still meets the baseline conceptual understanding of a perfectly biased actor: one that does not process information about its environment into its beliefs and therefore behaves independently of new or existing information.

There are several ascribed information processing tendencies that could render an actor as one perceived to be ideal-typically biased. First, an ascribed inability or unwillingness to receive information. Second, a belief that the actor will inevitably interpret whatever information it does receive in ways that are warped and inaccurate to such an extent that the conclusions it draws from information are as-if independent from the information itself. Third, a belief that the actor's beliefs are entirely fixed and will not update in response to novel information, no matter how apparently overwhelming the evidence. Theoretically, at their extreme, any of these flaws in the information processing phase of judgment would be sufficient to categorize an actor as ideal-typically biased. In practice, it is unlikely albeit not impossible that an actor would be perceived as being *completely* unable to receive, interpret, or update its beliefs in response to new information. However, these perceived processing flaws are not mutually exclusive, and when combined could quite conceivably yield an actor that observers expect to behave as-if perfectly biased. For example, members of the second Bush administration who subscribed to the 'Saddam the Madman' narrative did so in large part because they knew Saddam's advisors were reluctant to give him information that might displease him, and believed that Saddam's perception of risks was distorted by, among other things, his megalomaniacal nature and susceptibility to motivated biases.[67] It was not their expectation that Saddam would receive *no* information, nor that he would *never* interpret novel information accurately. Rather, the belief was that the information pipeline to Saddam was so unreliable and his interpretation of evidence or rational arguments sufficiently warped that nothing but the most universal language, namely that of force, could influence his actions.

---

[66] Weisiger described the resultant issue as a dispositional commitment problem. Weisiger, Alex. *Logics of war: Explanations for limited and unlimited conflicts*. Cornell University Press, 2013.

There are myriad characteristics that can move actors towards the 'biased' end of the perceived objectivity spectrum; any trait that increases the degree to which actors are seen as unable to receive, interpret, or update beliefs in response to new information could be a real or perceived source bias. Physiological conditions that actively hinders an actor's capacity to process information, for example blindness, deafness, or mental illness, are one potential source of perceived bias. In a similar vein, medicinal or recreational use of perception-distorting drugs

---

[67] Lake 2011; Pollack, Kenneth. *The Threatening Storm: What Every American Needs to Know Before an Invasion in Iraq*. Random House, 2003; Kaufmann, Chaim. "Threat inflation and the failure of the marketplace of ideas: The selling of the Iraq war." *International Security* 29, no. 1 (2004): 5-48.

(alcohol, narcotics, methamphetamines) is a potential physiological source of perceived bias, and in fact a relatively common issue with respect to state leaders.[68] However, it could also be the case that an actor does not have the material capability to receive or interpret certain types of information[69] or is reliant upon unreliable conveyers of information.[70] Ideological belief systems are another plausible source of perceived bias, be they political, religious, nationalistic or philosophical. Observers may expect counterparts to refuse to accept evidence that is contrary to their ideals[71] or to interpret all new information through a lens that supports their ideology.[72] Finally, certain personality traits may move an actor further towards the biased end of the perceived objectivity spectrum. Actors that are seen as emotional, temperamental, or lacking in self-control are likely to be seen as more biased because their passions are likely to dictate their information processing, whereas actors that are perceived to be calm and logical are likely to be seen as more objective.[73] Note that this is not intended as an exhaustive list of biasing traits, but rather illustrative examples.

---

[68] See for example Rose McDermott's book on Presidential Illness: McDermott, Rose. *Presidential leadership, illness, and decision making*. Cambridge University Press, 2007. John F. Kennedy was known to be on a cocktail of narcotics for his chronic health issues, Regan suffered from progressive dementia, and Nixon was drunk with such frequency and to such an extent that an aide infamously hid the nuclear football from him.

[69] Eg. the lack of a phone line or translators

[70] Eg. officials who hide information from a temperamental superior

[71] Eg. motivated reasoning

[72] Eg. confirmation bias

[73] Schelling's original description of apparently irrational leaders leans heavily on this perspective of emotion as counter to logic, though we should note that while this is a common perspective among policy makers as well as in scholarly discourse it reflects an outdated view of the link between emotion and judgment. Neruopsychological research pioneered the idea that affect and emotion were key to rational reasoning, and this has since been introduced to psychology writ large and begun tentatively breaking into IR. See McDermott, Rose. "The feeling of rationality: The meaning of neuroscientific advances for political science." *Perspectives on politics* 2, no. 4 (2004): 691-706.

Perceived Competence:

Perceived Competence denotes observers' beliefs about the action-plan construction phase of a counterpart's judgment process, and whether said counterpart is generally capable of devising plans that maximize its chances to advance its goals. Relative to perceived objectivity, perceived competence is more domain-specific at both the level of particular actors (it is possible to be perceived as a competent politician but an incompetent general) and at the conceptual level (perceived political competence means something different than perceived military competence). For the purposes of creating a coherent and manageable theory regarding either the causes or interstate consequences of variation in perceived competence, it is therefore necessary to specify a domain of competence to which this dimension of higher order beliefs about judgment applies.

Beliefs about the military-political competence of foreign leaders are most relevant to this project, and this domain of perceived competence primarily impacts threat perception and beliefs about relative military power. However, perceived competence captures the determinants of military capability and threat that are dispositional rather than material: namely the ability to devise optimal plans of action. It is necessary to separate these components of military capability because whether an actor has the skill to optimally employ the military resources at its disposal is a separate question from how impressive or underwhelming those military resources are on paper. A large and technologically advanced force poses less of a threat in the hands of an inept commander, while even rag-tag forces can pose a substantial threat to enemies when directed by skilled leadership.[74] All else equal, actors perceived as incompetent are expected to be less capable of achieving their goals than actors perceived as competent.

---

[74] See for reference the battle of Agincourt, the Vietnam War, the destruction of the Death Star by the Rebel Alliance at Yavin IV

More specifically, perceived competence denotes observers' beliefs about how a counterpart, on the basis of its beliefs about the situation, constructs possible military options for the furtherance of its military-political goals. This option construction can be at the tactical level, as in coordination of land and air forces in an acute military engagement, or the strategic level, as in broader force construction, movement or deployment, the targeted destruction of critical enemy facilities, and the coercive use of blockades, bombardments, or threats of escalatory punishment. Suboptimal tactical or strategic option construction leads actors to underperform relative to their raw military capabilities, and vice-versa.[75] The constituent components of perceived military-political competence include observers' beliefs about a counterpart's intellect and capacity for long term planning, military experience or expertise, and organizational cohesion.[76] As is the case with perceived objectivity, observers generally regard themselves as competent, but unlike with perceived objectivity it is not necessarily true that they regard themselves as an ideal-typical competent actor; it is entirely possible for an observer to believe a counterpart is *more* competent than itself, a condition I refer to as perceived *hyper-competence*.

The more *competent* an actor is believed to be, the less distance there is expected to be between an 'optimal strategy' and the options it actually constructs within the constraints of its capabilities and perceived situation. An ideal typically competent actor would be one that always construct options which make optimal use of its capabilities and is able to distinguish between optimal and suboptimal option-choices within the context of its beliefs. Ideal typical competent actors also have the organizational capacity to implement and maintain their chosen military-political plan of action unimpeded by dysfunction among their constitutive agents, such as that generated by corruption, factional infighting, bureaucratic lag, or the absence of standard operating procedures for communication among agents. In other words, an ideal typically competent actor is expected to behave like a rational actor in strategic choice games, with the only options it considers being the ideal ones as determined by the situation and the balance of capabilities. An observer that believes an adversary is competent would expect the adversary to consider the same political-military options itself would in a particular situation. Uncertainty about the options *considered* by a competent adversary therefore exists only to the extent that there is uncertainty about what it perceives to be the situation and balance of capabilities: this includes uncertainty or incomplete information about the material capabilities at the adversary's disposal. Uncertainty about the option that a competent adversary will *choose* exists to the extent that there is uncertainty about its goals and values: this includes uncertainty or incomplete information about the adversary's resolve and risk-tolerance.

---

[75] Biddle's theory of military efficacy centered on the *modern system* of force employment is a primary example of a theory that implicitly utilizes variation in competence, albeit not *perceived* competence, as an IV.

[76] Especially between military leaders in the field and heads of government. A disconnect between these, for example, would impede the capacity for planning or carrying out cohesive strategies to advance overall political because there would be a persistent divergence between operational and political metrics for the success of military operations. If commanders are planning for battlefield victory only and not thinking about broader political goals, while government leaders are thinking about broader political goals but giving insufficient direction to their commanders to realistically advance these, then the overall strategic military posture of a state is likely to be suboptimal for the advancement of its aims,

It is, however, possible for actors to believe counterparts are *hypercompetent*, which is to say more competent than themselves. An apparently hypercompetent adversary would be expected to construct options within its situational constraints that the observer had not thought of, effectively exceeding what the observer believed was the ceiling for optimal military-political tactics or strategy. Consequently, there is always some uncertainty about the options hypercompetent actors will consider, a constant possibility that they will surprise their adversary with brilliant military-political innovations that enable them to punch above their weight. In other words, higher levels of perceived competence, relative to lower levels, reduce uncertainty and increase perceived threat up to the threshold of perceived hypercompetence; past the hypercompetence threshold, which marks the point where an observer believes itself at a disadvantage in relative competence, higher perceived competence increases both uncertainty and perceived threat.[77]

Characteristics that can move actors towards the 'competent' end of the perceived competence fall into a range of categories: traits of the state or society, of particular organizations, or of specific individuals could all plausibly enhance the military-political competence of an international actor. States with cultures that prize discipline, organization, respect for authority, or other militaristic values are often expected to conduct themselves with exceptional prowess in military matters.[78] Ancient Sparta, Wilheim as well as Nazi Germany, and the Imperial Japan of WWII are all prototypical examples of this link between culture and perceived competence.[79] In a similar vein, militaries that have clear command structures, meritocratic systems of promotion, rigorous training and education standards, or otherwise display high levels or organization are for obvious reasons likely to increase perceived competence.[80] A clear and non-conflictual relationship between military and civilian government organizations is also an important source of perceived competence: whether this comes from institutionalized practices that dictate civilian control of the military or military dictatorship that renders civilian government irrelevant is unimportant. What matters is the absence of conflicting incentives, unclear command prerogatives, or other sources of civil-military conflict that could apparently reduce the coherency of the military-political decision-making process and thereby detract from the capacity to construct or implement optimal plans of action. Finally, certain personal traits of a state's leader or key military-political decisionmakers[81] can increase perceived competence: relevant experience and the reputation for competence that accompanies it is perhaps the most obvious manifestation of this and the one most extensively addressed in existing research.[82] For example, state leaders who have battlefield or command experience, or at least military training, are likely to be seen as more competent than those without: the same logic applies to officials who have substantial experience in international negotiating forums[83] or even those who have demonstrated apparent skill in domestic political contexts.[84] However traits based on apparent personality rather than observable experience or performance can also increase perceived competence: leaders viewed as particularly intelligent or cunning are also likely to be perceived as more competent, regardless of whether the ascription of these traits to the leader is due to past actions, media framing, calculated presentation of a persona by the leader in question, or some other reason.[85]

---

[77] Additionally in the case of an apparently hypercompetent adversary observers may be more inclined towards possibilistic, worst case thinking about the adversary's plans of action

[78] Note that I am not here subscribing to the argument, sometimes made in political science, that differences in culture explain actual differences in military political behavior or outcomes: this project does not take a stand on the true role of culture in affecting military behavior and/or competence. Rather, I am arguing that regardless of its true importance, culture plays an important role in shaping perceptions of military competence.

[79] Moreover, the German and Japanese cases demonstrate that this perceived link between cultural militarism and high military-political competence is not always justified.

[80] Biddle 2004

The more *incompetent* an actor is believed to be, the more distance there is expected to be between an 'optimal strategy' and the options it actually constructs within the constraints of its capabilities and perceived situation. An ideal typically incompetent actor would be one that is incapable or constructing options which make optimal use of its capabilities and is unable to distinguish between optimal and suboptimal option-choices within the context of its beliefs.

---

[81] Ex. Secretaries of State or Defense, theater commanders or other highly visible generals

[82] Horowitz, Allen, and Stam 2015.

[83] Eg. the UN, the G-20

[84] Eg. maintaining control in a dangerous domestic environment, achieving policy goals by overcoming domestic opposition through persuasion or maneuvering, constructing powerful domestic coalitions, or otherwise behaving like a savvy political operative.

[85] Indeed, governments with state media often devote substantial resources to framing state leaders as exceptionally intelligent, and foreign media with little else to go on frequently adopt these frames

Ideal typical incompetent actors also lack the organizational capacity to implement and maintain their chosen military-political plan of action: they are unable to consistently carry out their planned actions due to dysfunction among their constitutive agents, due for example to corruption, factional infighting, bureaucratic lag, or the absence of standard operating procedures for communication among agents. Put another way, not only are perfectly incompetent actors unable to construct or recognize optimal military-political plans of action, they also cannot consistently carry out whatever course of action they *do* decide upon. While at their extreme either incompetence in the construction or implementation of military-political action plans could alone render an actor incompetent, in practice it is not necessary for an actor to be considered highly incompetent that it be completely inept at *both* option construction and implementation of military-political action plans.

In the context of strategic interaction, apparently incompetent actors are expected to behave differently than a rational actor, even though they ultimately make choices among options in a perfectly rational manner. While an ideal typical incompetent actor is expected to choose between options on the basis of rational cost-benefit calculus, the only options it considers will be different from, and strategically suboptimal relative to, the ideal ones as determined by the situation and the balance of capabilities. Effectively, the options considered by incompetent actors are not exogenously determined by the situation, balance of capabilities or nature of the strategic game. An observer that believes an adversary is incompetent would expect the adversary to consider different political-military options than itself given a particular situation. A degree of uncertainty about the options *considered* by an incompetent adversary is therefore inherent to the belief that the adversary is incompetent, even if there is simultaneously certainty that whatever options the adversary considers will prevent it from punching at its weight. In brief, higher levels of perceived incompetence, relative to lower levels, increase uncertainty and reduce perceived threat.

Characteristics at the level of the state or society, of particular organizations, or of specific individuals could can all move actors towards the 'incompetent' end of the perceived competence. States with cultures that are seen as permissive or encouraging of deviance from authority, laziness, or overindulgence may be expected to perform especially poorly in military endeavors.[86] Ancient Athens, France in the WWI and WWII era, and modern day Greece all demonstrate the connection between cultural stereotypes and perceived incompetence.[87] By extension, militaries that lack clear command structures, have systems of promotion based on nepotism, corruption, or personal favoritism, lack rigorous training and education standards, or otherwise display low levels or organization are likely to decrease perceived competence. Opaque or conflictual relationships between military and civilian government organizations are also a source of perceived incompetence, since they can generate apparently incoherent military-political decision-making processes that are not conducive to constructing or implementing optimal plans of action.

At a more individual level, certain personal traits of a state's leader or key military-political decisionmakers[88] can increase perceived incompetence: a lack of experience, training, or relevant policy expertise is one important form of this. State leaders who have no history of military service may be seen as more likely to blunder in devising or executing military action-plans, due for example to their lack of practical or theoretical understanding of military strategy, overconfidence born of mental distance from the human costs of war, or ignorance of the nuances of military processes.[89] Less specifically but in a similar vein, state leaders who have only recently come to power are more likely to perceived as incompetent than those who have been in power for a long time, especially if the fledgling leader in question has never held government office; this can be traced to assumed unfamiliarity with policy-processes, but also to the naivety and inexperience broadly associated with ‘fresh meat.’ Finally, general personality traits not specifically tied to military-political experience or expertise rather may decrease perceived competence, whether these are ascribed to a leader as a result of personal behavior or media framing; leaders who are viewed as unintelligent, boorish, or inattentive to detail are also likely to be perceived as less competent.

---

[86] Note that I am not here subscribing to the argument, sometimes made in political science, that differences in culture explain actual differences in military political behavior or outcomes: this project does not take a stand on the true role of culture in affecting military behavior and/or competence. Rather, I am arguing that regardless of its true importance, culture plays an important role in shaping perceptions of military competence.

[87] Moreover, the German and Japanese cases demonstrate that this perceived link between cultural militarism and high military-political competence is not always justified.

[88] Ex. Secretaries of State or Defense, theater commanders or other highly visible generals

For the purposes of interacting *perceived objectivity* and *perceived competence*, I treat these dimensions as binary: the combination of these dimensions yields four ideal-types (Figure 2.1). These are objective/competent (*The Rational),* objective/incompetent (*The Fool),* biased/competent (*The Fanatic),* and biased/incompetent (*The Madman).*

Figure 2.1: Ideal Type Beliefs about Judgment:

		Perceived Competence	
Perceived Objectivity		**Competent**	**Incompetent**
	Objective	The Rational	The Fool
	Biased	The Fanatic	The Madman

---

[89] See Horowitz, Allan and Stam, *Why Leaders Fight*, for a theoretical treatment of leader experience in the military and in particular the role of combat experience, and a large N database of this historical trait as it relates to war initiation

The Rational:

The first ideal type is the *Rational*: the skilled player in chess, the experienced and attentive driver in chicken. Observers expect the *Rational* to accurately interpret information (a.k.a. to interpret information in the same way as the observer) and ably construct strategies to pursue their interests within situational constraints. The *Rational* is the type of actor typically represented in theoretical strategic interaction games; in models that explicitly operate according to the rational actor assumption all involved actors truly are the ideal-type *Rational,* but even most of those that relax this assumption contain an implicit assumption of ‘common knowledge,’ meaning actors believe their counterparts are the *Rational*.

The Fool:

The second ideal type is the *Fool*: the novice player in chess, the student driver in chicken. Observers expect the *Fool* to accurately interpret information but be incapable of constructing or implementing ideal action plans, or rejecting suboptimal ones, within situational constraints. In other words, for reasons unrelated to material resources, there is uncertainty about the *Fool's* capability to match beliefs and values to policies and action plans *under its own agency*.

Western perceptions of leaders in underdeveloped or former colonial territories, for instance with anti-communist south-east Asian states during the Cold war or in modern day Africa, are an excellent example of what it means for an observers to perceive a foreign counterpart as the *Fool.* US officials, and Ambassador Muccio in particular, derided Signman Rhee and indeed the entire South Korean military leadership as inept and therefore incapable of coherently commanding their own forces, who therefore required the guiding hand of the US to avoid stumbling into self-destruction. This was not a question of whether Rhee and his commanders saw the world objectively: rather they were seen as lacking the expertise to maneuver their forces, to maintain organizational cohesion, and generally to plan and execute policies to make the best possible use of their military capacity.

Likewise, leaving aside those who condemn George W. Bush as an evil warmonger with an insatiable thirst for oil (aka those who criticize his values) many[90] who criticize Bush for what became disastrous quagmires in Iraq and Afghanistan effectively say that Bush bungled the conflict because he was a *Fool*. Different critics cite different factors to demonstrate or explain Bush's managerial incompetence; he was unintelligent as demonstrated by the fact that he did poorly in a college that only admitted him because of his father, he was unprepared and unskilled because he was an erstwhile baseball manager never trained to be president, or he was such an inept and naïve executive that he effectively divested planning responsibilities to his inner circle of hawkish neocons.

To sum, the ascribed image common to leaders who are examples of the *Fool* is not one of 'irrationality' in the colloquial sense of madness as insanity; the *Fool's* perceived sub-rationality derives from ascribed inexperience, policy ignorance, or stupidity. These leaders' military failures seem due, simply put, to the fact that they do not know what they are doing. This impairment of judgment processes is outside the scope of even Schelling's fuzzy conception of rationality, and is a good example of the additional theoretical purchase we gain by thinking about perceived madness as one end of the broader spectrum of the rationality or irrationality of an actor's judgment.

---

[90] Particularly democrat and liberal partisans but also independents and some conservatives

The Fanatic:

The third type is the *Fanatic*: the master player looking at a different board than their opponent in chess, the skilled driver looking through broken glasses and a spiderweb cracked windshield in chicken. Observers expect the *Fanatic* to misperceive information and probabilities, but to nonetheless adopt and capably implement ideal strategic action plans within their warped view of reality. Put another way, observers cannot be certain the *Fanatic* will interpret situations in an unbiased way and so adopt beliefs about the state of the world that reflect 'objective' reality, which in turn means the *Fanatic* may be unaware of situational risks or hold beliefs that cannot be shifted by evidence or argument.

The *Fanatic* ideal type is exemplified in how some ‘classic’ mad leaders were seen by their adversaries. Hitler's angry tirades and apparently rabid nationalism caused British decisionmakers to question his sanity and recognize the possibility that Germany would launch a two-front war, since Hitler might not construe risks in a 'rational' way.[91] US officials worried that Khruschev would behave in rash or unpredictable ways because his emotions, or the alcohol he regularly drank in excess, inhibited his objective perception of reality. In other words, Kruschev was believed to be a *Fanatic* because he was temperamentally and, due to his alcohol abuse, physiologically incapable of seeing the metaphorical road on which he was driving. Decision-makers in the US government similarly thought Chairman Mao was prone to warped perceptions of reality, but unlike with Khruschev they pinned this distortion not on temperament or substance abuse but on Mao’s dogmatic subscription to communist ideology. The generally accepted view in US foreign policy circles was that Mao and his inner circle of advisors in the

---

[91] Keren Yahri-Milo. *Knowing The Adversary: Leaders, Intelligence, and Assessments of Intentions in International Relations.* Princeton University Press, 2014. p. 95-97.

CCP would not believe any information unless it came directly from the Kremlin or was otherwise ‘vetted’ by the international representatives of communism.

By contrast to the *Fool*, the images ascribed to leaders who fit the *Fanatic* type more closely resemble the colloquial conception of ‘irrationality’ as madness or insanity; the *Fanatic* acts in seemingly irrational ways because they hold what observers would consider a distorted view of the world. These leaders’ may make apparently self-sabotaging or self-destructive military-political decisions not because they hold the irrational goal of self-harm, but because decisions made according to a warped view of reality often (though not always) lead to unintended outcomes when they collide with objective reality. This impairment of judgment processes falls in the implicit scope established in Schelling’s fuzzy conception of rationality. Therefore, the difference between its theoretical impacts posited by this project and those posited by extant work offers a stark, clear illustration of where and why the expectations of these theories diverge and permits us to directly compare the theories’ respective veracity.

The Madman:

The fourth type is the *Madman*: the novice looking at a different board in chess, the student driver who cannot see the road in chicken, or at the ideal-type extreme the puppy playing chess or chicken. If the prospect of combining these strategic interaction games with these actors sounds ridiculous, this is because the *Madman* is characterized by its inability to engage in rational-seeming behavior; observers expect the *Madman* to both to see the world through a cracked glass and be incapable of constructing optimal, value-maximizing action-plans even within their own conception of reality. Even if the *Madman* makes a rational value trade-off in its final decision, the judgment processes that precede, and frame, that final decision are so fundamentally flawed that said decision will probably only be one that appears rational as a result of chance or luck. The uncertainty this creates may all but eliminate the prospect not only of diplomatic engagement but, in its most extreme forms, of strategic interaction in general. The *Madman's* beliefs cannot be ascertained by examination of the situation, nor can the options it will construct be deduced from its beliefs; this is, as Goffman describes, a "game against nature."

This combination of delusional and *dummkopf* is emblematic of the image often ascribed to a particular type of 'Rogue' leaders, branded as bumbling detached from reality who might be amusing were it not for their authority over military forces. Such 'mad' rogues include Mummar Qaddafi, Kim Jong-Il and Kim Jong-Un: many western elites labeled these individuals some version of 'crazy clown,' citing everything from their unhinged public speeches to their eccentric hobbies[92] and fashion choices.[93] A *Time Magazine* cover featuring Kim-Jong Il and bearing the heading "Greetings Earthlings" aptly captures the basic content of these *Madman* images: these leaders are so biased and incompetent that they can only be understood if one looks at them as something other than human.

Yet it is not just leaders of small, relatively underdeveloped, distant dictatorships that have been branded *Madmen*: for example, the widely held image of Hitler fits this type as evidence has emerged that he was a novice military strategist whose perception of reality was distorted by jingoistic nationalism and methamphetamine abuse. In the contemporary era, there is perhaps no more pertinent example of the ascription of a *Madman* image than how President Donald Trump is perceived by his critics. From US liberals and 'Never-Trump' republicans to officials in European states that are traditionally US allies, Trump's critics say that his narcissism distorts his view of any information into a weird 'fun-house mirror reflection' of reality at which Trump himself is the gravitational center. At the same time, they deride Trump as disinterested in the details of policy and incapable of directing the government of which he is ostensibly in charge, since he came up from a reality tv background and has no experience with or apparent interest in the details of governance. To summarize the composite image of Trump formed by the aggregation of these traits in the colorful language of Jon Favreau: "Trump's presidency has always been one-part despot, two parts delusion, two parts dip-shit."[94] Indeed, critics like Favreau often express a degree of gratitude that Trump is a *Madman,* since as incompetent as he is, he cannot effectively carry out his deluded policy goals.

---

[92] Kim Jpng-Il and Un are infamous for their love of Mario and Basketball, Kim-Jong Un in particular is frequently associated with his love of the accordion, an instrument that fairly or not carries connotations of absurditity

[93] Qaddafi was very well known for his flowing gowns, elaborate umbrellas and ostentatious sunglasses, while the Kim dynasty has become synonymous with the dictatorial pantsuit.

Because the default for most strategic interaction theories is that actors perceive counterparts as the *Rational*[95]*,* introducing an actor that believes its counterpart is a *Fool, Fanatic,* or *Madman* is a significant shift from the prism through which we usually theorize interstate interactions. This departure from baseline theoretical models seems likely to cause significant alterations in the preferences, beliefs, or behavior of actors across a wide range of international issue areas. It is of course possible, if not cogently plausible, that actors treat leaders they are believe are narcissistic morons (*Madman*) identically to leaders they believe are clear-eyed strategic geniuses (*Rational*) so long as the structural position[96] of said leaders is the same. However, whether and to what extent there is a primacy of structure over beliefs about judgment[97] is a theoretical question with which IR has insufficiently grappled,[98] and an empirical question that IR has neglected to tackle at all with respect to any of the myriad areas in which this question is relevant. In order to construct a parsimonious theory, maintain a manageable project, and speak as directly as possible to the extant body of work on the 'rationality of irrationality,' I choose to focus on how variation on the spectrum of beliefs about judgment impacts what I term the *diplomatic bargaining preferences* of actors. Below, I specify the dimensions of this dependent variable in the abstract and outline observable implications of variation on said dimensions; I then explicate the mechanisms connecting beliefs about judgment and diplomatic bargaining preferences.

---

[94] *Pod Save America*, "Your President of Lawlessness and Disorder," June 4, 2020.

[95] This is evident whether looking at Fearon 1995 or Jervis 2019

[96] Power, geography, etc…

[97] As many canonical theories of international politics would us to expect

[98] Due at least in part to the difficulty of structured theorization about the impact of intangible, unobservable, or difficult to observe variables such as perceptions, beliefs. According to Waltz this difficulty extends so far as to make theorization regarding the impact of individual-level factors (leaders, personality, etc…) and even domestic governmental factors prohibitively difficult. While myriad authors have since registered their vehement disagreement with this point and posited theories that incorporate these factors, the path-dependent influence of this

Diplomatic Bargaining Preferences: Definition and Scope:

I examine how actors' beliefs about an adversary leader's judgment impact their *diplomatic bargaining preferences*, and therefore their approaches to crisis and conflict diplomacy. Diplomacy is "the nonviolent and negotiated pursuit of state interests through the communication and exchange of information, even if the threat of coercion, either economic or military, might be part of the process."[99] Information is the main currency of diplomacy, but one-sided information *gathering* does not constitute diplomacy. Active information communication is central to diplomacy, usually in the setting of interstate talks between leaders or their representatives. During these negotiations, actors communicate information to a counterpart to alter its beliefs, and thus behavior, to shift the distributive outcome of any settlement in a favorable direction. 'Talks' and 'negotiations' are here used interchangeably to avoid the attachment of any artificial judgments as to the value or effectiveness of the

---

initial rejection of non-structural variables by Waltz and his contemporaries or disciples remains apparent in the discipline. It is thus both an avenue of criticism against which this project must be prepared to respond, and an alternative argument it must address.

[99] Brian C. Rathbun, 2014, p. 12

communicative engagement to which these terms refer.[100] Essentially, they connotate active engagement in diplomacy as defined above, even if fighting continues during the period for which talks are ongoing.

Diplomatic Bargaining Preferences: Dimensions:

Diplomatic Bargaining Preferences are composed of two dimensions each of which has multiple observable implications for behavior. The first dimension, what I term *Persuasive Preference*, denotes whether an actor prefers purely informational or coercive methods of persuasion. The second dimension, what I term *Escalation Preference*, denotes the extent of an actor's willingness to escalate, resume, or start a conflict by exiting diplomacy.

Persuasive Preference denotes the degree of an actor's reliance on various methods of persuasion, which is to say of getting a counterpart to alter its behavior: argument, coercion, or violence. Violence refers to actions which actively harm the counterpart until either it voluntarily adopts the assailant's desired behavior or, because of the damages it has suffered, loses the ability to behave in undesired ways. Violence is the only method of persuasion that does not necessitate some form of bilateral communication. *Coercion* seeks to alter the target's cost-benefit calculus for certain actions by, drawing upon some existing power, explicating some punishment or reward in which said actions will result.[101] Threats to use force unless the adversary acts in a desired way, to use force if the adversary behaves in an undesired way, and the promise of reward for good behavior are all examples of coercion.[102] While coercion does not necessitate the same level of iterative dialogue as *argument*, it does require at least some form of bilateral communication, even if it is something as superficial as the sporadic trading of threats through twitter. *Argument,* by contrast, operates through rational persuasion: it seeks to alter the beliefs of its target by introducing causal information or framing known information in a novel way.[103] Actors may communicate information on any subject, though it tends to be relevant to the situation: their interests, capabilities, beliefs, intentions, or some evidence about the environment. This information need not be true, and could be communicated as hard evidence or framed as argument: as an example of the latter, people may explain the reasons why they hold particular interests in an attempt to secure said interests in a settlement. What makes *argument* distinct from both coercion and violence is that it does not draw upon power, express a conditional intent to use power, or otherwise involve mobilization of material resources. For this reason, *argument* tends to be less costly than coercion, whether measured by resources spent, credibility risked, or damage to the relationship between the negotiating parties.[104]

---

[100] Mastro, Oriana Skylar. *The Costs of Conversation: Obstacles to Peace Talks in Wartime*. Cornell University Press, 2019.

101 Robert A Pape 1996; Schelling 2008; Art, Robert J., and Patrick M. Cronin 2003.

[102] Art and Cronin 2003; Jentleson and Whytlock 2006; Pape 1996;

This does not mean that *argument* is entirely costless *or,* importantly, an immediately available method that actors can use the instant they decide to do so; some investment of time, energy, and resources may be necessary to create the requisite infrastructure for the iterative discussions foundational to argument-based persuasion. Examples of such infrastructure include: a dedicated channel for easy communication (such as the famous ‘hotline’ connecting the Kremlin to the White House set up in the aftermath of the Cuban Missile Crisis), certain multilateral international institutions which afford space for interstate discussions,[105] and the forum for face-to-face talks provided by visits, dinners, and ‘summits.’ While these *are* relatively low-cost, they must still be created by actors that have decided to engage in diplomacy via iterative argument, whether these are current or (in the case of more ‘permanent’ communication infrastructures like institutions) former government officials. We therefore cannot take their existence, and by extension the opportunity to engage in diplomacy via iterative argument, as a given.

---

[103] Grobe, Christian. "The power of words: Argumentative persuasion in international negotiations." *European Journal of International Relations* 16, no. 1 (2010): 5-29.

104 Kirgis, Paul F. "Bargaining with Consequences: Leverage and Coercion in Negotiation." *Harv. Negot. L. Rev.* 19 (2014): 69

[105] Though historically these or of suspect utility as a forum for diplomacy in acute crises or ongoing conflicts, which tend to call for more specific infrastructures exclusive to the actors involved in the conflict and possible third party mediators.

Escalation Preference denotes whether an actor prefers to pay the requisite costs of escalation or de-escalation; the starkest terms in which to think about this is the price of peace vs. the price of war. A preference for de-escalation indicates that an actor is willing to concede to adversary demands (succumb to sticks), or to offer costly benefits to the adversary (offer carrots) in order to attain a negotiated settlement or at the very least maintain diplomacy and avoid intensification of military conflict. The greater the preference for de-escalation, the greater the costs an actor is willing to pay relative to its optimal outcomes. Put another way, actors with a higher preference for de-escalation place a higher value on negotiating a cessation to or pause in conflict regardless of the content of the settlement. Conversely, a preference for escalation indicates that an actor prefers to stand firm in the face of adversary demands and abstain from offering costly benefits to the adversary even if these actions make a settlement impossible to achieve and inevitably intensify military conflict. The greater the preference for escalation, the lower the costs an actor is willing to pay to achieve a settlement or prevent the breakdown of diplomacy; in effect actors with a high preference for escalation are not that concerned with the potential costs of starting, continuing, or escalating military conflict.

As the label suggests, the scope of *Diplomatic Bargaining Preferences* is limited to preferences, attitudes, and beliefs. This caveat is important to note, since there are many reasons why actors might adopt behavior that does not match their preferences. Actors may lack the material capability to implement their preferred policies or otherwise face structural barriers to behaving in line with their preferences: an actor may prefer coercion yet lack the military capability to make credible threats. Alternatively, the behavior of counterparts may make impede acting on preferences: an actor that prefers *argument* may find this preference frustrated if the adversary is unwilling to talk, while an actor that prefers coercion may pushed by allied third parties to open negotiations. In sum, variation in an actor's diplomatic bargaining preferences will not necessarily produce variation in its behavior. Regardless of these potential exogenous impediments, however, it is not only plausible but probable that an actor's diplomatic bargaining preferences will shape its observable behavior during interstate crises and conflict. I explicate the observable behavioral implications of variance in diplomatic preferences before hypothesizing the causal relationship between my independent variable (*beliefs about judgment*) and dependent variable *(diplomatic bargaining preferences*), and explicating the mechanisms driving these theorized effects.

Beliefs about Judgment and Diplomacy: A Theory of Systematic Preference Shifts:

Having defined my independent and dependent variables, I now explicate my theory of how actors' beliefs about the judgment of adversary leaders impact their diplomatic bargaining preferences in international crises or conflicts. By adversary, I refer to a foreign counterpart[106] with divergent or opposing interests with which the observer is in open or tacit conflict: this scope excludes allied or third party states as well as non-state actors that may play a role in international conflicts. When actors engage adversaries in diplomacy they are seeking a negotiated settlement that, without initiation or continuation of violence, satisfies as many of their own and as few of their adversary's interests as possible.

---

[106] Almost always a state

By ‘leaders,’ I refer to the individual or collection of individuals in charge of policy: this usually means the titular ‘head of state[107]’ but can include a broader circle of high-ranking officials.[108] Especially in the case of states where there is less centralization of final decision-making authority, the ‘leader’ could be as expansive as the entire governing institution, though empirically in these cases actors often refer to adversary leadership as a personified monolith (e.g. ‘the Kremlin,’ ‘the Chinese,’ ‘the Politburo’) rather than individuals. My focus on leaders excludes from explicit consideration beliefs about the judgment of foreign mass publics, or categories such as race, religion, and culture that actors often use as simplifying heuristic for categorizing foreign peoples. However, beliefs about the judgment of foreign publics can influence beliefs about the judgment of their leaders; the two may even be indistinguishable in the case of observers who are unable or unwilling to think about a particular state as anything but an undifferentiated monolith.[109] In brief, my theory focuses on leaders, but the distinction between beliefs about foreign leaders and their publics is not guaranteed to be clear-cut.

In brief, I argue that beliefs about the judgment of adversary leaders shape actors’ diplomatic bargaining `preferences in ways exogenous to structure, interests, or personality; as Schelling suggested, perceived rationality and irrationality do indeed impact interstate crises and conflict. However, contrary to Schelling’s thesis, the psychological mechanisms that I argue drive this impact suggest that leaders perceived as various types of irrational are unlikely to experience improvement in their bargaining prospects because of a *Fool, Fanatic,* or *Madman* image. First, when actors believe adversary leaders are biased, this moves their *Persuasive Preferences* away from argument and towards coercion or violence: the degree of ascribed bias determines the magnitude of the preference shift. Second, when actors believe adversary leaders are incompetent, this moves their *Escalation Preferences* towards escalation and away from de-escalation: again, the degree of perceived incompetence determines the magnitude of the perceived shift. Thus, the ideal typical *Madman* is all-else-equal more likely to find its adversaries have a preference for coercive escalation, the observable behavioral implication of which is ultimatum bargaining: this is the polar opposite of the low cost bargaining successes that the rationality of irrationality suggests a madman image delivers.

---

[107] E.g. The President, Prime Minister, Supreme Anointed Leader, Queen, etc…
[108] E.g. Foreign Minister, Secretary of State, Supreme Allied Commander, Generals
[109] As Jervis posits is the general pattern amongst decision-makers and policy officials at the levers of state foreign policy.

Perceived Objectivity, Informational Efficacy, and Persuasive Preferences:

Because information is the key 'currency' of diplomacy,[110] it seems intuitive that beliefs about how counterparts perceive or interpret information, and incorporate new information into their beliefs, would have a major influence on preferences in diplomacy. I argue that the *Persuasive Preferences* of actors who believe an adversary leader is biased will move towards coercion and violence, while the *Persuasive Preferences* of actors who believe an adversary leader is objective will move towards argument. Put simply, actors are more inclined to threaten or use force against delusional leaders than against apparently reasonable leaders. This relationship is driven primarily by the effects of perceived bias vs. objectivity on the expected persuasive efficacy of complex vs. simple information: I refer to this value as the perceived *complexity cut-off* for persuasive information. Perceived bias lowers the complexity cut-off and thereby the expected persuasive efficacy of complex information, while perceived objectivity raises the complexity cut-off and therefore the expected persuasive efficacy of complex information. Before discussing why this happens and positing explicit hypotheses, it is necessary to establish what makes complex and simple information distinct and why this distinction matters when we consider the logic of their respective use in interstate diplomacy.

---

[110] Rathbun 2014

Simple information is the primary informational mode of persuasion through coercion and violence. It constitutes novel material information regarding relative power (e.g. the revelation of hidden military forces), threats of material action ('I will bomb you,') or information that otherwise draws upon power or the intent to use it. In other words, simple information has a material foundation that makes it more universally comprehensible: this material foundation usually comes in the form of military, economic, territorial, or even political resources. These material foundations can be revealed as implicit leverage, promised as an incentive for desired behavior (e.g. delivery of desired capital or resources, inclusion in an international body, or their destructive use can be threatened as a punishment if desired behavior does not manifest. In a case where the disputed issue is a piece of territory, persuasion through simple information could be a threat to take the territory by force, a threat to inflict punishment on the adversary if the territory is not conceded, or an offer to buy the territory.[111] The threats and trades that constitute simple information are the primary currency in most formal bargaining models of interstate diplomacy, crises, and conflict bargaining.

Complex information is the primary informational mode of persuasion through argument. It comprises the kind of 'pure information' fundamental to ideal-typical rational argument: for example, presentation of evidence, interpretive frames for existing information, or subjective opinions and philosophies. Usually persuasion through complex information operates by convincing a counterpart that its interests are better served by the persuader's desired behavior, but without attaching novel material considerations to artificially shift the adversary's valuation of whatever issues are at dispute. In a case where the disputed issue is a piece of territory, persuasion through complex information might seek to convince the adversary that the territory actually is not valuable ('its natural resources are spent,') or that it should belong to the persuader because of ancestral rights, ethnic and linguistic composition, or similar reasons. The ethnically and historically grounded arguments Hitler used to justify the legitimacy of his initial expansions are an excellent example: this was a use of complex information to persuade counterparts that these areas *should* be part of Germany.

---

[111] Basically this is the model of persuasion expressed in the coercive equation put forward by Robert Pape in *Bombing to Win*:

Simply put, constructive dialogue is complicated, while coercive carrots and sticks are simple. This logic does not imply that simpler information is more effective, but rather that when the complexity-cut-off is low an actor can trust that simple but not complex information will convey to the counterpart the consequences of shifting their behavior. "If you do not give me what I want I will hurt you" is an easy message to understand and so somewhat universally effective at getting the intended point across. By contrast, "this land is sovereign German territory not only historically but based on the manifest desires of the majority German population that lives there and was forcibly annexed in an unjust war settlement" is much more complicated. The persuasive efficacy of that piece of information depends upon shared worldviews, convergent interpretations of history and of novel information, common frames of reference for concepts like sovereignty or justice, actors with beliefs that are open to change, and a host of factors the existence of which cannot be universally assumed.

Perceived bias creates expectations of divergent ways of thinking and interpreting information, thereby undermining the aforementioned factors that increase the expected efficacy of complex information. The higher the level of bias observers ascribe to an adversary leader, the lower the complexity-cut-off for comprehensible information and the more likely said observer is to rely on simple information to affect desired changes in the leader's behavior. Thus, perceived bias shifts *Persuasive Preferences* towards coercion.

H1: Actors that believe an adversary leader is biased will have a *Persuasive Preference* for coercion rather than argument; the greater the bias ascribed to the adversary leader, the greater the magnitude of the shift toward coercion.

This coheres with psychological research on the role of bias-perception in creating conflict spirals in intergroup and interpersonal interaction.[112] Biased others may be expected to hold fixed beliefs or interpret information in flawed ways, rendering pure or complex information an unreliable tool for influencing behavior through beliefs. If belief divergence is a barrier to settlement, actors who see their adversary as biased are unlikely to expect the sharing of complex information sharing lead to settlement.

Research in both interpersonal and intergroup settings empirically demonstrates this complexity cut-off mechanism for the impact of perceived bias on approaches to bargaining. People hold more pessimistic attitudes about the plausibility of achieving compromise or understanding through iterative dialogue with people they believe are biased, as they do not expect novel information or well thought out arguments to change the beliefs of biased counterparts. Indeed, people often prefer to not even *attempt* to reason or share information with those they perceive as biased. The majority of research in this area has been conducted in the context of policy debates between US partisans on opposite ends of the political spectrum:

---

[112] Emily Pronin, Kathleen Kennedy, and Sarah Butsch 2006; Kathleen Kennedy and Emily Pronin 2008; Gilovich, Thomas, and Ross 2015.

studies consistently demonstrate several behavioral patterns across issue areas from climate policy to abortion law. Subjects presented with a counterpart that disagrees with their policy opinion respond differently to this divergence depending on what they are told about the counterpart's partisan identification. Subjects informed the counterpart shares their partisan identification are more likely to conclude the counterpart is misinformed or underinformed, and to express an expectation that the policy disagreement can be resolved through information provision and rational discussion. By contrast, subjects informed that the counterpart subscribes to an opposing partisan identification are more likely to conclude the counterpart is irrationally devoted to their opinion.[113] When people explain issue-specific divergences of policy opinions as a product of their 'opponent's' biases, they are less likely to prefer dialogue-based means of resolving the disagreement, since they believe information provision and rational discussion would be fruitless.[114]

Research by scholars including Gilovich, Ross, and Pronin demonstrates that the consequences of bias-perception extend beyond 'civil' political disagreements and into violent intergroup conflict.[115] Compared with adversaries they believe are rational/objective, people prefer violent to non-violent means of conflict resolution with adversaries they perceive as biased. For example, Pronin conducted an experiment where subjects saw descriptions of terrorist decision-making as either rational/objective or biased/irrational, then were asked their preferences for violent vs. non-violent methods of counterterrorism. The results showed "when terrorists were depicted as biased and irrational... participants were more likely to advocate military action and threats against terrorism and less likely to advocate diplomacy. This effect was mediated by perceptions of terrorists' capacity for reason" rather than either terrorists' intentions or positive or negative feelings towards terrorists.[116]

---

[113] Benforado, Adam, and Jon Hanson. "Seeing bias: Discrediting and dismissing accurate attributions." *Ideology, psychology, and law* (2012): 453. Ehrlinger, Joyce, Thomas Gilovich, and Lee Ross. "Peering into the bias blind spot: People's assessments of bias in themselves and others." *Personality and Social Psychology Bulletin* 31, no. 5 (2005): 680-692.

114 Robert J. Robinson, Dacher Keltner, Andrew Ward, and Lee Ross. "Actual versus assumed differences in construal:" Naive realism" in intergroup perception and conflict." *Journal of Personality and Social Psychology* 68, no. 3 (1995): 404.

115 Thomas Gilovich and Lee Ross. *The Wisest One in the Room: How You Can Benefit from Social Psychology's Most Powerful Insights*. Simon and Schuster, 2016.

To sum, subjects do not anticipate the kind of reasoned discussion implicit in words like diplomacy, debate, and dialogue to be a viable persuasive tool with apparently biased counterparts. Yet the same subjects are more likely to advocate or rate as effective methods of persuasion through simple-information, such as monetary incentives and threats of harm, against counterparts they perceive as biased. Therefore, the empirical record does not support Schelling's implicit contention that the mechanism through which perceived bias affects conflict bargaining is by making it seem more difficult, or even impossible, to affect desired change in the behavior of apparently biased actors. Instead, extant work suggests that perceived bias affects approaches to conflict bargaining by shifting expectations about the types of information that have a reasonable chance of affecting desired changes in the behavior of apparently biased counterparts; it alters the tools required for the task of persuasion without necessarily impacting the degree of difficulty involved in the task.

Conversely, perceived objectivity creates expectations of convergent ways of thinking and interpreting information, thereby increasing the expected persuasive efficacy of complex information. The higher the level of objectivity observers ascribe to an adversary leader, the higher the complexity-cut-off for comprehensible information and the more likely said observer is to first attempt to use complex information to affect desired changes in the leader's behavior.

---

[116] Emily Pronin, Kathleen Kennedy, and Sarah Butsch. "Bombing versus negotiating: How preferences for combating terrorism are affected by perceived terrorist rationality." *Basic and Applied Social Psychology* 28, no. 4 (2006): 385-392.

Thus, perceived objectivity shifts *Persuasive Preferences* towards argument. When belief divergence is a barrier to settlement, actors who see their adversary as objective are likely to attribute this to missing rather than misinterpreted information, and so expect complex-information sharing to lead towards settlement by moving beliefs towards a point of convergence.

H2: Actors that believe an adversary leader is objective will have a *Persuasive Preference* for argument; the greater the objectivity ascribed to the adversary leader, the greater the magnitude of the shift toward argument.

Even before we incorporate variation in perceived competence, **H1** and **H2** have clear observable implications for the approaches to interstate diplomacy we can expect actors to advocate.

Actors who believe an adversary leader is biased should be extremely skeptical of iterative diplomatic dialogue; there are several ways we can expect this coercive *Persuasive Preference* to manifest in the behavior of actors who hold it because the ascribe bias to the adversary. First, these actors should express reservations about or outright opposition to the creation of any new communications infrastructure, based on their expectation that it is pointless or counterproductive to arrange the opportunity for argument-based rational dialogue. They will push back against face-to-face meetings, the organization of formal summits, or the establishment of new electronic or institutional channels. If this infrastructure for iterative communication either already exists or is established despite concerted opposition, for example if arrangements are made for formal negotiations, these actors will press for blunt, materially based approaches to persuasion. This includes real or threatened cost-infliction: for example, economic or diplomatic sanctions, blockades, destruction of infrastructure, and use of force against adversary military or civilians. Conversely, it includes real or promised material benefits: for example, monetary incentives, territory grants, prisoner swaps, and troop withdrawals. In brief, actors who believe the adversary is biased will advocate using the paradigmatic 'carrots and sticks' of rationalist bargaining models, with the precise balance and form of carrots and sticks depending on what is feasible given the balance of material assets in the conflict at hand. They will likely take this stance before any dialogue takes place, which is to say before the costless approach has been attempted, but especially if dialogue is 'dragging' or considerations of relative material power apparently favor their own side.

By contrast, actors who believe an adversary leader is objective should be optimistic about and supportive of iterative diplomatic dialogue; there are several ways we can expect this *Persuasive Preference* for argument to manifest in the behavior of actors who hold it because they ascribe objectivity to the adversary. First, these actors should push for the creation of new communications infrastructure, based on the belief that dialogue is less costly than coercion and their expectation that engagement in an argument-based rational dialogue would enable the sides to reach agreement on the best path forward for their respective interests. They will advocate face-to-face meetings, the organization of formal summits, or the establishment of new electronic or institutional channels for dialogue. In essence, they prefer to create maximum space for dialogue because they believe rational information sharing will move the adversary's beliefs, or even conception of its own interests, such that it opts to alter its behavior without being coerced or forced into doing so. During talks, actors who prefer persuasion through argument should advocate open communication with adversary leadership. Of course, the nature of iterative diplomatic dialogue makes the specific contours of the communications they are likely to advocate difficult to predict, as these are dependent on the actors involved and the issues at hand.

In general, they should be more likely to advocate approaches that answer questions of *why* as well as *what:* for instance why do they hold particular interests, why do they think the adversary should recognize and grant these interests in a settlement, and why do they think a settlement in line with their own interests would be best for all parties involved?[117] By extension, they will oppose immediate resort to the material instruments of persuasion that coercively inclined actors are likely to advocate, as these may poison the metaphorical well of diplomacy via rational dialogue: argumentatively inclined actors should prefer to let diplomacy take its course.

To sum, actors who believe the adversary is biased will oppose complex dialogue in acute conflict-pausing negotiations, and advocate an indirect (a.k.a. the language of violence) rather than direct (the language of dialogue) approach to conducting whatever ongoing bargaining conversation is occurring in the course of conflict.[118] By contrast, actors who believe the adversary is objective will advocate complex dialogue in acute negotiations, as well as conducting ongoing bargaining conversations through explicit communication rather than simply the implicit communication of fighting.[119]

---

[117] These are some of the characteristics of what contract negotiation theory calls "value creating" negotiations: negotiations wherein participating parties find creative ways to expand the total value of the metaphorical pie they are disputing, rather than simply agreeing on a way to divide a pre-existing pie. Rathbun used this frame when describing the contours of rational, iterative diplomacy: see Rathbun 2014.

[118] In other words, they will be satisfied with the theater of war as the only bargaining table and uses of force as the principle method of communication.

[119] Which is to say, they will not be satisfied with ongoing bargaining only in the form of conflict as bargaining, a now widely used theoretical intervention to the bargaining model of interstate conflict that posits war itself is essentially communication at the bargaining table. For actors with a *Persuasive Preference* for argument, this is not untrue, but it is not sufficient to meet their preferred approach to ongoing bargaining which they should advocate involve both the implicit communication of violence or other costly signals and the explicit communication of dialogue and other so called costless signals.

Perceived Competence, Costs of Conflict, and Escalation Preferences:

Variation on the perceived competence dimension impacts actors' *Escalation Preferences* through their expectations regarding the costs and risks of resuming or escalating conflict; perceived incompetence moves preferences towards escalation, while perceived competence moves preferences towards de-escalation. A principle purpose of persuasion in the context of diplomacy is to obviate the need to fight a conflict to its conclusion; the incentive to reach diplomatically negotiated settlements comes from the expected costs of continued conflict.[120] Fighting is inherently harmful to all participants, so state leaders should prefer to achieve aims nonviolently and weigh any potential benefits they could attain through violence against the expected costs of the violence itself.[121] Actors' *Escalation Preferences* should all else equal be more de-escalatory if they expect the costs of conflict to be high, and escalatory if they expect the costs of conflict to be low. While relative material power influences how severe or negligible these costs are expected to be, it is less determinative than extant theory suggests. Importantly, the costs of conflict are also significantly impacted by how competently or incompetently each side utilizes the raw resources at its disposal.

Experimental evidence supports this general contention. Research suggests that the level of threat people perceive from a 'cold' or hostile counterpart is dramatically influenced by the level of competence they ascribe to the counterpart in question.[122] Perceived competence in these studies is indicative of a counterpart's expected ability to advance its aims, so when these aims are at odds with those of the observer (as is definitionally the case with adversaries) an apparently competent counterpart poses a greater threat than an apparently incompetent one. Provided they believe they have the capacity to do so, people actively attempt to impede competent adversaries, though absent that capability the preference for active interference is often subordinated to the practical necessity of grudging accommodation. By contrast, people tend to dismiss or neglect incompetent adversaries; while people see the goals of such counterparts as at odds with their own, the possibility that such incompetent others will achieve their aims is dismissed as low and so not a threat to take seriously.

---

[120] See Fearon 1995 on the inherent costs of conflict and why this should logically lead rational actors to settle differences through bargaining.

[121] There are exceptions to the contention that violence and conflict are always costly or reduce the overall value of resources available to both sides. For example, actors might enjoy violence and value it for its own sake, or more commonly derive important domestic political benefits from violence. In these cases actors are not so much as weighing the potential benefits of violence against the inherent costs but rather the inherent and potential benefits of violence against the risk of a conflict ending in their loss.

[122] Fiske, Susan T., Amy JC Cuddy, Peter Glick, and Jun Xu. "A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition." *Journal of personality and social psychology* 82, no. 6 (2002): 878. Cuddy, Amy JC, Susan T. Fiske, and Peter Glick. "The BIAS map: behaviors from intergroup affect and stereotypes." *Journal of personality and social psychology* 92, no. 4 (2007):

The interstate corollaries for these interpersonal effects of perceived competence are obvious, their implications intuitively persuasive, and suggest that beliefs about the military-political competence of adversary leadership shape actors' *Escalation Preferences* in ways exogenous to material capabilities or the balance of interests.

Incompetent adversaries are likely to bungle their strategy, misuse resources, and generally make mistakes: conflict with these adversaries is therefore a threat that can be more easily dismissed. Actors facing apparently incompetent adversaries will be more ready to walk away from the negotiating table, willing to make fewer accommodations to the adversary to achieve a settlement, and generally less desperate to end conflict. Because fighting for everything they want is less costly than sacrificing some of what they want to avoid the costs of fighting, actors

---

631. Cuddy, Amy JC, Susan T. Fiske, and Peter Glick. "The BIAS map: behaviors from intergroup affect and stereotypes." *Journal of personality and social psychology* 92, no. 4 (2007): 631.

who believe their adversary is incompetent are likely to have more escalatory *Escalation Preferences.*

H3: The more incompetent observers believe an adversary leader to be, the greater their preference for escalation, and the lower the costs they will pay for de-escalation before reverting to escalation and *de-facto* abandoning negotiations

Competent adversaries are likely to conduct themselves capably in a military conflict, making optimal use of the resources at their disposal: such conflict is thus a potentially serious threat that actors should seek to avoid unless they are certain of success, even at the cost of accommodating some of the adversary's interests. Actors facing apparently competent adversaries will be more eager to come to the negotiating table, slower to walk away from negotiations, willing to make more accommodations to the adversary to achieve a settlement, and generally more desperate to end or at least avoid escalating conflict. Because fighting for everything they want, even if successful, is likely to be more costly than accepting only a piece of their desired 'pie' to end violent conflict, actors who believe their adversary is competent are likely to have de-escalatory *Escalation Preferences.*

H4: The more competent observers believe an adversary leader to be, the greater their preference for de-escalation, and the greater the costs they will pay for de-escalation before reverting to escalation and *de-facto* abandoning negotiations

There are limited observable implications we can draw for interstate diplomatic bargaining from **H3** and **H4** alone, since the policies used in escalatory and de-escalatory bargaining respectively depend on the implementing actor's *Persuasive Preferences*. However, we can make some basic behavioral predications from **H3** and **H4.** Actors who ascribe incompetence to adversary leaders and therefore prefer an escalatory approach should generally dismiss the military threat posed by, and the urgency of reaching a settlement with, the adversary. During internal deliberations, these actors will express confidence that the adversary can be beaten in the field, that escalation of conflict can be managed, and by extension that there is no need to make extensive concessions to the adversary in the course of negotiations. During negotiations, they will be more likely to ignore adversary attempts at argument, to scoff at the seriousness of any threats or demands made by the adversary, and to advocate abandoning negotiations if progress is slow or a settlement costly. By contrast, actors who ascribe competence to adversary leaders and therefore prefer a de-escalatory approach should take seriously the military threat posed by the adversary and treat negotiating a settlement as an urgent goal. During internal deliberations, these actors will express uncertainty that the adversary can be militarily defeated, at least at acceptable cost, or that the escalation of conflict would be manageable: they will therefore advocate making concessions or offering costly incentives to the adversary in the course of negotiations. During negotiations, they will be more likely to take seriously any threats or demands made by the adversary, and to oppose withdrawal from negotiations even if progress is slow or a settlement costly.

Who is the Observer?: The Theoretical Links Between the Public and Political Elites:

To this point, the theory has not explicitly distinguished between members of the public and the political elites who are directly responsible for setting diplomatic policy or carrying out negotiations, which begs the question: can the same theory of individual preference shifts and behavior apply to both elite decisionmakers and members of the public? A longstanding criticism of political psychology, and experimental methods in particular, is that findings at the individual level cannot be expected to apply to elites because of systematic differences between these groups.[123] Whether the posited mechanism is self-selection into elite circles by certain types of individuals or that membership in the elite inculcates specific modes of thinking through socialization or education, the resulting implication is the same: separate theory is needed at the individual and elite levels.[124] In contrast, a number of recent studies show little distinction between ‘mass public’ and ‘elite’ samples across a range of experimental conditions, so the extent and scope of an ‘aggregation problem’ remains unclear and should be specifically addressed for each theory.[125] Nevertheless, there are several reasons I expect my hypotheses to apply to both elites and members of the public.

First, evidence from historical and contemporary cases suggests that leaders, particularly in democratic states, monitor public opinion and use it to inform their decisions in international crises, from the opening or cessation of diplomacy to the initiation, escalation or termination of military conflict.[126] President Eisenhower’s brinksmanship in the 1958 Taiwan crisis was tempered when he realized the US public did not consider defending Quemoy and Matsu to be worth the costs of war. Johnson’s Great Society initiative, and eventually his entire presidency, was crushed by unceasing attacks on his approach to Vietnam that turned the public against a military approach. Whatever the wisdom of this common historical narrative, Johnson evidently found it persuasive: he subsequently remarked that foreign military endeavors are lost at home rather than abroad, a notion later formalized in the Powell Doctrine, which explicitly enshrines widespread public support as necessary for the use of force.[127] Therefore, understanding the general public's support for diplomacy versus force, and how they form expectations about the efficacy of these foreign policy approaches, is of considerable importance in understanding elite choices and outcomes in interstate diplomacy and crises.

---

[123] Powell 2017

[124] Hafner-Burton, Emilie M., D. Alex Hughes, and David G. Victor. "The cognitive revolution and the political psychology of elite decision making." *Perspectives on Politics* 11, no. 2 (2013): 368-386. Mintz, Alex, Steven B. Redd, and Arnold Vedlitz. "Can we generalize from student experiments to the real world in political science, military affairs, and international relations?." *Journal of Conflict Resolution* 50, no. 5 (2006): 757-776.

[125] Mintz, Alex. "Foreign policy decision making in familiar and unfamiliar settings: An experimental study of high-ranking military officers." *Journal of Conflict Resolution* 48, no. 1 (2004): 91-104; Fatas, Enrique, Tibor Neugebauer, and Pilar Tamborero. "How politicians make decisions: A political choice experiment." *Journal of Economics* 92, no. 2 (2007): 167-196; Linde, Jona, and Barbara Vis. "Do politicians take risks like the rest of us? An experimental test of prospect theory under MPs." *Political Psychology* 38, no. 1 (2017): 101-117.

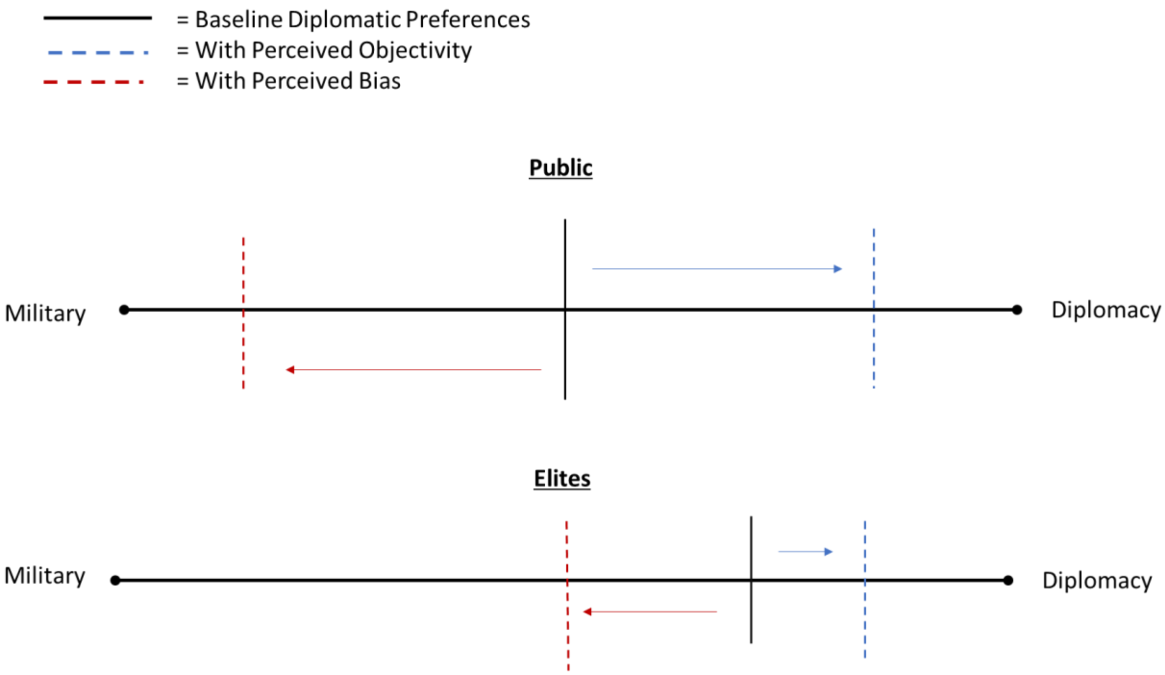
[126] Kertzer 2016; Christensen, Eben J., and Steven B. Redd. "Bureaucrats versus the ballot box in foreign policy decision making: An experimental analysis of the bureaucratic politics model and the poliheuristic theory." *Journal of Conflict Resolution* 48, no. 1 (2004): 69-90.

Second, this is a theory of how variation in beliefs about judgment *move* diplomatic preferences: it does not claim to determine exactly where on the spectrum of diplomacy versus force an individual's preferences fall. Whatever an individual's baseline diplomatic preferences, there is no a priori reason to expect the belief that an adversary's judgment is a particular type to shift the preferences of elites in a *different direction* than the preferences the public: the same belief should move diplomatic preferences of elites and the public in the same direction. This is not to say the theory predicts that elites and the public will have identical attitudes towards diplomacy versus force; it is possible, and indeed likely, that in any given situation of interstate conflict elites and mass publics will have different baseline diplomatic preferences. To list just a few differences that could contribute to this divergence, elites have more information about strategic factors, likely account for a wider range of different variables in their thinking, and have incentives distinct from those of the public. I illustrate the logic of this claim in Figure 2.2 below.

---

[127] Weinberger 1984

Figure 2.2: Public vs. Elite Shifts


= Baseline Diplomatic Preferences
= With Perceived Objectivity
= With Perceived Bias
Public
Military
Diplomacy
Elites
Military
Diplomacy

. Figure 2.2 underscores several basic points. First, variation in beliefs about the judgment of adversaries move the diplomatic preferences of individuals in the same direction whether those individuals are elites or members of the public. Second, the baseline diplomatic bargaining preferences of elites may differ from those of the public, since when forming persuasive and escalation preferences populations consider different information, weight the same information in distinct ways, and operate according to divergent incentives. Third, and relatedly, the magnitude of the shift may be smaller for elites given the wider array of factors contributing to their diplomatic bargaining preferences. However, the extent of this difference is an empirical question, and given the difficulty of causal inference in international relations the direction of causal effects is almost always more important than their size. Finally, the aggregation of the aforementioned considerations suggest that elites and members of the public who share beliefs about an adversary's judgment may nonetheless end up at different spots on the spectrum of diplomatic preferences.

As a hypothetical illustration, consider a situation where State A finds itself in an acute crisis with State B after State B encroaches on State A's territory. The leaders of State A have access to classified information suggesting the military capabilities of State B are more advanced, and thus armed conflict would be costlier, than is widely known. Between this information and concern about the electoral consequences of starting a war, State A's leaders are more pro-diplomacy than its public, which wants State B to retract its invasion but is apathetic about how this is accomplished. However, a leader comes to power in State B who is widely perceived to be biased: this makes the leaders of State A pessimistic about the prospects of successful diplomatic de-escalation such that they are effectively neutral on the question of diplomacy versus force. Meanwhile, the public in State A, who give less thought to the costs of conflict and believe force is the only language the adversary understands, become actively pro-force. The possible consequences of this shift, from diplomatically inclined leaders of an apathetic public to leaders with neutral diplomatic preferences under public pressure for military action, are obvious.

Observable Implications, Caveats and Conditions of the Theory:

A number of predictions about state behavior can be derived from applying **H1-H4** to both the individual elites who are directly responsible for setting as well as enacting state policy, and individual members of the mass public whose foreign policy attitudes influence elite decision-making and thereby indirectly shape state policy. Below I discuss these predictions for each of the ideal-type beliefs about adversary judgment (*Rational, Fool, Fanatic, Madman)* with reference to real world examples. I then discuss scope conditions before directly comparing these predictions with those explicated or implied by the rationality of irrationality thesis, as well as several other prominent alternative arguments.

While as detailed above the hypothesized preference shifts apply to both elites and the public, the observable behavioral implications of these shifts will differ somewhat between these groups due mainly to their respective sophistication regarding the details of foreign policy and interstate bargaining. Relative to the elites actual designing or implementing policy, we can expect that the behavior of individual members of the public will change in ways that are general and non-specific, in line with the average individual's grasp of the details of foreign policy. For example, while most members of the public distinguish between military and diplomatic methods and so we can draw observable implications about their behavior along this dividing line, they are less likely to see meaningful distinctions between various manifestations of coercion.

Bargaining with The Rational:

Against adversary leaders who are seen as *The Rational*, state policy should resemble a serious effort at dialogue in which state officials are committed to the traditional give and take of diplomacy as practiced between rational, respectful statesmen and women. Mass publics should generally be more likely advocate diplomatic engagement with and oppose war against states they believe are led by *Rational* types, as these states are serious adversaries that must be carefully contended and can feasibly be reasoned with: the public is less likely to support hostility against such states before and perhaps even if the path of diplomatic dialogue has been exhausted. Both due to and independently of these public attitudes, state elites should prefer to use iterative argument through complex information as primary language through which to conduct diplomacy with the *Rational* type. While a shift away from dialogue is always feasible, these elites should use simple-informational methods of coercion and carrots only sparingly or as a last resort given the certain costs of carrots and sticks and the potential costs of escalating conflict with a dangerous opponent.

With respect to specific behavior, states in a real or potential conflict with the *Rational* should engage in an all-hands-on-deck effort at diplomacy. We should see efforts to either communicate an interest in conducting or to actually conduct diplomatic dialogue through a wide range of plausible channels. For example, direct outreach from diplomatic personnel to their counterparts (perhaps through a dedicated embassy if one is available), conversations organized on the sidelines of larger international forums (e.g. the UN) or through third party officiators, or verbal and written communication between representatives of relevant military, intelligence, or other non-diplomacy specific agencies on either side. Moreover, we should see attempts to organize specific forums for dedicated diplomacy, almost always first between relevant representatives and personnel but perhaps with the eventual goal of a face-to-face meeting between leaders and even a formal diplomatic summit.

During diplomatic negotiations, we should expect state officials interacting with *Rational* leaders or their representatives to support a genuine dialogue in both internal deliberations and formal negotiations. They will be more likely to speak about shared interests that can be achieved through cooperative effort, talk more in terms of reaching an understanding than forcing an agreement, and oppose steps they fear could destabilize negotiations. Elites should speak about the adversary in respectful terms that suggest perceived objectivity and competence: for example, ‘practical,’ ‘reasonable,’ or ‘intelligent.’ Such descriptors should be particularly common when elites are making the case for the persuasive efficacy of argument, expressing optimism about the prospects for diplomatic dialogue, or otherwise trying to convince skeptics that their preferred approach (de-escalatory diplomatic argument) is best.

Many well-known instances of détente were arguably examples of the state behavior we would expect when elites and/or the public adopt the belief that the adversary leader is rational, whether this represents an evolution in their perceptions of longstanding adversary leadership or occurs due to a change in the leadership of either the observer or adversary state. Margaret Thatcher's impression of Gorbachev as a more rational kind of Soviet leader, one still very much devoted to and produced by his own system but nonetheless a reasonable man and someone she famously remarked she "could do business with," was a major factor in the UK's shift towards a frank and open diplomatic dialogue with Moscow to seek areas of cooperation for mutual benefit. This preceded, but ultimately mirrored in many ways, the logic of the notoriously anti-Communist Regan administration's shift towards argument and dialogue-based diplomacy with the Soviet Union in the aftermath of Gorbachev's rise to power in 1985. Despite the favorable impression Regan's domestic advisors and international allies had of Gorbachev as more reasonable and less dogmatic than his predecessors, Regan maintained his belief that the USSR under the new Soviet Premier was an intractable and enemy. Yet five weeks after Gorbachev rose to power Regan noted, in contrast to his position with Gorbachev's predecessor, that "we should work hard to establish channels directly between Gorbachev and me through quiet diplomacy,"[128] a paradigmatic example of the prediction that actors facing *Rational* leaders will create new communication infrastructure to permit iterative dialogue in pursuit of more "constructive relations."[129] At the 1985 Geneva summit, this shift in Regan's perception of Gorbachev was apparently solidified: the face-to-face meeting helped humanize Gorbachev to the President. Regan left Geneva convinced that while Gorbachev's worldview was at odds with his own, the Soviet Premier was neither delusional nor unreachable; what is more, Regan believed that he had the capacity to use his powers of persuasion to affect change in Gorbachev and through him the USSR. This belief that Gorbachev was *Rational,* and Regan's resultant preference for diplomatic dialogue, informed US policy toward the USSR to the end of Regan's presidency and beyond. The negotiating processes through which the US and USSR voluntarily came to agreement on the mutually binding arms reduction and control treaties that are held up as the gold-standard of détente are ideal-typical instances of the de-escalatory diplomatic dialogue actors prefer when facing apparently rational adversaries.

---

[128] Farnham, Barbara. "Reagan and the Gorbachev revolution: Perceiving the end of threat." *Political Science Quarterly* 116, no. 2 (2001): 225-252.

[129] Farnham 2001

Bargaining with The Fool:

Against adversary leaders who are seen as *The Fool,* state diplomatic policy should constitute an opportunistic attempt at dialogue in which officials strive to advance their state's interests 'on the cheap' by taking advantage of a gullible, inexperienced foreign counterpart. Mass publics should generally be less concerned about conflict with states they believe are led by *Fool* types, meaning they may be more supportive of escalation in the event of a salient conflict of interests, though with the exception of especially jingoistic individuals we would expect most people to prefer at least trying diplomatic dialogue before coercion if only because the former is less costly. State elites should prefer to use complex information as the primary language through which to conduct diplomacy with the *Fool* type, as it represents an opportunity for great gain at minimal cost. Like with the *Rational*, a shift away from dialogue to coercion if the former is proving fruitless is plausible. Yet this coercion should be mostly or entirely stick and no carrot, and representatives should be more open to walking away from the negotiating table, since there is less incentive to make concessions, pay costs, or exercise patience to avoid conflict escalation with *Fools*. Essentially, *Fools* are more pests to be ignored or opportunities to be exploited than enemies to be feared.[130]

With respect to specific behavior, states in an adversarial relationship with the *Fool* should engage in opportunistic and in some sense casual diplomacy: officials may not think much before starting a dialogue or, conversely, ending one. Overall, we should see less active or urgent outreach to the *Fool* than to the *Rational*, since the prospect of conflict with the *Fool* is not so concerning as to create an overriding existential incentive to engage in diplomacy. However, officials who are especially confident of their capacity to manipulate the *Fool* should press for the opening of communication channels in much the same way as officials bargaining with the *Rational*. Moreover, even officials who are not sufficiently certain of the adversary's susceptibility to manipulation to be the first-movers in opening dialogue should be open to offers by the adversary or entreaties by third parties to open a dialogue: what, after all, do they have to lose from engaging in diplomatic argument with an actor who by all appearances is sufficiently objective to accurately interpret complex information?

During diplomatic negotiations, we should expect state officials interacting with *Fools* or their representatives to speak about shared interests, extol the value of cooperation, and proffer respect and praise for the adversary in face-to-face interactions: all of this, however, is more likely than with any other ideal-type to be at least somewhat disingenuous. Put another way, officials will speak as if committed to achieving a mutually beneficial settlement through diplomacy as a way of maximizing the advancement of their own interests at minimal cost. In internal deliberations, the same officials who promote cooperation and sing the praises of their counterparts in interstate communications are likely to take a far more derisive and dismissive view of the *Fool*. They may describe the adversary using words like ‘novice,’ ‘green,’ ‘untested,’ or ‘idiotic,’ particularly when 1) discussing the prospect of using argument to convince the adversary to behave in desired ways or 2) playing down the risks of abandoning negotiation and escalating conflict if the adversary does not change its behavior in response to rational arguments.

---

[130] A corollary in extant theory are so-called barbaric nations in Hermann et. Al's original 1987 Image Theory paper.

While the case is too current for us to have access to detailed archival records, insider accounts and journalistic reporting on the Trump presidency suggest that there are many foreign leaders who believe Trump is a *Fool* and approach diplomatic bargaining with him accordingly. Leaders like Vladamir Putin and Xi Jinping seem to treat diplomacy with Trump or his representatives as an opportunity to, through complex-informational methods that avoid the need for costly material incentives or punishments, convince a leader too inexperienced with and disinterested in the details of policy to know better to act in accordance with their own goals. Put another way, to quote former national security advisor John Bolton, Putin “thinks he can play Trump like a fiddle” because he does not see the man sitting across the metaphorical (or in some notable cases literal) table from him as a strategic or intellectual equal. Putin allegedly swayed Trump to support Venezuelan President Maduro over Juan Guaido, the opposition leader backed by official US policy, by comparing Guaido to Hillary Clinton. Erdogan, the President of Turkey, reportedly used a single phone-call of flattery to dissuade Trump from intervening on behalf of the Kurds when Turkey began a brutal campaign against the US allied Kurdish forces and their civilian dependents. These may not seem particularly complicated, but both these methods fit the definition of complex information as operating without a material foundation: there was no penalty, monetary or otherwise, that Putin was introducing to convince Trump to *de facto* pull his backing for Guaido, for example. While these foreign leaders who reportedly see

Trump as a *Fool* are willing to talk given the expected prospect of significant gain at minimal cost, they have also been quite willing to walk away from negotiations on such substantive issues as trade, arms control, and the Syrian Civil War. They do not take Trump's threats seriously, to such an extent that Erdogan reportedly threw the President's threat in the trash,[131] and apparently see little incentive to make costly concessions for the purpose of achieving major settlements.[132] Whether or not the above accounts turn out to be accurate or an incomplete reflection that fuller archival evidence eventually casts into doubt, they illustrate the opportunistic dialogue we would expect states dealing with a *Fool* to engage in.

Bargaining with The Fanatic:

Against adversary leaders who are seen as *The Fanatic*, state policy should resemble what we traditionally think of as coercive diplomacy;[133] a cautious trading of threats, reassurances, or promised concessions wherein actors work to advance the interests of their state as best as possible without moving up the escalation ladder or unnecessarily 'rocking the boat.' Mass publics should generally be more likely to oppose diplomatic engagement with and advocate the threat and use of force against states they believe are led by *Fanatic* types, as these states are serious adversaries that cannot feasibly be reasoned with. The public is more likely to support hostile measures (sanctions, force) against such states before diplomatic dialogue has been exhausted, though given the danger of unchecked escalation with a competent opponent they are also likely to support concessions if necessary to avoid war. Independently of and in response to these public attitudes, state elites should prefer to use coercion conducted through simple, materially based information as the primary language through which to conduct diplomacy with the *Fanatic* type, whether coercive information is conveyed with actions alone or through formal communication channels that could also host argument-based diplomacy. Despite the inherent costs of carrots and sticks relative to purely informational dialogue, elites should express this preference for a coercive approach both before any attempt at diplomatic dialogue and during any dialogues that occur despite their opposition. Yet a preference for coercion stemming from its expected efficacy relative to dialogue should not be confused with a reckless or jingoistic attitude towards conflict bargaining. Because escalation is an outcome to be avoided if possible when dealing with a competent leader, we should see elites trying to calibrate a mix coercive sticks and concessionary carrots to keep the conflict manageable, though the absence of open dialogue associated with coercive persuasive preferences increases the likelihood of failures of escalation management, due for instance to miscommunication.[134]

---

[131] Erdogan reportedly threw Trump's self-described 'very powerful and strongly worded letter' warning him against attacking Kurdish forces into the trash.

[132] A sensible stance considering Trump's proclivity to claim victory and walk away happy regardless of the content of agreements.

With respect to specific behavior, states in a real or potential conflict with the *Fanatic* should try to bludgeon or bribe their way to a de-escalatory settlement without sparking further conflict. Elites who believe an adversary leader is a *Fanatic*, or represent constituents who hold this belief, should be hesitant to use existing communication infrastructure to engage in iterative diplomatic dialogue and instead recommend using them as a means of persuading the adversary to change its behavior by conveying simple information. For example, if there are already back channels between the states, elites treating the adversary as a *Fanatic* should push for using these to issue threats such as the use of economic sanctions or military force, or to promise concessionary moves such as economic relief and troop withdrawal. These same elites should likewise question the utility of, or outright oppose, organizing novel communication infrastructure such as face-to-face meetings and diplomatic summits. Particularly in the contexts of ongoing crises or conflicts, this opposition is likely to be posited through critique of formal dialogue as signaling weakness, coddling the enemy, or buying the adversary time for maneuvering at the strategic (e.g. arms building, outreach to third parties requesting aid) or tactical (force movement and repositioning) levels. However, elites who believe the adversary is a *Fanatic* may acquiesce to opening formal negotiations if they see this as necessary to de-escalate, pause, or at least prevent further escalation of the conflict. In such cases where the start of negotiations rather than their content is what is important for de-escalation, we should witness elites engaging in the apparently paradoxical behavior of accepting or supporting negotiations while expressing pessimism about the prospect of achieving anything meaningful through a dialogue based in rational argument.

---

[134] Miscommunication is a too often inappropriately used term in IR scholarship, but here it is appropriate since the basis of claiming miscommunication as a reason for unintentional conflict escalation is specifically the absence of open discussion: when communication is lacking miscommunication is almost inevitable. Generally however it is the opinion of this author that the miscommunication and miscalculation explanations are too often used as catch-alls for a much wider range of occurrences that fall outside of a rational actor spectrum, and that their ubiquity in scholarship is less a function of their empirical validity or theoretical utility than of the need to collapse complex psychological or plausible irrational phenomena into a rational actor model of the world. For example, Mearsheimer and many of the realist thinkers who followed Waltz claimed that bad or foolish decisions were a function of miscalculation due to bad information or miscommunication of information: in effect miscommunication is being used by these scholars to obviate the introduction of human irrationality to state behavior through uses of words like misperception.

During any negotiations that take place, we should expect state officials interacting with *Fanatic* type leaders or their representatives to 'bargain hard' through material threats or incentives, but not so hard as to intentionally escalate conflict. During internal deliberations, they will express pessimism about the persuasive efficacy of a mutually respectful rational dialogue, and argue that, in simple and somewhat stereotypically simple terms, 'force is the only language the adversary understands.' We should see these elites make the case for the exclusive efficacy of simple coercive information by using words like 'delusional,' 'deranged,' 'blind,' or ‘emotional’ to characterize the adversary leader’s perception of reality as biased and therefore not susceptible to influence through argument or other forms of complex information. By extension, these elites will press for using negotiations as a forum to convey material threats, which could target the adversary’s political, economic, or military interests, as long as they believe the distribution of power and interests puts their side in a position where such threat-making is plausibly credible. Given the potential danger of uncontrolled escalation against a *Fanatic*, however, we should expect these elites to caution against and attempt to avoid making coercive threats recklessly, and to take seriously any threats the adversary makes in kind. Furthermore, if threats fail, the same danger that we would expect to induce caution in the issuance of threats may likewise induce these elites to, however reluctantly, advocate paying the necessary costs to make concessions to at least some adversary demands in order to keep the conflict from spinning out of control in a way far more costly than the de-escalatory concessions in question.

The British appeasement of Hitler at Munich, and the logic behind it, is an excellent illustration of the conciliatory manifestation of the observable behavior my theory expects of states facing a *Fanatic*. In 1938, British Ambassador Henderson wrote that “Hitler’s sense of values is so abnormal that argument seems powerless… his capacity for self-deception and his incapacity to see any point which does not meet his own case are fantastic.”[135] This suggests not only the belief that Hitler was a *Fanatic*, but also supports my theory that the ascription of bias to an adversary leader (Hitler) will undermine actors’ (Henderson) faith in the persuasive efficacy of rational argument. The British, with the wounds of WWI still fresh and facing the prospect of conflict with an apparently competent German leader, saw the prospect of escalation and war as unacceptably costly, so how did they approach persuading a biased leader to halt his expansion short of plunging Europe into another continental war? Through the simple-informational tool of conceding the Sudetenland; Chamberlain thought that absent the threat of force, which he was unwilling to employ, escalation was inevitable if Hitler's demand was not met.

---

[135] Ascher, *Was Hitler a Riddle?*, 73

We can look at US policy towards Khruschev during the Cuban Missile Crises for an example that shows a more typical mixture of carrots and sticks. Khruschev was doubtless seen as a biased leader, though there was disagreement in US decision-making circles on the extent to which Khruschev's perception of reality was distorted: some thought his bias was principally ideologically derived, while others suggested his emotional outbursts and alcohol abuse indicated he was even less attached to objective reality than a 'typical' Soviet leader. However, Kennedy had a healthy respect for Khruschev's competence, particularly after the Vienna summit when Kennedy left feeling like Khruschev had "just beat the hell out of me," and generally accepted the idea that Khruschev did not want a nuclear war. This, again, generally fits the description of the *Fanatic:* looking at a different road, but a capable driver with no desire to crash. When faced with the infamous Cuban Missile Crisis, Kennedy's administration to use the simple-informational tools my theory would expect, and with the predicted level of caution, to persuade his *Fanatic* counterpart to de-escalate. The imposition of a naval 'quarantine' around Cuba, and the threat of air strikes or other military action against the Soviet missile installations already on Cuba if Khruschev did not agree to remove them, are classic examples of the careful use of coercive threats we would expect of a state facing a *Fanatic*. Likewise, the Kennedy administration privately offered to remove US Jupiter missiles from Turkey in exchange for the removal of Soviet missiles from Cuba, adding positive material weight to its persuasive efforts: a carrot dangled to entice Khruschev to behave as the US desired. Notably, the sticks and carrots detailed above were conveyed in large part through private channels between the superpowers, Kennedy and Khruschev had already met face-to-face at Vienna, and the two were in personal contact through a long series of letters (the latter's writings were later described as 'rambling' by Kennedy and members of his administration). In other words, the infrastructure for complex diplomatic dialogue was at least feasibly in place, but to the extent it was used to bargain during the crisis it was for the purpose of conveying simple rather than complex information. While Kennedy and his advisors thought Khruschev's most basic interest of war avoidance was in line with their own goals, they evidently did not think of engagement in rational argument as the go-to method to influence the behavior of the Soviet leader, since they skipped it entirely.

Bargaining with The Madman:

Against adversary leaders who are seen as *The Madman*, the observable implications of my theory for state behavior are somewhat simpler: states should abstain from diplomacy entirely or use it for the purpose of issuing ultimatums. Whereas the coercive diplomacy I predict will be practiced with *Fanatics* is cautious and seeks to avoid escalation, the coercive ultimatums I predict will be practiced with *Madmen* are bellicose declarations of an intent to escalate if all demands are not met. The public is more likely to support hostile measures (sanctions, force) against such states before diplomatic dialogue has been exhausted, and more likely than with any other adversary type to support further escalation until the adversary gives in even if this means beginning or continuing a 'hot' military conflict. Elites should advocate the imposition of hostile escalatory measures, rather than for example the recalibration of demands or the rethinking of material threats, if the adversary does not take the initial ultimatum seriously enough to alter its behavior in response to the threat alone.

Elites who believe they are facing a *Madman* should oppose the use of existing communication channels or the opening of new forums for face-to-face interaction for argumentative dialogue, or even as an independent means of achieving de-escalation. They may use these channels to convey the terms of their ultimatum, or even advocate a face-to-face meeting if they see an in-person ultimatum as more credible, but we can usually expect such communications to be more unilateral than interactive.

The ultimatums issued by elites facing *Madmen* are likely to be purely based on threatened cost-imposition: all stick and no carrot, in other words. This may mean sanctions and military force, but empirically it is also likely to mean threats or attempts to kill the adversary leader or otherwise remove them from power: achieving behavioral change by replacement rather than persuasion. The logic of a stick-only persuasive approach to *Madmen* is simple: when dialogue is ineffective and therefore not available or appealing as a costless approach to persuasion, there are only so many paths to achieving ideal outcomes at low cost. The give-and-take of traditional coercive bargaining almost always involves some concessions to the adversary, but because the cost of fighting an incompetent *Madman* is low de-escalation through concessionary carrots is an unattractive option. In sum, the combination of coercive *persuasive preferences* and escalatory *Escalation Preferences* logically implies a ‘my-way-or-the-highway’ approach to conflict bargaining. During internal deliberations, elites advocating this approach should describe the adversary leader with words or colorful descriptors that simultaneously connote delusion and idiocy, for example: “mad dog,” “insane,” “crazy clown,” or “lunatic.” The argument that these labels and the patterns of thinking they denote dictate an ultimatum based-approach may, more-so than the other ideal-types, take on a normative as well as logical aspect. Put another way, we might expect elites to pitch their preferred approach not only by claiming that the only effective way to change the behavior of a *Madman* is to force them to change but also that such a hardline is moral, so making concessions is somehow unethical as well as materially unnecessary.

The Regan administration's confrontations with Muammar Gaddafi serve as an excellent illustration of both the ascription of a *Madman* image to an adversary and the logic through which this image leads states to use ultimatums or force to shift that adversary's behavior. Regan publicly branded Gaddafi "the mad dog of the Middle East," which mirrored private assessments even predating his administration: US diplomats thought Gaddafi was "increasingly losing his grip on reality," Regan wrote in his memoirs that Gaddafi was an unpredictable fanatic, and the CIA believed he used hallucinogenic drugs which warped his view of reality and inhibited his strategic decision making, a belief that informed an official CIA assessment that "Qadahfi is not controllable." In these assessments, we can see that US perceptions of Gaddafi's mental derangement also fed the conclusion that he was a wholly incompetent military and political leader. Because Gaddafi could not be reasoned with, the Regan administration opted in crises with Libya to skip diplomacy entirely and engage in immediate military confrontations to contain or reverse behavior they deemed undesirable: this evinces not only the expected inefficacy of complex information but also the willingness to escalate conflict. For example, in response to the 1986 La Belle disco bombing, the US launched Operation El Dorado Canyon, airstrikes targeting Libyan military and terrorism-related facilities. Evidence later emerged that not only was this operation an attempt to control Gaddafi's behavior by reducing his military capacity: it was also hoped that the operation would kill Gaddafi. Killing and removing a leader to alter their behavior, rather than engaging in any attempts at persuasion through dialogue *or* coercion, is a perfect encapsulation of the most extreme manifestation of the preference for ultimatum bargaining I expect states to adopt in conflict with an apparent *Madman*.

Intrastate Disagreement and its Implications:

An important caveat for my theory is that there can be intrastate divergence on beliefs about the judgment of an adversary leader among actors with direct or indirect influence over a state's approach to diplomatic bargaining. The implications for state behavior explicated above all speak in terms of the individual elites or members of the public within states rather than treating the state as a unified monolith; we may see intrastate clashes over the best approach to diplomatic bargaining between actors who disagree about whether the adversary is a *Rational, Fool, Fanatic,* or *Madman.* In the event of such conflict, my theory does not attempt to posit a robust prediction regarding which set of diplomatic bargaining preferences will be expressed in state behavior. Ultimately, the outcome in these cases will depend upon factors such as the distribution of domestic political power and the cogency with which actors on either side of the dispute make their case to persuadable actors within the state.

As an illustrative example of this intrastate disagreement in action, we can look at the US approach to Iran regarding the issue of its nuclear program in the years after George W. Bush: the Obama administration's choice to enter negotiations over a freeze of Iran's nuclear program and the eventual organization of the Joint Comprehensive Plan of Action (JCPOA), the opposition this effort at diplomacy faced from the American public and political elite, and the Trump administration's almost immediate reversal of course in withdrawing from the JCPOA. In part due to the 2013 election of now President Hassan Rouhani, who came to power on a platform of relative moderation and reformist vision that was a distinct departure from his more radical predecessor, key decision-makers in the Obama administration started to think of Iran's leadership as objective. Though there were never any illusions that Iran had become a morally ‘good’ actor or abandoned geopolitical interests that were fundamentally at odds with US interests, the election of Rouhani served as an inflection point for the Obama administration to raise its expected complexity cut-off for informational persuasion of Iran. Rouhani seemed a more reasonable leader and so the persuasive preferences of many US elites shifted away from coercion and towards argument; they advocated moving from crippling economic sanctions to formal diplomacy as the method of persuading Iran to agree to a negotiated settlement on its nuclear program.

Yet this effort faced significant opposition from both political elites and members of the public who continued to express the belief that Iranian leadership was the *Fanatic* type. This camp held firm to the position that Iran’s leaders were deluded by, for example, theocratic dedication to Islamic fundamentalism, and therefore could be reliably persuaded only through the coercive language of pain. Notably, there was also a smaller but extremely vocal coterie of thinkers and policymakers who expressed beliefs that, in effect, Iran under the ayatollahs fit the *Madman* type and that therefore a regime-change ultimatum was the best and only way to change the nation’s behavior on its nuclear program.[136] While this internal opposition did not doom the pursuit of agreement through diplomatic argument, it meant that the effort cost the Obama administration more political capitol, generated a discourse of doubt and condemnation surrounding the eventual agreement, and prevented the negotiated agreement from being ratified as a treaty and therefore obtaining cross-administration staying power. All these factors contributed to the speed and ease with which Obama’s successor Trump, surrounded by advisors and receptive to voters who believed Iranian leadership was a *Fanatic* or *Madman,* scrapped the

---

[136] Bolton for example said that avoiding conflict with Iran was irrational: Fitzpatrick, Mark. "Iran and Mr Bolton." *Survival* 62, no. 4 (2020): 39-46.

JCPOA. In sum, the internal disagreement over US policy, and the relatively rapid progression of that policy from crippling coercive sanctions to argumentative negotiation to the abandonment of diplomacy, appears to have been inextricably linked to the beliefs of decision-makers and their constituents about the rationality, or lack thereof, of Iranian leadership. Of course, this variance in beliefs about the judgment of Iranian leadership was not solely responsible for determining US policy: this brings us to some of the other caveats and limitations of my theory that it is important to keep in mind. Most of these were previously introduced in the discussion of the theory's constituent variables, but below I briefly reiterate them to outline their meaning in light of the theory's observable implications for state behavior.

Power, Perception, and other Caveats:

First, beliefs about adversary judgment are the sole determinant of diplomatic bargaining preferences: structural and dispositional variables remain important.[137][138] In terms of *Persuasive Preferences* for example, some actors have internal personality traits that make them far more inclined toward dialogue. They may engage in motivated reasoning that enhances their expectations about the persuasive efficacy of complex information, or even advocate argument-based diplomacy on principle absent the expectation that it will be effective: we may sometimes observe this directly when actors simultaneously acknowledge that the prospects of persuasion through argument are poor and nonetheless argue it should be attempted as a matter of principle.

Similarly, *Escalation Preferences* are not solely determined by perceived competence or incompetence: the principle mechanism through which beliefs about the competence of adversary leadership is one of threat perception and the expected costs of escalation, but there are a multiplicity of material variables that also feed into those factors.[139] Most obviously, escalation against leaders who command immense military power or nuclear weapons is likely to be considered costly even if said leaders are grossly incompetent.

---

[137] Richard K.Herrmann and Jonathan W. Keller 2004.

[138] Joshua D. Kertzer and Brian C. Rathbun. 2015

Second, diplomatic bargaining preferences shape choices and so outcomes but are not the only variable that does so. The impact of diplomatic preferences is independent of structural variables, but these structural variables play a key role especially insofar as they determine the material or political capacity of state actors to realize their preferences through action. For example, actors that have a persuasive preference for coercion may lack the requisite military power to make threats or the economic influence needed to impose sanctions: in such cases, observable state behavior will also depend on escalation preferences. Actors with a non-actionable preference for coercion and a preference for de-escalation may be forced to turn to a concessionary approach, or to try to ignore the adversary and hope for the best if they lack not only the capacity to impose coercive costs but also the capacity to offer carrots. International pressure from third parties may induce actors with actionable coercive preferences to attempt a formal negotiation through rational dialogue, though they will probably do so unenthusiastically. Too often empirical research in IR treats state behavior as a proxy for state preference, but the two are importantly distinct and should be thought of as such: the non-deterministic influence of diplomatic bargaining preferences on state behavior is a good illustration of both the existence of the distinction and the need for scholars to take it seriously.

The actions and choices of the adversary also matter with respect to whether an actor's diplomatic preferences are realized in state behavior: in diplomatic negotiations it takes two to tango.[140] Actors who believe their adversary is a *Rational* type, and accordingly prefer de-escalatory diplomatic dialogue, may attempt to engage the adversary in communications through various channels only to find that the adversary is not open to dialogue. In such situations when attempts to build the infrastructure for more formal diplomatic negotiations are initially rebuffed, we often see actors continuing hostile actions while maintaining attempts to reach out: this is, in effect, a use of coercion to persuade the adversary to engage in diplomatic dialogue. This is different from engagement in coercive diplomacy to shift an adversary's position on an eventual negotiated settlement; coercion in this situation is being used to get negotiations started so a process of more rational dialogue can hopefully take over, rather than to affect the outcome of said negotiations. These behaviors are not mutually exclusive but nonetheless are meaningfully distinct, and my theory predicts them under different combinations of adversary attitudes towards negotiation and beliefs about adversary judgment. While coercion to persuade adversaries during negotiation and affect the terms of an ultimate settlement should be observed against apparent *Fanatics*, coercion to start negotiations should be observed against apparently *Rational* adversaries who are initially closed to dialogue.

---

[139] Walt, Stephen M. *The origins of alliance*. Cornell University Press, 1990.

Third, beliefs about judgment exist in the mind of the observer. I therefore make no claims about the 'true' type of either the observer or the subject, though I assume observers treat their own perceptions and judgments as a baseline for how rationality.[141] I do not theorize as to whether observers are 'correct' in their beliefs about their own judgment or that of their counterparts: whether observers *should* hold particular beliefs about the judgment of counterpart. Rather this project investigates how the beliefs observers *do* hold impact their diplomatic preferences, focusing on adversaries to control for affective confounders such as warm/cold feeling.[142] This is not an 'irrational' theory, however;[143] people have reason to adjust their behavior in accordance with how they believe an adversary thinks, even if these beliefs are or may be inaccurate.

---

[140] Rathbun 2014, Skylar-Mastro 2019
[141] Pronin 2008, 2006

Diplomatic Bargaining Preferences: Alternative Explanations in Existing Work:

Only Schelling's rationality of irrationality and the work of a small number of contemporary scholars responding to it specifically provide alternative arguments about the relationship between actors' diplomatic bargaining preferences and their beliefs about the rationality of adversary leaders. As detailed in the discussion of motivations for the conceptual and empirical contributions of this dissertation, Schelling did not explicitly differentiate among types or degrees of perceived irrationality, nor consider the effects of perceived irrationality in clear relation to those of perceived rationality so as to construct a coherent theory. Because Schelling lumps all plausible denotations and connotations of perceived irrationality under a single heading, we cannot directly compare his predictions to the predictions of my theory with respect to the particular dimensions of beliefs about adversary judgment as conceptualized in this project. Yet his theory as stands predicts that the perception of adversary irrationality will generate diplomatic bargaining preferences and behaviors that are close to the polar opposite of those posited by my theory. According to Schelling actors facing apparently irrational adversaries should refrain from attempts at coercion or deterrence through the threat of cost-imposition, be more likely to seek de-escalation, and be more willing to make concessions to adversary demands. In other words, because irrational actors are heedless of costs, they can only be placated through appeasement, so the perception of adversary irrationality will push actors' *Diplomatic Bargaining Preferences* towards de-escalation and the use of simple information in the form of concessions but not threats. The contention that perceived adversary irrationality induces actors to use simple information to affect adversary behavior is the only prediction Schelling's rationality of irrationality has in common with my theory, but even in this respect the difference between the theories is quite stark. Whereas Schelling predicts capitulation without coercion, I predict that perceived bias in particular will make actors more likely to advocate the use of coercive and deterrent threats: determining which of these theories reflects empirical reality has massive implications for the accuracy of the widely ingrained view that an irrational image is an international asset.

---

[142] See Cuddy, Amy JC, Susan T. Fiske, and Peter Glick 2008; Richard K. Hermann, James F. Voss, Tonya YE Schooler, and Joseph Ciarrochi 1997; Richard K. Hermann 2017

143 McDermott 2004; Mercer 2005; Rathbum, Kertzer, and Paradis 2014

My dissertation is not the only piece of contemporary scholarship to engage with or challenge Schelling's rationality of irrationality, though the body of work that engages is quite small and the contingent of research that actually challenges is almost non-existent. Most notably, Roseanne McManus began research in this area around the same time as this author, conducting an empirical test of the relationship between leaders' level of perceived madness impact the favorability of their crisis bargaining outcomes.[144] Because her methodology, theoretical position, and dependent variable differ significantly from my own our combined work holds a potent potential for complementarity in the production of knowledge on an important international topic. Though this makes her work less of an alternative and more of an additive argument, I nonetheless cover it in this section given its importance as the exception to the rule that IR scholars have not empirically or, in any serious way, even theoretically engaged with the rationality of irrationality.

---

[144] McManus, Roseanne W. "Crazy Like a Fox? Are Leaders with Reputations for Madness More Successful at International Coercion?." *British Journal of Political Science* (2019): 1-19.

McManus uses Lexus Nexus' database of media sources to construct a Large-N dataset measuring the degree of irrationality ascribed to leaders, and tests the effect variation in degree of perceived madness so defined has on the likelihood of leaders experiencing success in deterrence and international crisis bargaining. Based on the number of media mentions of a given leader using particular words associated with madness, McManus classifies whether leaders are 'a little mad' or 'very mad:' she hypothesizes and demonstrates within the context of her empirics that leaders perceived as 'a little mad' may experience better crisis outcomes, while those perceived as 'very mad' will have worse outcomes and are likely to have adversaries initiate crises against them more often. She posits a narrow albeit cogent theoretical mechanism for these hypotheses, namely that apparently mad leaders cannot credibly reassure their adversaries that they will not use force if certain conditions are met: in other words, they suffer from what Weissiger would call a 'dispositional commitment problem'[145] the severity of which increases with their level of apparent madness. This work is a major step forward, but there are several issues with it that my own, itself not free of issues, addresses. First, the measurement strategy for the IV (perceived madness) is interesting and independently valuable as a dataset but flawed for the purpose it is intended to fulfill. Particularly given media incentives for sensationalization, it is in no way clear that media narratives about foreign leaders are representative or determinative of the beliefs of either elites or the mass public, at least one of which would need to hold a particular perception of a foreign leader's madness in order to affect state policy. Second, McManus uses a unidimensional conception of perceived madness, which is a problematic approach both theoretically and empirically for the same reasons as with Schelling's original theory. Third, the dispositional commitment problem mechanism does not readily lend itself to extrapolation into issue areas outside of classic deterrence and crisis bargaining frameworks. It is not clear for example how it would apply to the conduct of interstate diplomacy whether between adversaries, neutrals or third parties: put another way it is a valid mechanism but its scope of potential applicability is narrower than that of my theory.

---

[145] Weissiger 2013

Beyond Schelling and those immediately responding to his rationality of irrationality thesis, IR scholarship provides myriad alternative explanations for variation in diplomatic bargaining preferences and behavior that are not linked to perceptions of adversary judgment and/or rationality. Alternative explanations for variation in diplomatic preferences offered by extant IR scholarship fall into two broad camps. First, structural explanations that look to material capabilities, interests, or other elements of the international situation in which actors exist. Second, dispositional explanations that examine how the internal traits of states or the individuals within them influence diplomatic preferences, choices, styles, or outcomes in ways that a purely structural approach does not account for.

Structural, Material, and Rationalist Arguments:

While diplomacy is a central element of international relations, the predominant trend in IR scholarship has been to treat war, which is by far a more popular subject, as the breakdown of bargaining. Diplomacy is treated as "automatic, unproblematic, and ultimately unimportant."[146] If the strong always get their way at the expense of the weak, or if states with closely aligned interests always cooperate, then the occurrence of diplomacy is unimportant. If the nature of diplomacy is determined by material variables, if those with power always employ coercive leverage and those without always attempt reasoned dialogue, then diplomats lack agency and the content of diplomatic interactions is epiphenomenal to structure. In other words, a majority of IR scholarship takes the structural approach to explain variation in diplomatic preference, choice, and outcomes. Rationalist IR scholarship is particularly explicit that diplomatic communication "is cheap and uninformative"[147] because of incentives to misrepresent information.[148] When diplomacy is treated in this way, theories about variation in preference for diplomacy effectively become the inverse of theories about variation in preference for war. As this is one of the most thoroughly covered topics in political science, a full review of the range of arguments posited to explain variation in preference for force would be impractical here, though in the case study presented in Chapter Four I consider some of their implications and predictions in more detail to provide a point of comparison for the implications of my theory. However, structural explanation that treats diplomacy as the absence of war, from structural theories that stress the balance of power or interests, to those that stress the perception of hostile intentions as key to stoking a spiral of conflict, to bargaining theories, incorporate the role of how adversaries are believed to think in shaping preference for force.

---

[146] Brian C. Rathbun 2014, p. 1. James Fearon, "Signaling versus the balance of power and interests: An empirical test of a crisis bargaining model." *Journal of Conflict Resolution* 38, no. 2 (1994): 236-269. James Fearon, "Rationalist explanations for war." *International organization* 49, no. 3 (1995): 379-414. Erik Gartzke, "War is in the Error Term." *International Organization* 53, no. 3 (1999): 567-587.

Another vein of rationalist IR scholarship neglects diplomatic bargaining preferences not by treating it as the absence or inverse of war, but rather by treating bargaining processes as an assumed constant occurring in the background of interstate conflict. Particularly in theories centered on the bargaining model of war, states are treated as constantly being at the bargaining table as soon as conflict over some disputed interest is being considered or carried out: counterparts will strike a bargain to end conflict once structural and informational factors produce convergence in their beliefs about the probable course of conflict.[149] This assumes away the question of when and why states choose not to talk, or just as importantly to abstain from or exit negotiations; empirically this is an issue where we observe significant variation, and different choices in this area are one of the more important observable implications of variation in *Diplomatic Bargaining Preferences* as defined in this project.

---

[147] Rathbun 2014, p. 13.
[148] James Fearon 1995

More recent scholarship that fits within the broader contours of material-structural theory has problematized and sought to explain variation in state actors' decision to talk, or not talk, during conflicts or crises. Skylar Mastro in particular makes a cogent case for theorizing the choice to approach the bargaining table independently, as it is a distinct choice that carries distinct advantages and costs.[150] Mastro's 'costly communications thesis' posits that officials will be open to talks when they think the adversary lacks the power or will to escalate conflict to an unacceptably costly level if it takes the proposal to talk as a sign of exploitable weakness, in effect centering her explanation on the canonical variable of the 'balance of power and resolve.' Additional work looks to the independent benefits states draw from opening diplomatic dialogue: for example, the opportunity to collect intelligence, to pause fighting, to gain international approval. All this research represents important advancements in our understanding of a still understudied question, to which my theoretical and empirical work contributes.

Psychological and Dispositional:

Recently there has been a notable increase in work that takes a psychological approach to the topic of diplomacy, challenging IR's neglect of this subject by moving outside a structural or material lens. This includes research on topics ranging from emotions[151] and empathy in diplomacy,[152] to the importance of face-to-face diplomacy[153] and personal impressions,[154] to the study of diplomacy as a "practice."[155] This work challenges the notion that diplomacy's occurrence or content are structurally determined, as well as the contention that variation in the occurrence and content of diplomacy are exogenous to structure determining international outcomes. Most of this work narrowly focuses on a single aspect of diplomacy or why discrete diplomatic interactions do or do not result in a negotiated settlement rather than theorizing diplomacy in broader terms. Indeed, much of the work restricts itself to making the case that a previously neglected variable, such as personal impressions gleaned from facial expressions, *can* plausibly impact conduct and outcomes in interstate diplomatic bargaining, rather than when or how these variables will impact diplomacy in systematic, generalizable ways.

---

[149] Fearon 1995; Reiter 2003

[150] Skylar-Mastro 2019

One notable exception is the work of Brian Rathbun, who posits a more general theory of variation in diplomatic preferences, or styles as he terms it. Rathbun defines three *diplomatic styles*: *coercive bargaining* (Rationalist)[156], *pragmatic statecraft* (Realist)[157], and *reasoned dialogue* (Liberal).[158] Drawing on experimental evidence that, even in the same structural setting, individuals negotiate in different ways, Rathbun links systemic variation in diplomatic style to two attributes of decision-makers that feature prominently in numerous theories of individual and social psychology: social value orientation and epistemic motivation. He hypothesizes that these traits determine diplomatic styles in ways meaningfully exogenous to alternative explanatory factors.[159] The principle contribution of behavioralizing diplomacy as Rathbun and his peers do is that if non-structural variables influence diplomatic styles, they therefore affect the emergent spirit of negotiations and so have the potential to impact international outcomes.

---

[151] Todd H. Hall, *Emotional diplomacy: official emotion on the international stage*. Cornell University Press, 2015.

[152] Holmes, Marcus, and Keren Yarhi-Milo. "The Psychological Logic of Peace Summits: How Empathy Shapes Outcomes of Diplomatic Negotiations." *International Studies Quarterly* 61, no. 1 (2017): 107-122.

[153] Wong, Seanon S. "Emotions and the communication of intentions in face-to-face diplomacy." *European Journal of International Relations* 22, no. 1 (2016): 144-167, Holmes, Marcus. "The force of face-to-face diplomacy: mirror neurons and the problem of intentions." *International organization* 67, no. 4 (2013): 829-861.

[154] Hall, Todd, and K. E. R. E. N. YARHI-MILO. "The personal touch: Leaders' impressions, costly signaling, and assessments of sincerity in international affairs." *International Studies Quarterly* 56, no. 3 (2012): 560-573, Yarhi-Milo, Keren. *Knowing the Adversary: Leaders, Intelligence, and Assessment of Intentions in International Relations*. Princeton University Press, 2014.

[155] Pouliot, Vincent. *International security in practice: the politics of NATO-Russia diplomacy*. Vol. 113. Cambridge University Press, 2010.

[156] Rathbun 2014, "States use threats and exploit their leverage to pressure other states to concede... Diplomacy is a game of high stakes poker in which states have no incentive to show their cards or believe the cheap talk of others."

[157] Rathbun 2014, "Here the far-sighted diplomat focuses on securing the most vital of interests, conceding on issues of less importance to avoid creating unncessary conflict... good diplomacy is chess rather than poker"

[158] Rathbun 2014, "Diplomacy is a process of augmentation in which state representatives aim to persuade others of their point of view while listening closely to their claims as well."

However, a notable shortcoming of existing behavioral work on diplomacy, which Rathbun's work epitomizes, is a focus on the internal traits of decision-makers that neglects relational variables, a considerable omission given the interactive nature of diplomacy. In Rathbun's formulation, decision-makers have an ingrained preference for one diplomatic style over others that does not vary by counterpart until diplomatic communication actually takes place. Once communication occurs decision-makers may alter their behavior in accordance with what they experience as their counterpart's diplomatic style, though their underlying preference for a particular diplomatic style remains the same. There are several problematic assumptions in this story. First, diplomats' initial beliefs about their counterpart's diplomatic style are either accurate (if their epistemic motivation is high) or the product of a self-projection heuristic (if their epistemic motivation is low). In sum, when decision-makers' beliefs about their counterpart's diplomatic style differ from said counterpart's 'true' type, it is because they are operating under the naïve assumption that others share their diplomatic style. This neglects alternative models for how people intuit the intentions, beliefs, and minds of others: for example, heuristic reliance on outgroup stereotypes.[160] Furthermore, diplomats accurately perceive the diplomatic style of their counterpart once diplomacy has commenced: iterative interaction causes prior beliefs about a counterpart's style to converge with objective reality. Yet whether this actually happens is an empirical question: it is for example possible that strong priors are resistant to updating.

---

[159] He does not however focus on theorizing this 'obliging' style, as he argues that both it and the “rosy-eyed idealists” inclined to practice it will be relatively rare in international politics,

Finally, none of the psychological or dispositional theories of diplomacy address any potential impacts on diplomatic preferences of real or perceived variation in psychological characteristics of counterparts, which is to say of perceived variation in how counterparts think. Some of the work in this camp, such as Rathbun’s, examines the role of beliefs about how others think about diplomacy, but not how they think in general: how they make judgments, process information, and construct options. Regardless of the diplomatic style to which a decision-maker is predisposed, would they attempt reasoned dialogue with an adversary they believed was intrinsically unreasonable? Existing work leaves us no way to know, but this omission of heterogenous beliefs about judgment is the rule rather than an exception in international relations scholarship.

Conclusion:

International Relations scholarship too often takes a simplistic view of perceived irrationality or rationality and has not yet robustly interrogated the consequences of this perceptual phenomenon, even in the issue area, diplomatic and coercive bargaining, to which Schelling most famously applied it. In this chapter I explicated the concept of *beliefs about*

---

160 Ames and Mason 2012

judgment and presented a typology of ideal-type beliefs about the judgment of adversary state leaders based upon variation on two dimensions: perceived objectivity and perceived competence. These dimensions combine into four ideal-types into which beliefs about adversary judgment can fall: *the Rational, the Fool, the Fanatic*, and *the Madman.* After defining my dependent variable, *diplomatic bargaining preferences*, I detailed the theoretical mechanisms linking my IV to my DV. Evidence suggests that ascriptions of bias undermine the expectation that information sharing will lead to convergent beliefs, and ascriptions of incompetence reduce perceived threat. I described the two primary reasons why these theoretical mechanisms apply to individuals in both the mass public and elite decision-making circles: the broad applicability of the proposed psychological mechanisms, and the role of public opinion. I then unpacked the observable implications of variation in these preferences with reference to real-world examples: what does diplomatic bargaining look like when conducted by an actor who prefers coercion without compromise versus dialogue with compromise? Sixth, I discussed alternative arguments in order to illustrate how the theory presented in this chapter expands upon our understanding of diplomatic bargaining and the role of leaders in international politics.

In the chapters that follow, I empirically test my theory against the predictions of alternative arguments. I use a mix of experimental and case study methodology to examine the veracity of my theory at the public and elite level, respectively. Chapter Three uses three novel survey experiments to investigate how variation in beliefs about adversary judgment shape the diplomatic bargaining preferences of the mass public. The first survey was fielded through Amazon’s Mechanical Turk (MTurk) and centered on a hypothetical nuclear crisis between the US and the imaginary state of Mykonara, the second survey was fielded through MTurk but centered on the real-world crisis between the US and North Korea, and the third survey was fielded on a nationally representative sample and again used Mykonara. Subjects were told about the crisis then randomly presented with between zero (control) and two opinion articles about the adversary leader which described the leader as biased or objective, and competent or incompetent. Following the treatments, subjects were asked a series of questions to ascertain their preferences and expectations for how the US should approach diplomatic bargaining with the adversary. The results from all three surveys provide strong support my theory, particularly the hypothesized effects and mechanisms of perceived bias; for example, in all three surveys subjects assigned the 'biased' article were significantly less likely to advocate diplomacy, or expect it to be effective, compared to subjects in the 'objective' or 'control' group. This chapter ends by addressing the 'levels of analysis; problem by discussing the pros and cons of these experimental findings as a means of making inferences about the behavior of political elites.

In Chapter Four, I test my theory at the elite level through a case study of US decision-making during the Korean War, which permits me to trace variation in elites' diplomatic bargaining preferences towards three adversaries and across time in the context of a single conflict in which a ceasefire was the immediate aim of US policy. I focus on the initial ceasefire negotiations at Kaesong and Panmunjon in 195, though I also draw evidence from 1946 through the end of war under the Eisenhower administration and examine in some detail US discussions surrounding three additional noteworthy periods: the North Korean invasion that began the Korean War, the entrance of Chinese forces into the war, and Eisenhower's threat of nuclear devastation as a means of inducing a ceasefire. The case supports the veracity of my theory at the elite level. Most US decision-makers saw the Soviets as responsible statesman who either could be reasoned with, or at the very least had to be accommodated to avoid the devastation of a third world war. Officials were willing, even eager, to engage in iterative diplomatic dialogue with the Soviets in pursuit of a ceasefire, and evinced readiness to make some concessions to Soviet interests. However, these same officials thought diplomatic dialogue with the ‘irrational’ Chinese would be fruitless and concessions to their incompetent leadership unnecessary. It took immense pressure from a range of UN coalition members to bring the US to the negotiating table at Kaesong, and once at that table US officials mainly issued threats and ultimatums to Chinese representatives as opposed to engaging in an iterative dialogue. In short, US beliefs about the rationality of the Soviets, Chinese, and North Koreans shaped their diplomatic bargaining preferences in ways prominent alternative arguments fail to predict.

In Chapter Five, I use experimental methods to probe the sources of beliefs about leaders’ judgment to provide an empirical response to two major critiques of my theory’s use of these beliefs as an independent variable. First, that the beliefs follow policy positions rather than the other way around. Second, that they are epiphenomenal to other variables that shape diplomatic bargaining preferences. I use a choice-based conjoint survey experiment to probe the plausibility of a wide range of potential explanations suggested by existing work. The conjoint results suggest that while material variables are important, leaders' ideology and personal behavior also carry significant weight in subjects' inferences about their rationality, though the conjoint design does not provide leverage for causal inference. A second survey experiment, which utilizes a multivariate factorial design, creates a more controlled setting in which to test the causal impact of a leader's race, ideology, personal behavior, and state economy on subjects' perceptions of their rationality and competence. The results provide strong support for the hypothesis that leaders with speaking styles that seem 'angry' or 'eccentric,' and leaders that are ideologically distant, are far more likely to be perceived as irrational than other leaders who possess equivalent material capability and make identical policy decisions. The centrality of identity and costless behavior in inferences about foreign leaders' rationality suggests such inferences are exogenous to other important factors in IR, and so must be taken seriously as an explanatory variable.

Chapter 3: An Experimental Test of the Consequences of Perceived Irrationality

I begin my empirical investigation by conducting an experimental test of my theory's predictions as they pertain to the foreign policy attitudes of individual members of the public. To reiterate, I argue that individuals' preferences and choices in international conflict are influenced by their beliefs about the rationality of adversary leaders' judgment, and that this influence is exogenous to structural or dispositional factors. First, people are less likely to attempt persuasion via reasoned argument or the non-violent achievement of negotiated settlement with actors they believe are biased. Second, people are less inclined to make concessions to, and more open to conflict with, actors they believe are incompetent. Finally, people may abstain from all forms of negotiation or diplomacy with those they believe are not objective or competent enough to be capable of sound judgment.

I test this argument using a novel pair of survey experiments, which are ideally suited to ascertaining the causal impact of variation in perceptions and beliefs, fielded on members of the US public. The survey experiments employ a factorial treatment design: subjects are randomly assigned to read some combination of articles that describe the leaders of a foreign adversary as either objective or biased and either competent or incompetent. After this they answer an array of questions about their preferences for diplomacy versus force as a means of resolving an ongoing crisis with the adversary state. To minimize bias from priors, the first survey uses an imaginary state called 'Mykonara.' To conduct a hard test of the hypotheses, the second survey uses the highly salient real-world case of North Korea. The results from both surveys provide strong support for my arguments, suggesting that beliefs about the rationality of adversaries play an important role in shaping diplomacy and are therefore a variable IR scholarship should take seriously.

The research described in this chapter is necessitated by the absence to date of any serious scholarly treatment of the international political role of people's beliefs about the judgment of their foreign counterparts. Much ink has been spilt in political science scholarship on how beliefs about the desires and intended actions of foreign counterparts shape international politics; in fact, these are some of the oldest, best established debates in the IR canon. The knowability or unknowability of intentions is a primary driver of differential predictions about international outcomes between offensive realists,[161] defensive realists[162] and neoliberal institutionalists.[163] Indeed, most of 'the Intentions Debate'[164] comes down to disagreement over 1) what methods, if any, states use to determine or constrain the intentions of foreign counterparts, 2) whether states make worst case assumptions or probabilistic inferences about the intentions of foreign counterparts if said intentions are inherently unknowable. A core component of offensive realism, for example, is the contention that in a dangerous world where the intentions of others are inherently uncertain, states must ubiquitously assume the worst about counterparts' intentions to avoid being caught off guard and potentially eliminated by a foreign aggressor.[165] The resultant need to guard against the possibility of attack by every other state in the international system drives states to maximize power in an effort to ensure their survival. By contrast states in a defensive realist world make probabilistic inferences about the intentions of counterparts, and so need not achieve regional hegemony to feel secure; once they have *sufficient* relative power to believe it is highly unlikely they will be attacked and destroyed, they can be satisfied with the status quo distribution of power. In short, because offensive and defensive realism disagree about whether states think probabilistically or possibleistically about the inherently uncertain intentions of counterparts, these theoretical traditions make significantly divergent predictions about state behavior.

---

[161] Mearsheimer 2001, Rosato 2015

[162] Jervis 2019, Kydd 2007, Walt 1990, Glaser 1997 and 2010

[163] Keohane 2005, Lipson et al 1991

[164] Rosato 2015, Glaser 2010, Mearsheimer 2001, Schweller 1994

[165] Mearsheimer 2001

Actors' beliefs about the desires of foreign states play a similarly crucial role in foundational IR theory, particularly with respect to the logic and likelihood of international negotiation and cooperation versus escalation and conflict. Put another way, the core distinction between many ontological perspectives on crucial issues of IR can be traced to whether they contend states believe their counterparts are security seekers or power maximizers, status-quo or revisionist, or similar conceptual distinctions that boil down to the question of what states believe their counterparts want.[166] Whether, why, when, and even how states are expected to cooperate with their counterparts depends in large part on perceived compatibility of interests.[167] Cooperation is easiest when goals are compatible, plausible when they overlap in at least some areas,[168] and difficult or impossible if goals are believed to be incompatible and therefore gains zero-sum.[169] Assertions about the likelihood of war or security competition depend not only on whether states are in reality satisfied security seekers or relentless power maximizers, but where they believe their counterparts fall on this spectrum.[170] Explanations for crisis escalation and prescriptions for preventing conflict depend on this same distinction; threats and arms building may deter a power maximizer, but are likely to create conflict spirals when employed against security seeking adversaries.[171]

---

[166] Waltz 2010 and Jervis 2019 (Security Seeker), Mearsheimer 2001 (Power maximizer), Schweller 1994

[167] Keohane 2005

[168] Thus creating the possibility for mutually beneficial trade-offs and value-added interaction

[169] Jervis 2019

[170] Schweller 1994, Glaser 2010

In short, IR scholarship has not merely addressed the international consequences of variation in actors' beliefs about the desires and intentions of their counterparts, it has done so exhaustively. Indeed, these forms of mind-reading are a ubiquitous component of IR theory, and the above examples are an illustrative but by no means exhaustive indication of their importance in the field.

Yet surprisingly, IR scholarship has neglected to rigorously theorize or test the international consequences of variation on the other principle dimension of mind-perception; actors' beliefs about how their counterparts *think*. In some ways, this is a comprehensible neglect: the subject is a complicated one that can be difficult to conceptualize or measure. The minds of others are inherently impossible to know with certainty. We cannot know how others see the world or what they believe, what others want, or what others are planning to do. We could imagine what we would believe, want or do in a counterpart's shoes, but this sort of naïve self-projection often leads us astray.[172] Of course, we could always ask someone what they believe, desire, or plan to do, but how can we be sure they will tell us the truth? Moreover, how can we even be sure they are being truthful with *themselves*? Beliefs can be subconscious, desires can change or be repressed, and whatever plans an actor has for future behavior, actions are inherently uncertain until the moment they are taken, meaning people cannot be sure how they themselves will act in any given situation. Mind reading, in other words, is impossible.

---

[171] Jervis, Robert. "Deterrence, the spiral model, and intentions of the adversary." In *Psychology and the prevention of nuclear war*. 1986.

[172] Ross 2015; Pronin 2002; Pronin 2007

Yet this issue is not a theoretical or empirical black-hold; scholarship in social and cognitive psychology has unpacked implications of bias-perception,[173] competence assessments, and general beliefs about how others think for interpersonal and intergroup interaction. Impossible though mind reading may be, people not only attempt it, but frequently form strong beliefs about the beliefs, desires, and intentions of others: extant research consistently shows that, accurate or no, these beliefs inform actors' subsequent interactions with the counterparts to which they pertain. However, this project is the first to apply mind-reading to the interstate level.

The rest of this chapter proceeds as follows. First, I provide a summary of existing experimental evidence on how beliefs about how counterparts think impacts actors' preferences and choices in interpersonal and intergroup interactions. Second, I describe the logic of using a survey experiment to build upon this existing work and re-state my hypotheses regarding how individuals' beliefs about the way foreign adversaries perceive their environment or formulate strategies impact preferences and choices in interstate crises. Third, I outline hypotheses for alternative explanations based on the extant research on interstate diplomacy described in Chapter One. Fourth, I describe the design for an experimental test of these hypotheses utilizing a pair of survey experiments. Fifth, I present and discuss the results from Study 1. Sixth, I present and discuss the results from Study 2. I conclude with a summary discussion of results.

Existing Experimental Evidence:

Extant experimental research serves as an invaluable basis for hypothesizing how variation in beliefs about the objectivity, competence, or rationality of adversary judgment might impact individual preference in interstate diplomacy. I outline this evidence below with a focus on demonstrating why existing empirics cannot be extrapolated to an international context, and thereby illustrating the need for the novel data collection and analysis I present in this chapter.

---

[173] Pronin 2002, 2007, 2009

Bias Perception:

Research in both interpersonal and intergroup settings has empirically demonstrated numerous potential consequences for behavior of perceived bias in a counterpart's capacity for rational judgment. People hold more pessimistic attitudes about the plausibility of compromise or understanding with people they believe are biased, as they do not expect novel information or arguments to change the beliefs or behavior of biased counterparts; indeed, people often prefer to not even *attempt* to reason or share information with those they perceive as biased.[174] The majority of research in this area has been conducted in the context of policy debates between US partisans on opposite ends of the political spectrum: studies consistently demonstrate several behavioral patterns across issue areas from climate policy to abortion law. Subjects presented with a counterpart that disagrees with their policy opinion respond differently to this divergence depending on what they are told about the counterpart's partisan identification. Subjects informed the counterpart shares their partisan identification are more likely to conclude the counterpart is misinformed or underinformed, and to express an expectation that the policy disagreement can be resolved through information provision and rational discussion. By contrast, subjects informed that the counterpart subscribes to an opposing partisan identification are more likely to conclude the counterpart is irrationally devoted to their opinion. When people explain issue-specific divergences of policy opinions as a product of their 'opponent's' biases, they are less likely to prefer dialogue-based means of resolving the disagreement, since they believe information provision and rational discussion would be fruitless.[175]

---

[174] Ward, Andrew, L. Ross, E. Reed, E. Turiel, and T. Brown. "Naive realism in everyday life: Implications for social conflict and misunderstanding." *Values and knowledge* (1997): 103-135. Sammut, Gordon, Frank Bezzina, and Mohammad Sartawi. "The spiral of conflict: Naïve realism and the black sheep effect in attributions of knowledge and ignorance." *Peace and Conflict: Journal of Peace Psychology* 21, no. 2 (2015): 289. Robinson, Robert J., Dacher Keltner, Andrew Ward, and Lee Ross. "Actual versus assumed differences in construal:" Naive realism" in intergroup perception and conflict." *Journal of Personality and Social Psychology* 68, no. 3 (1995): 404. Chambers, John R., Robert S. Baron, and Mary L. Inman. "Misperceptions in intergroup conflict: Disagreeing about what we disagree about." *Psychological Science* 17, no. 1 (2006): 38-45. Keltner, Dacher, and Robert J. Robinson. "Imagined ideological differences in conflict escalation and resolution." *International Journal of Conflict Management* (1993).

Research by scholars including Gilovich, Ross, and Pronin demonstrates that the consequences of bias-perception extend beyond 'civil' political disagreements and into violent intergroup conflict.[176] Compared with adversaries they believe are rational/objective, people prefer violent to non-violent means of conflict resolution with adversaries they perceive as biased. For example, Pronin conducted an experiment where subjects saw descriptions of terrorist decision-making as either rational/objective or biased/irrational, then were asked their preferences for violent vs. non-violent methods of counterterrorism. The results showed "when terrorists were depicted as biased and irrational... participants were more likely to advocate military action against terrorism and less likely to advocate diplomacy. This effect was mediated by perceptions of terrorists' capacity for reason" rather than either terrorists' intentions or positive or negative feelings towards terrorists.[177]

However, this research is inadequate as a representation of interstate relations for several reasons and so cannot hold as empirical evidence that this effect would manifest in interstate conflict. First, it deals with an adversary framed as a diffuse group defined by tactics and goals (terrorists) rather than a cohesive entity also defined in political and territorial terms (a state). Second, Pronin does not control for threat perception as an alternative potential mediator. Third, conflict with a non-state actor is not theoretically or empirically comparable to conflict with an adversary state: a multiplicity of variables at play in the latter are of subdued or null importance in the former. These include but are not limited to material capacity and the resulting risks of conflict, the implications of either diplomacy or violence for relationships with third party states, and in all probability a longer history of relations.

---

[175] Robert J. Robinson, Dacher Keltner, Andrew Ward, and Lee Ross. "Actual versus assumed differences in construal:" Naive realism" in intergroup perception and conflict." *Journal of Personality and Social Psychology* 68, no. 3 (1995): 404.

[176] Thomas Gilovich and Lee Ross. *The Wisest One in the Room: How You Can Benefit from Social Psychology's Most Powerful Insights*. Simon and Schuster, 2016.

[177] Emily Pronin, Kathleen Kennedy, and Sarah Butsch. "Bombing versus negotiating: How preferences for combating terrorism are affected by perceived terrorist rationality." *Basic and Applied Social Psychology* 28, no. 4 (2006): 385-392.

Competence Assessments:

On the subject of competence, existing research suggests that competence assessments primarily impact the effort individuals put towards hindering or facilitating the goals of counterparts.[178] Competence assessments have less independent impact on whether people seek to help or hinder the goals of others, which is more immediately shaped by other attitudes and beliefs, principally the dimension of hot/cold feeling towards the other. Effectively, whether people actively attempt to hinder or help others depends primarily upon their affective attitude towards (how much they like/dislike) a given counterpart. Yet whether efforts to aid or hinder the goals of the other are passive or active is primarily a function of perceived competence: on how capable the counterpart is believed to be of actually advancing its goals. People are more likely to actively engage counterparts they believe are competent, and to consider the goals of these counterparts when making their own decisions. By contrast, people tend to neglect counterparts they believe are incompetent and give little thought to the goals of incompetent counterparts. with compatible goals or accommodate the goals of counterparts they believe are competent, and more likely to neglect the goals of counterparts they believe are incompetent.

---

[178] Amy JC Cuddy, Susan T. Fiske, and Peter Glick. "The BIAS map: behaviors from intergroup affect and stereotypes." *Journal of personality and social psychology* 92, no. 4 (2007): 631.

Experimental evidence points to several patterns in the behavioral consequences of various combinations of stereotype content on the dimensions of affect (warm/cold) and competence assessments (competent/incompetent).[179] First, people tend to actively cooperate with or attempt to advance the goals of admired (warm/competent) counterparts. Second, people usually either benignly neglect or adopt a paternalistic approach to pitied (warm/incompetent) counterparts, as such actors seem incapable of advancing their best interests on their own.[180] Third, provided they believe they have the capacity to do so successfully, people usually actively attempt to impede the interests of envied (cold/competent) counterparts. Absent the perceived capacity to meaningfully hinder the goals of a competent/disliked counterpart, the preference for active interference is often subordinated to the practical necessity of grudging accommodation and whatever passive resistance is plausible. Finally, people tend to manifest contemptuous neglect for counterparts they dislike and perceive as incompetent: while people see the goals of such counterparts as at odds with their own interests, the possibility that such incompetent others will actually achieve their aims is dismissed as a threat that need not be taken seriously.

There are several issues with using the experimental findings outlined above to generate hypotheses for specific issue areas, such as behavior in international crises. First, the general behavioral approaches these findings suggest people take on the basis of perceived warmth and competence are far from clear even in the abstract. For example, the behavioral dimension associated with perceived competence is fairly vague: ‘passive facilitation’ is used interchangeably with ‘association,’ and ‘passive harm’ is used interchangeably with ‘neglect;’ the connection between these is not exactly intuitive.

---

[179] Cuddy, Amy JC, Susan T. Fiske, and Peter Glick. "The BIAS map: behaviors from intergroup affect and stereotypes." *Journal of personality and social psychology* 92, no. 4 (2007): 631. Fiske, Susan T., Juan Xu, Amy C. Cuddy, and Peter Glick. "(Dis) respecting versus (dis) liking: Status and interdependence predict ambivalent stereotypes of competence and warmth." *Journal of social issues* 55, no. 3 (1999): 473-489. Lee, Tiane L., and Susan T. Fiske. "Not an outgroup, not yet an ingroup: Immigrants in the stereotype content model." *International Journal of Intercultural Relations* 30, no. 6 (2006): 751-768.

[180] Eckes, Thomas. "Paternalistic and envious gender stereotypes: Testing predictions from the stereotype content model." *Sex Roles* 47, no. 3-4 (2002): 99-114.

Second, due in no small part to this opacity in the abstract, it is unclear what would be the practical manifestation of these behavioral approaches in real world contexts. For example, for the purposes of this project the most relevant counterpart types are cold/competent and cold/incompetent. Yet what precisely is meant by ‘passive facilitation and active harm,’ the approach people are said to adopt towards cold/competent counterparts, and how does this differ from passive neglect and active harm, the approach people tend to take towards cold/incompetent counterparts? Exactly what these behaviors look like seems highly specific to the issue or situation at hand. In the context of bargaining games, one of the closest corollaries of interstate diplomacy, ‘passive facilitation and active harm’ seems to mean actively negotiating and taking the counterpart’s interests seriously (association/passive facilitation) while working towards an outcome that minimizes the gains of the counterpart (active harm).

There are, however, insufficient empirical bases to assert that these behavioral impacts of competence assessments uncovered by existing work should be expected to hold at the interstate level. Image Theory *can* be interpreted as applying research on the behavioral impact of competence stereotypes to IR, albeit indirectly. Image theory proposes that images of the other are constructed on the basis of three characteristics: “perceived relative power, perceived culture, and perceived threat or perceived opportunity that a subject believes another actor represents.”[181] Various combinations of these characteristics generate images of the other as being, for example, an enemy, ally, colony, or degenerate barbarian: each of these images denotes a set of beliefs about the other's intentions, the cohesion and sophistication of its elites' decisionmaking, and overall capabilities. Enemies are high capability with hostile intent, allies are high capability with friendly intent, colonies are low capability with neutral or friendly intent, barbarians are low capability with hostile intent. These images facilitate decision-making and so shape behavior by simplifying a complex world.[182]

---

[181] Herrmann, Richard K, James F Voss, Tonya YE Schooler and Joseph Ciarrochi. “Images in international relations: An experimental test of cognitive schemata." *International Studies Quarterly* 41(3) (1997): 403-433

Yet whereas the competence dimension of stereotype content captures characteristics like skill, intelligence, and other dispositional determinants of capability, Image theory’s treatment of the 'capabilities' dimension emphasizes material determinants of capability.[183] This is a major issue for anyone seeking to use Image Theory work to unpack how competence assessments shape interstate interaction, as the behavioral impact of perceived competence cannot be separated from the behavioral impact of perceived material capabilities.

In sum there are major holes, which I seek to fill, in our understanding of when and how perceived irrationality, and beliefs about judgment, matter in international politics. In addition to bridging the gap between IR and prior research on the consequences of bias perception and competence assessments, a principle contribution of this project is its interaction of competence assessments with bias perception. Despite competence assessments and bias perception being theoretically adjacent, as both refer to beliefs about the judgment of others, prior work has yet to investigate the behavioral impact of simultaneous variation in these variables. Further research is therefore necessary to understand how, and indeed whether, these dimensions of *beliefs about judgment* have a different effect on behavior when considered in tandem than in isolation

---

[182] Jervis 1976, Cottam 1977, Herrmann 2013.

[183] Military and economic resources, for example, as opposed to the skill or knowledge necessary to deploy these to their maximal effect.

Why Survey Experiments and why Domestic Public Preferences?:

This purpose of this chapter is to provide microfoundations for the impact of beliefs about adversary rationality on diplomatic preferences, which necessitates the choice of an empirical context. I select public opinion as the domain in which to experimentally test my theory. Public opinion research is not the only feasible research method through which to establish the psychological microfoundations of my theory. Other plausible approaches include laboratory experiments on a sample of elites or field experiments conducted on government officials, whose diplomatic preferences are more immediately associated with the foreign policy of a state. However, there are both theoretical and methodological reasons to focus on the mass public.

First, evidence from historical and contemporary cases suggests that leaders, particularly in democratic states, monitor public opinion and use it to inform their decisions in international crises, from the opening or cessation of diplomacy to the initiation, escalation or termination of military conflict. President Eisenhower's brinksmanship in the 1958 Taiwan crisis was tempered when he realized the US public did not consider defending Quemoy and Matsu to be worth the costs of war. The administration of George W. Bush was intently attentive to the public's attitude towards both the broader war on terror and the American military presence in Iraq and Afghanistan; in 2001 alone, the administration spent a million dollars on polling, though President Bush made a show of disavowing the idea of conducting foreign policy by Gallup polls.[184] Public claims to the contrary aside, Bush (and indeed any democratic leader) had ample reason to account for public opinion in conducting foreign military and diplomatic policy; public disillusionment with the 'forever wars' in Iraq and Afghanistan was a major factor in delivering

---

[184] Kertzer 2016, p. 50

Congress to the democratic party in 2006 and has been cited as a major factor in the contemporary subsumption of 'Bush-era' republicanism under a more isolationist, populist strand of conservatism.

The mechanisms through which public opinion impacts foreign policy are complicated, however: the 'foreign policy by polling' and 'electoral consequences' pathways outlined above are some of the more direct ways in which the mass public can influence elite decisions in international politics, but are not an exhaustive list of causal pathways. For example, domestic groups that oppose the leader and elites currently in power are attentive to public attitudes on international issues, and utilize popular dissatisfaction with existing foreign policy to impede the agenda of the current executive or usher in a change in government leadership. This is precisely the strategy adopted by the US democratic party in designing its electoral messaging for the 2006 midterm elections, but examples of domestic opposition capitalizing on public pushback to particular foreign policy agendas abound in US history. Johnson's Great Society initiative was crushed by unceasing attacks on his approach to Vietnam, while Truman found his domestic options limited and his political prospects ultimately doomed by a consistent drumbeat of accusations that he was soft on Communism and unable to handle the escalating situation in East Asia. Indeed, Johnson subsequently remarked that foreign military endeavors are lost at home rather than abroad, a notion later formalized in the Powell Doctrine which explicitly enshrines widespread public support as necessary for the use of force.[185] Therefore, understanding the general public's support for diplomacy versus force, and how they form expectations about the efficacy of these foreign policy approaches, is of considerable importance in understanding elite choices and outcomes in interstate diplomacy and crises.

---

[185] Weinberger 1984

The public opinion context is also attractive because it permits me to test a theory of the link between beliefs about adversary judgment and diplomatic preferences using experimental methods without the need for infeasible expenditures of resources or time. The basic logic of experiments is intuitive: manipulate elements of the world whether in information provided in a survey (as in this chapter) or in a laboratory setting and observe how participants respond to different versions of the world. This gives researchers working in social scientific contexts unparalleled potential for control, and hence the ability to engage in causal inference without concerns about endogeneity or causal pathways running in unanticipated directions. Outside of an experimental context, it is difficult to disentangle the multiplicity of factors that may feed into individuals' preferences for diplomacy and war either in a general sense or in particular interstate crises. If individuals favor military force even in the presence of power parity, is this because they are uniquely insensitive to the costs of war or overoptimistic about the prospects for success? Or is it because they have internalized the belief that the adversary only understands the language of force? Experiments offer a distinct advantage in answering this question: they permit me to model international political situations directly while manipulating only the central variables of my theory. By presenting subjects with identical information the structural factors of a crises and the actors involved, with the exception of those elements being manipulated by the researcher, I can rely on random assignment to ensure that differences across subject groups are due to the experimental manipulations rather than potential confounders.

Hypotheses:

Before describing the design of my experiments, I first reiterate and further specify the hypotheses they are designed to test. I argue that beliefs about adversary judgment shape individuals' diplomatic preferences in ways exogenous to structure, interests, or personality.

These beliefs shape the general preference for diplomacy vs. force via the mechanisms of their impact on individual expectations about 1) the efficacy of *pure information* versus *coercive leverage* for influencing adversary beliefs and thereby behavior and 2) the costs of a breakdown in negotiations. Each of the following sets of hypotheses refers to the hypothesized effect of the specified treatment relative to the control treatment on the relevant dimension of beliefs about adversary judgment (*perceived objectivity* and *perceived competence*, respectively)

First, I hypothesize that individuals who believe adversary leaders are biased will 1) be less likely to advocate diplomacy versus force 2) expect diplomacy to be less effective relative to force as a means of crisis resolution and 3) expect pure information or argument to be less effective negotiating tools than coercive leverage. This is in line with psychological research on the role of bias-perception in creating conflict spirals in intergroup and interpersonal interaction.[186] Biased others may be expected to hold fixed beliefs, rendering pure information an ineffective tool for influencing behavior through beliefs. Alternatively, they may be expected to have unfixed beliefs but interpret information in flawed ways, rendering pure information an unreliable tool for influencing behavior through beliefs. In short, biased actors cannot be relied upon to understand or shift their beliefs in response to information that exceeds a low level of complexity and non-universality. Therefore, biased actors cannot be persuaded to alter their behavior through rational argument and complex information sharing. Thus, individuals who believe adversary leaders are biased will be pessimistic about and unsupportive of crisis resolution strategies to which these mechanisms are central.

---

[186] Emily Pronin, Kathleen Kennedy, and Sarah Butsch. "Bombing versus negotiating: How preferences for combating terrorism are affected by perceived terrorist rationality." *Basic and Applied Social Psychology* 28, no. 4 (2006): 385-392. Kathleen Kennedy and Emily Pronin. “When Disagreement gets Ugly: Perceptions of Bias and the Escalation of Conflict.” *Personality and Social Psychology Bulletin* 34, no.6 (2008). 833-848. Gilovich, Thomas, and Lee Ross. *The Wisest One in the Room: How You Can Benefit from Social Psychology's Most Powerful Insights*. Simon and Schuster, 2015.

H1A: Subjects who receive the *biased* treatment will be likely to express lower expected efficacy of diplomacy and persuasion through argument relative to military force and coercive leverage
H1B: Subjects who receive the *biased* treatment will be likely to express lower advocacy of diplomacy and persuasion through argument relative to military force and coercive leverage
H1C: Subjects who receive the *biased* treatment will be likely to express lower expected likelihood that adversary leaders can be persuaded by reasonable arguments or understand the risks of their actions.

Second, I hypothesize that individuals who believe adversary leaders are objective will 1) be more likely to advocate diplomacy versus force 2) expect diplomacy to be more effective relative to force as a means of crisis resolution and 3) expect pure information or argument to be more effective negotiating tools relative to coercive leverage. Objective others are expected to hold fluid beliefs that are responsive to new information, and to interpret information objectively, making pure information a potentially effective tool for influencing behavior through beliefs. When belief divergence is a barrier to settlement, individuals who believe adversary leaders are objective are likely to attribute this divergence to missing rather than misinterpreted information. They will therefore expect information sharing and rational argument to lead towards settlement by moving beliefs towards a point of convergence, and be optimistic about and supportive of crisis resolution strategies to which these mechanisms are central.

H2A: Subjects who receive the *objective* treatment will be likely to express higher expected efficacy of diplomacy and persuasion through argument relative to military force and coercive leverage

H2B: Subjects who receive the *objective* treatment will be likely to express higher advocacy of diplomacy and persuasion through argument relative to military force and coercive leverage
H2C: Subjects who receive the *objective* treatment will be likely to express higher expected likelihood that adversary leaders can be persuaded by reasonable arguments or understand the risks of their actions.

Third, I hypothesize that individuals who believe adversary leaders are competent will 1) be more likely to advocate concessions to avoid a breakdown in negotiations and 2) be less likely to advocate military methods relative to diplomatic methods. Higher perceived competence increases the costs of a breakdown in negotiations by lengthening the shadow of the future, as competent actors are seen as unlikely to decline or disappear. Furthermore, all else equal competent adversaries are costlier to fight and harder to defeat: individuals who think adversary leaders are competent will see conflict with the adversary as a greater threat.

H3A: Subjects who receive the *competent* treatment will be likely to express higher advocacy of making significant concessions to the foreign adversary in order to avoid the breakdown of negotiations
H3B: Subjects who receive the *competent* treatment will be likely to express lower advocacy of military methods relative to diplomacy
H3C: Subjects who receive the *competent* treatment will be likely to rate the adversary as more threatening and expect military conflict with the adversary to be costlier

Fourth, I hypothesize that individuals who believe adversary leaders are incompetent will 1) be less likely to advocate concessions to avoid a breakdown in negotiations and 2) be less likely to advocate diplomatic methods over military methods. Higher perceived incompetence decreases the costs of a breakdown in negotiations by shortening the shadow of the future, as incompetent actors are seen as peripheral players that are likely to disappear from the international arena due to mismanagement, subjugation, or defeat. Furthermore, all else equal, incompetent adversaries are less costly to fight and easier to defeat: individuals who think adversary leaders are incompetent will see conflict with the adversary as a lesser threat.

.

H4A: Subjects who receive the *incompetent* treatment will be likely to express lower advocacy of making significant concessions to the foreign adversary in order to avoid the breakdown of negotiations
H4B: Subjects who receive the *incompetent* treatment will be likely to rate the adversary as less threatening and expect military conflict with the adversary to be less costly

Interacting these hypotheses generates several predictions regarding the link between ideal-type beliefs about adversary judgment and diplomatic preferences. Individuals who receive the interacted *Rational* treatment (*objective/competent*) should be most likely to prefer diplomacy and expect it to be effective. Individuals who receive the interacted *Fool* treatment (*objective/incompetent*) should be more likely to expect diplomacy to be effective, but less likely to advocate concessions to prevent escalation if diplomacy breaks down. Individuals who receive the interacted *Fanatic* treatment (*biased/competent*) should be less likely to expect diplomacy to be effective, but more likely to advocate concessions to prevent escalation if diplomacy breaks down (in line with ‘madman theory’). Finally, individuals who receive the interacted *Madman* treatment (*biased/incompetent*) should be least likely to prefer diplomacy or expect it to be effective, and most likely to advocate abstention from negotiations entirely.

Alternative Hypotheses:

Extant research on interstate diplomacy and conflict provides several clear alternative explanations to my hypotheses for individual variation in interstate diplomatic preferences. These fall into several broad categories. First are structural explanations: the balance of power, threat, and interests. Second are domestic political explanations: individuals should simply espouse diplomatic preferences in line with the policy position of their political party. Third are explanations centered on endogenous variation in individual preferences, such as hawkishness/dovishness, or personality, notably social value orientation. I elaborate on each of these below, and in the design of the experiment describe how I plan to test or control for each.

Structural explanations endogenize variation in diplomatic preferences to differences in relative power, along with the willingness to use said power in the crisis at hand, between the actors. The stronger an actor is relative to its adversary, the more willing it should be to use force, the less willing it should be to engage in diplomacy in order to reach a compromise settlement, and the more it should rely on coercion as opposed to reasoned argument in the context of negotiations. The weaker an actor is relative to it adversary, the less willing it should be to use force, the more willing it should be to engage in diplomacy, and the more it should rely on reasoned argument as opposed to coercion in the context of negotiations. Effectively, the strong will take what they want and the weak will suffer as they must.

An alternative but still structural explanation for variation in diplomatic preferences centers on the balance of interests; in brief, the higher the stakes for an actor in a crisis, the more willing it should be to use force and the less willing it should be to compromise. The balance of interests is less materially grounded, and therefore harder to operationalize, than the balance of capabilities, but there are some common proxies for superiority in the balance of interests. Namely, states that are defending the status quo, that are fighting to protect territory, or that otherwise have an existential stake in the issues at hand should have an advantage in the balance of interests. Unlike with the balance of power explanation, there is no *a priori* theoretical reason to expect the balance of interests to have a systematic impact on the use of coercion versus rational argument in the context of any negotiation.

Domestic political affiliation provides an alternative model of individual diplomatic preferences, one based not on structural factors but on partisan cues from political elites. Public opinion research suggests that the primary driver of differences in foreign policy opinion among members of the mass public is individuals' political party and ideology; people will espouse whatever foreign policy position coheres with their party line as signaled by elites in the media and political spheres. In the context of American politics republicans and conservatives are more likely to support the use of force abroad, while democrats and liberals are more likely to support international diplomacy and negotiations, though this is not an ironclad rule and historically has varied based on shifting positions of party elites or unique international political contexts. This domestic political explanation for variation in individuals' diplomatic preferences seems plausible given how little a majority of Americans know, care, or think about Foreign policy. Lacking incentives to acquire detailed information or think through policy issues, the majority of Americans may simply turn to the simplifying heuristic of political party and so engage in this sort of reflexive, cognitively shallow position taking.

In contrast to structural and political theories, dispositional theories posit a variety of individual level traits, in the form of belief systems or personality, that drive variation in diplomatic preferences in ways exogenous to structural or political variables. Belief system level explanations include the following: highly nationalistic individuals are more likely to favor force than cosmopolitan individuals, and hawkish individuals are more likely to favor force than dovish individuals.

Likewise, there are a variety of existing personality-based explanations for variation in diplomatic preferences in an interstate context. First, individuals who are high in social value orientation and epistemic motivation should be most likely to favor diplomacy as reasoned dialogue, while high epistemic pro-selfs are unlikely to advocate compromise but likely to advocate diplomacy as long as they expect their opponent to do the same, low epistemic pro-selfs are unlikely to favor diplomacy, and low epistemic pro-socials are likely to favor diplomacy and be quick to compromise. Second, individuals who score high in measures of empathy and perspective taking are likely to favor diplomacy compared to those who score low in empathy. Third, individuals who score high on measures of authoritarian personality or social dominance orientation are more likely to advocate force than individuals who score low on these same measures.

Finally, perhaps the simplest alternative explanation for variation in individual diplomatic preferences is also the most intuitive; the more negatively individuals feel about an adversary, the more likely they are to advocate military force against it, and vice-versa. There are two principle reasons why this affective explanation for heterogenous diplomatic preferences is one this project must take very seriously. First, positive and negative feelings towards others could plausibly be correlated to the perceived objectivity and competence; it may be that individuals feel more positively about counterparts they believe are objective than about counterparts they believe are biased, and more positively about counterparts they believe are competent than counterparts they believe are incompetent. Second, the affective explanation's posited mechanism for heterogeneity of diplomatic preferences runs directly counter to this project's hypothesized mechanisms regarding the expected efficacy of information sharing. Contrary to that cognitively grounded hypothesis, the affective argument is that people are more prone to use force against counterparts they dislike and more prone to engage in dialogue with counterparts they like, regardless of cognitively driven variation in the relative expected efficacy of these approaches. In other words, because warm/cold feeling towards a counterpart is potentially correlated with the IV (beliefs about judgment), and because the posited mechanisms stand in stark contrast, this alternative hypothesis is one that must be robustly accounted for in the research design of this project.

I test my hypotheses against these alternative explanations using a pair of survey experiments, deemed Study 1 and Study 2. Study 1 presents subjects with a fictional crisis involving the leaders of an imaginary country. Study 2 replicates the design of study 1, but increases the realism of the task by substituting the fictional crisis with the US/North Korea nuclear crisis. The use of a real foreign adversary constitutes a 'hard test' of the hypotheses; given widespread media coverage of North Korea, Kim Jong Un, and the US-North Korean nuclear crisis, there is a high probability that respondents' prior beliefs would weaken the impact of experimental treatments. Below for Study 1 and 2 respectively, I first detail the design then present and discuss the results.

Experimental Design: Study 1:

Study 1 was fielded through Amazon's Mechanical Turk (MTurk) on a sample of 1,119 subjects[187] (accounting for responses deemed ineligible) as well as a nationally representative sample of 977 subjects recruited through SSI. To incentivize attention and effort, all subjects were informed they would be ineligible to receive compensation if they failed one of several attention checks. See Appendix A for full survey design.

The survey employs a factorial treatment design. Respondents first read identical descriptions of an ongoing crisis with the imaginary state of Mykonara, which is under the leadership of Supreme President Ekon Omari. The use of a hypothetical foreign state is intended to minimize, though not eliminate, bias from prior information. Descriptions of Mykonara's capabilities, intentions, culture, and history are all kept constant across subjects. This design controls, albeit potentially imperfectly, for subjects' perception of relative material capabilities between the US and the foreign state**.**

---

[187] Subjects were allowed to participate only once and paid a base rate of $1.00.

Respondents are randomly assigned to either: 1) Read one article excerpt about the adversary's objectivity *or* competence, 2) Read two article excerpts, one about the adversary's objectivity and one about its competence, or 3) Read nothing (control). The article excerpts were formatted to resemble a generic opinion article with a "Foreign Affairs" heading[188] There are two versions of the '*objectivity*' article (objective/biased) and two of the '*competence*' article (competent/incompetent): respondents are randomly shown one, but never both, versions of articles to which they are assigned. (See Appendix for full treatment texts). There are thus nine total treatment groups including the control.

The treatment is not designed to affect subject responses by providing novel information. It is unrealistic to expect the use of a hypothetical state to eliminate the impact of priors.[189] By backing descriptions of adversary judgment with quotations, detailed examples and other vivid information, the treatments aim to prime existing information or beliefs by manipulating their salience.

An important caveat is that this design should somewhat mute the difference between the impact of perceived *competence* and *incompetence* (**H3, H4**) even if competent adversaries are perceived to be more capable than incompetent ones. The crisis presented is between the US and a foreign state to increase its salience for the sample population, but this makes the scenario one in which structural factors overwhelmingly favor the US. Given the massive military gap between the US and all other states, perceptions of relative capability are unlikely to shift substantially even if an adversary has incredible strategic skill: competence only goes so far.[190] In other words, the way in which this design controls for perceptions of relative capability substantially increases the likelihood that subjects will behave in accordance with the power-driven alternative argument**.**

---

[188] I chose this publication as one that evokes minimal or no partisan associations.

[189] Subjects presented with a hypothetical country tend to either consciously attempt to figure out what the 'real' country is or subconsciously substitute beliefs and information they have about an existing country in order to fill in gaps.

The '*biased'* treatment emphasizes: 1) the leader is delusional, with beliefs that have little grounding in reality, 2) the leader's narcissism makes him overconfident and resistant to having his views challenged, 3) the leader is a dogmatic ideologue and dismisses evidence at odds with his ideology, 4) the leader is temperamental and lets emotion dictate his beliefs, 5) the leader's advisers are yes-men who will not challenge his beliefs or tell him things he does not want to hear. The '*objective'* treatment emphasizes: 1) the leader is objective and holds realistic beliefs, 2) the leader is open to alternative views, 3) the leader is a pragmatist, 4) the leader does not allow emotions to dictate his beliefs, 5) the leader's advisors will challenge his views and give him unfiltered information. The '*incompetent*' treatment emphasizes: 1) the leader has no military skill or experience, 2) the leader has a history of self-destructive military actions and threats, 3) the leader is a tyrant with no governing skill, 4) the leader has weakened his power through ill-conceived purges, 5) the military is disorganized, untrained, corrupt, and led by generals promoted for loyalty rather than experience. The '*competent'* treatment emphasizes: 1) the leader has military experience and is a skilled commander, 2) the leader has a history of savvy military actions and threats, 3) the leader is a tyrant skilled in governance, 4) the leader consolidated power through necessary purges 5) the military is organized, highly trained, and led by expert generals.

---

[190] This tradeoff is acceptable given that while there is substantial research on the effects of perceived capability differences, there is a total absence of empirics on perceived objectivity

Treatments were double-blind coded to check that they 1) conveyed the intended information about the adversary's judgment and 2) did not contain extraneous information that could influence subject responses. For example, it was important that treatments not convey different information about adversary capabilities or generate systematic variation in warm/cold feeling toward the adversary.

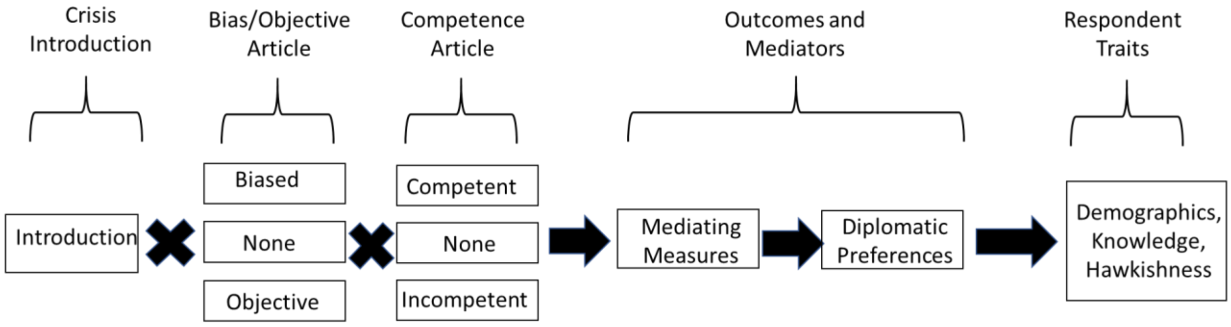
Post-treatment, respondents answer an array of questions on: 1) attitudes and beliefs about the adversary, 2) preferences for and expected efficacy of diplomatic vs. military methods of crisis resolution, 3) preferences for and expected efficacy of leverage vs. dialogue within the context of negotiation, 4) support for major concessions to avoid escalation if negotiations breakdown. Most of these are multiple choice questions using 7-point Likert-type scales[191] (See Appendix for full questions). The mediating measures aim to identify mechanisms and account for some of the alternative hypotheses. They ask whether respondents think the adversary can be persuaded by reasonable arguments (**H1C, H2C),** how great a threat they think the adversary poses **(H3C, H4C)**, how important they think the issues at stake in the crisis are to the US and how they feel toward the adversary on a feeling thermometer. If my hypothesized mechanisms are correct, I would for example expect mean ratings of the adversary's reasonableness, but not the feeling thermometer or the importance of the crisis, to vary systematically between treatment groups, specifically the *objective* and *biased* group.

Finally, the survey administers measures of respondent characteristics to account for the remaining alternative hypotheses. These start with standard demographic questions including political party and political ideology, which additionally albeit imperfectly proxy for social value orientation, social dominance orientation, and empathy.[192] Respondents then receive questions to measure traits that could drive systematic variation in results, such as political knowledge, nationalism/cosmopolitanism, and hawkishness/doveishness. To avoid survey fatigue, measures of empathy and social dominance orientation were necessarily minimalistic as opposed to the comprehensive question batteries used in experiments for which these traits are of central importance as either principle or alternative hypotheses. While political party and ideology provide secondary though imperfect proxies for these traits, the survey instrument is therefore admittedly unable to comprehensively account for any impact of these respondent traits on the dependent variables of interest (See Figure 3.1 for Survey Flow).

---

[191] See appendix for question response wording

Figure 3.1: Survey Flow


Crisis Introduction
Bias/Objective Article
Competence Article
Outcomes and Mediators
Respondent Traits
Introduction
Biased
None
Objective
Competent
None
Incompetent
Mediating Measures
Diplomatic Preferences
Demographics, Knowledge, Hawkishness

Responses were analyzed using both the original 7-point Likert scales and a binary variable generated for ease of interpretation that was coded to equal 1 if a response was anti-diplomacy and a 0 if a response was pro-diplomacy. Indifferent responses are coded as “1” in these binary variables,[193] as several considerations make the *absence* of support for diplomacy a notable result. First, a majority of Americans believe peace is best secured through diplomacy rather than strength:[194] in line with this observation the mean and median response is pro-diplomacy across most DV questions. Second, the MTurk samples skewed liberal, democratic, educated, and female, all demographic groups that are less likely to support military force.[195] Finally, military action against a nuclear armed adversary carries extremely high costs and risks relative to diplomacy. The effect of each treatment condition was analyzed both independently of (*main effects*) and interacted with the other treatment condition to which respondents were assigned.[196]

---

[192] Kertzer and Rathbun 2014.

[193] For an alternative version of the binary variables that instead omits indifferent responses, see Appendix

Results 1: Beliefs About Judgment and Diplomatic Bargaining Preferences in a Hypothetical Scenario:

Are people's diplomatic preferences impacted by beliefs about an adversary's judgment, or do their preferences reflect their disposition, attitude towards the adversary, and the balance of capabilities and interests? Tables 3.1 and 3.2 present descriptive data from the Study 1 MTurk sample: results from each binary DV are reported by both main and interacted treatment conditions. Figure 3.2 presents a visualization of the efficacy measures from these tables as confidence interval plots.[197] (See Appendix for analysis of Likert variables and full regression tables).

---

[194] *Pew Research Center:* "Trust, Facts and Democracy" (Foreign Policy): 2017

[195] *Pew Research Center*. Mary-Kate Lizotte. "Investigating the Origins of the Gender Gap in Support for War." *Political Studies Review* (2017)

[196] For example, the independent/main analysis of *Biased* includes any respondent in one of the following treatment combinations: *Biased/Control, Biased/Competent, Biased/Incompetent*. These independent groups had an N of 355-380 in the Mykonara survey and 442-446 in the North Korea Survey. By contrast, the interaction analysis of *Biased* includes only respondents that received *Biased/Control*: the interaction groups had an N of 116-130 in the Mykonara survey and 147-159 in the North Korea survey.

[197] See Appendix for this table with all treatment groups.

Table 3.1: Descriptive Results For Study 1, MTurk (Main Treatments)

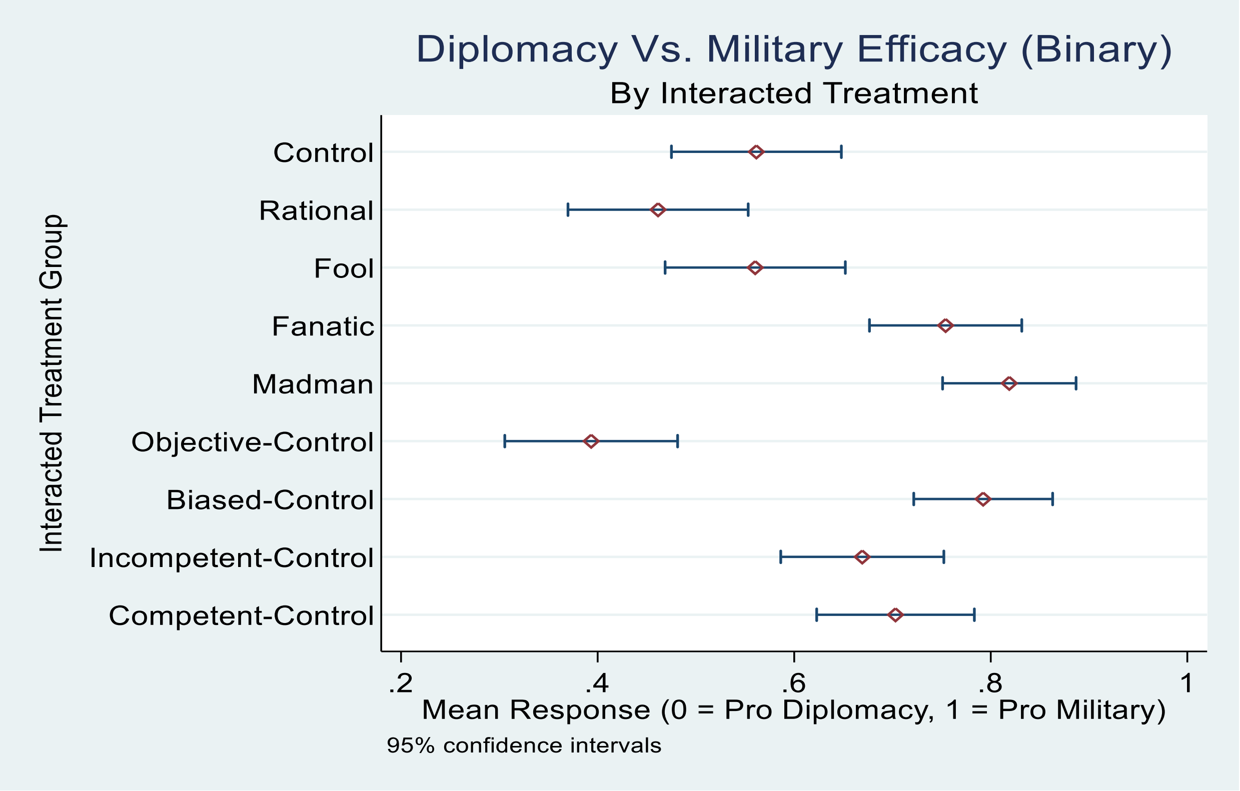
Treatment Group	**Binary Dependent Variable** % of Group coded ‘1’ (Anti-Diplomacy)				
	Diplomacy vs. Military (Preference)	*Diplomacy vs. Military (Efficacy)*	*Diplomacy (Preference)*	*Diplomacy (Efficacy)*	*Leverage vs. Dialogue (Efficacy)*
Objective (N=355)	28.1%	47%***	19.6%*	47%***	53.8%***
Biased (N=379)	41.7%***	78.9%***	34.5%**	71.7%**	81.2%***
Incompetent (N=370)	38.4%**	68.6%**	31%	61.6%	72.9%**
Competent (N=367)	34.6%	64.3%	30.2%	64.8%**	71.9%*
Control (N=130)	20.7%	56.1%	22.3%	53.8%	66.1%

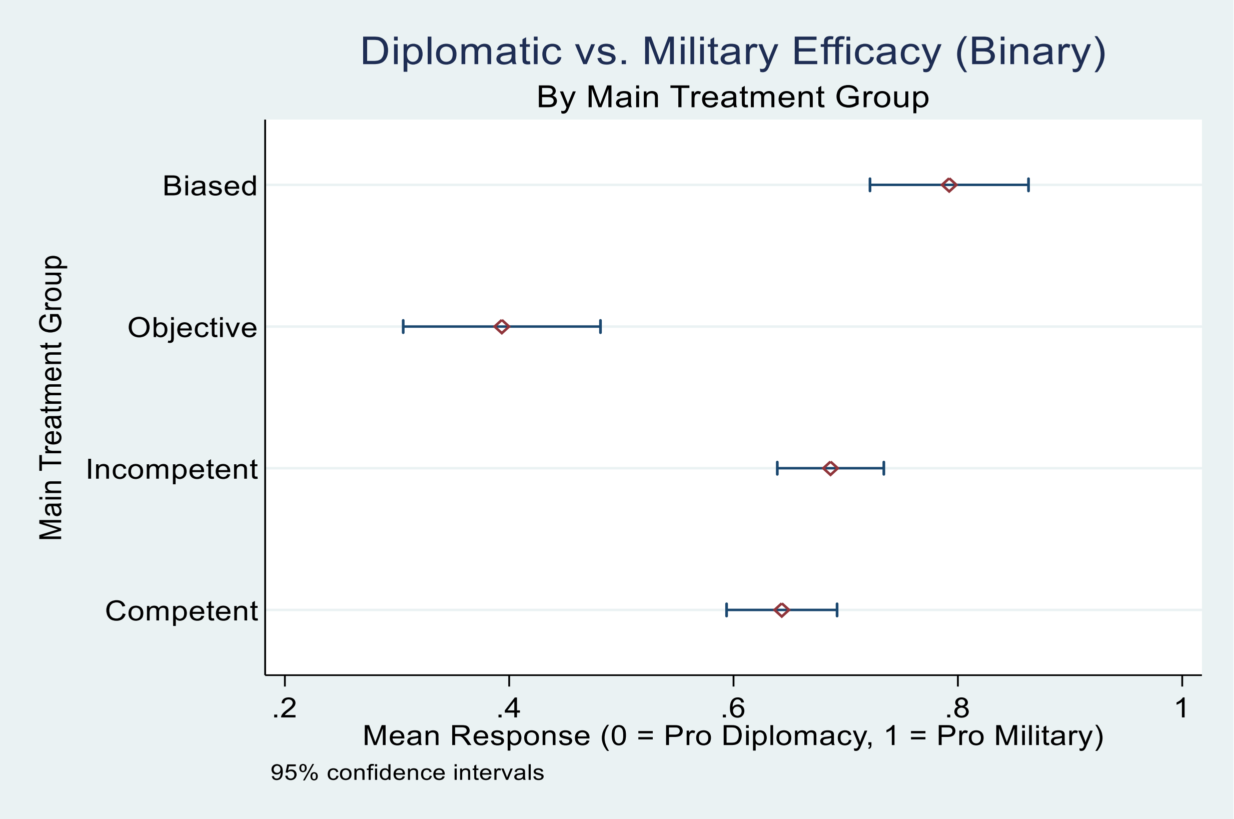
Table 3.2: Descriptive Results For Study 1, MTurk (Interacted Treatments)

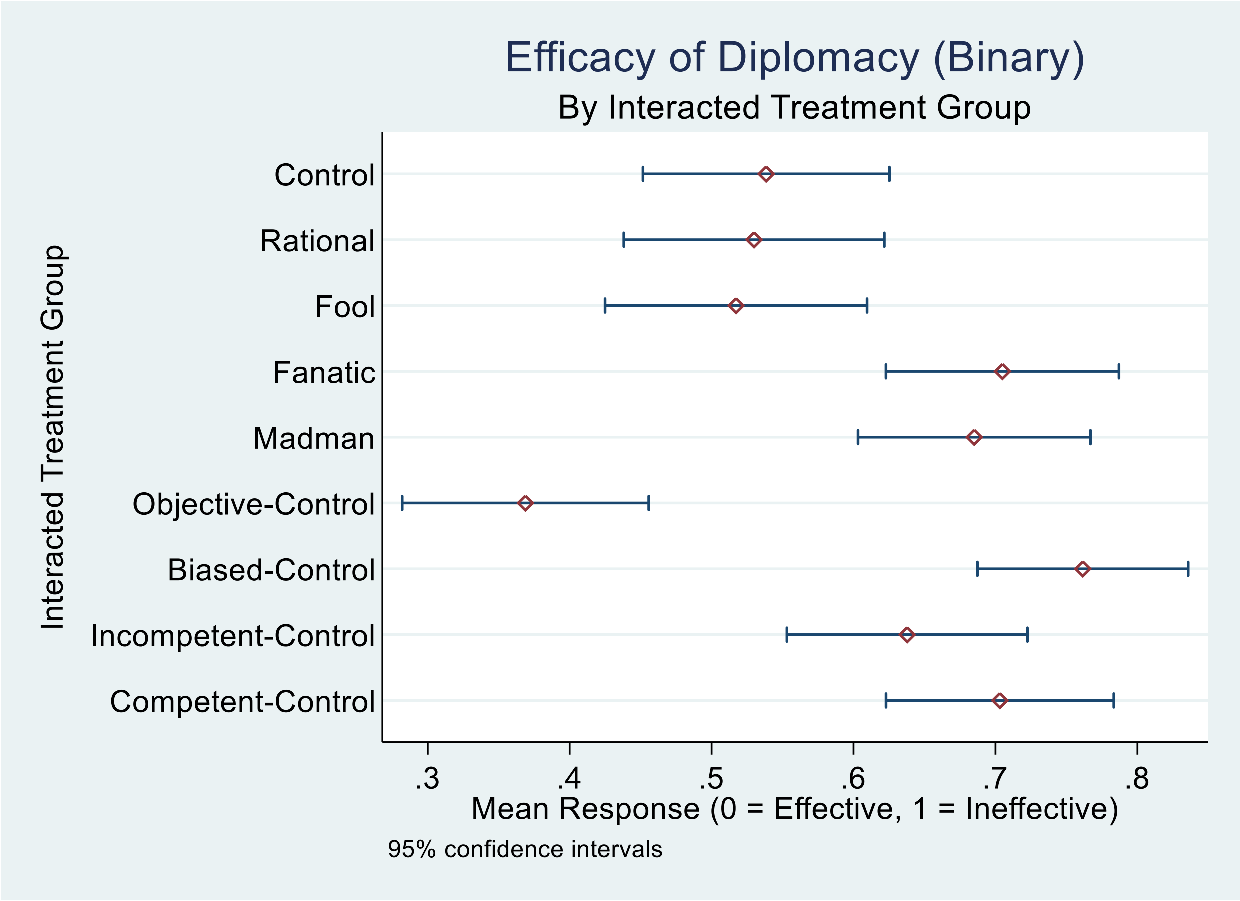
Treatment Group	**Binary Dependent Variable** % of Group coded ‘1’ (Anti-Diplomacy)				
	Diplomacy vs. Military (Preference)	*Diplomacy vs. Military (Efficacy)*	*Diplomacy (Preference)*	*Diplomacy (Efficacy)*	*Leverage vs. Dialogue (Efficacy)*
Madman (N=127)	44.9%***	81.9%***	34.6%*	68.5%**	81.1%**
Rational (N=117)	31.6%	46.1%	25.6%	52.9%	58.9%
Fool (N=116)	26%	56%	21.5%*	51.7%	58.6%
Fanatic (N=122)	41%***	75.4%**	34.4%*	70.5**	82.8%**
Control (N=130)	20.7%	56.1%	22.3%	53.8%	66.1%

Significance Key: *= p ≤ 0.05, ** = p ≤ 0.01, *** = p ≤ 0.001
Significance reported from logit with “*Control*” as constant. See Appendix for all interactions.

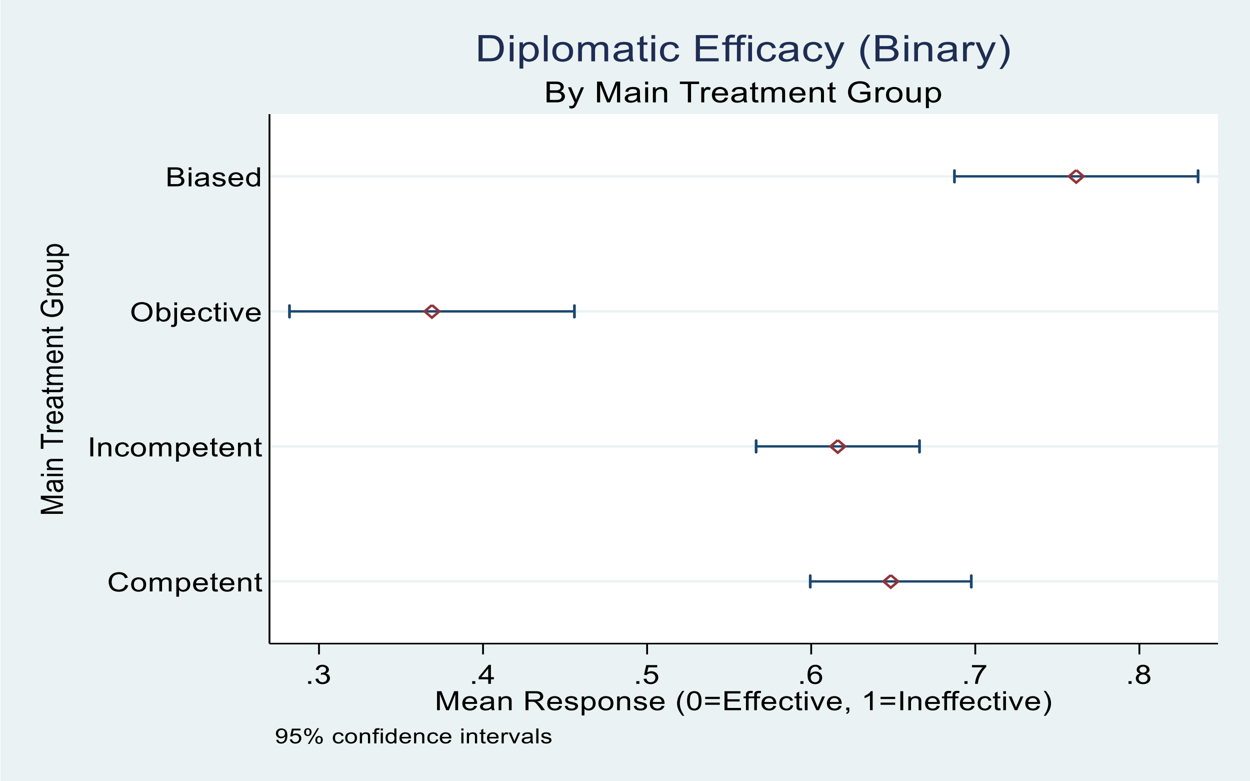
Figure 3.2: Mean and confidence intervals on Binary Dependent Variables:







Efficacy of Diplomacy (Binary)
By Interacted Treatment Group
Interacted Treatment Group
Control
Rational
Fool
Fanatic
Madman
Objective-Control
Biased-Control
Incompetent-Control
Competent-Control
.3
.4
.5
.6
.7
.8
Mean Response (0 = Effective, 1 = Ineffective)
95% confidence intervals



Diplomatic Efficacy (Binary)
By Main Treatment Group
Main Treatment Group
Biased
Objective
Incompetent
Competent
.3
.4
.5
.6
.7
.8
Mean Response (0=Effective, 1=Ineffective)
95% confidence intervals

The results strongly support H1A and H1B. 41.7% of subjects in the *Biased* condition, and 44.8% from the *Madman* interaction, preferred military to diplomatic methods, roughly double the *Control* (20%) and *Objective* (28%) groups **(H1B, H2B)**.[198] *Biased* (35%) also had the highest proportion of respondents that did not support diplomacy considered on its own rather than as the alternative to military action, each a substantial increase relative to the proportion from *Objective* (22%) and *Control* (22%).

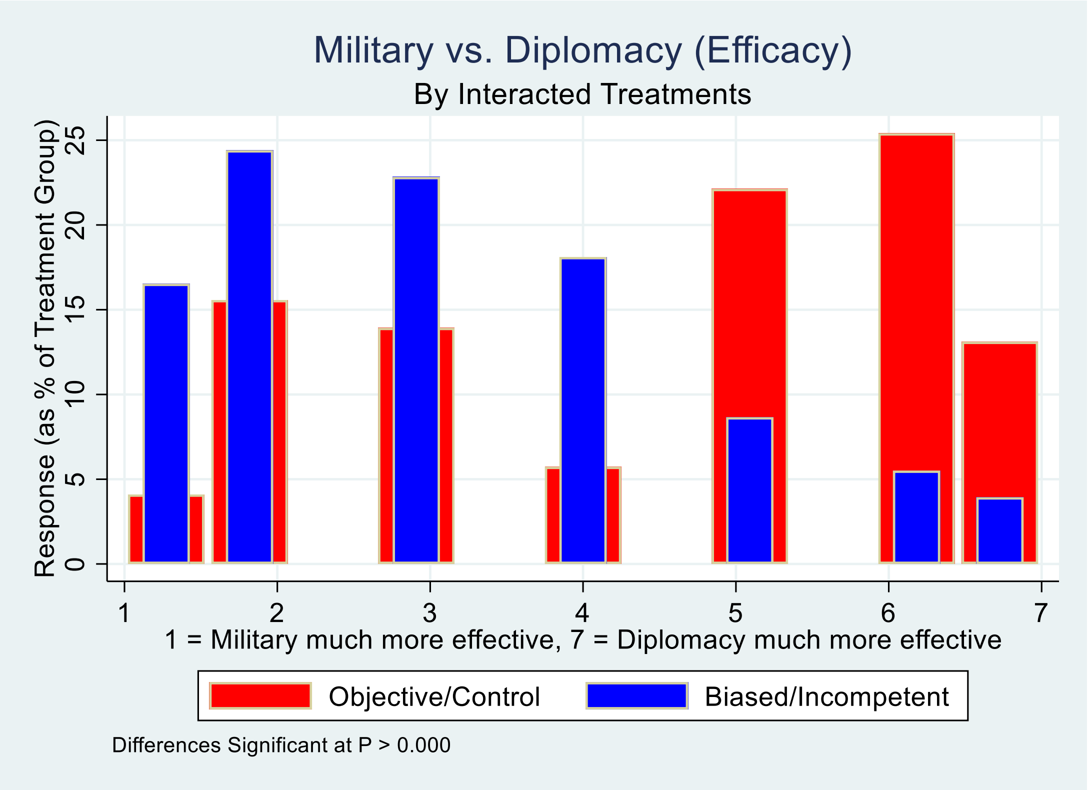
Furthermore, the *Biased* treatment made subjects much less likely to think diplomacy would be effective, independently or contrasted with military action **(H1A)**. 78% of the *Biased* group and 82% of the *Madman* group thought military methods would be more effective than diplomatic methods, relative to 47% of the *Objective* **(H2A)** and 56% of the *Control*.[199] This pattern held when respondents were asked how likely negotiations were to resolve the crisis without contrasting diplomacy to military methods: 28% of the *Biased* group and 22% of the *Madman* group thought negotiations would be effective. By contrast 53% of the *Objective* and 63% of the *Objective/Control group* thought negotiations were at least "somewhat likely" to be effective **(H2A)**, while the *Control* group was close to evenly split: 46.2% thought it was at least "somewhat likely" negotiations would be effective. Figure 3.3 illustrates the difference in expected efficacy of military vs. diplomacy between the *Objective* and *Biased* treatments, as well as the *Madman* and *Objective/Control* treatments, on the original Likert scale (the hypotheses predict the latter two interactions will manifest the greatest difference in their impact on this measure).

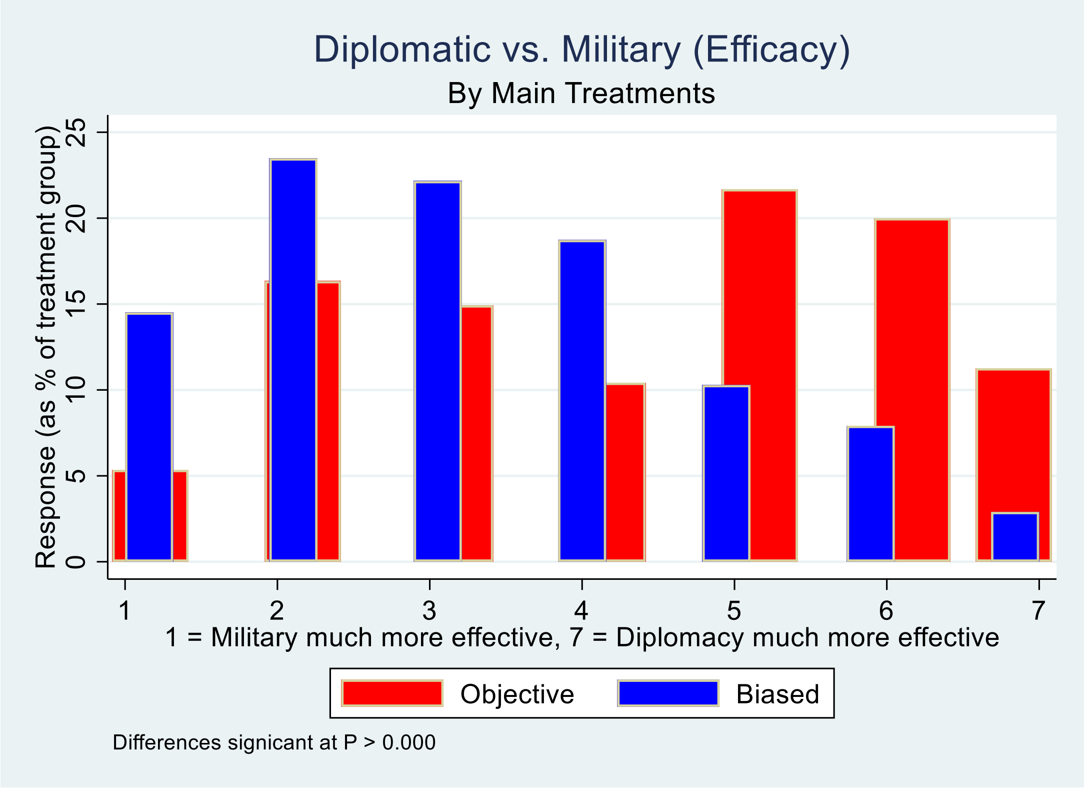
---

[198] Regression on the Likert scale showed a 0.7 point difference in coefficients of *Objective* and *Biased*

[199] Regression on the Likert scale showed a 1.1 point difference in the coefficients of *Objective* and *Biased*

Figure 3.3: Histograms of Expected Efficacy of Military vs. Diplomacy:





Strengthening support of H1A and H1B, The *Biased* treatment had a similar effect on respondents' estimates of the efficacy of leverage vs. dialogue. 33% of the *Control* group thought dialogue would be more effective, compared to 46% of subjects in the *Objective* treatment, and 53% in the *Objective/Control.* By contrast, 19% of subjects that saw the *Biased* article (18% of the *Madman*) thought dialogue would be more effective, less than half the proportion from the *Objective* group.[200]

Results also provide some support for H4B; the *Incompetent* group had the highest proportion of respondents who thought the breakdown of negotiations and resultant escalation was preferable to concessions (66%), an increase of 17% over the *Control* (49%). Only the *Incompetent* treatment had a statistically significant effect on this measure, whether on the binary or Likert scale.

Finally, mediating measures support the theory that variation in beliefs about judgment influences diplomatic preferences through people's conception of their adversary's reasonableness, not their warm or cold feelings towards the adversary. The *Objective* group on average thought it was "very likely" adversary leaders could be persuaded by reasonable argument, while the *Biased* group on average thought it was "very unlikely:"[201] close to a 3-point difference in coefficients on a 7-point scale **(H1C, H2C)**. By contrast treatment effects on measures of alternative explanations were substantively small and few were statistically significant; on the feeling thermometer only *Objective* and *Incompetent* met the minimal significance standard, and each generated at most a 5-point shift on a 100-point feeling thermometer.

---

[200] Regression on the unaltered Likert scale shows a 1 point difference in the coefficients of *Objective* and *Biased*

[201] Mean and 95% CI from Likert regression: *Objective* (-1.06, -1.48/-0.64)***, *Biased* (1.53, 1.12/1.95)***

To increase confidence in the results, and account for the possibility of systematic bias in the MTurk respondent pool, Study 1 was administered to a nationally representative sample of 977 US adults collected through SSI.[202] The results from this sample resemble those from the MTurk sample, and strengthen support for the primary hypotheses, especially **H1** and **H2**; the directional impact of each treatment was the same for each DV, and the effect sizes were similar (Table 3.3 presents descriptive results on the binary DVs). Of the *biased* treatment group, 42.7% preferred military methods **(H1B)** compared to 26.3% of the *objective* group **(H2B)** and 33.5% of the *control*. 79.9% of the *biased* treatment group thought military methods would be more effective than diplomatic methods **(H1A)**, versus 54.9% of the *objective* group **(H2A)** and 64.9% of the *control*. Finally, 83.3% of the *biased* group thought coercive leverage was a more effective negotiating tool than rational persuasion, compared to just 65.3% of the *objective* group and 77.6% of the *control*. The principle ways the SSI results differ from MTurk are: mean responses were less pro-diplomacy, there were a greater proportion of 'neutral' responses, and the *objective* treatment more frequently had a statistically significant effect across DVs.

---

[202] The subjects were balanced by age, education, party ID, geography, sex, and other demographic factors. 773 subjects were screened out for failed attention checks by the time the survey reached the target threshold of 975 acceptable responses.

Table 3.3: Descriptive Results For Study 1, SSI (Main Treatments)

Treatment Group	**Binary Dependent Variable** % of Group coded '1' (Anti-Diplomacy)				
	Diplomacy vs. Military (Preference)	*Diplomacy vs. Military (Efficacy)*	*Diplomacy (Preference)*	*Diplomacy (Efficacy)*	*Leverage vs. Dialogue (Efficacy)*
Objective (N=308)	**26.3%***	**54.9%*****	**32%**	**58.77%*****	**65.3%*****
Biased (N=288)	**42.7%**	**79.9%****	**43%****	**78.4%***	**83.3%**
Incompetent (N=322)	**36.3%**	**72.9%*****	**38%**	**74.22%****	**79.8%**
Competent (N=302)	**35.7%**	**74.5%*****	**34.4%**	**74.5%****	**77.1%**
Control (N=134)	**33.5%**	**64.9%**	**32.8%**	**64.9%**	**77.6%**

Significance Key: *= $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$
Significance reported from logit with "*Control*" as constant. See Appendix for all interactions.

In sum, with and without individual controls, both samples support the hypothesized impact of perceived bias and objectivity on: preference for diplomatic versus military methods, expected efficacy of diplomatic versus military methods, and expected efficacy of dialogue versus leverage. Furthermore, though the design likely muted the impact of perceived competence and incompetence, both samples show that perceived incompetence reduces individuals' support for concessions to adversaries **(H4A)**.

Experimental Design: Study 2:

Study 2 was fielded through MTurk on a sample of 1,524 subjects. Study 2 uses an identical design to Study 1, but increases the realism of the crisis scenario by naming the adversary state as North Korea and replacing Supreme President Ekon Omari with Kim Jong-Un. To maximize salience and create a hard test of the hypotheses, Study 2 was administered nine days before the Singapore summit between President Donald Trump and Kim Jong-Un.

As with Study 1, the treatments in Study 2 are not intended to affect subject responses by providing novel information. It is unrealistic to expect subjects to have no prior information about North Korea, but the provision of vivid information can increase the salience of narratives about North Korea to which subjects could already have been exposed.[203] Indeed, given the prevalence of North Korea in the American media environment at the time Study 2 was fielded, the respondent pool for Study 2 is likely to, relative to an equivalent sample taken at a different, contained a higher proportion of individuals exposed to a multiplicity of narratives regarding Kim Jong Un's capacity for rational judgment, or lack thereof.

Fielding this companion study offers several advantages. First, replicating any study on a new sample (subjects from Study 1 were excluded from Study 2) increases confidence in the results. Additionally, making the same predictions in two studies that use the same conceptual design and analysis strategy should increase readers' confidence that my hypotheses were not formulated after seeing results.

Second, using a real, highly salient international adversary in the context of a potential crisis situation that was receiving blanket news coverage at the time of the study[204] poses the hardest possible test of the hypotheses. This design makes it far more likely that responses will be influenced by attitudes, beliefs, and information that the experiment cannot control. In particular, given the partisan charge recently injected into the issue of US policy towards North Korea, this design provides ample opportunity for partisan or ideological beliefs to express themselves and overpower the effects of the treatments. In other words, Study 2 should produce more externally valid results, though we would expect small effect sizes relative to Study 1.

---

[203] Subjects presented with a hypothetical country tend to either consciously attempt to figure out what the 'real' country is or subconsciously substitute beliefs and information they have about an existing country in order to fill in gaps.

[204] The survey was fielded nine days before President Trump met Kim Jong Un for a summit in Singapore on June 12, 2018

Results 2: Beliefs About Judgment in the US/North Korea Nuclear Crisis:

The results from Study 2 underscore the results from Study 1 and provide additional support for the primary hypotheses. Across DV measures the treatments moved respondents in the same direction as in Survey 1, and while effect sizes were (expectedly) smaller even minor effects are substantively interesting in such a hard test. Differences in effects of contrasting treatments were substantively and statistically significant across DVs. Yet relative to Study 1 significance was somewhat muted in regression analyses that simultaneously compared all treatment groups to the control, suggesting that most respondents (unsurprisingly) had stronger priors about North Korea than an imaginary state. However, there were some notable exceptions to this, elaborated below.

Supporting H1B and H2B, treatments had a notable effect on whether respondents preferred diplomatic or military methods, a striking finding considering the dire consequences of conflict. 29% of the *Objective* and 25% of the *Rational* group preferred military methods or were indifferent **(H2B)**, in contrast with 35% in *Biased* and 40% in *Madman* who said they preferred war to diplomacy **(H1B)**: a 15% gap between the opposite ideal-types. If anything, these results bias the divergence in treatment effects downwards since the Likert results from *Madman* and *Biased* skew heavily towards the highest preference for military methods, while those from the *Objective* treatment group skew towards the highest preference for diplomacy, but this skewed distribution is not captured by the binary measure.

The *Biased* treatment made subjects much less likely to think diplomacy would be effective in contrast with military action. 64% of subjects in the *Biased* group and 68% in the *Biased/Control* thought military methods would be more effective than diplomatic methods **(H1A)**, relative to 53% of the *Objective* group and 50% of the *Objective/Control* **(H2A).**[205] Again, these percentages bias the treatment effects downwards since the Likert results from *Biased* and *Biased/Control* skew heavily towards the highest expected efficacy of military methods, while those from the *Objective* treatment group skew almost as heavily towards the highest expected efficacy of diplomacy. The treatments had a directionally comparable impact on expected efficacy of coercive leverage versus rational dialogue: *Biased* reduced the proportion of subjects who thought dialogue would be more effective (33%) while *Objective* (41%) increased it.

Of particular interest, *Biased* (64%) and especially *Fanatic* (71%) increased the proportion of subjects who agreed it was better to make major concessions to North Korea than to have negotiations break down **(H3A)**. These results, which are highly significant on both the binary and Likert scales, represent a substantial increase over *Madman* (57%), *Objective* (58%) and *Incompetent* (55%). This supports the interaction of H1A, H3A and H4A, and provides empirical evidence for aspects of the original madman theory.

Finally, as in Study 1 mediating measures in Study 2 suggest that my hypothesized mechanisms rather than plausible alternative explanations are driving the treatment effects on the principle dependent variables. Subjects in the *biased* treatment group were substantially less likely to believe that North Korean leadership could be persuaded by reasonable arguments or understood the risk of their actions **(H1C)**, while subjects in the *objective* treatment group expressed the opposite belief **(H2C)**. By contrast, results on the feeling thermometer did not show substantively or statistically significant differences between treatment groups, with subjects universally expressing cold feelings towards North Korea: unsurprisingly, in other words, most American adults dislike North Korea. Therefore, we can conclude that variation in affect is not driving the results in Study 2 with even greater certainty than in Study 1, which showed substantively larger, though still small and statistically insignificant, differences between treatment groups in the feeling thermometer.

---

[205] Other main effects and interactions moved in the same direction as in the MY survey, but the main effect of *Biased* and the *Biased/Control* interactions were by far the most statistically significant relationships.

Discussion:

I now discuss the implications of the results, and consider potential challenges to drawing inferences about interstate relations on the basis of the survey stemming from the design of the experiment, the use of a mass public sample, and other plausible threats to validity.

Results from the surveys strongly support the hypothesized impact on diplomatic preferences of both independent and interacted beliefs about judgment. Respondents who believe an adversary is biased are substantially less likely to advocate diplomacy, or to expect diplomacy and dialogue to be effective, than those who believe an adversary is objective. Conversely, respondents who believe an adversary is biased are more likely to advocate military methods of dealing with an adversary, and more likely to expect military might and leverage to be effective, than those who believe an adversary is objective. Furthermore, respondents who believe an adversary is competent are much less likely to prefer military methods and more likely to advocate concessions than those who believe an adversary is incompetent. This is despite perceived competence making no difference in expected efficacy of military methods: respondents believed force would be equally effective against either competent or incompetent adversaries, but that fighting a competent opponent would be costlier.[206] These relationships are present when holding constant the balance of capabilities, adversary intentions, adversary regime type, and how people feel about the adversary.[207] In sum, higher order beliefs about how an adversary *thinks* are a crucial component of diplomatic preferences: neither structural nor individual factors are deterministic.

The findings are especially notable because they constitute empirical evidence for the perceptual foundations of 'madman theory,' in addition to my theory of how beliefs about judgment impact diplomatic preferences. The results suggest that individuals are more likely to advocate concessions to adversaries they believe to be unreasonable provided they *also* believe said adversaries are competent. Since this effect was *more* rather than *less* pronounced in Survey 2, it passes the hardest test of the theory, and does not seem to be an artifact of abstraction or low stakes. In other words, at least among the public, it appears that people are more likely to make concessions to capable opponents they believe are unreasonable, 'crazy,' or irrational.

It is notable that both surveys found difference in treatments' effect sizes between 'preference' and 'efficacy' DVs. This distinction seems to validate the theoretical basis of the design: advocacy of policies is correlated with but distinct from the belief that said policies will work. It is worth considering the implications of this divergence between 'preference' and 'efficacy' for our ability to draw valid inferences about international politics from the experimental results. Recent studies show little distinction between 'mass public' and 'elite' samples across a range of experimental conditions. Therefore, if a substantial proportion of respondents advocated diplomacy over military force despite believing diplomatic methods would be less effective than military ones, should we expect policymakers to form preferences and by extension make choices in a similar way?

---

[206] This interpretation is supported by the relationship between *Competent/Incompetent* treatments and measures of perceived threat from the adversary and expected costs of a conflict: *Competent* increased both threat and expected costs

[207] There was no systematic difference in feeling thermometer scores between treatment groups: most people do not like states described as evil, longtime adversaries of the US, and they particularly hate North Korea.

Perhaps, but there are two crucial reasons to believe expected efficacy would inform the policy advocated by elites in positions of real decision-making power more heavily than it did the policy preferences of the sample population for this study. First, subjects in this study were presented at most two options when asked their preferred policy approach, and so advocated diplomatic overtures they expected to fail because the only alternative was war. Decision-makers are not so artificially constrained: they can look for alternatives and otherwise try to innovate policies they believe *will* be effective, in essence *creating* a menu of options that reflects their expectations. For example, if dialogue is not expected to be effective, there are numerous alternatives (sanctions, non-recognition, pressure brought to bear through a third-party ally of the adversary) that fall far short of war. Second, the subjects in this study lacked any real decision-making authority or responsibility and indeed were not even asked to imagine themselves in the role of a decision-maker: they therefore had more freedom to express a policy preference based on principle, morality, or criterion other than the likelihood of success. By contrast, assuming decision-makers at least attempt to be instrumentally rational in pursuit of national interests, probability of success should be their primary criterion for policy advocacy, especially in a world where devising alternatives or doing nothing are options. In sum, treatment impacts on expected efficacy of policies is probably at least as theoretically relevant as advocacy of policies, if not more so, for the purpose of using this study's individual-level results to explain or predict interstate behavior.

Do the treatments themselves pose a threat to validity, or otherwise weaken the link between the theory and the experiment? Vignette or article-style treatments may be artificially powerful enough to inflate findings, but despite these and similar concerns I believe my results remain informative. Concerns about overestimation are valid but should not be prohibitive: given the difficulty of causal inference in IR, scholars "should be more concerned about mapping the existence and sign of effects rather than their strength."[208] Moreover, there are three primary reasons to believe my estimates were actually *smaller* than their real-world counterparts. First, inherent components of MTurk surveys (low stakes, an inattentive or non-expert audience, abstract Likert measurements for which the difference between a '1' and '3' is inconsistent between subjects) may weaken treatment effects. Second, Survey 2 was included precisely to add realism and introduce barriers to treatment effects; strikingly, the results from this hard test, especially the existence and sign of effects, were very similar to those from Survey 1.

Finally, my theory concerns how individuals' diplomatic preferences are impacted by their *internalized beliefs*, not exposure to discrete information or opinions about an adversary. Beliefs are difficult to accurately measure let alone control; the use of opinion articles as an imperfect proxy for different beliefs about the adversary's judgment is a concession to inherent constraints of survey experiments. While it is plausible that these treatments overestimate the effect of reading an article, they would still underrepresent internalized beliefs, especially in the NK survey which created ample opportunity for subjects to bring their own beliefs to their responses. It is also likely that the treatments underestimate the impact that more detailed, targeted narratives about adversary leaders espoused by authoritative domestic elites over an extended period can have on the public's diplomatic preferences. In other words, these treatments should if anything *underestimate* the size and significance of the real-world corollaries for which they are proxy and with which the experiment is concerned; individuals who hold particular beliefs about the judgment of an adversary should be more rather than less likely to behave in accordance with the corresponding results. This logic of underestimation holds irrespective of *why* individuals hold particular beliefs about the judgment of a foreign leader: whether they internalized said beliefs through independent consideration of the foreign leader's actions and identity (as in Chapter 5), exposure to narratives that could be called propagandistic (for which the treatments from Survey 1 and 2 proxy even more directly, albeit still imperfectly, than they do internalized beliefs), or some other mechanism. Therefore, even artificially strong or salient treatments do not pose a threat to inference, as they bias size and significance of results *down* relative to the real-world issue of interest.

---

[208] Renshon, Jonathan, Allan Dafoe, and Paul Huth. "Leader influence and reputation formation in world politics." *American Journal of Political Science* 62, no. 2 (2018): 325-339.

Conclusion: Implications and Next Steps:

The implications for international politics of the findings presented in this chapter fall into three broad categories; implications for public behavior that can be directly drawn from the results, implications for elite behavior that can be plausibly inferred from the results but require further testing to verify, and implications for the interaction between elites and the public in the area of international diplomacy as influenced by beliefs about the rationality of foreign leaders' judgment. Below, I discuss each of these in turn, then outline my plan to explore the latter two in Chapter 4 through an in-depth case study of US decision-making during the Korean War.

Public Behavior:

The most direct conclusions we can make from these results concern members of the mass public, and the causal link between their beliefs about the judgment of foreign leaders and their diplomatic preferences. As discussed earlier in this chapter, there is both a compelling theoretical logic and substantial empirical evidence which suggest that the preferences of the mass public, particularly in democratic states, can play a major role in shaping elite foreign policy decision-making and thereby the behavior of the state as a whole.

Both Study 1 and Study 2 suggest that members of the US public are liable to significantly alter their diplomatic preferences based on what they are told about whether a foreign leader is objective or biased and, to a lesser extent, competent or incompetent. The effects of information about foreign leaders' rationality expressed themselves at all levels of public diplomatic preferences: from the degree to which individuals advocate negotiations, support concessions, or expect diplomacy to be effective relative to force. In brief, even minor amounts of information about the rationality of an adversary leader's judgment can significantly shift public attitudes towards war.

The results of survey 1 and 2 do not permit us to ascertain whether the source of these beliefs significantly effects the magnitude of their impact on individuals' diplomatic preferences. In other words, would the magnitude of the effects identified in Survey 1 and 2 change based on whether an individual arrived at their beliefs through independent consideration, exposure to a small amount of information, or repeated exposure to a particular narrative posited, for example, by government officials, the media, or both. While as discussed above it is likely the magnitude of the treatment effects found in survey underrepresent the real world corollaries for which the treatments proxy, this is not confirmable within the context of the experiments. We cannot say whether an individual whose belief that an adversary is irrational arose from a value-added interaction between their own personal observations and an elite narrative that supported said observations would become *more* anti-diplomacy than an individual who internalized the same belief (the adversary leader is biased) through one but not both of these pathways. This question presents an avenue for future research.

Surveys 1 and 2 also do not permit a direct analysis of how public beliefs about the judgment of adversary leadership interact with structural factors, as considerations of experimental control and realism necessitated controlling for variables like relative power in the design of the experiments. The adversary state described in these surveys represent a class of states (nuclear armed non-great powers) with the potential to inflict significant harm on the US in the event of military conflict: given the high costs of war we would expect diplomatic preferences to be very difficult to shift in crises with such adversaries. It is, however, true that an adversary that was relatively closer to power parity (such as Russia or China) might be a harder test, and not one these survey instruments can directly compare with a less powerful but still extremely dangerous adversary. An experiment that factorially manipulates descriptions of rationality and military capability should be conducted in the course of future work.

Even with these limitations, however, the results in this chapter significantly advance our understanding of public attitudes towards war. First, it suggests that elites could use narratives about adversary rationality as a tool to manage public attitudes towards war, whether to generate support for a desired conflict or increase support for planned negotiations and de-escalation of a crisis. However, as discussed earlier in the chapter and detailed below in the 'Interaction Between Elites and the Public' section, the possibility of competing narratives and the role of domestic opposition make it seem doubtful that democratic leaders can usually utilize this narrative tool to shift public preferences in whatever direction they want.

Elite Behavior:

What inferences, if any, can we make about the behavior of political elites on the basis of the results presented in this chapter? There are endemic issues of external validity when extrapolating conclusions from a study about broader populations than the specific subjects of the study in question. In particular, many political scientists express skepticism that experimental research conducted on randomly selected members of the public is informative with respect to the behavior of elites, due an array of systematic differences between elites and the mass public. Frequently cited examples of these systematic differences include: the knowledge and expertise elites acquire through their position, the selection effects involved in determining the type of individuals that become political elites, and the stakes elites versus members of the public have in the policy decisions about which experiments inquire. Effectively, skeptics contend that these systematic differences between elites and the mass public create an inescapable ‘levels of analysis’ problem for researchers seeking to make inferences about elite behavior from experiments conducted on members of the public.

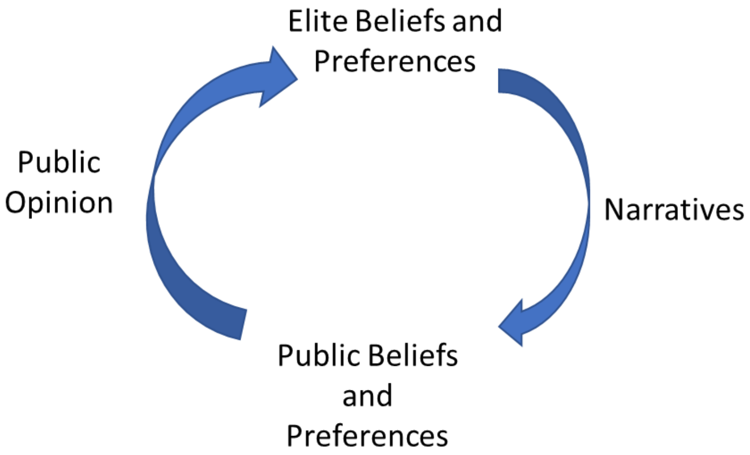
While the levels of analysis problem is one that must be taken seriously, substantial evidence undermines the argument that elites are universally distinct from members of the public: therefore, the scope in which this skepticism is warranted is an empirical question that depends to a large extent on the nature of both the question being asked and the mechanism of action behind any hypothesized behavior. While theories with mechanisms based in knowledge, experience, of incentives may need to make separate predictions about political decisionmakers and the people they represent, evidence suggest that more foundational biases, emotions or belief systems do not exert a systematically distinct impact on the behavior of elites versus the public. In particular, the directional impact of these mechanisms on behaviors of interest tends to be the same for elites and the public: for example, a shift in emotional state that makes one group more pro-war than its baseline is likely to move the other group in the same direction, even if the final preference for war differs markedly between the two groups. In other words, the severity of the levels of analysis problem for a given theory depends on both its mechanism of action and whether it is concerned with the direction of behavioral shifts or final behavioral outcomes.

The hypotheses presented and tested in this chapter are concerned with the directional impact of variation in beliefs about the judgment of an adversary on the diplomatic preferences of individuals. Systematic differences in information, expertise, and incentives may and in all likelihood do generate divergence in the *baseline* diplomatic preferences of elites and the public. Yet elaborated in Chapter 1, there is no theoretical reason to expect beliefs about the objectivity or competence of adversary leaders to shift the diplomatic preferences of elites and the mass public in different directions relative to their baselines. It is plausible that elites and the public who share beliefs about the judgment of an adversary leader will shift their preferences for diplomacy versus force in the same direction relative to their respective baselines, even if they end up at different absolute preferences for diplomacy versus force, In other words, we could infer from these results that elites who believe adversary leaders are objective will be more likely to advocate diplomacy (**H1)** than elites who believe adversary leaders are biased **(H2)**, and elites who believe adversary leaders are competent will be more likely to seek talks and make concessions (**H3)** than elites who believe adversary leaders are incompetent (**H4**). This is not a claim of strong causal proof: rather the experimental evidence presented in this chapter serves to refine and probe the plausibility of my theory at the elite level. More targeted empirical examination is needed for a robust causal test of the theory's validity with respect to elites: this is the purpose of the detailed Korean War case study presented in Chapter 4.

The Interaction Between Elites and the Public:

Finally, these results have implication for the interaction between elites and the mass public in relation to beliefs about the judgment of adversary leaders and attitudes towards diplomacy versus force. In brief, they directly support the first part (the elite-to-public side) of the cyclical relationship between elites and the public described in the theory chapter and re-illustrated below.

Figure 3.4: The Elite/Public Cycle:



The substantial differences between the expressed diplomatic preferences of treatment groups, particularly *objective* versus *biased* but also *competent* versus *incompetent*, suggest that information about the judgment of adversary leaders can have a drastic impact on public attitudes towards diplomacy and war. This holds even in extremely salient, high stakes crises charged with partisanship, all of which we would expect to mitigate the impact of information about an adversary leader's judgment on individuals' diplomatic preferences. Considering that the treatment articles provided to subjects were brief and credited to unfamiliar sources, it is probable that the effect of information about adversary judgment on individuals' diplomatic preferences would be even greater when that information is: highly detailed, provided by known or trusted sources, conveyed over a long period of time, made part of the national conversation, or all of the above. Therefore, the results in this chapter directly imply that propagandistic narratives about the judgment of foreign leaders put forth by political, media, or policy elites can shape the diplomatic preferences of the mass public.

The validity of this conclusion is, importantly, not conditioned on the motivation or identity of propagandizing elites. Narratives propagated by elites about the judgment of a foreign leader should exert the same impact on public policy preferences regardless of whether they are a reflection of elites' genuine beliefs, a strategic choice, or a simplifying hyperbole used out of what could charitably be called laziness. In other words, elites do not need to believe their propaganda for it to shape the attitudes of the mass public. Moreover, the propagandizing elites do not need to have a particular institutional or partisan affiliation for the narratives they posit to influence mass public attitudes towards diplomacy and the use of military force. The elites mentioned in the treatment conditions were not given a partisan affiliation and were tied, albeit loosely, to the academic, media, and foreign policy communities. Of course, further research would be needed to conclusively ascertain the impact of either allied or adversarial relationships between the partisan affiliations of propagandizing elites and members of the public. Yet while existing research provides some support for the intuitive hypothesis that individuals are more likely to incorporate the claims of partisan allies into their beliefs, this relationship is weaker than one might expect, especially when the claims lack an explicit policy position.[209]

Put simply, when it comes to the judgment of adversary leaders, the words of elites are not just sound and fury: they matter, especially in light of a public that on average lacks independent information on or interest in foreign affairs. It is probable, for example, that the longstanding narrative which frames the Iranian Ayatollahs as fanatical madmen has reduced US public support for diplomacy with Iran. Conversely, the move from describing Soviet leaders as irrational ideologues to characterizing Gorbachev as “a man we can do business with” helped increase public support for more robust diplomatic negotiations with the USSR.

These shifts in public opinion should be expected to influence the foreign policy choices of political decisionmakers concerned with maintaining popular support, whether said shifts are the product of strategic propaganda by those same elites or the result of organic, uncontrollable narratives. Thus, political elites may find their policy options constrained by the stories they or their domestic opponents tell the public about the rationality of foreign adversaries. This suggests a mechanism, public opinion, through which elite claims about the judgment of foreign adversaries may shape foreign policy even if said beliefs are not internalized by decisionmakers. It is, however, important to recognize that this is an inference drawn from existing research rather than a causal claim derived from the experimental results presented in this chapter.

Next Steps:

---

[209] In other words, elite narratives are more likely to be accepted across partisan affiliations if they make claims about the judgment of a foreign adversary without advocating a specific policy position

The results presented in this chapter provide strong support for the contention that the public's beliefs about the judgment of foreign leaders impact their diplomatic preferences, regardless of the ultimate source of these beliefs.[210] Yet they do not directly suggest that these shifts in public diplomatic preferences impact the foreign policy decisions of political elites: the latter relationship is a plausible inference drawn from related empirical research.

More targeted investigation is therefore needed to assess the validity of the inference that public opinion shapes the diplomatic preferences and choices of elites, and in turn to confirm or reject what I theorize as a cycle of influence between elite narratives about the rationality of adversaries, public opinion, and elite choices in foreign policy. It is for this purpose that I now proceed to an in-depth case study of the preferences and decisionmaking of US political elites during the Korean War. This enables me to directly examine multiple questions of interest, two of which stand out. First, how do elites' internalized beliefs about the judgment of foreign adversaries shape their diplomatic preferences in the context of an ongoing conflict? Second, to what extent are the diplomatic preferences and choices of elites shaped by public opinion, and by extension what role do elites themselves play in shaping public attitudes through propagandistic narratives about the judgment of adversaries?

---

[210] As elaborated in Chapter 5, we should expect public beliefs about the judgment of adversaries to form based on some mix of the personality traits of individual observers, the behavior and identity of adversary leadership, and 'propagandistic' narratives posited by elites.

Chapter 4: The Korean War:

Do the diplomatic bargaining preferences of foreign policy elites move in line with the predictions of my theory, and are these expressed in state behavior within the confines of the theory's scope conditions? I have argued that the theory crosses levels of analysis and therefore applies to both elites and the public, and that there are two primary through which the results of the experiment presented in Chapter 3 can be extrapolated to make inferences about the beliefs and behavior of elites. First, elites are people: there is no reason to expect that if they have internalized a set of beliefs about the judgment of an adversary leader, their expectations about the efficacy of information or the threat of conflict will move in different directions than what we observe in members of the public. Second, particularly in democracies, elites are subject to political pressure from the public, whose attitudes towards diplomacy and force are (as shown in Chapter 3) heavily influenced by beliefs about the objectivity and competence of adversary leaders.

In this chapter, I subject this primary argument (that my theory applies to elites) and the two proposed mechanisms of effect to testing through a case study of US decision-makers in the Korean War. There are several reasons why this case is an ideal one in which to examine the theory's veracity at the elite level. First, US Korean war policy affords an opportunity to examine the impact variation in beliefs about adversary judgment has on diplomatic preferences in three adversarial dyads – US/USSR, US/China, and US/North Korea – within the context of the same armed conflict. This avoids selecting for a particular value of the independent variable, and poses a fuller test of the theory than case studies focused on prototypical 'mad leaders' such as Hitler. Second, it allows within-case comparison over time of variation in US beliefs about the judgment of, and diplomatic preferences towards, three distinct adversaries in the same geostrategic context. This permits a degree of control by, at any given time, holding constant numerous plausible confounding variables; these include US strategic interests, the balance of power, the strategic as well as tactical situation, the identity of US decision-makers, and the domestic political situation. In other words, key structural and dispositional factors are constant across adversary dyads.[211] Finally, there is an abundance of primary source material, which is crucial for examining perception or beliefs outside an experimental context; I examine statements and correspondence of US military and civilian policy-makers to ascertain their beliefs about the judgment of adversaries, though I do not interrogate the origins of these beliefs.[212]

In particular, I examine the diplomatic bargaining preferences and choices of US decision-makers in the lead up to and during the ceasefire negotiations that began at Kaesong and eventually moved to Panmunjon. I focus on explaining the divergence in US diplomatic bargaining preferences towards the USSR versus Communist China. I find that a deeply ingrained set of beliefs about the leadership of these adversaries, which held that the Soviets were a "reasonable adversary" and China the "irrational foe,"[213] drove a substantial divergence in US conduct of diplomacy that is not convincingly accounted for by alternative structural or dispositional arguments. US officials were willing to engage in iterative diplomatic dialogue with the Soviets in pursuit of a ceasefire, evinced readiness to make some concessions to Soviet interests, and were measured in their use of coercive threats. However, these same officials thought diplomatic dialogue with the Chinese would be fruitless and concessions to their leadership unnecessary: it took immense pressure from a range of UN coalition members to bring the US to the negotiating table, and once there US officials mainly issued threats and ultimatums to Chinese representatives. While I also examine the US attitude towards North Korea, the sheer neglect with which the US regarded North Korea means this portion of the case is less extensive.

---

[211] Because the theory is fundamentally relational, it is important to hold constant as many purely internal or external variables as possible, though of course experimental levels of control are impossible to achieve in archival work.

[212] I examine causes of variation in beliefs about judgment, such as adversary ideology, race, status, rhetoric, and past or current behavior, in a related paper.

[213] Rosemary Foot. *The Wrong War: American Policy and the Dimensions of the Korean Conflict, 1950–1953*. Cornell University Press, 2019. P. 27

The rest of this chapter proceeds as follows. First, I provide a brief background on the Korean War, the overall course of the conflict, and the lead-up to the Kaesong and Panmunjon talks. Second, I describe US interests and goals in Korea at the time in which I am interested, outline some of the thinking behind these goals, and use them to frame what I consider the four most puzzling divergences in US diplomatic bargaining preferences and conduct between China and the USSR. These are: 1) Why did the US initially oppose formal and informal negotiations with the Chinese but seek them with the Soviets? 2) Why did the US engage the Soviets in dialogue that occasionally gave way to coercion while skipping directly to coercion and ultimatums with the Chinese? 3) Why did the US entertain concessions to the Soviets but not the Chinese? 4) Why did the US contemplate direct action against the Chinese homeland but recoil at the prospect of any action that might further escalate armed conflict with the USSR? Third, I discuss prominent theories which fail to predict these differences, consider theories that provide alternative explanations for the divergences, and explain why I find these alternative hypotheses incomplete or unsatisfactory. Fourth, I detail US beliefs about the judgment of Soviet leadership and explain how these shaped the diplomatic bargaining preferences and choices of US elites engaging with Soviet counterparts. Fifth, I detail US beliefs about the judgment of Chinese leadership and explain how these shaped the diplomatic bargaining preferences and choices of US elites engaging with Chinese counterparts. Sixth, I probe the plausible use of my theory to explain why the US largely ignored North Korea in thinking about and conducting diplomacy. I conclude with a discussion of the additional light my theory sheds on the Korean War, plausible extensions, and implications for future work.

The Korean War: A Brief Timeline of The Forgotten Quagmire:

Before the word ‘quagmire’ become synonymous with Iraq, Afghanistan, or Vietnam before them, there was the “forgotten war” in Korea: a bloody conflict dragged on for two years by stalemated armistice talks, and a war that has still not formally ended to this day.

At the beginning of the Cold War, the US and USSR engaged in an ultimately doomed attempt to reach an agreement on the status of the Korean Peninsula. The principal aim of US policy in the Far East was to prevent the communization of the region, and US officials in the sought to establish a Korea that was 1) democratic and 2) unified, a goal that was ostensibly to be accomplished by allowing the Korean people to determine their own form of government in free and fair elections held under UN observation. From 1946-47 the US pursued this goal through negotiations, in the forum of the US/USSR Joint Commission on Korea, but these efforts ended in failure and the establishment of the 38th parallel dividing the communist North from the nominally democratic South. The governments of North and South Korea, however, maintained the goal of unifying the Peninsula under their respective rule, and tensions were high.

These tensions exploded into conflict with a surprise North Korean invasion across the 38th parallel in late June 1950. The North Koreans quickly overwhelmed the unprepared and outmatched South Korean defense forces, pushing them to and beyond Seoul. The tide was turned only through aggressive military intervention by UN coalition forces, comprised mainly of US soldiers and under the command of the now infamous General MacArthur.[214] Weeks of bitter fighting, and a daring amphibious landing of US troops behind enemy lines at Inchon, reversed the South's territorial losses and culminated in a triumphant MacArthur driving past the 38th parallel with the intent to push the enemy to and beyond the Yalu river at the northern border of the Peninsula. However, the Chinese Communist Party made good on its threat to intervene in force if it looked like US forces were set to cross the river, and in late October the intervention of the Chinese army again reversed the course of the war. By mid-1951, the belligerents were effectively back where they had started, facing off against some version of the 38th parallel with little major territorial shifts in the direction of either side: truce talks began at Kaesong on July 10, 1951. Yet failed ceasefire negotiations dragged the conflict out for two more years before an armistice was finally reached on July 27, 1953.

---

[214] [214] Schnabel, James F. *United States Army in the Korean War: Policy and Direction, the First Year*. Vol. 4. United States Center for Military History. Government Printing Office, 1961.

In the end, some three million Korean lives (half or more of them civilian) were lost, along with roughly thirty-six thousand Americans and at least six hundred thousand Chinese soldiers.[215] Stalemated negotiations bear a substantial portion of the responsibility for this massive death-toll: most of the casualties occurred during 'meat-grinder' style fighting along the 38th parallel that did little to advance the interests of either side. Explaining why these negotiations came to such an intractable halt is a task that should interest scholars of political science and history alike.

US Interests and Capabilities in Korea:

The US had been "prepared to stay in Korea… until an agreement could be reached with the Russians,"[216] Yet after negotiations failed, there was no appetite for attempting more aggressive measures: “Unlike the Soviet Union the United States attached little importance to Korea as a strategic area.”[217] Few in US policy making circles Korean unification was worth spilling substantial blood or treasure for, and military planners were opposed to treating the peninsula as an area of military importance.[218]

---

[215] Bruce Cumings, *The Korean War*.

[216] Schnabel, James F. *United States Army in the Korean War: Policy and Direction, the First Year*. Vol. 4. United States Center for Military History. Government Printing Office, 1961. Pg 28.

Korea was considered of secondary importance even in NE Asia, itself of secondary import to Europe: the country’s strategic significance paled in comparison to that of its neighbors, and President Syngman Rhee’s corruption and apparent incompetence made him a problematic ally. Moreover, a war for Korea was one the US was ill-equipped to fight, at least without undermining the defense of Europe. Initially the principle concern of US officials was direct engagement with Soviet forces, as geographical constraints alone ensured the regional balance would favor the USSR. However, conflict with both North Korea and mainland China also became sources of significant concern once these areas had fallen firmly under communist party control, in 1948 and 49 respectively. North Korean forces significantly outmatched their Southern counterparts and would require substantial US investment to defeat: MacArthur reported that South Korea had no hope of maintaining its status as an independent nation under the protection of its own security forces in the face of the Northern threat.[219] Meanwhile China’s numbers meant it could overwhelm US/UN forces in the area or at least tie them down indefinitely in the event of a land war. US officials believed that they could fight and win a war with either North Korea or China absent direct USSR intervention, but this would have necessitated diverting forces from Europe.[220]

---

[217] Schnabel pg. 7
[218] Schnabel pg. 28
[219] Schnabel pg. 30
[220] Foot 2019, Schnabel pg. 27

In sum, from the very beginning a plurality of US officials wanted a democratic and unified Korea, but felt this was not worth fighting for, and that the national interest along with the constraints of relative capability bargaining preferable to fighting. The JCS felt "that Korea [was] of little strategic value to the United States and that any commitment of to… military force in Korea would be ill-advised."[221] However the definitive military and political 'write-off' of Korea came when the JCS asserted that no security guarantee should be made to Korea, and in April 4, 1948 Truman approved a policy that stated "The United States should not become so irrevocably involved in the Korea situation such that… [any] action… in Korea could be considered a 'causus belli' for the United States."[222] Indeed, in an infamous speech Acheson announced this to the world by leaving Korea out of what he described as the zone of interests vital to the US.[223]

When North Korean forces surprised the US by launching an invasion across the 38th parallel, and Chinese forces later surprised the US by joining the fight, the preference of US policy makers was to end the conflict before it could further escalate, rather than devote the necessary resources to attain a military victory. The primary change from the pre-invasion stance, in which Korea was written off completely, was that the US should not concede the Peninsula completely to the communist aggressors: in other words, they wanted a ceasefire and return to the status quo of the 38th parallel.[224] Moreover, US officials were certain that avoiding a general war was in the mutual best interest of all involved parties, a convergence of interests that was absent from the goal of unifying Korea under democratic governance.

---

[221] Schnabel pg. 50
[222] Schnabel pg. 50
[223] Schnabel pg. 51
[224] Schnabel pg. 73

This shift reflected a changed calculus of Korea's symbolic rather than strategic importance. The Truman administration was operating according to the CIA's conclusion that "The invasion of the Republic of Korea by the North Korean Army was undoubtedly undertaken at Soviet direction and Soviet material support is unquestionably being provided,"[225] and that generally speaking the actions of both China and North Korea were controlled by the Kremlin and therefore extensions of its global strategy. Under that assumption, famed sovietologist George Kennan told the President that "the symbolic significance of South Korea's preservation is tremendous:"[226] Truman and his advisors agreed. Even if the Communist invasion did not signal Moscow's readiness to engage in all-out-war with the US and its western allies, it represented a challenge to the will of the non-communist world that had to be met. Domination of the Korean Peninsula was unnecessary to meet this challenge, but the status quo of demarcation along the 38th parallel had to be restored to show the Communist world that the US would resist such aggression.

MacArthur's stunning success briefly gave US officials grander ideas of establishing full control over Korea, dealing a devastating blow to the Communist image and cause. However, China's entrance in force quickly disabused most US decision-makers of this notion. The US was militarily overstretched and outmatched on the ground, and while initially there was some talk of US air superiority making up the difference this hope was curtailed when intelligence reports identified the formation of a Soviet "volunteer" air force that was enhancing the air power of Communist forces.[227] By spring of 1951, "national intelligence estimates indicated that the Communists had the ability to '(a) expel the US from Korea, (b) ensure successful defenses of the Chinese mainland against any force.'"[228] Faced with the prospect of being at best tied down and at worst beaten by an enemy whose numbers were far superior, a status quo ceasefire once again became the primary goal of US policy.

---

[225] Cumings p. 14, Foot 2019
[226] Hastings p. 75
[227] Foot p. 136

Yet US officials manifested drastically different diplomatic bargaining preferences, and took very different actions, in how they advanced this overarching goal when dealing with the Chinese versus the Soviets. Below I describe the difference in US diplomatic bargaining preferences towards China versus the USSR, and why in light of the expectations of prominent extant theories this divergence constitutes a puzzle in need of explanation.

Same War, Divergent Diplomacy: The Principle Puzzles in the US Approach:

On June 23 1951, the Soviet representative at the UN, Malik, proposed the arrangement of truce talks between the belligerents in the Korean War; US officials welcomed the idea of negotiations with the Soviets but were much more dubious about the prospect of including the Chinese. Prior to Malik's proposal, US officials had treated the prospect of formal negotiations with the Chinese with skepticism at best or outright opposition at worst: this did not change after Malik's proposal. Bringing the US to the negotiating table with the Chinese Communists took the combined forces of the prospect of an endless war on the peninsula and the political pressure exerted by a majority of the states in the UN coalition, from the UK to Norway and Canada, which advocated diplomatic engagement with China as a means to de-escalation. While the US agreed to truce talks beginning at Kaesong on July 10 1951, the spirit with which the US delegation approached these negotiations was hardly one of diplomacy between reasonable equals: they sought to threaten and bludgeon China into agreeing to a settlement in line with US interests, and paid little regard to Chinese goals or perspectives. Put in terms of this project's dependent variable, US officials' *diplomatic bargaining preferences* were de-escalatory and mostly argumentative towards the USSR but coercive and mostly escalatory toward China. Considering that the Korean War was a single conflict in which the US had one overarching set of primary goals and, however incorrectly, saw the Communist adversaries as more monolithic and less distinct, this divergence begs explanation.

---

[228] Foot p. 136

Specifically, there are four puzzles in US diplomatic bargaining preferences towards the Soviets versus the Chinese during the armistice negotiations. These are: 1) Why did the US initially oppose formal and informal negotiations with the Chinese but seek them with the Soviets? 2) Why did the US engage the Soviets in dialogue that occasionally gave way to coercion while skipping directly to coercion and ultimatums with the Chinese? 3) Why did the US entertain concessions to the Soviets but not the Chinese? 4) Why did the US contemplate direct action against the Chinese homeland but recoil at the prospect of any action that might further escalate armed conflict with the USSR?

1) *Coercive vs. Argumentative (communication infrastructure, pre-formal negotiations):* A majority of US officials were reluctant about, or outright opposed to, the idea of opening negotiations with the Chinese, and there were few if any channels for bilateral communications between the US and the CCP during the course of the fighting. Whatever communication or bargaining took place in the background of conflict almost always went through third parties, usually representatives of the Indian government. By contrast a majority of US officials stressed the need for negotiations with the USSR during internal deliberations, and there were numerous

channels for bilateral communications between US officials and their Soviet counterparts.

2) *Coercive vs. Argumentative (during diplomacy)*: The US engaged in little to no open dialogue with the Chinese (at one low point even refusing to speak to their delegation for 2 hours and 11 minutes): most officials saw little persuasive value in attempting to talk to the Chinese or understand their interests. Internal debates on how to persuade the Chinese to accept a settlement more in line with US interests centered on identifying the most effective material, military, and economic mechanisms through which to pressure the Chinese. During negotiations, the US primarily sought to influence Chinese behavior by threatening or enacting economic sanctions, the destruction of civilian and military infrastructure, and even the use of nuclear weapons. Many officials wanted to abandon even this aggressive coercive diplomacy and move to simple ultimatums. This was in sharp contrast to the spirit of US diplomacy with the USSR, which comprised a mixture of rational dialogue and coercive trading of threats and rewards. Both in formal settings and, more often, through established backchannels, US officials negotiating with their Soviet counterparts did so through what often looked like an open dialogue between adversarial but rational and, at least superficially, mutually respectful counterparts. While most US foreign policy officials thought it was desirable to negotiate with the Kremlin from a position of strength, US officials began the majority of these negotiations with an argument-based approach. They would however shift to coercive diplomacy if persuasion through rational dialogue failed, and in these cases having started negotiations from a position of strength was particularly important.

3) *Escalatory vs. De-escalatory (concessions):* US officials gave little serious consideration to the use of carrots or major concessions to the Chinese, but both of these negotiating tools featured prominently in deliberations about and manifested policy towards the USSR.
4) *Escalatory vs. De-escalatory (conflict intensification):* With few exceptions, US officials expressed willingness to escalate the military conflict with China, albeit to varying levels of intensity: some drew the line at bombing power plants along the Yalu and in Manchuria, while the most jingoistic wing wanted a full-scale ground invasion into the Chinese mainland. Indeed, the extension of hostilities into the Chinese mainland was a major locus of US strategic deliberations during the war throughout the period of negotiations. Yet even the most bellicose officials were extremely hesitant to take any action they thought might produce escalation with the Soviets, a worst-case scenario they referred to as "a third world war." Indeed, one of the most important factors in discussions of the wisdom of escalating to the Chinese mainland was whether this would draw the USSR further into war.

Alternative Theories and Arguments:

What are the most relevant theoretical expectations that render the above divergences in US diplomatic bargaining preferences puzzling, and why are the most plausible explanations for this puzzle outside of my own theory unsatisfactory? Many theories that inform our expectations in this area are ill suited to explaining or predicting differences in an actor's diplomatic preferences between adversaries but within the same conflict. This is because they center on explanatory variables that do not obviously vary between adversaries, such as power distribution, outcomes of fighting during the course of conflict, interests, and internal characteristics of the actor whose preferences and actions they are seeking to explain. Furthermore, extant arguments that could account for differences in diplomatic bargaining approaches at the level of adversarial actors generate predictions or explanations that are mismatched with the empirics of the Korean War case.

The Expectations of Structurally Determined Diplomacy and Traditional Bargaining:

The traditional conception of diplomacy as epiphenomenal to power, structure, and interests holds that diplomatic decisions merely reflect rather than influence or determine international political outcomes, and that these outcomes are traceable to material variables such as hard military power.[229] In this theoretical perspective, the strong do what they will and the weak suffer as they must: stronger actors will conduct 'diplomacy' through threats and ultimatums because they can do so credibly, and moreover can afford to make good on those threats if they do not get what they want. Weaker actors will pursue diplomatic discussions and make concessions because they have no hope of coercive or military success. The empirical issue with this perspective in the Korean War case is immediately apparent: the US was not materially positioned to bludgeon China into a ceasefire: indeed US officials' awareness that a protracted conflict with China would be unacceptably costly was a principle reason they reverted to the limited war aim of a status-quo ceasefire shortly after China's forces crossed the northern border. If relative power determines diplomatic choices, then at least in the theater of the Korean War it is unclear why the US would be so bearish on the prospect of even talking to the Chinese, let alone engaging them in a formal negotiation conducted through rational argument.

---

[229] Rathbun 2014

The traditional bargaining approach likewise does not adequately explain US diplomatic bargaining preferences prior to and during the start of talks at Kaesong. It is obvious why the assumption of traditional bargaining models that talks between belligerents are a constant background feature of conflict is unconvincing in this case, but moving beyond that strawman we can use the logic of traditional bargaining models to derive some general hypotheses, which also prove insufficient for the case at hand.

The basic expectation for US diplomatic bargaining preferences derived from the bargaining approach to war is that the thumping UN coalition forces suffered when China entered the war should have moved US diplomatic bargaining preferences significantly towards de-escalation and at least somewhat towards argument.[230] Because combat setbacks caused US officials to downgrade their estimates of eventual victory, they should have been willing to enter negotiations with and entertain concessions to the Chinese to facilitate a ceasefire.[231] In short, by the time US officials acknowledged they could not conquer the Peninsula at acceptable cost and risk, there was little reason not to seek a negotiated settlement through the least inherently costly methods possible (namely, argumentative dialogue). Needless to say, US officials did not abide by these predictions: even after major military setbacks they were resistant to approaching the bargaining table with the Chinese and once pressured there by UN allies maintained a thoroughly coercive approach to negotiation.

A related hypothesis from the bargaining approach is that the fighting between China and the US should have revealed enough information to bring their expectations about the likely outcome of continued conflict towards a point of convergence.[232] When both sides realized they probably could not achieve an absolute victory (and archival evidence from the Chinese side suggests that Mao at the time shared that conclusion) they should have come to shared expectations about the range of agreements acceptable to both sides and been willing to engage in the bargaining processes necessary to arrive at one of these bargaining equilibriums. Yet while the sense of a shared interest in a ceasefire among all belligerents was expressed by US officials in both intrastate and interstate deliberations, these same officials remained opposed to negotiations and dialogue with the Chinese. Put another way, the apparent existence of shared interests revealed by ongoing combat was evidently insufficient to convince the US that dialogue conducted through complex information was a feasible means through which to reach a negotiated settlement in which the shared interest of both sides was achieved. By contrast, US officials were open to negotiated dialogue in pursuit of shared interests with the Soviets, whose forces had not met their own in combat or revealed any of the information conveyed by fighting conceived as part of the bargaining process of war.

---

[230] Skylar Mastro 2019
[231] Reiter 2003
[232] Fearon 1995, Reiter 2003

Adverse Signaling and the Costly Communication Thesis:

Skylar Mastro's costly communications thesis starts from the basic and intuitive contention that proposing diplomatic negotiations can be costly if adversarial actors interpret this as a sign of weakness in line with the canonical deterrence model of conflict.[233] The logic of the costly communications thesis is that whether actors will be closed to diplomatic negotiations depends on 1) whether they think it is likely that an open diplomatic posture will result in *adverse inference*, whereby the adversary concludes that openness to negotiations signals weakness of power or resolve and 2) whether the adversary has the *strategic capacity* to escalate conflict to an unacceptably costly level if it makes an adverse inference of weakness. The contours of this dependent variable are not perfectly aligned with *diplomatic bargaining preferences*, but we can infer what dynamics we would need to observe for the costly communications thesis to explain the divergences in US diplomatic bargaining preferences. In sum, US officials would have had to think that either China's likelihood of *adverse inference*, *strategic capacity,* or both were 'high,' while simultaneously thinking that one or both of these dimensions were 'low' with respect to the Soviets.

---

[233] Skylar Mastro 2019

There are several reasons this characterization of US officials' thinking, and its use as an explanation of their diplomatic bargaining preferences, is unconvincing. First, US officials worried extensively about the possibility of adverse inference with Moscow, far more so than they did with the CCP. Indeed, the obsession with maintaining a reputation for resolve[234] and the fear that Stalin had green-lighted both the North Korean and the Chinese invasions because he had inferred weakness and irresolution from US diplomatic posture that was a principle factor motivating a more robust defense of the Korean Peninsula.[235] Yet despite these concerns US officials apparently believed it was worthwhile to engage the Soviets in ceasefire negotiations.

Second, if the demonstration of resilience through fighting is the primary method for reducing the likelihood of adverse inference, then why would the fighting in Korea have reduced this more for the Soviets than the Chinese? Both Peiping and Moscow bore witness to the US fighting in Korea, China had the metaphorical front-row seat, and the operating perspective of the vast majority of US officials was that there was little meaningful separation between the two. If US officials believed the decisions and beliefs of Mao's government were effectively dictated by the Kremlin, why would they also draw substantially different conclusions about the likelihood of adverse inference by Peiping versus Moscow? Third, the USSR's latent power and capacity for military escalation in the event of adverse inference exceeded China's: in fact while US officials were eager to avoid the costs of uncontrolled military escalation with China, they were largely optimistic about their prospects if this came to pass, that is unless and until the USSR entered the conflict in force.

---

[234] While these concerns were very real and drove major aspects of US cold war policy, the veracity of these concerns have since been called into question. This is true both of the narrow question of whether Stalin and Mao thought the US was unresolved because of its Korea policy (see Mercer 2014), and the more general question of whether reputation for resolve matters in international politics to the extent decisionmakers at the time thought. an extensive scholarly interrogation of the importance of face so conceived including the work of Mercer (1996), Kertzer, Yahri-Milo, and many others, culminating in a denunciation of the "cult of reputation," though like much good social science this has not been the final word and work has since been done re-examining the importance of reputation.

[235] Mercer 2014

Personality, Individual Beliefs, and Internal Characteristics:

Arguments about how the personalities and belief systems of individual US decision-makers shaped their diplomatic bargaining preferences help to account for the existence of disagreement and debate within US policy-making circles. However they do not provide a deterministic explanation for US diplomatic bargaining preferences, nor can they account for the divergence in the approaches identical officials advocated taking towards China versus the USSR. Indeed, most top US officials "knew very little about China,"[236] lacked policy experience with or priors about dealing with Chinese leadership, and more or less exported their operational code for dealing with the Soviets to dealing with the Chinese in order to fill in the gaps.

The highest ranks of US government held a worldview, shaped in no small part by the experience of World War II, that it was necessary to negotiate from a position of strength, that to back down in the face of aggression signaled weakness and brought dishonor, and that adversaries should always be treated with skepticism if not outright distrust. This included President Truman, Secretary of State Acheson, and Assistant Secretary of State for Far Eastern Affairs Dean Rusk. Truman believed that Moscow in particular "respected nothing but force" and "only in the face of force would they talk and negotiate," a conclusion informed by his perspective that Soviet leaders were a "wily adversary" prone to risk taking.[237] Truman extended this distrust and determination that negotiation could occur only with at least a background threat of force to the Chinese leadership. Acheson largely shared and therefore reinforced Truman's beliefs. Area experts in the influential Policy Planning Staff (PPS) occasionally expressed slightly more nuanced perspectives with respect to their adversary of interest; George Kennan was a vocal advocate of recognizing Moscow's concern for its own security and its basically cautious approach to foreign policy, but by 1950 the PPS was staffed mostly by officials whose foreign policy operational code resembled that of Truman and Acheson.

---

[236] Foot p. 33

Yet despite their misgivings Truman and Acheson came to seek and accept the idea of negotiations with the Soviets while still dismissing diplomacy with the Chinese as a fruitless endeavor, even if international pressure eventually pushed them to reluctantly approach the bargaining table. Why US civilian officials' essentially ubiquitous distrust of communism and subscription to a deterrent model of conflict or crisis diplomacy would have produced differences in diplomatic bargaining preferences between the communists in Russia and those in China is, in brief, unclear.

The same is true of most arguments that, in effect, apply individual psychology and personality theory at the level of aggregated state actors. Theories like Rathbun's, which use state-level factors such as the partisan orientation of government to proxy for personality traits that predict variation in diplomatic bargaining preferences, do not explain why the same government with the same partisan makeup manifests different diplomatic bargaining preferences towards different adversaries in the structurally identical context constituted by a single shared military conflict. This is not to say that these arguments are wrong, or that they are not useful for understanding state officials baseline diplomatic styles, but they only tell us so much. In line with Rathbun's theory, the Truman administration was less openly jingoistic and exclusively in favor of coercion, ultimatums, and force than the Republicans who were in the minority during his presidency as well as the Republican president, Eisenhower, who succeeded him. However, this does not explain the different diplomatic styles advocated toward China and Russia within the Truman administration. At a more general level, the partisan and ideational composition of US government during the Korean War seems to have influenced military and diplomatic policy: Eisenhower's militaristic republican administration was considerably more willing to exert massive military pressure than was Truman's.

---

[237] Foot p. 32-3

The Explanatory Utility of Beliefs About Judgment:

I argue that my theory sheds invaluable light on the divergent diplomatic bargaining preferences US officials expressed toward China and the Soviet Union prior to and during the 1951 truce talks, and that it also accounts for some of the internal disagreements between US officials. Officials who fell more on the side of what can broadly be termed 'hardline' diplomatic policy toward either China or the Soviets also ascribed greater degrees of bias and/or incompetence to the adversary in question, while the reverse was true for officials on the other side of the argument.

In brief, most US officials' view of Soviet leadership closely approximated *the Rational* type, though not an ideal-type *Rational* as few saw the Soviets as completely unbiased:

hardliners saw the Soviets in a way that better fits the *Fanatic* type. By contrast, most US officials saw China's leadership as the *Madman*, but did not think Mao was so incompetent as to negate the dangers posed by China's vast material power. Some hardliners, many of whom were in military leadership, viewed Mao's government as a *Madman* type so thoroughly incompetent that its military resources posed a negligible threat. While I do not focus on North Korea, the near-universal view of its leaders among US elites was one that fit the *Madman* type, a fact that sheds some light on why throughout the war US officials barely considered the prospect of negotiations with North Korea extensively enough to ignore the prospect.

Figure 4.1, below, summarizes these beliefs about judgment, the diplomatic bargaining preferences expressed by officials who held them, and the associated choices we observed in state behavior. In the section that follows I separately examine each adversary in detail to trace the link between beliefs and preferences, and discuss the additional factors outside the scope of my theory that impacted whether and how these preferences manifested in state behavior.

Figure 4.1: Summary of Beliefs, Preferences, and Behaviors:

Adversary	**Beliefs about Judgment**	**Diplomatic Bargaining Preferences**	**Overall Policy**
USSR	-*Rational* (majority) -*Fanatic* (hardliners)	-*De-Escalatory Dialogue* (majority) -*Coercive Diplomacy* (hardliners)	Ongoing Negotiations, Backchannel Dialogue
China	-*Madman*	-*Coercive Diplomacy* (majority) -*Ultimatum diplomacy and war* (hardliners)	Non-concessionary coercive diplomacy
North Korea	-*Madman*	*No Negotiation*	Neglect

Communist China: The Irrational Foe:

American diplomatic preferences towards China were impacted by two primary beliefs about the judgment of Communist Chinese leadership. First, the Communist Chinese viewed the world through an ideological prism that warped their conception of reality, rendering them almost hopelessly biased. Second, the Communist Chinese were incompetent, and thus nowhere near as threatening as their raw capabilities would indicate. These beliefs became calcified when the Chinese, defying American predictions of what a rational actor would do, intervened in October 1950, and heavily influenced how US policy for getting China to end the conflict. US officials were extremely skeptical of direct negotiations with the Chinese; they thought that while China *might* respond to force and other forms of simple information, argument and complex information would be ineffective, and seeing little need to make concessions to China saw little reason to open negotiations before the Chinese had suffered sufficient damage that they would be forced to accept US terms. While unlike with North Korea the US did eventually enter into negotiations with China, a plurality of US representatives saw these negotiations as a forum to make threats and dictate terms rather than to hold a dialogue.

Perceived Bias and Persuasive Preferences:

The pervasive view among US officials was that China's leaders were extremely biased because of their fervent devotion to communism and tendency to perceive reality through an emotional lens. "The tragic aspect of the China situation," according to the State Department's Charles Burton Marshall, was that "The Chinese under Peiping are… supporting the side which is against their own interests," as the Chinese Stalinists in their immeasurable "ignorance… [thought] no word is believable but that of Moscow. They believe everything from that quarter. They are foreclosed from believing anything else."[238] Marshall was certain that "the mad course followed by Peiping at the behest of Moscow" would bring devastation to China's government and people.[239] Yet while US representatives thought it would be "one of history's most tragic paradoxes" if China's blind and emotional leaders precipitated a general war,[240] they also held no expectation that China's leaders could be made to come to their senses and see that they were blindly waging. a war that furthered no interests but those of Moscow as long as the indoctrinated Communists remained in power.

This ascription of bias was if anything crystallized by China's October invasion and reports from the front of Chinese forces fighting with irrational, inhuman barbarity and disregard for their own lives.[241] While US diplomats hoped the Chinese leaders would come to their senses,[242] they thought this was highly unlikely as China's leaders and populace seemed to be in the "full flush of revolutionary fervor" and that "Peiping's fevered eyes" were incapable of separating fact from ideologically warped fiction. China's initial military success worked to strengthen US commitment to these beliefs: representatives of UN member states who were talking more directly to the Chinese told their US counterparts that the Chinese were "drunk with success and increasingly difficult to deal with."[243]

---

[238] FRUS, Korea and China 1951 volume 2, S/P Files: Lot 64 D 563
[239] FRUS, Korea and China 1951 volume 2, S/P Files: Lot 64 D 563
[240] FRUS, Korea and China 1951 volume 2, S/P Files: Lot 64 D 563, Document 83
[241] Hastings

US officials' perception of bias in Chinese leadership shifted their *Persuasive Preferences* toward coercion and away from argument. They expected an argumentative, dialogue-based approach to persuasion to be ineffective, and thought that only the simple language of force could feasibly persuade the Chinese to change their behavior. One of the factors US officials cited most frequently in expressing skepticism of iterative dialogue with China was the irrationality of its leadership, particularly as manifested in the complete subservience of their independent thought to communist dogma and the word of the Kremlin. This was true in both civilian and military circles: Harriman had the "feeling that negotiations were impossible,"[244] and General Bradley later concurred that direct communication with China was "not practical or desirable."[245] George Kennan underscored both this preference and its source in a memo he sent to Secretary Acheson, reporting the Chinese were "excited, irresponsible people, on the consistency of whose reactions there could be no reliance" and therefore impossible to persuade with rational arguments or measured dialogue**.**[246]

---

[242] FRUS, Korea and China 1951 volume 1, 320/1-651: Telegram, Document 25
[243] FRUS, Korea and China 1951 volume 1, 691.93/1-2251: Telegram, Document 86
[244] FRUS, Korea and China 1951 volume 1, 102.201/5-2951, Document 305
[245] FRUS, Korea and China 1951 volume 1, 795.00/6-2851, Document 373
[246] FRUS, Korea and China 1951 volume 1, 795.00/5-2551, Document 327

Truman and Acheson, already inclined by their personal worldviews to be skeptical of the need to understand adversaries' interests and inclined to negotiate more through force than argument, likewise expressed the belief that China's leaders were biased and were in turn extremely skeptical of any proposal for negotiations. This effectively doomed the prospects of the camp of so-called 'Titoist' officials who advocated open engagement with China as a means of opening a rift between the CCP and the Soviets: Acheson shot this prospect down, saying that Mao and his party had demonstrated through their invasion that they were in the grip of "sheer madness;"[247] because they would never simply see reason they had to learn "by bitter experience what it was like to be Russia's ally."[248] Truman, meanwhile, saw the Chinese as he saw many Communist leaders: as "unstable and unpredictable,"[249] and in this case on an ideologically driven rampage: the President was quite intractable in his resultant belief that military force rather than diplomatic dialogue was necessary to achieve any kind of acceptable settlement with China.[250] Indeed, Truman reflected this position not only in internal deliberations but in his private diary, where he came down on the side of negotiating ultimatums that went to such an extreme as "an end to hostilities in Korea or the complete destruction of China."[251]

Top military officials were, if possible, even less subtle in expressing what they thought about the delusional biases of Chinese leadership and its implications for the efficacy of argumentative diplomacy. The Joint Chiefs of Staff dismissed the idea of negotiations as dialogue by asserting that the communist Chinese "understand only what they wanted to understand and… to sit down with these men and deal with them as representatives of an enlightened and civilized people is to deride one's own dignity."[252]

---

[247] Foot p. 81
[248] Foot p. 48
[249] Foot 32
[250] Foot p. 175-6
[251] Foot p. 176

In sum, because of what they perceived as China's insusceptibility to reason or logic, US officials believed that only military pressure, and to some extent economic pain, could alter the mind of Chinese leadership.[253] In other words, while a ceasefire was thought to be in the interests of all parties, US policy reflected the belief that only economic and military leverage would convince the Chinese 1) that it was in their interests to stop fighting[254] and 2) that they should acquiesce to a negotiated settlement establishing a ceasefire wherein Korea remained divided along the 38th parallel. Put another way, US officials believed it pointless to engage in a dialogue with China to negotiate terms until such time as "the Chinese nation should be able to restablish rational control of its policy."[255] Notably, this conclusion about the inefficacy of dialogue and the converse efficacy of simple informational threats was reached before the US had made any attempt to negotiate in the more costless forum of pure information as dialogue and argument.

Perceived Incompetence and Escalation Preferences:

The general US aversion to a costly (relative to dialogue) military approach was further mitigated by a general belief that Chinese leaders were, in Ridgway's words, "monumentally stupid,"[256] a group of total novices in international affairs. In light of this apparent incompetence, the threat posed by continuous conflict with China was believed to be less severe than the material balance suggested, and in turn the need to compromise was less urgent despite the low strategic priority placed on Korea. In brief, the perception of military political incompetence in Chinese leadership shifted officials' *Escalation Preferences* away from de-escalation despite the overarching goal of ceasefire. Even otherwise cautious decision-makers advocated a stick-only approach to coercive diplomacy. The US officials who were most convinced of Chinese military political incompetence were confident that, despite what raw military capabilities suggested, escalated conflict with China posed a threat that was not just low but negligible: these hardliners pressed not only for escalation but full-scale military invasion of China.

---

[252] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 7: Telegram, Document 496

[253] Foot 2019, Hastingns 2010

[254] See Acheson's telegram for delivery to the UK: FRUS, Korea and China 1951 volume 1, 795B.5/4-3051: Telegram, Document 266

[255] FRUS, Korea and China 1951 volume 2, S/P Files: Lot 64 D 563, Document 88

[256] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 118: Telegram, Document 613

The perception of Chinese incompetence pervaded US thinking, from the lowest soldiers looking at their counterparts in the Chinese military to the highest-level officials looking at their counterparts in Peiping. Notably, there was an unmistakably racial component to this belief in the universal incompetence of 'gooks' and 'chinks,' directed not only at the North Korean and Chinese adversaries but the South Korean allies.[257] That the ascription of incompetence to Chinese leaders was at least partially driven by racial stereotyping may account for the ease with which many adopted it in the face of minimal evidence, and moreover why the disastrous misjudgment that Chinese forces would be incompetently led proved so sticky in the face of China's military successes.[258]

MacArthur's derision of Chinese military competence was perhaps the most extreme, and probably therefore the best-known example of American perceptions of Chinese incompetence. MacArthur had been infamously confident of his ability to outmaneuver and annihilate the Chinese, particularly given what he and his generals derided as the dumb decisions constantly made by Chinese military leadership, and called Chinese military status as a "myth" that was easily exploded.[259] MacArthur thought, put simply, that "chinamen cannot fight,"[260] and expressed confidence to John Foster Dulles that he could handle the war "with one arm tied behind my back:"[261] he was a persistent advocate of extending the war into the Chinese mainland both before and after Chinese forces pushed the US back from the Yalu**.** Of course, MacArthur was something of a unique figure: his bombast, narcissism, burning hatred of 'Orientals,' and proclivity for self-important delusions were legendary in their own right. To extrapolate from his beliefs about Chinese competence to the US government as a whole would therefore be suspect, since it is possible that these merely reflected his inflated sense of his own competence. With respect to this personality-oriented perspective, however, it is worth noting that MacArthur's perception of adversary incompetence did not extend to the Soviets.

---

[257] Hastings p. 70

[258] Cumings p. 14

Moreover, MacArthur's belief that the Chinese were incompetent extended beyond him and outlasted his term of service. As quoted above, Ridgway likewise derided the military skill of China's leaders, and Truman subscribed to the perspective that China's political and military prestige were "dangerously exaggerated" and had to be deflated so that UN member nations would stop pressing for concessions or direct dialogue**.**[262] A common conclusion of early historiographies of the Korean War was that it was always viewed as a limited conflict and MacArthur's jingoism was an aberration, but it has since become clear that there were many other influential proponents of war expansion, optimistic about US capacity to deal with a backwards guerilla band under incompetent leadership.[263] During the MacArthur hearings, the

---

[259] FRUS, Korea and China 1951 volume 1, 795.00/2-2351, Document 141
[260] Cumings p. 26
[261] Cumings p. 14
[262] FRUS, Korea and China 1951 volume 1, 795.00/1-1351: Telegram, Document 61
[263] Foot p. 25

MacArthur claimed that the JCS had supported his views on the need to expand the war into China, a point the JCS publicly disputed but acknowledged in their secret testimony before the joint committee.[264] While the war remained relatively limited, it is worth noting the seriousness with which a full escalation was contemplated, and by extension the persistence with which US officials deflated the threat posed by China independent of the USSR.

Public Opinion on China:

Beyond internal jockeying, another apparent cause of elite movement towards behaving as if Chinese leaders were biased and incompetent is that this was a narrative successfully espoused to the US public. The contrast between China as the "irrational foe" and the USSR as the hostile, implacable, but still cautious and responsible opponent was baked into public discussion of US/China/USSR relations from an early stage.[265] Some of the most determined advocates of this image were military leaders who, apart from the apparent fact that many of them fully believed this narrative, were dissatisfied with the resources allocated to the East Asian theater and seeking to ramp up public pressure on the Truman administration for a more forceful posture on the Korean Peninsula. MacArthur in particular was infamous for publicly denouncing the administration's insufficient commitments to the Far East and "appeasement of Communists in Asia,"[266] an argument that he backed with inflammatory rhetoric about the barbarity of the Asiatic foes with whom US forces were engaged.

In a similar vein, Truman's domestic political opposition was determined to push the image of China as an out of control "fundamentalist, revolutionary power"[267] that could not be reasoned with and had to be militarily punished. This narrative helped them to undermine support for Truman's foreign and domestic policies through a ruthless drumbeat of accusations that Truman and his administration were soft on communism and in particular the Asiatic states that subscribed to the ideology and so effectively become willing slaves of the Kremlin. The more ubiquitously accepted this narrative became, the more public attitudes towards negotiations with China soured: by 1951 polling showed roughly 68% of the public opposing negations and favoring force.[268] The Truman administration was both attentive and, in policy terms, responsive to these public attitudes. Fear of political backlash, and the need for Truman to demonstrate that he was militarily tough on the apparently expansionist communist satellite, further reduced the Truman administration's openness to the prospect of talks with the Chinese.

---

[264] Foot 141

[265] Rosemary Foot, p.27

[266] Foot p. 70

[267] Foot, p. 28

Preferences in Practice: US Behavior Towards China Before and During Truce Talks:

In line with my theory's predictions about diplomatic bargaining preferences and choices with a militarily powerful *Madman*, most US officials expressed a preference for some form of escalatory coercive diplomacy. In sum, they were hesitant to enter talks, and once negotiations began, they rebuffed attempts at iterative dialogue and relied on the simple, universal language of threats to persuade the other side.

Before the truce talks, US officials were extremely skeptical of any discussion of formal negotiations, or even informal talks to build a better relationship with China, and dragged their feet on any kind of ceasefire talks for months after the official goal of US policy had reverted to the attainment of a status quo ceasefire. Despite entreaties coming from both their adversaries (the Soviets) and their allies (UK, Sweden, Canada, other UN allies), US officials made little to no effort to create infrastructure for bilateral communication and indeed actively sabotaged early U.N. efforts to organize ceasefire talks. During the first of the meetings between President Truman and British Prime Minister Atlee, Acheson summarized Atlee's view on ceasefire negotiations as follows: "At this time we had no choice but to negotiate with the Chinese. These negotiations would, of course, extend beyond Korea and it was certain the price the Chinese demanded would be Formosa, a seat in the U.N., and recognition."

---

[268] Foot p. 181

While the UK may have accepted both the prospect of negotiations and the likely need for concessions, the US was having none of it. American officials worked to ensure that the UN ceasefire resolution was one the Chinese would find unacceptable, essentially poisoning the waters by refusing to include any mention of the various issues the Chinese delegation had raised as central: the withdrawal of the seventh fleet from Taiwan, international recognition of its claim to a UN seat, and withdrawal of foreign troops from Korea. Acheson was particularly averse to holding out the prospect of a CCP seat at the UN as a bargaining card, saying that it would be the "worst possible" idea to negotiate with the communists at this time and in this way.[269] While the US did sign on to the UN resolution for the sake of maintaining allied unity, they did so "in the fervent hope and belief that the Chinese would reject it," and China did not disappoint them.[270]

That simple rejection, however, was not enough: the US then introduced a UN resolution officially branding the CCP as the aggressor party in the Korean war, the political equivalent of the metaphorical 'nuclear option.' Once this resolution passed, the US was determined to sever all diplomatic relations with China, impose major sanctions, and get as many UN states as possible to follow suit. Opposition from more cautious UN members was intense, but US determination proved greater and the aggressor resolution passed on February 1: the US opposition to dialogue and concessions, and its preference for coercion, had at least temporarily been realized.

---

[269] Hastings 181
[270] Foot p. 111

This position did not prove sustainable: the costs of the conflict increased as Chinese land power was ‘secretly’ reinforced with Soviet air power, and this along with intensifying pressure from UN allies brought Truman and his top officials to reconsider their total opposition to talks. When the Soviet ambassador Malik formally proposed talks on June 23, 1951, the US accepted, and on July 10 the truce talks began at Kaesong.

However, while the US had agreed to talks, American officials’ total opposition to an actual dialogue remained extremely strong: the US adopted the methods of ‘threat only’ coercive diplomacy more or less the moment the truce talks began, before making any attempt at persuasive argument. Operating under the belief that “the Chinese Communists understand strength only”[271] and “in these discussions, the language of diplomacy is inappropriate and ineffective,”[272] American representatives made heavy use of leverage in an attempt to extract concessions without giving anything in return. Chinese delegates made occasional attempts to convey the military or political rationale for their bargaining positions, but the US flatly refused to engage in such discussions; American officials were adamant that the UN “would not negotiate between logic and illogic, reasonableness and unreasonableness.”[273] Memos sent from the negotiating team frequently summarized the remarks of Nam, China’s representative, in such terms as “typically illogical,”[274] “irrational,”[275] or as indicating a “disregard for facts.”[276] How, the officials wondered, could they possibly engage with reason and logic these fervent communists who "understand only what they want to understand?"[277] The answer was simple: the could not, and in effect, they did not. Perhaps the symbolic low point in the talks came when the delegations simply sat in silence for two hours and eleven minutes, refusing to utter a word to their counterparts: after all what was the point?[278]

---

[271] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 4: Telegram, Document 493
[272] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 7: Telegram, Document 496
[273] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 99: Telegram, Document 842
[274] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 1: Telegram, Document 492
[275] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 22: Telegram, Document 506
[276] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 24: Telegram, Document 539

Instead, the US sought to bring unbearable military, economic and political pressure on China by, along with a plurality of the international community, imposing crippling sanctions and threatening an escalation to all-out war if China did not agree to a settlement on US terms. In Ridgway's terms, the US bargaining approach had to be one that contained "more steel and less silk:"[279] this reflected a widely held belief that only economic pain or "military suffering could alter the enemy mind:" Truman was particularly wedded to achieving a settlement through means of force rather than dialogue.[280] The real internal debate amongst US officials with respect to how best to persuade the Chinese was between economic sanctions, naval blockade, bombings, and ground invasion: neither a more concerted attempt to understand China's interests nor any real concessions were truly part of the discussion. Indeed, negotiations stalemated for two years over the issue of the repatriation of communist POWs: China wanted all prisoners repatriated, while the US wanted to provide the option of voluntary repatriation. Without dismissing the importance of freedom of choice for foreign POWs, we can acknowledge that two years of additional war over this issue is difficult to comprehend given its tangentiality to primary US geostrategic interests. Yet rather than talk in the face of Chinese recalcitrance US officials dug in, to the point that there was in both the Truman and Eisenhower administration serious consideration of using nuclear weapons to break the stalemate.

---

[277] FRUS, Korea and China 1951 volume 1, Lot 55D128: Black Book, Tab 7: Telegram, Document 496

[278] Hastings p. 232

[279] Hastings p. 220

[280] Foot p. 175

In other words, US policy makers resorted to measures that were costly in terms of both resources and political capitol because they believed dialogue was a fundamentally ineffective persuasive tool where China was concerned.

The USSR: The Reasonable Adversary:

American diplomatic bargaining preferences towards the USSR were shaped by several principle beliefs about the thinking and judgment of Soviet leaders. First, they were no friends of the West or misunderstood seekers of security and global harmony. These "were men of limitless ambition and an ideology that brooks no rival system of thought:"[281] they would take every opportunity to expand their power and extend the grip of their communist ideology if they thought they could get away with it. Second, they were fundamentally rational: objective in most areas, extremely competent, and quite cautious especially where the prospect of World War III was concerned. As long as the threat of force was in the background to disabuse the Soviet leaders of the prospect of an easy win, it was possible and desirable to negotiate with them on the basis of rational dialogue.

This basic set of beliefs about the judgment of the Kremlin shaped how US officials approached diplomacy with their Soviet counterparts throughout the course of the war: with the prospect of force always in the background. US officials engaged their Soviet counterparts in both formal and, especially, informal dialogues, conducting the latter through an every expanding series of backchannels through which the superpowers worked 'behind the curtain' to avoid unchecked escalation while maintaining their rivalry.[282] While no one that rational dialogue could persuade the Soviets to support a unified and democratic Korea, most were optimistic that these would be an effective means of achieving tactic cooperation in pursuit of mutual interests; namely, the institution of a ceasefire and the avoidance of escalation to general war. While in reality US officials were unwilling to instigate World War III over Korea, they were willing to take an aggressive approach to diplomacy with China in large part because they believed the Soviets were likewise averse to a third world war.

---

[281] FRUS, Korea and China 1951 volume 2, Lot 64 D 563: S/P Files, Document 2

[282] Austin Carson. "Facing off and saving face: covert intervention and escalation management in the Korean War." *International Organization* (2016): 103-131.

Perceived Objectivity and Persuasive Preferences:

In the period of the Korean War, America's appraisal of the "actions of Moscow and [Peiping] in Korea established the basis for distinguishing between the USSR as its 'reasonable adversary' and China as its 'irrational foe.'"[283] US officials believed that Soviet leadership was committed to an expansionist communist agenda, but they did not believe communist ideology biased the Kremlin's judgment, at least not to anywhere near the same extent as the thought it warped the perspectives of North Korea and China. Put another way, US foreign policy circles saw Soviets propagation of communist ideology as a strategic means of spreading Moscow's influence, rather than evidence of fanatic dogmatism. Truman saw Stalin as a "wily adversary"[284] and an opportunistic realist.[285] While he thought that "only in the face of force would [Stalin] talk and negotiate,"[286] he also believed that under such conditions Stalin was sufficiently reasonable and cautious that fruitful negotiations could take place.

George Kennan is perhaps the official whose beliefs best encapsulated, and worked to cement in the Truman administration's thinking, this view of the Soviets as an objective but intractable adversary. In a series of memos reporting his views of and interactions with Soviet officials, Kennan expressed confidence that the Soviets were fundamentally reasonable. Kennan wrote that the Soviets, unlike the "hysterical and childlike" Chinese, "were reasonable, businesslike, and, after all, the real power to deal with,"[287] and that the US could deal directly with the Kremlin because they "took a serious and responsible attitude toward what they conceived to be their own interests."[288]

---

[283] Foot p. 27
[284] Foot p. 32
[285] Foot p. 97
[286] Foot p. 33

Put another way, the Soviets had an objective understanding of what would and would not advance their interests, and the majority consensus in US civilian and military government was that chief among these was the avoidance of general war. To maintaining an ongoing convergence of expectations about what would risk versus avoid a general war, US frequently engaged Soviet officials in backstage conversations to clarify US interests, intentions, and the causes for which they would commit further military resources to Korea. This is especially noteworthy given the open animosity between the superpowers, and underscores both the importance the US placed on escalation avoidance and the belief that bilateral communication was vital to keep fighting at an acceptable level of intensity.

This is not to say that Kennan, or a majority of US officials, thought that the Kremlin was wholly unbiased, especially when it came to comprehending Western ideas; referring to the same officials whose reason he had lauded, Kennan lamented that the Soviet perspective gave them a distorted view of the US that "was more pathetic than sinister."[289] All the careful dialogue and intelligently constructed arguments in the world would not bridge the plethora of disagreements between the US and USSR.

---

[287] FRUS, Korea and China 1951 volume 1, 795.00/3-1751: Document 167
[288] FRUS, Korea and China 1951 volume 1, 795.00/5-2551: Document 327
[289] FRUS, Korea and China 1951 volume 1, 795.00/5-2551: Document 327

However, US officials did not expect these distortions of Soviet perspective to prevent dialogue between the adversaries from advancing mutual interests. Officials acknowledged that the Soviets would never accept a unified or democratic Korea; the divergent "ideas represented by the two systems [US and the USSR] made reconciliation impossible."[290] Yet while there "not the slightest possibility of extracting a capitulation in Korea from… the Russians, the possibility of a deal… does exist since the present situation in Korea is unsatisfactory to both the Soviet Union and the United States. There would appear to be a mutuality of interest."[291] Kennan held frequent conversations with Yakov A. Malik, the Soviet representative to the UN: he thought that through these open and rational dialogues "we were able to ascertain that there was some identity of view between our two governments,"[292] and that the superpowers could likewise engage in a dialogue to come to agreement on a settlement in Korea that advanced these shared interests. In other words, he thought that *complex information* would be an effective persuasive tool: negotiations could serve as a forum for the adversaries to clarify their interests through dialogue and achieve a settlement in the interests of both sides. This is a stark contrast to the US view regarding the persuasive efficacy of complex information with China's leaders: they identified the same shared interest in a ceasefire, but thought that Mao and the CCP were so biased that only the languages of threats and pain would make them realize their interest in an armistice.

Perceived Competence and Escalation Preferences:

Furthermore, US officials held the Soviets' political and military competence in high, even exaggerated, regard: "while in international affairs the CCP were bigots and novices, the kremlin are seasoned practitioners" and had to be taken seriously.[293] Indeed the prevailing narrative in the Truman administration was that the Kremlin was the puppet-master pulling the strings of China and North Korea, that without Soviet direction and aid its Asiatic satellites would be unable to militarily engage the US[294], and it was likewise "inconceivable" the North Korean or Chinese invasions were carried out without instruction from the Kremlin.[84]

---

[290] FRUS, Korea and China 1951 volume 2, S/P Files: Lot 64 D 563: Document 2
[291] FRUS, Korea and China 1951 volume 1, 795.00/3-1751: Document 167
[292] FRUS, Korea and China 1951 volume 1, 795.00/5-2551: Document 311

In no small part because they perceived the threat of military escalation with the USSR as high, US officials were willing to make concessions to Moscow in the interest of halting the conflict and avoiding escalation. Even when China posed a more severe and immediate threat to American forces in Korea than the USSR, US officials thought compromise was necessary only with the latter, as a settlement not backed by Russia would be one "built on sand."[295] Indeed, one of the main concerns quashing the voices of those advocating for military measures against the Chinese mainland was fear of the escalatory reaction this might provoke from the Soviets.[296]

It was a near universal opinion in US policymaking circles that allowances must be made for concessions to the Soviets; Kirk, the ambassador in the USSR, was adamant that the door not be closed on compromises to Moscow's demands.[297] Johnson adhered to this stance, on a number of occasions harshly criticizing proposed policies that he thought took modus vivendi with the USSR off the table: while he knew a non-communist Korea was an impossibility, he was also convinced that reaching an understanding with the Soviets was not only desirable but necessary for the sake of US interests.[298] Where the Soviets were involved, the Joint Chiefs were adamant that “great care should be used, in putting forward a negotiating position, not to allow talks to break down”[299] in light of the suffering this would inflict on the American people.

---

[293]

[294] FRUS Korea 1950, "Period of the UN Offensive: September 16-November 28," 795.00/10-1050

[84] Schnabel p. 68-9

[295] FRUS, Korea and China 1951 volume 1, 795.00/3-1751: Document 167

[296] Foot

[297] FRUS Korea 1950, "Period of the UN Offensive: September 16-November 28," 795.00/11-1950

[298] FRUS, Korea and China 1951 volume 1, 795.00/4-651, Document 208

This is not to say that American perceptions of the Soviets as competent were the sole reason they manifested compromising diplomatic preferences towards their primary adversary: the material capabilities of the USSR were also a major factor. However, as the case of China demonstrates, material threat alone is insufficient to explain preference for compromise with an adversary: observers also need to think the adversary has the strategic competence to effectively employ the raw capabilities at its disposal.

Public Opinion on the USSR

The Truman administration faced substantial obstacles to iterative diplomacy with the USSR as a result of public opinion. While the American public was no more eager for a major war than their government, but their anti-communist attitudes were extremely strong, fanned by republican legislatures who sought to portray Truman as soft on communism in order to undermine his domestic standing. Whatever the republican minority truly thought about the judgment of Soviet leadership, they publicly demonized the Kremlin as not only adversarial but entirely unreasonable, the sort of fanatic individuals with whom it was impossible to hold a fruitful conversation. In their narrative if Truman or his representatives were speaking with the Soviets it was not a rationally constructed diplomatic approach to an adversary that might be swayed by logic, but a capitulation to America’s enemy by a president who was either too weak to oppose the Soviets or sympathetic to their cause. In short, the US public expressed diplomatic bargaining preferences reflective of a belief that Soviet leadership was the *Fanatic*, rather than the *Rational*.

---

[299] FRUS, Korea and China 1951 volume 1, 795.00/6-3051, Document 391

Preferences and Public Opinion in Practice:

In line with my theory's predictions about diplomatic bargaining preferences and choices with an initially unreceptive *Rational*, most US officials expressed a preference for using force to demonstrate the necessary resolve to bring the Soviets to the table for a rational diplomatic dialogue, in this particular case to pursue the shared goal of maintaining a limited war. However, the US public's broad acceptance of the narrative that the Soviets were a fanatical foe that neither could nor should be reasoned with created major domestic pressure against diplomatic negotiations: this worked to push most of the talks pursued by the Truman administration behind closed doors.[300]

The divergence between the US public and the US government regarding beliefs about Soviet judgment and resultant diplomatic bargaining preferences posed a problem for the Truman administration when Malik proposed truce talks on June 23rd. By that point, both the civilian and military officials in US government had concluded that a ceasefire was both desirable and the best possible outcome the US could achieve in Korea at anything resembling an acceptable cost. Moreover, there was unanimous agreement that 1) Soviet buy-in would be necessary for a lasting ceasefire[301] and 2) the Soviets were the only adversary in the conflict with the requisite rationality of judgment to seek or accept a ceasefire without having to be coerced into doing so.

---

[300] Carson 2016

[301] FRUS, Korea and China 1951 volume 1, 795.00/4-651, Document 208

However, Congressional republicans demanding a public explanation for why Truman was letting the "crimson clique in the State Department"[302] run Korea policy and bow down to the Soviets made it politically untenable for Truman to pursue diplomacy or détente with the Kremlin. Instead, in the theater of American domestic politics republicans and democrats "were fighting to show their uncompromising attitude towards Communism"[303] by deriding ongoing negotiations and promising escalation to end the war.[304] This public opinion backdrop proved a major constraint that dampened the political will of the Truman administration to act on its diplomatic bargaining preferences towards the USSR, at least in an open forum. While this public opinion backdrop obviously did not entirely prevent Truman and his administration from engaging the Soviets in ongoing dialogue, it moved the negotiations behind closed doors as a prophylactic measure against domestic backlash.

During the truce talks at Kaesong and Panmunjon, therefore, the US primarily conducted diplomatic dialogue with the Soviets on the backstage[305] while simultaneously negotiating with the Chinese on the main stage. This dynamic seems to have limited the potential to achieve a ceasefire settlement with the party, the USSR, that the US 1) believed could be rationally negotiated with and 2) was willing to take conciliatory measures towards to avoid escalation. US officials were focused on finding a ceasefire settlement that the Soviets would accept, but in the end this focus may have been misplaced: the Kremlin kept a degree of distance from the formal talks so as to maintain some deniability. Ironically, US representatives were far less focused on ensuring the acceptance of their direct negotiating counterparts, the Chinese, due in no small part to the pervasive but ultimately incorrect assumption that China's leaders were Soviet puppets.

---

[302] Foot p. 107
[303] Foot p. 183
[304] Hastings p. 315
[305] Carson 2016

A Note on North Korea: The Neglected Adversary:

In 1945 when Korea was first divided, US officials did not believe there were marked differences between the judgment of North and South Koreans. American officials had always been skeptical democratic reforms could be implemented by the people of North *or* South Korea, who they saw as "not interested in politics… [and] illiterate," certainly not the sort of "educated enlightened electorate" that could implement democracy on its own.[306] Yet at least initially, they held North Korean leaders in high regard. An assessment of Kim Il-Sung reads:

> "Kim is brilliant… calm and self poised; when he deals with business he grasps and settles the point immediately. He is a great leader."[307]

Moreover, the North Korean population was thought to be sufficiently reasonable that if given the opportunity to hold elections, they could be persuaded to choose democracy over communism, and implement this governmental reform under the guidance of US/UN trusteeship. Put another way, they believed the North Korean population was incompetent, but not particularly biased.

However, throughout the Commission negotiations, US observers grew increasingly convinced that, aided by the Soviet's propaganda and their "puppets"[308] in the North Korean communist party, communist ideology had permeated North Korea and subsumed whatever capacity the population had for independent thought. By the time the US-USSR commission officially fell apart US observers were certain that North Korea was thoroughly lost to communism, and skepticism that outreach could prove effective transformed into certainty. Indeed a majority of US officials ceased to think of the North Korean government (and population) as distinct agents capable of independent thought or actions. Intelligence reported that "No Soviet Satellite… had been molded to the political likeness of the USSR as had North Korea:"[309] and US officials had "no doubt that North Korea… was a typical Soviet Puppet State."[310] Kim Il Sung, along with the rest of North Korean leadership, were now viewed as incompetents who "except for loyalty and subservience… possess few qualifications for the responsibility of high government and party office."[311]

---

[306] National Archives, Records of the XXIV Corps Historical Division: US-USSR: Transcripts AG Radio May 1947, 17 April 1947

[307] National Archives, Records of the XXIV Corps Historical Division: North Korean Personalities, Box 69

[308] National Archives, Records of the XXIV Corps Historical Division: North Korea today, Box 57

The pervasive US belief that the North's government and population were biased *and* incompetent was entrenched by the narrative that emerged to explain the surprise North Korean invasion of the South in June 1950. US officials thought that to risk a general war against the militarily superior US, North Korean decisionmakers had to be either blinded to reason by the communist indoctrination, wholly subservient to the will of the Kremlin, or both. These beliefs deeply influenced the policies US officials advocated and expected to be effective in pursuing an end to the conflict; the idea of direct diplomacy with North Korea was dismissed as futile and unnecessary, even as US officials approached their Soviet counterparts in search of a negotiated settlement.

US officials unanimously concluded that it was "inconceivable" that the North Korean invasion was carried out without aid and instruction from the Kremlin;[312] in other words, the invasion was a Soviet initiative.[313] If Soviet and not North Korean interests were driving conflict, then Soviet interests had to be addressed to bring the conflict to a halt. US policy was therefore to ignore the incompetent puppet and look to negotiating a compromise with the puppeteer: the assumption was that North Korea lacked agency or distinct interests and would not stop fighting without orders from the Kremlin.

---

[309] National Archives, Records of the XXIV Corps Historical Division: North Korea today, Box 57

[310] FRUS Korea 1950, "Period of the UN Offensive: September 16-November 28," 795.00/10-1050

[311]

[312] FRUS Korea 1950, "Period of the UN Offensive: September 16-November 28," 795.00/10-1050

[313] Schnabel p. 68-9

US officials further believed that direct dialogue was rendered pointless by the biased judgment of the North Koreans; logic and common sense would be lost on a people and government so thoroughly indoctrinated by communism. Intelligence summaries in the aftermath of the invasion stated unequivocally that "North Korean leaders were not exposed to the moderating influence of truth but… only violent propaganda from Communist sources. An appeal to reason therefore would not have affected this course of action [the invasion]."[314] This is noteworthy because US officials believed any reasonable actor would conclude war was not in the interests of North Korea, but did not think the North could be persuaded of this. It was not, in short, that the US saw no convergence of interests, but that it believed it would be impossible to convey the existence of this convergence point to the other side. Between the belief that North Korea was an incompetent puppet and that dialogue was ineffective, US officials did not even default to explicit coercion: diplomatically, they simply ignored North Korea.

These diplomatic preferences manifested in behavior that contributed to the prolongation of conflict: despite the ostensibly low costs of dialogue, no serious attempt was made to engage North Koreans directly in order to halt the conflict or drive a wedge between the USSR and its satellite. Indeed, US representatives viewed communications and peace feelers from the North Korean government "with complete contempt,"[315] and strongly opposed proposals from UN members to directly approach North Korea to drive a wedge between the USSR and its satellite.[316] While US officials never vowed to veto direct dialogue with North Korea through the offices of the UN, they remained convinced that such a move was folly and cautioned against or threw roadblocks in its path at every turn.

---

[314] FRUS Korea 1950, "Period of the UN Offensive: September 16-November 28," 795.00/10-1050

[315] FRUS, Korea and China 1951 Part 1, 357.AK/4-2151: Telegram, Document 251

[316] FRUS, Korea and China 1951 Part 1: 357.AD/4-1151: Telegram, Document 225, 357.AD/4-1151: Telegram, Document 226, 357, AD/4-1151: Telegram, Document 227

Conclusion:

The case evidence presented in this chapter suggests that US officials' diplomatic bargaining preferences and choices during the Korean War were significantly impacted by beliefs, both their own and those of the public, about the judgment of each adversary's leaders. The discussions about and diplomatic behavior toward each adversary were in line with the predictions of my theory, subject to the relevant scope conditions.

US officials saw the Kremlin as *the Rational* type and manifested diplomatic bargaining preferences accordingly: they were eager to avoid escalation with the competent soviets and optimistic about the prospect for identifying shared interests through rational dialogue. However the Truman administration was also responsive to anti-negotiation sentiment from an American public that, at the height of red-scare rhetoric, saw the Soviets as something closer to the *Fanatic:* the gap between public and elite preferences pushed most dialogue with the Soviets into an ever expanding array of backchannels.

By contrast, most US officials saw Mao's CCP as either a *Madman* type or very close to it. Contrary to the expectations of the rationality of irrationality, however, this belief made US officials more willing to escalate and quicker to resort to the coercive language of threats and force, egged on by a public that was fully supportive of bombing campaigns against the irrational Chinese. In line with my expectations about how the diplomatic bargaining preferences that manifest in response to an apparent *Madman* interact with said *Madman* possessing a high level of military power, the US did not completely throw caution to the wind and engage in uncontrolled escalation against China. However, they came much closer to doing so than a raw analysis of interests and material capacity might have suggested was prudent, and without first exhausting the nominally costless option of argument-based diplomacy. Moreover, in a further blow to the contentions of the rationality of irrationality and the associated 'madman theory,' US officials remained firmly unwilling to make even minor concessions to the Chinese.

The examination of evidence presented here problematizes the argument that the US neglected negotiation and compromise due to reputational concerns, or because it had a preponderance of power that made diplomacy unnecessary. If these factors were the sole or primary driver of diplomatic bargaining preferences, we would expect their impact to be evident across adversaries, to shape US diplomatic preferences towards the perceived ultimate aggressor and not merely its puppet states. The purchase granted by these explanations in accounting for heterogeneity of diplomatic bargaining preferences across adversaries but within a single conflict and time period is therefore unsatisfactory. By contrast, considering US beliefs about the judgment of its adversaries seems a powerful explanation for what is otherwise a puzzling divergence of preferences and behavior. This further illustrates the importance of incorporating decisionmakers beliefs about how others *think* into our theories of international politics.

The analysis in this chapter also raises questions that present opportunities for future research: I detail these in this dissertation's Conclusion but outline them here in order to motivate the empirical research in Chapter Five and to provide an honest reflection on what the case leaves unanswered or does not satisfactorily address.

First and foremost, this case does not answer the question of where beliefs about judgment come from: what factors influence them, how do they change and update, and why do we observe disagreement both between elites and the public and within political elites in a given state? The case discussion hinted at some possible answers: race and ideology, for example, were frequently linked to assessments of adversary competence and bias, but this link is correlational rather than causal and lacks satisfactory scope conditions. While the significant divergence between adversaries in how the US pursued the same general policy (ceasefire) weakens the counterargument that beliefs about judgment follow from policy preferences and not the other way around, it would be incorrect to say that the evidence in this case cogently eliminates that possibility. More research in this area is needed, and in Chapter Five I present a pair of experiments designed to test the null hypothesis and probe a range of variables that could influence beliefs about the judgment of adversary leaders.

Second, this case raises but does not address a longstanding question that is often raised in response to theories of state behavior and preferences in which the actors are individual elites: namely, how do elites aggregate into state policy? Because the primary DV of interest for my theory is elite preferences and not state behavior, this question remains outside the predictive or explanatory scope of my theory. While I advanced some narratives about preference aggregation, these were not specific and should not be taken as an attempt to answer a question this project does not ask. However, future researchers might consider using beliefs about adversary judgment and their impact on diplomatic bargaining preferences as a lens through which to empirically examine various theories of preference aggregation, since it provides an alternative avenue of examination to more traditional models of bureaucratic and institutional competition and could therefore shed additional light on this area.

Finally, this is only a single case study, and in the future I and other scholars should look to bring in additional cases for the purposes of conducting a comparative case study test of the theory. Several of the leaders used as illustrative examples of my typology in Chapters One and Two would make very interesting subjects for more in depth investigation, though to avoid cherry picking it may be wise to go beyond the boundaries of these more obvious examples. Apart from further case studies of interstate conflict, it would be informative to study intrastate conflicts, either between the state and non-state actors or among non-state actors vying for control of the state, in order to see whether the theory applies to and sheds light on the dynamics of insurgency, terrorism, and civil war.

Chapter 5: Madness in the Eye of the Beholder: How Individuals Form Beliefs about the Rationality of Foreign Leaders

We're all mad here. – Chesire Cat, Alice in Wonderland

What shapes peoples' beliefs about the *minds* of foreign leaders, from their objectivity and competence in particular issue areas to more all-encompassing questions about their rationality and overall mental capacity? Actors' preferences and choices in any strategic interaction depend in no small part on how they anticipate their counterparts will behave, and how are expected to behave is inextricably linked to how they are expected to *think.* For example, leaders may be less optimistic about the prospect of de-escalation with adversaries that seem to hold inflated views of their own strength, skill, or smarts, and are therefore prone to overestimate their chances of military victory while underestimating the costs of conflict. Likewise, actors may take a different approach to trade arrangements with leaders who have erratic preferences, short time horizons, or are generally prone to poor economic judgment whether due to ignorance, disinterest, overconfidence, or some other trait. Understanding why these beliefs about foreign leaders' *judgment*, about how their minds work, form and update is key to understanding a variety of interstate interactions.

However, existing international relations research provides neither theory nor empirics to help us attain this understanding. Idiomatic wisdom holds that actions speak louder than words, but we live in a world in which neither facts nor actions necessarily speak for themselves. In other words, while it is possible that actors' beliefs about the judgment of their counterparts reflects the objective reality of how those counterparts think, it is also possible to imagine a world in which these beliefs are anything but objective, or one in which they are merely rhetorical window-dressing that exert no causal impact on important choices in interstate relations. Assessing the relevance of theories regarding the international impact of perceived rationality or irrationality, from Schelling's 'rationality of irrationality' or the colloquial 'madman theory,' requires attaining a robust understanding of what determines these perceptions in the first place. Without taking that crucial first step, we cannot confidently dismiss the 'null hypotheses' that beliefs about the minds of foreign counterparts are 1) inconsequential artifacts of the variables (material or otherwise) that *actually* shape international politics or 2) post-hoc rationalizations for policies, as opposed to a factor that actually shapes policy preferences or choices. Put another way, we cannot currently be certain whether expressed beliefs about the judgment of adversaries are, with respect to material and structural variables, epiphenomenal and therefore inconsequential, or exogenous and therefore independently impactful.

In this chapter, I examine a number of distinct variables that could plausibly influence people's beliefs about the judgment of foreign leaders, and demonstrate that central to these beliefs are two variables usually neglected or dismissed as inconsequential by theories of international politics; namely leaders' personal 'costless behavior,' from speaking style to physicality to temperament, and ideology. First, people are more likely to perceive ideologically distant leaders as being biased, unaware of risks, or generally irrational. Second, people are more likely to ascribe irrationality to leaders who (due to their style of speech or gesticulation) seem dispositionally temperamental or erratic, while the opposite is true of leaders whose personal behavior makes them seem calm and collected. These effects hold regardless of whether we control for material factors (e.g. military and economic power) or the intentions and attitudes of either the leader (e.g. whether the foreign leader is an enemy or ally, whether or not they harbor hostile designs) or the observer (e.g. whether the observer has warm or cold feelings towards the foreign leader). This consistency is significant because it shows that, at least amongst members of the public, beliefs about the rationality of foreign leaders' judgment are informed by variables exogenous to structure or affect. This, in turn, implies that any impact these beliefs exert on international politics cannot be dismissed as either an epiphenomenal artifact of other explanatory variables or simply a rationalist process of reputation assignment based on objective assessments of a counterpart's costly behavior.

The rest of this chapter proceeds as follows. First, I draw on extant literature to detail three distinct pathways through which we might expect observers to form and update beliefs about the judgment of foreign counterparts; a rationalist behavior-based pathway, a pathway based on identity and personal behavior, and a structurally deterministic pathway. I describe a range of variables that fit within each of these pathways, and outline the mechanisms linking these variables to inferences about the judgment of foreign leaders. Second, I detail a novel conjoint experiment that I use as a hypothesis-generating plausibility probe: its purpose is to ascertain how members of the public weigh these variables in assessing the rationality of foreign leaders when all of the variables are presented simultaneously and none are controlled for. Third, I briefly discuss the results of this hypothesis-generating conjoint experiment, which suggest that a foreign leader's political ideology and personal behavior carry significant weight in observers' assessments of their rationality (full results of the conjoint are presented in Appendix B). Fourth, I detail a multivariate factorial survey experiment that I designed to build on these findings by testing the causal effect of the leader-level variables which the conjoint results implied carried weight in subjects' assessments of a leader's rationality. This second experiment controls for material variables and a leader's policy decisions, and presents subjects with randomized vignettes describing leader-level variables of interest: race and culture, ideology, and personal behavioral style. Fifth, I discuss the results of the second experiment, which show that differences in foreign leaders' political ideology and personal behavior have strong and statistically significant causal effects on 1) observers' assessment of their overall rationality, competence, or judgment and 2) how likely it is that the leader was rational to initiate a particular act of interstate aggression. People are more likely to attribute irrationality to ideologically dissimilar leaders, and to leaders whose behavior is temperamental, odd or otherwise does not fit the model of a calm statesman. Put another way, variation in leader-level variables causes observers to form very different beliefs about the rationality of leaders whose material power, policy actions, and intentions are identical. I conclude by discussing the implications of these findings for using variation in beliefs about leader rationality, competence, and judgment as an Independent Variable in theories of international politics.

How Beliefs About Judgment Form: Explanations from Extant Literature:

Roughly speaking, there are three distinct pathways through which we might expect observers to form and update beliefs about the judgment of foreign counterparts. First, a rationalist pathway wherein inferences about the minds of counterparts are based purely on their costly behavior, so for example observers will perceive adversaries as irrational when those adversaries act irrationally. This rationalist pathway coheres with classical theories of reputation formation, in which past actions and patterns of behavior form reputations that alter the way actors are treated in strategic interactions. Second, an identity-based pathway wherein inferences about the minds of counterparts are based not just on the costly actions of those counterparts, but who those counterparts are as well as their costless behavior. Third, a structurally deterministic pathway in which inferences about the minds of counterparts are a product of attitudes and intentions that arise from material factors, for example branding a weak adversary as a dangerous madman to justify intended aggression ('Saddam the Madman') or a strong rival as a rational statesman to justify an apparently necessary détente ('Uncle Joe' Stalin). Such structurally determined inferences may arise due to motivated reasoning (self-justification) or the need to sell policy positions to third parties. Either way, however, a world in which the structurally deterministic pathway is dominant is one in which inferences about the judgment of foreign counterparts do not meaningfully shape international politics, since they are epiphenomenal to causally important material variables.

Mirror Imaging and The Null Hypothesis:

Importantly, none of the three pathways outlined above and detailed below capture the null hypothesis explicitly or implicitly subscribed to by a wide array of IR scholarship. In sum, much of the extant IR literature takes the position that there is no variation in the beliefs international political actors hold about the rationality, competence, or judgment of their foreign counterparts. Most of the literature that subscribes to the null hypothesis does so explicitly by simply neglecting altogether the subject of perceived rationality in international politics.

However, theories of state 'misperception,' from the foundational work of Jervis onward, posit the null hypothesis as a primary explanatory mechanism for many otherwise puzzling behaviors in international politics. The misperception approach explicitly contends that the decisionmakers responsible for state policy, along with members of the public, always believe their counterparts will interpret information and formulate plans of actions as they themselves would. Put another way, no matter what, states as composite actors inevitably fall prey to the mirror imaging fallacy and so believe their counterparts are rational. This misplaced belief in turn helps drives everything from the classic security dilemma to high-risk crisis behavior. Modern works inspired by canonical misperception research do not all explicitly treat mirror imaging as inevitable or theoretically central, but they all at least implicitly accept the inescapable ubiquity of mirror imaging in international politics. Put simply, canonical and current works on misperception in international politics treat mirror imaging, and mutual expectations of common knowledge, as a constant in international politics.

The failure of either foundational or recent misperception literature to address variation in beliefs about the judgment of counterparts is at first surprising; after all, the principle mechanisms through which this literature explains otherwise puzzling outcomes or choices in international politics revolve around the knowledge that the state actors are not universally rational. A seemingly obvious and important extension of this assertion would be to, theoretically and empirically, examine when the mutual expectation of rationality between state actors is, justifiably or not, undermined; surely if scholars can perceive irrationality in state actors, state actors can perceive irrationality in their counterparts.

Structural Determinism:

While the basic contention that identity matters in reputation formation, and the inferences people make about the traits of counterparts more generally, is hardly counterintuitive, there is a surprising dearth of theoretical or empirical work on this topic in international relations scholarship. A significant proportion of IR research, especially the canonical texts written from the mid to late Cold War period that still constitute a jumping off point for much contemporary research, dismisses identity outright by explicitly ‘black-boxing’ states; in other words, assuming states both are and see each other as actors that a not meaningfully differentiated by anything except structural factors (power, size, location, etc…).[317] In the most purist of these structural approaches, even past and current actions are irrelevant in shaping actors' beliefs about their international counterparts: all that matters is what counterparts *could* do in light of structural variables, not what they have done or are doing, much less who they are or how they behave in costless issue areas.[318]

---

[317] Waltz, Kenneth N. *Theory of international politics*. Waveland Press, 2010.

In effect, if the structural determinism explanatory mechanism is correct, we should expect beliefs about the judgment of adversary leaders to be driven by variation in their material capabilities. Therefore, relative military and economic power would be the most immediate drivers of variance in perceptions of leader rationality, with people imputing rationality or lack thereof based on the policies these material variables drive them to advocate. While unlike the null hypothesis both expressed and internalized beliefs about the judgment of foreign counterparts may vary if the structural determinism pathway is accurate, these beliefs will not independently impact preferences or choices in international politics. The structural explanation is agnostic as to how material variables, and the policy preferences they induce, determine actors' internalized or expressed beliefs. Motivated reasoning, cognitive biases such as the 'halo effect,' and audience manipulation are an illustrative but not exhaustive list of plausible operative mechanisms.

It is illuminating to think about these mechanisms in the context of the classic story where actors develop a preference for preventive military action against an adversary state that is rising in relative material power. Actors may persuade themselves that an adversary with growing material power is irrational to convince themselves that negotiation and peaceful coexistence is un-likely and therefore strengthen their internal case for preventive hostilities, or vice-versa. For example, Chamberlain desperately wished to avoid conflict with a rising Germany and so had strong motivation to perceive Hitler as rational; his beliefs were driven by material variables and their implications rather than Hitler's policy actions or personal traits. Unconscious cognitive biases may cause actors to impute negatively connotated characteristics, of which irrationality may be one, to counterparts they think of as increasingly dangerous or those they expect to be in conflict with. Finally, if state decision-makers wish to make the case for an intended conflict against a rising adversary to either domestic or international audiences, they may brand the leader of the rising power as irrational in order to influence audience attitudes. Such attempted audience manipulation was apparently a principle reason the George W. Bush administration devoted substantial effort to pushing the 'Saddam the Madman' narrative, though this does not necessarily mean no one in the administration bought into the metaphorical snake oil they were selling.

---

[318] E.g. Mearsheimer, John J., and Glenn Alterman. *The tragedy of great power politics*. WW Norton & Company, 2001.

Behavior-Based Beliefs:

These pure structural approaches are an extreme, however: most IR work incorporates the idea that state actors update their beliefs about international counterparts in response to said counterparts' actions. A majority of existing scholarship in some fashion assumes states think probabilistically rather than possibilistically about the traits, goals, or likely actions of counterparts, and treats actions as a primary informer of assigned probabilities. Put another way, most research assumes states look not only at what their counterparts are capable of doing, but what their counterparts have done.

However, this research has tended to focus on how past actions inform beliefs about the values or desires of counterparts. For example, a great deal of work has been conducted on some iteration the impact state actions have on others' beliefs about their relative aggressiveness, tolerance for cost, and willingness to carry through on threats; all of these fall under the heading of desires, or the relative value a counterpart attaches to the advancement of particular aims.

To the extent that political science has addressed how actions inform beliefs about the judgment processes of counterparts, it has done so implicitly and without empirical underpinnings. For all the apparent interest generated by Schelling's conception of the rationality of irrationality, the question of how state leaders come to be perceived as irrational has never been rigorously interrogated. While scholars, policymakers, and journalists have debated how leaders can credibly act irrationally, this debate is premised on an unspoken assumption that there are particular behaviors and policy choices that are self-evidently irrational, and the principle barrier to acquiring a reputation for irrationality is the fact that observers believe such a reputation is strategically advantageous and so may see irrational actions as performative rather than genuine. In this model actors would for example conclude an adversary was biased when a sufficient proportion of its behavior suggested this was the case: in simple terms actors believe an adversary is irrational when that adversary acts irrationally. There are no filters through which an adversary's actions are interpreted: behavior speaks for itself and its meaning does not depend upon identity of the actor or the observer. It is this combination, the informational primacy of costly actions and the absence of heterogeneity across actors in the inferences about counterparts made on the basis of a particular action, that most neatly characterizes the 'costly-behavior' based pathway of belief formation about the minds of counterparts.

The observation of unexpected policy behaviors that seem situationally inappropriate ("*I* would not have done that"), is under this theoretical model the most likely impetus for observers to update their beliefs about the subject's judgment.[319] Put another way, unexpected or ambiguously rational foreign policy decisions, such as picking a fight with a more powerful rival or voluntarily pulling out of mutually beneficial international institutions, are the most probable acute cause of ‘negative’ (aka. towards incompetent or biased) updating of beliefs about a foreign leader’s judgment.

Identity, Costless Behavior, and Leader Level Variables:

Leader-level variables from identity to ideology to appearance might function as heuristic prisms through which observers interpret and draw inferences from the words and actions of others. Because traits like personality, thought processes, and judgment are not directly observable, beliefs about these traits in counterparts are liable to be particularly reliant upon non-behavioral factors.

It is not difficult to find examples of the subjective meaning of words and actions, along the tendency to interpret them through leader-level variables like identity, appearance and ideology. Words that make one speaker seem sincere may make another seem duplicitous: a profession of moral regret from the gruff, perpetually scowling Richard Nixon may sound less sincere than the same profession from the more smooth-talking, fresh-faced Bill Clinton or the unfailingly sincere George W. Bush. On the flip side, what some observers find endearing others may find off-putting: some people may be inherently sympathetic to moralistic rhetoric from their president while others reflexively dismiss it as insincere politicking. Finally, relational factors such as partisan affiliation may make Democratic observers more likely to infer sincerity or positive characteristics from the words of Bill Clinton than Republican observers, and vice-versa. These anecdotal observations are all examples of the subjectivity of 'costly actions', and the corresponding role of identity and 'costless behavior,' in inferential processes about the intentions or moral character of others.

---

[319] This project therefore takes the situational expectations based position on attribuition, characteristic for example of Tomz 2007.

Political science does not entirely neglect the influence of leader-level variables, or state identity, on the beliefs actors form about their counterparts. Identity, whether conceived as a simple friend/enemy dichotomy or a component of more nuanced areas (ex. culture, ideology, status, religion, or race), plays a central role in a range of IR theories regarding how observers make inferences about interstate counterparts. A more recent wave of research, often conducted through examination of face-to-face diplomacy, examines how certain stylistic aspects of a foreign leader's personal behavior can impact the inferences actors make about said leader's internal traits.[320] Both verbal and somatic behavior constitute a rich source of vivid evidence from which observers draw conclusions about the traits, and explain the actions, of others.[321] Note that this refers not to the *content* of an actor's behavior (ex. what they say or do) but rather what we might call the *style* of an actor's behavior: how they look and sound independent of what they are actually doing. Facial expressions, hand gestures, even posture are all components of physical behavior that inform people's impressions of others. Myriad stylistic components of speech likewise affect impression formation: volume, cadence, tone, and pitch are all examples of ways speech can impact perceptions independent of the content of the speech in question.

---

[320] Keren Yarhi-Milo,. *Knowing the adversary: leaders, intelligence, and assessment of intentions in international relations*. Vol. 146. Princeton University Press, 2014. Hall, Todd, and Keren Yarhi-Milo. "The personal touch: Leaders' impressions, costly signaling, and assessments of sincerity in international affairs." *International Studies Quarterly* 56, no. 3 (2012): 560-573. Holmes, Marcus, and Keren Yarhi-Milo. "The psychological logic of peace summits: How empathy shapes outcomes of diplomatic negotiations." *International Studies Quarterly* 61, no. 1 (2017): 107-122. Holmes, Marcus. "The force of face-to-face diplomacy: Mirror neurons and the problem of intentions." *International organization* (2013): 829-861.

321 For example an observer that perceives a counterpart's gestures and tonal fluctuation as indicative of erraticism may, for example, more readily attribute future military action to delusional expectations or the equivalent of self-defeating tantrums. By contrast the cool, collected performance of a leader may make observers more inclined to search for transient situational explanations for apparently risky, surprising, or strategically suspect actions. Indeed, decisionmakers convinced of an adversary leader's strategic acumen are more likely to think apparently bungled military actions are a strategic ruse than a genuine flop: US analysts were convinced the Soviet's vulnerable missile deployments were a scapegoat meant to distract from the *true* arsenal.

However, these theories focus overwhelmingly on explaining variation in actors' beliefs or inferences about the values and desires of counterparts; identity is used to explain the formation of beliefs about what others want, but not the formation of beliefs about how others think. Prominent illustrations of this trend come from the extant literatures on threat perception, trustworthiness, intentions, and reputation. Some iteration of identity is a major variable in many theories of threat perception, and in a similar vein how observers formulate beliefs about the aggressiveness, 'greediness,' hostility, or general military intentions of foreign states. Indeed, some of the most influential (if controversial) works in international relations advance the thesis that cultural and ideological differences are *the* deterministic factor driving not just perceived but actual hostile intentions between states in the international system (the implicit suggestion being that states *should* base assessments of threat and hostile intentions on the identity of counterparts).[322]

Theories of reputation formation that incorporate identity (even conceived in the broadest possible terms of ingroup/outgroup) take a similar approach; a notable strand utilizes identity as an explanatory variable for why actors make divergent inferences about the resolve of distinct counterparts whose behavior is similar or identical.[323] While resolve is something of an umbrella concept and so poorly defined,[324] it frequently includes factors such as cost tolerance, credibility when threatening or promising to fight, and valuation of the issues or stakes of a given dispute. The mechanism through which each of these are posited to shape resolve is reducible to the ‘values/desires’ component of decisionmaking: for example, cost tolerance is a function of the military and economic price an actor is willing to pay to advance a particular aim. The above examples illustrate that existing research has almost exclusively concerned itself with the role of identity in shaping inferences about the goals rather than thought processes of international counterparts.

---

[322] Eg. Huntington, Samuel P. "The clash of civilizations?." In *Culture and politics*, pp. 99-118. Palgrave Macmillan, New York, 2000.

[323] Mercer, Jonathan. *Reputation and international politics*. Cornell University Press, 2010.

[324] Kertzer, Joshua D. "Resolve, Time, and Risk." *International Organization* 71 (2017)

The Importance of Minds, not Goals:

These are theoretically distinct, and have different practical implications than, a theory wherein perceived variation in greediness or cost tolerance is driven by how an actor is expected to interpret information and understand the world. Attention to the link between identity and beliefs about adversary rationality permits us to think more thoroughly about the mechanisms through which leader-level variables, identity, and costless behavior may drive international political behavior. For example, there are different theoretical and empirical implications if variations in perceived greediness due to leader-level variables are driven by, for example, how actors are expected to perceive the status quo based on all the information available to them. The same is true if identity-driven attributions of insurmountable resolve actually work through the expectation that an adversary leader will reflexively dismiss or refuse to believe reports that their soldiers were being defeated on the battlefield.

In brief, judgment-based perceptions have distinct implications from their values-based counterparts. It is therefore problematic that linkages between structure, costly behavior, and identity or costless behavior have not been rigorously theorized or tested. Moreover, this neglect is surprising and indeed counterintuitive in light of how casually identity is linked to objectivity and competence, not just in informal conversation or journalistic writings but also in the throwaway remarks of scholarly texts. For instance, there is a tendency to cite actors' political positions as evidence, or a result, of their relative objectivity and rationality; people commonly claim counterparts they identify as political rivals are biased or irrational, whether because the counterparts' political ideology warps their view of reality or because only those with a warped view of reality would subscribe to the rival ideology in the first place. Put another way, those who disagree with our politics either generally view the world through cracked glasses, or they will interpret the world through partisan lenses. Yet like so many 'conventional wisdoms' about human behavior, this linkage has not been rigorously theorized in political science research, let alone operationalized and empirically tested.

Outside of political science, a range of academic work acknowledges that people can perceive counterparts' judgment processes as different from their own; this research provides overwhelming support for a thesis that finds scant purchase in political science; actors' identity as well as actions shape the inferences counterparts draw about their judgment. At the most general level, how observers interpret behavior that departs from their rational baseline is strongly influenced by whether they regard a counterpart as an ingroup or an outgroup; this is the "fundamental attribution error." The presence of salient ingroup/outgroup divisions prompts observers to draw dispositional rather than situational inferences from others' behavior.[325] Applied to inferences about the minds and judgment of counterparts, the fundamental attribution error implies that observers of will, on the basis of the same ambiguously rational actions, make inferences about the judgment of outgroups versus the information available to ingroups.

A range of research in psychology suggests that ideological divergence especially increases the likelihood of bias attribution. The probable mechanisms driving this relationship seem intuitive; particularly in ‘self-selected’ dimensions of identity such as partisan affiliation, a counterpart’s membership in an outgroup may be perceived as both cause and/or consequence of biased perceptions.[326] The expectation that ideological convictions are a source of bias is hardly without merit: considerable evidence corroborates the biasing effect of ideology on the judgments of both laymen and experts,[327] though individuals seldom recognize their own ideologies as a source of impediments in judgment.[328] The relationship between bias attribution and ideological divergence is also traceable to people’s naïve realist conviction that their own ideology is the objectively ‘correct’ one, and that biased judgment is therefore a prerequisite for anyone to adopt a worldview at odds with their own. In other words, counterparts are biased because they are part of an ideological outgroup, and they are part of an ideological outgroup because they are biased.

---

[325] For general evidence, see the psychology literature on the 'fundamental attribution error,' notably Gilovitch and Ross 2015. For a pioneering application to international politics, see Mercer (1996).

The primary issue area in which the relationship between ideological divergence and bias attribution have been investigated is political ideology in American Politics. People are more likely to ascribe bias to counterparts who express political beliefs with which they disagree. For example, subjects told a counterpart voted for a candidate they opposed were significantly more likely to rate as ‘high’ the likelihood that the counterpart’s vote was driven by bias. Similarly, subjects told a counterpart opposes them on some policy position, for example abortion or climate change, are more likely to attribute that position to the counterpart’s biased view of reality than those told a counterpart agrees with them or is undecided. The impact of ideological rivalry on intergroup bias attribution is not merely an artifact of domestic partisan divisions: it also manifests in the context of more ‘general’ ideologies, such as nationalism. Notably, this suggests the relationship between ideological division and bias attribution holds for intergroup ideological divides that are not inherently in opposition. Nationalism is not necessarily an active opposition to a single other nation in the way that American Republicanism is a rejection of the Democratic party and vice-versa. Rather it is as an unerring devotion to one nation above, and in opposition to, all others equally.

---

[326] Gilovitch and Ross 2015

[327] See for example the seminal work of Phillip Tetlock, *Expert Political Judgment: How Good is it? How Can We Know?* Princeton, NJ: Princeton Univeristy Press, and Richard K. Herrmann "How Attachments to the Nation Shape Beliefs About the World: A Theory of Motivated Reasoning." *International Organization* 71, no. S1 (2017): S61-S84.

[328] Gilovitch and Ross 2015, Robinson, Keltner, Ward and Ross 1995

To summarize, this third potential pathway suggests that leaders’ behavior does not speak for itself: it is usually interpreted through heuristics linked to costless behavior and components of identity such as group membership, physical appearance, and ideology.[329] This does not mean that the actions of foreign leaders are irrelevant in a world where costless actions and leader-level variables are important. However, non-behavioral variables strongly and systematically impact how actions are interpreted, so observers may make divergent inferences about the objectivity or competence of different leaders who behave in objectively identical ways on the international stage.

Methods Outline:

To test the empirical microfoundations of the theoretical pathways detailed above, I use a pair of novel survey experiments. The first is a choice-based conjoint experiment that examines the full range of plausible variables associated with each pathway: this first survey is a hypothesis-generating plausibility probe and should not be confused with an experiment fielded for the purpose of causal inference. The second, which builds upon the first, is a multivariate vignette-based factorial experiment that controls for variation in structure and policy choices to focus on the role of leader-level variables that the choice-based conjoint suggests are influential in the formation of individual beliefs about the rationality of foreign adversaries.

---

[329] As with all heuristics, there is nothing inherently 'wrong' about this approach, which is often adaptive and reflective of real linkages between the observable trait and the unobservable trait it is used to infer.

The choice-based conjoint is ideally suited to simultaneously assessing the relative weight a variety of dissimilar variables play in belief formation and updating; it is a tool often used in marketing research to, for example, determine an ideal combination of product traits. Subjects are shown eight pairs of foreign leaders, with each leader generated from a random combination of nine traits each of which has multiple possible values: political ideology, religion, race, government type, speaking style, personal behavior, economic advancement, military power, and gender attitudes. Subjects are told that each leader has issued a military threat against a neighboring state, and for each pair of leaders are asked to choose the one they think was less likely to have been rational to use the threat of military force in order to advance their goals.

This conjoint experiment allows me to examine the relative importance and directional impact of a wide variety of variables in shaping individuals' beliefs about the rationality of foreign leaders, in a way that would be prohibitively difficult with a more traditional or tightly controlled design. However, the sheer number of plausible combinations, reliance on repeat sampling, and the plausible lack of realism stemming from the thin descriptions of foreign leaders, all reduce the conjoint's utility as a tool of causal inference, though it remains useful for the purposes of hypothesis generation and refinement. Moreover, because it uses a binary choice between leaders based on a generalized idea of 'rationality' as a dependent variable, it cannot account for the multi-dimensional conception of 'judgment' adopted in this project.

To address these shortcomings, I pair the conjoint experiment with a second experiment that employs a multivariate (4 x 4 x 3 x 3) factorial design with vignette-style treatments.

Subjects read a description of a longstanding territorial dispute between two states, the militarily weaker of which has recently attacked its rival; they are then told about the President of the weak aggressor state. The description of the President is randomized on four dimensions: 1) race and name, 2) political ideology, 3) personal behavior represented through reputation and speaking style, and 4) degree of economic and cultural advancement experienced during their leadership. Subjects are then asked an array of questions about the rationality of both the president's military aggression (for example was it rational? How likely do you think it was the president understood the associated risks?) and the president in general (how likely do you think it is the president bases his beliefs on reasonable interpretations of evidence? How generally competent and knowledgeable do you think the president is?) In brief, this design moves from a general to a targeted approach with respect to independent variables: it focuses on the most theoretically important variables and uses (relative to the conjoint) long-form descriptions of their ideal types as treatments and showing each subject only one version of each variable. Conversely, it incorporates multiple operationalizations of the dependent variable, beliefs about a foreign leader's judgment, some of which focus on specific dimensions of the dependent variable while others examine it holistically. This allows for observation of more nuanced variation in treatment effects.

Why Start with Survey Experiments in the Mass Public Context?:

Why are novel survey experiments the best available method for investigating what variables influence beliefs about the judgment, rationality, and thinking of foreign leaders? After all, there are numerous avenues of research that could plausibly generate empirical evidence to shed light on the question at hand, such as interviews, laboratory experiments, or public opinion surveys without experimental treatments. However, for the purposes of ascertaining the impact of a variety of variables on public beliefs about the judgment of foreign adversaries, survey experiments offer a variety of advantages relative to these alternative methodologies.

First, surveys are the gold standard for ascertaining beliefs, examining the factors that contribute to them, and testing when and how they update. Second, relative to methods like interviews and laboratory research that require greater investment of time and resources in each subject, surveys are a cheap way to sample a high number of subjects from a wide range of demographic groups. Introducing randomized treat conditions into a survey therefore permits researchers to gather a high volume of data points from which to make causal inferences about relationships between independent and dependent variables of interest. While survey experiments are by necessity more superficial in the kind of information they can convey or gather than are laboratory experiments or interviews, this limitation coheres with the nature of the hypotheses being tested, since in the context of public opinion they are fundamentally about how members of the public form or update beliefs about the judgment of foreign adversaries based on superficial information.

There are numerous reasons why the question of public belief formation is one we should address. First, as previously discussed, public attitudes have significant implications for the attitudes and behavior of elite decisionmakers, particularly in democratic systems of government such as the United States. Second, the question of how people assess the minds of outgroups as encapsulated in the mind and/or judgment of the group leader is bigger than IR. It has implications for social psychology, anthropology, and evolutionary psychology on the mechanisms driving group cohesion and conflict, to name just a few disciplines that may be interested in the findings of this research. IR is a conceptual and empirical black hole with respect to this question, but these other disciplines have produced a range of hypotheses on this issue that are competing, contradictory, and/or simply do not speak to each other at all. While the surveys detailed below adopt an IR frame, they are also the first empirical work of which this author is aware that attempts to simultaneously assess all of these hypotheses. Therefore, the findings of this work may serve as a jumping off point for understanding intergroup mind-perception in myriad contexts.

Furthermore, if we are aiming to harness the causal inferential power of experimental methods, then we must begin with samples of the mass public even if we are ultimately more interested in the beliefs and decisions of the political elites that most directly determine a state's foreign policy. Investing the necessary time and resources to obtain an elite sample, and to re-configure experimental design parameters to fit such a population, makes sense only after replications and extensions on public samples have increased our confidence in the experimental parameters, given a strong sense of plausible findings, and sufficiently narrowed the variables of interest such that a small number can be manipulated to maintain the validity of findings from the extremely small population sizes that characterize elite samples. By permitting meaningful testing of an otherwise infeasibly high number of variables, a conjoint design serves as a kind of 'first crack' at narrowing the field of variables. While in this dissertation I use a case study to test at the elite level the theories confirmed and refined via experiments at the level of the public, a clear avenue for future research will be experimental research on an elite sample.

None of the above should be taken to imply I believe survey experiments are flawless, nor that surveys alone are an ideal or even sufficient method for addressing the issues with which this chapter is concerned. Indeed, it is because I am cognizant of the limitations of survey research that I ultimately employ a mixed methods approach. Survey experiments *are* an ideal tool for addressing the hypotheses specific to public opinion. However, they are less well suited to ascertain whether public beliefs about the rationality of adversary leaders impact the beliefs and/or behavior of political elites and vice-versa. The relevant political elites are a population that is expensive to access and difficult to conduct informative survey experiments on even if resources do not present a barrier; survey experiments therefore cannot provide more than inferential evidence beyond the public side of the hypothesized cycle.

Survey 1: Choice-Based Conjoint Design:

Comprehensive consideration of extant IR and psychology research points to very different variables that may shape observers' inferences about the judgment of interstate adversaries. Each of these offers an empirical implication that can be subjected to empirical testing. However, testing many competing hypotheses against each other poses methodological challenges, especially when the hypotheses in question do not necessarily speak directly to each other and therefore cannot be treated as either mutually exclusive or complimentary, nor simultaneously tested with a single coherent manipulation. For reasons of statistical power which force most traditional experiments to focus on manipulating only a small number of factors at a time, such difficulties are somewhat difficult to overcome, and leave us with no clear answer to one of the central questions this project is interested in. Namely, when presented with a plethora of factors that might directly or indirectly influence perceptions of a counterpart's objectivity or competence, what variables weigh most heavily on the inferences observers make, and conversely what factors are ignored or subsumed?

To address this question, I use a conjoint experimental design, which is perfectly suited for addressing the question in which I am interested: one that centers on how observers weight numerous factors that are not necessarily related nor mutually exclusive in forming and updating beliefs about foreign leaders. In my conjoint experiment, subjects are shown randomly-generated profiles for two foreign heads of state, **A** and **B**, and told that each of these leaders has recently threatened a neighboring rival with military force unless the rival in question hands over a piece of disputed territory. Subjects are then asked which leader they think was more likely to have been irrational in attempting to advance their territorial ambitions with the threat of military force. I then repeat the exercise seven more times, each involving an independently randomized pair of foreign leaders. All of the treatments are straightforward, consisting of a single sentence or less, and the levels of the categories were chosen either because there were obvious but limited options, because the literature suggested certain categories (eg. Democracy vs. dictatorship), or to highlight certain contrasts in which I was particularly interested (eg. Bombastic leaders vs. quiet ones). Of course as with any design choice, the way these treatments were phrased has consequences that must be acknowledged in consideration of the results; in choosing to make some subtle differentiations between levels in the category of personal behavior (passionate vs. angry) it is entirely plausible the impacts of this variable were downplayed relative to categories in which the distinctions between the levels were stark and clear, such as military power (powerful vs. weak) or economic development (rich vs. poor).

The most puzzling question which a conjoint design permits us to address is: how do people weight a range of distinct factors when asked to assess the rationality of two foreign aggressors? Conjoint designs carry substantial advantages in the area of statistical power, as they permit me to simultaneously manipulate the myriad factors discussed above despite the often unconnected, contradictory, or seemingly redundant nature of the theoretical predictions they provide. Conjoint surveys free researchers from the power considerations that constrain factorial experimental designs by making a small number of assumptions that are either guaranteed by the design or can be empirically verified in analysis of the results, permitting us to pool observations across choice tasks. The overall result of the conjoint experiment design is that we can estimate the effect of each of a wide range of treatments – in this set up referred to as Average Marginal Utility Effect (AMUE) – with a relatively small number of subjects. The AMUE denotes the average difference in the probability of being seen as more likely to have been irrational in threatening a rival when comparing two different attribute levels – for example an Authoritarian leader vs. a Moderate leader – where the average is taken over all possible combinations of other leader/country variables.

Sample**:**

The primary conjoint study was fielded on 2,000 American adults recruited through Amazon Mechanical Turk (MTurk) in June 2019. A secondary conjoint study, identical to the primary one but for the inclusion of an 'expert opinion' variable included to assess the role of elite narratives in informing public beliefs, was fielded on 500 American adults recruited through MTurk in June 2019.[330] Because the primary and secondary conjoint studies are near-identical, in the 'Conjoint Design' section below I do not describe them separately; for separated analysis see Appendix **C.**

Conjoint Design:

---

[330] The sample size, while optimal for meaningful calculation of AMUEs, renders this secondary study more of a plausibility probe. Because the expectation, strongly supported by the results, was that particularly in the context of an experiment elite narratives would be of overwhelming importance: if people are going to parrot back the information conveyed to them by experts, the low stakes and attention of an experimental setting makes that context one in which it would be fairly surprising if most people did not reflexively match their answers to the elite narrative variable so as to minimize effort and reduce cognitive load. Therefore, investing resources in a larger sample size would have been of limited value in a standard choice based rather than ratings based conjoint design.

After the consent process, subjects saw an introductory text explaining that state leaders often attempt to advance their territorial and political aims by threatening rivals with military force unless their demands are met, despite the fact that such threats are a risky proposition that can easily backfire. Participants were then told they would be presented with a series of pairs of foreign leaders, with identifying names and details omitted, who had recently threatened a neighboring state with military force unless the neighbor in question handed over a piece of disputed territory. Subjects were told that for each pair, they would be asked to assess which of the two leaders was more likely to have been irrational in risking a threat of military force to advance their territorial aims. Subjects were then presented with the first scenario, which randomly generated attributes for two foreign leaders that had recently threatened their neighbors, and asked to choose the leader profile (A or B) they thought was more likely to have decided to threaten force through a decisionmaking process that could be characterized as irrational. The full list of treatments is presented in **Table 5.1**: within these treatments, 'leader-traits,' 'leader-behavior,' and 'state-level' variables are separated. Each treatment was randomized completely independently of other treatments: the randomized value assigned to 'political ideology' had no effect on the randomized value assigned to 'religion' or 'military power,' for example. A pair of sample profiles are depicted in **Figure 5.1.**

Table 5.1: Conjoint Study Treatments:

	Variable Name:	**Shared Starting Text**	**Variable Levels**
Leader Level (Traits):			
	A) Political Ideology	*The Leader…*	*-is a political moderate* *-is a far-right conservative* *-is a far-left liberal* *-is a socialist*

			-is an extreme nationalist *-is an Authoritarian*
	B) Religion	*The Leader…*	*-is a devout member of the Christian faith* *-is a devout member of the Islamic faith* *-is strictly secular and non-religious*
	C) Government	*The Leader…*	*-is an absolute Dictator* *-is the head of an autocratic, single party state* *-heads a Democratic government, and his election is widely seen as legitimate* *-heads a 'fake democracy:' he was supposedly elected but the elections are widely considered illegitimate*
	D) Race	*The Leader…*	*-is Arab* *-is White* *-is Hispanic* *-is Asiatic* *-is Black*
Leader Level (Behavior):			
	E) Behavior Reputation	*The Leader is known for being…*	*- temperamental, strange and eccentric* *- calm, but strange and eccentric* *- passionate, strange and eccentric* *- temperamental and uptight* *- calm and uptight* *- passionate but uptight*
	F) Speaking Style	*The Leader's public speaking tends to be…*	*- erratic and delivered in a loud, overstated tone* *- erratic and delivered in a quiet, understated tone* *- organized and delivered in a loud, overstated tone* *- organized and delivered in a quiet, understated tone*
	*G) Expert Opinion**		*- Many experts have expressed doubts about the leader's state of mind* *- Many experts have expressed confidence in the leader's intellect and praised his mode of thinking* *- Some experts have expressed confidence in the leader's intellect and praised his mode of thinking*

			- A small number of experts have expressed confidence in the leader's intellect and praised his mode of thinking *- Some experts have expressed doubts about the leader's state of mind* *- A small number of experts have expressed doubts about the leader's state of mind* *- The leader does not have a notable reputation among experts*
State Level:			
	H) Economy	*The Leader's state has…*	*- an economy that is advanced and rich* *- a mid-level economy: it is neither particularly rich nor poor* *- an economy that is backwards and poor*
	I) Military Power	*The Leader's state…*	*-has a military that is relatively powerful* *-has a military that is about as power as the military of the threatened neighbor* *-military is relatively weak compared to the threatened neighbor*
	J) Culture, Women's Rights	*Women in the state…*	*- have the right to own property, work outside the home, choose who they marry and freely speak their mind* *- cannot own property or work outside the home, but may choose who they marry and freely speak their mind* *- have the right to own property and work outside the home, but cannot choose who they marry and can be formally punished for freely speaking their mind* *- cannot own property or work outside the home, are not free to choose who they marry, and can be formally punished for freely speaking their mind*

Figure 5.1: Sample Pair of Leader Profiles:

Leader A	Leader B
The leader is a far-right conservative	The leader is an extreme nationalist
The leader is a devout Christian	The leader is a devout Muslim
The leader heads a Democratic government, and his election is widely seen as legitimate	The leader is an absolute Dictator
The leader is Asiatic	The leader is Black
The leader is known for being temperamental, strange and eccentric	The leader is known for being calm and uptight
The leader's public speaking tends to be erratic and delivered in a loud, overstated tone	The leader's public speaking tends to be organized and delivered in a quiet, understated tone
The leader's state has an economy that is advanced and rich	The leader's state has an economy that is backwards and poor
Woman in the state cannot own property or work outside the home, are not free to choose who they marry, and can be formally punished for freely speaking their mind	Women in the state have the right to own property, work outside the home, choose who they marry and freely speak their mind
The leader's state has a military that is relatively powerful compared to the threatened neighbor	The leader's state military is relatively weak compared to the threatened neighbor

Participants were then presented with seven additional randomly-generated pairs of foreign state/leader profiles, such that each respondent performed the choice task a total of eight times. Either before or after the main study, participants answered a series of demographic questions, including gender, age, education, Party ID, political ideology, religious affiliation and political knowledge. The demographic section also includes batteries of questions designed to measure some plausible dispositional drivers of the results, mainly nationalism, international cosmopolitanism, and hawkishness. To keep the survey tool at a manageable length, these dispositional batteries are more cursory than the full slate of questions that are used in experimental designs for which the traits in question are theoretically central.

It is worth stressing that the list of treatments, as large as it may be at 8 factors with 3 to 6 levels each, is still not exhaustive of factors that plausibly play minor or even major roles in inferences about the judgment of foreign counterparts. There are a multitude of factors that play some part in at least tangentially related areas of IR literature, and especially the literature on reputation formation, which could be usefully incorporated into future work. A leader's past actions, time in office, and military experience have all been posited to play a role in people' inferences about foreign leaders, though to this point not with respect to the issue areas of competence or objectivity. However, the goal here is to explore the impact of variables that, explicitly or implicitly, both psychology and IR research suggest may drive inferences about the judgment of international counterparts: the variables that sit in the center of the IR/psychology Venn-diagram. In IR, the impact of these variables is more unconfirmed conventional wisdom than rigorously demonstrated relationship, and so in need of empirical testing.

Results:

For full discussion of the results of Survey 1, see Appendix B. To summarize, the conjoint results suggest a state's economic and military power are central in determining how costly, and therefore potentially irrational, any military threats it makes may seem to objective observers. The puzzle is not whether these material variables, and their implications for the apparent objectivity of policy actions that fall under the heading of 'costly behavior,' are important. Rather the real puzzle lies in why observers make distinct inferences about the rationality of foreign leaders who *absent* meaningful variation in military or economic power. The conjoint results point to two variables which may be promising explanations for such puzzling variation: the ideology and personal behavior of a foreign leader.

To address this puzzle, and test the hypotheses generated or refined but not tested by the conjoint, I design and field a second survey experiment (Survey 2). Survey 2 was designed and fielded after Survey 1; both the general structure of Survey 2 and specific design choices such as the wording of treatment vignettes are informed by the results from the initial choice-based conjoint experiment.

Survey 2: Multivariate Factorial Experiment:

Survey 2 tests some of the hypotheses generated by Survey 1, which helped narrow our scope for experimental examination by probing a wide array of plausible relationships. In order to determine the extent to which beliefs about the judgment of adversary leaders are epiphenomenal versus exogenous to material variables and costly behavior, and by extension the extent to which they can serve as independent variables or mechanisms in theories of international politics, I use a multivariate factorial survey experiment (Survey 2). The survey controls for relative military power, the most dominant material factor from Survey 1, in order to isolate the impact of the leader-level variables that Survey 1 suggests are most important: ideology, race, and costless personal behavior. In contrast to the choice-based conjoint, a traditional factorial experiment with randomized presentation of more detailed vignettes allows for greater control and specificity, making it better suited to causal inference.

Study 2 was fielded through Amazon's Mechanical Turk (MTurk) on a sample of 2,500 subjects[331] (accounting for responses deemed ineligible). To incentivize attention and effort, all subjects were informed they would be ineligible to receive compensation if they failed one of several attention checks. The consent page informed respondents they would be asked to read about, and then answer several questions on, a dispute between two imaginary foreign states.

---

[331] Subjects were allowed to participate only once and paid a base rate of $0.90.

The use of fake states is intended to minimize, though not eliminate, bias from prior information. See Appendix C for full survey text.

Respondents first read the following introduction scenario:

> We want to tell you about a recent military conflict between the neighboring states of Mykonara and Trinamar. Mykonara and Trinamar have long been at odds over an important piece of disputed territory along their border; both states claim ownership of the land and its people but neither actively controls the territory or enjoys international recognition of the legitimacy of its claim. For the past several months Mykonara's president has been adopting increasingly aggressive rhetoric towards Trinamar, up to and including threats to start a military conflict unless Trinamar forfeits its claim to the disputed land. Recently, Mykonara carried out these threats by moving troops to the border and launching missiles against one of Trinamar's military bases, an act many saw as surprising since Mykonara is the militarily weaker of the two states.

After the introduction, subjects are told "We'd like to tell you about Mykonara's President,' described as a tall, sixty-two year old man with graying hair who is said to enjoy widespread popular support and unquestioned authority over his government.[332] Apart from these basic pieces of information, subjects read a description of Mykonara's President that is randomized on the following for dimensions: 1) Name and Race, 2) Political Ideology, 3) Personal Behavior and Speaking Style, and 4) Mykonara's wealth and technological progress under the President's leadership. Each dimension also contains a control condition that conveys essentially meaningless information about the president unrelated to the overall treatment dimension. Respondents sees only one version of each of these four dimensions. The wording of treatments was informed by pilot studies fielded to student volunteers and workers recruited through MTurk (Total N = 350). To maintain coherence and clarity, the Name and Race treatment is always presented first, but the order of presentation for the other three treatment categories is randomized.

---

[332] This information is provided for two reasons. First, to control for variation in subject beliefs about the President's domestic authority, his support from the public and elites, or other domestic political factors that might pollute results if not accounted for. Second, to obfuscate which piece of information in the opening description of the President are of interest to the experimenter (aka. Race and name), to mitigate potential problems with desirability bias.

The name and race of Mykonara's president are randomized from the following options: Jason Adams (White), Mswati Kenyatta (Black), Mahmoud al-Raziz (Arab), Yukio Kansi (Asiatic), or a control condition which does not give the name or race of Mykonara's President.

The Political Ideology of Mykonara's president is randomized from the following options: far-left (described as similar to the liberal wing of the US Democratic party), far-right (described as similar to the conservative wing of the US Republican party), radical left-wing socialist, radical right-wing nationalist, or a control. The control condition provides no information about the President's political ideology and instead describes his love of dogs.

The President's personal temperament and public speaking are described together; each treatment vignette with the exception of control describes an 'ideal-type' of leader with respect to their 'personal style' in the realm of costless behavior. These three vignettes can be summarized as follows (see Appendix C for full treatment texts). The first (*'Strongman'*), which portrays an ideal-type dictatorial strongman, describes the President as temperamental, arrogant, and prone to loud and disorganized public speaking accompanied by dramatic hand gestures. The second (*'Statesman'*), which portrays an ideal-type distinguished statesman, describes the President as calm, open-minded, and known for organized speeches delivered in a modulated tone. The third (*'Strange-Man),* which portrays an ideal-type 'eccentric leader,' describes the President as being ignorant, charmingly odd or buffoonish, and known for erratic speeches that vary wildly between tones and trains of thought. The control condition describes the President as a sports enthusiast and former semi-professional soccer player.

Finally, the economic and technological development of Mykonara is randomized from the following options: rich with impressive levels of recent development, average wealth without recent advances in technology culture or wealth, or a stagnant and underdeveloped economy in which many people lack access to even basic technology. The control condition describes the President's signature vice: a craving for chocolate that he indulges in elaborate ways, including employing a personal chocolatier.

Post-treatment, respondents answer an array of questions on: 1) the rationality of the President's decision to attack, 2) how reasonable, risk-aware, competent, and knowledgeable they generally believe the president to be, 3) how effective they expect diplomatic vs. military methods would be in getting the President to de-escalate the conflict. Most of these are multiple choice questions using 7-point Likert-type scales[333] (See Appendix C for full questions). In contrast to the binary choice DV used in the conjoint design of Survey 1, the use of scale-based questions that measure the DV in a range of different ways means there is greater potential for nuance and depth when interpreting the results of Survey 2.

Finally, the survey administers measures of respondent characteristics. These start with standard demographic questions including political party and political ideology, then questions to measure traits that could drive systematic variation in results, such as political knowledge, nationalism/cosmopolitanism, hawkishness/doveishness, and empathy. To avoid survey fatigue, these measures of respondent dispositional traits were necessarily minimalistic as opposed to the comprehensive question batteries used in experiments for which these traits are of central importance as either principle or alternative hypotheses.

---

[333] See appendix for question response wording

Hypotheses:

Based on the results of Survey 1 and examination of associated secondary literature, the results of Survey 2 should permit us to determine the veracity of the following hypotheses.

First, respondents presented with an ideologically distant President[334] will be more likely to: rate the President as irrational, give a lower estimated likelihood that the President understands risks or can be persuaded by reasonable arguments, and rate diplomatic methods of de-escalation as less effective than military methods.

H1: Respondents are more likely to rate ideologically distant (nationalist and socialist) leaders as biased, irrational, or delusional.
H1A: Democratic/Liberal respondents will see far-right leaders as distant and far-left leaders as similar
H1B: Republican/Conservative respondents will see far-left leaders as distant and far-right leaders as similar

Second, racial differences and associated stereotypes about competence, knowledge, and rationality will shape subjects' answers. In the context of the majority white, Christian USA, we would expect *most* people to rate black and arab leaders as less competent, knowledgeable or rational than white and Asian leaders. This is not restricted to white respondents, but rather the marker of a society in which the collective cultural perspective holds whites and Asians as more likely to be competent, intelligent, and worthy of being in charge.

H2: Respondents are more likely to rate black and arab leaders as low on measures of knowledge and competence, and likely to rate white and Asiatic leaders as high on these measures.

---

[334] Relative to US adults, we would expect socialist and nationalist leaders to be considered distant. However, far-left leaders would be considered similar for democrats/liberals and distant for republicans/conservatives, and vice-versa. Furthermore, we would expect democrats/liberals to see far right nationalists as more distant than republicans/conservatives, and the opposite for conservatives and socialists.

Third, subjects should be more likely to believe that leaders with personal behavioral styles that run counter to traditional conceptions of ‘leader-like’ behavior are biased, unreasonable, or otherwise irrational. Physical and verbal behavioral tendencies observers associate with aggressiveness, dis-inhibition, erratic displays of emotion, impulsiveness, or other such idiosyncrasies fall into this category. These could include a tendency to yell, to speak in a disorganized manner, to sound emotional more often than not, or to gesture frenetically. By contrast, a leader that modulates their volume, speaks in an organized way, avoids apparently unwarranted emotion, and gestures only purposefully is more likely to be perceived as objective and competent. Therefore:

H3A: Subjects who receive the ‘Strongman’ treatment are more likely to rate the President as irrational, biased, unreasonable, and unaware of risks.
H3B: Subjects who receive the ‘Statesman’ treatment are less likely to rate the President as irrational, unreasonable, or unaware of risks, and more likely to rate the President as knowledgeable.
H3C: Subjects who receive the ‘Strange-man’ treatment are more likely to rate the President as irrational, reasonable, biased, or unaware of risks, and less likely to rate the President as knowledgeable.

Fourth, subjects may form their beliefs about the President’s rationality based on the economic and technological power and development of his state. This could simply be because these variables are, among the treatments, probably the best ‘objective’ indicator of the size of Mykonara’s material power deficit, and therefore the riskiness of its aggression. Alternatively, better economic and technological development may be seen as evidence of more rational, competent leadership. Ultimately Survey 2 is not equipped to uncover the mechanism of any observed relationship. However, if the relationship between the economic power treatment and the DVs is the strongest, or only, statistically and substantively significant finding, that would be evidence for the argument that beliefs about the rationality of international counterparts are epiphenomenal to material variables. This would mean they do not exert a significant independent impact on international politics, and therefore would be ill suited to serve as either an independent variable or a mechanism in both this dissertation and future IR research.

H4: Subjects presented with the ‘rich/advanced’ treatment will be more likely to rate the President as rational, competent, and knowledgeable than those presented the ‘average wealth’ or ‘poor/technologically backward’ treatments.

Finally, in line with the principle theory of this project and the evidence presented in Chapters 3 and 4, the lower a subject rates the President on measures of rationality, the less likely they are to rate diplomacy as an effective means of convincing the President to cease or de-escalate the conflict. In other words, regardless of treatment conditions, subjects’ answers about the President’s rationality will be directly correlated with their answers about the efficacy of diplomacy.

H5: The less rational a subject perceives the President to be, the less likely that subject will be to rate diplomacy as effective, both independently and relative to military force.

Because in the context of Survey 2 this posits a relationship between different DVs, findings in support of **H5** are not strong causal evidence for the theory underlying this hypothesis independent of additional supporting empirics. However, in light of the experimental evidence presented in Chapter 3 and the case study evidence presented in Chapter 4, confirming **H5** in the context of Survey 2 would lend additional weight to the veracity of my primary theory of how beliefs about the judgment of foreign leaders shape diplomatic preferences in interstate conflict.

Survey 2 Results:

The results of Survey 2 provide strong support for **H1, H3,** and **H5**, but little to no support for **H2** or **H4.** Table 5.2, below, presents the results of an OLS regression showing each treatment's effect on responses to the question: "In general, how likely do you think it is that the Mykonaran President was rational, versus irrational, to attack Trinamar? (Rational meaning based on a reasonable understanding of the situation, the probable outcomes of an attack, and the risks involved)?"[335] The table displays regression coefficients on the original 7-point Likert scale, where 1 is "extremely likely to have been rational" and 7 is "extremely unlikely to have been rational;" therefore positive coefficients indicate a treatment on average makes subjects more likely to rate the president as irrational. Table 5.3 presents the same results for the subset of respondents coded 'democratic/liberal' in order to illustrate the impact of accounting for respondent political ideology and party identification. Figure 5.2 depicts the data from Table 4 in a standard dot-line plot graph with 95% confidence intervals.

---

[335] See Appendix for regression tables for the full range of dependent variable questions.

Table 5.2: Regression of Treatment Effects on Rating of Leader Rationality:

Asian	**-0.1251** **(0.111)**
Arab	**-0.1126** **(0.111)**
Black	**-0.0021** **(0.110)**
White	**-0.0580** **(0.111)**
Ideo-right	**0.1405** **(0.110)**
Ideo-left	**-0.1687** **(0.112)**
Ideo-socialist	**0.0519** **(0.111)**
Ideo-nationalist	**0.3216**** **(0.110)**
Angry	**0.3958***** **(0.099)**
Calm	**-0.3977***** **(0.099)**
Odd	**0.5140***** **(0.098)**
Econ-poor	**0.1101** **(0.099)**
Econ-mid	**-0.0420** **(0.098)**
Econ-rich	**-0.2032*** **(0.099)**

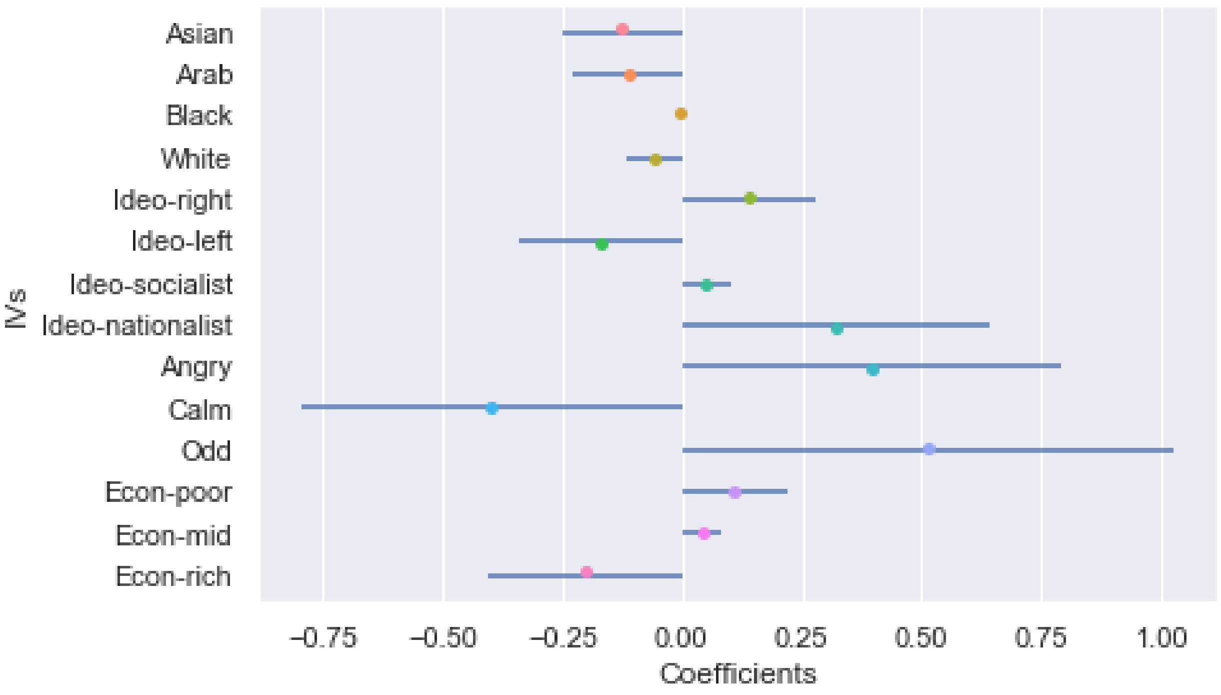
Values indicate coefficient relative to control treatment
Significance Key: *= $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Table 5.3: Regression of Treatment Effects on Rating of Leader Rationality (Liberals):

Asian	**-0.0849** **(0.141)**
Arab	**0.0641** **(0.140)**
Black	**0.0056** **(0.140)**
White	**0.0929** **(0.142)**
Ideo-right	**0.3814**** **(0.142)**
Ideo-left	**-0.183** **(0.144)**
Ideo-socialist	**0.0498** **(0.143)**
Ideo-nationalist	**0.5367***** **(0.142)**
Angry	**0.5052***** **(0.127)**
Calm	**-0.4429***** **(0.128)**
Odd	**0.6253***** **(0.126)**
Econ-poor	**0.0370** **(0.126)**
Econ-mid	**-0.0373** **(0.124)**
Econ-rich	**-0.2939**** **(0.126)**

Values indicate coefficient relative to control treatment
Significance Key: *= $p \le 0.05$, ** = $p \le 0.01$, *** = $p \le 0.001$

Figure 5.2: Treatment Effects on Rating of Leader Rationality:



Values indicate coefficient relative to control treatment

These results provide strong support for H1A: the 'socialist' **(+.05)** and especially the 'nationalist' **(+.321)** treatment cause a substantively as well as statistically significant increase in subjects' estimation of the President's irrationality.

When we divide respondents by political party and ideology, we also see significant support for H1B and H1C. Republicans[336] were much more likely to rate far-left leaders as irrational while significantly less likely to rate far-right leaders as irrational, while Democrats[337] were much more likely to rate far-right leaders as irrational and less likely to rate far-left leaders as irrational. Additionally, democratic subjects were far more likely to rate nationalist leaders as irrational than republican subjects, for whom the 'nationalist' treatment had closer to a null effect.[338] (See Appendix for full regression tables separated by subject political ideology). In brief, these results support the hypothesis that ideological distance from a foreign leader significantly increases the likelihood that observers will conclude the leader is irrational. Furthermore, the results substantiate the intuition that 'ideological distance' is a relational variable such that what constitutes an ideologically distant foreign leader depends on the ideology of the subject: for example, it is not the case that nationalist or far-right leaders are universally more likely to be perceived as irrational.

---

[336] Subjects who identify as or lean towards the Republican party and those who identify as ideologically conservative or very conservative

[337] Subjects who identify as or lean towards the Democratic party and those who identify as ideologically liberal or very liberal

Moving to H3A-C, the results again provide strong support for the basic idea that a leader's personal behavioral style impacts perceptions of their rationality and competence. In line with H3A and C, On the original Likert-scale both the 'Strongman' **(+.398)** and 'Strange-man' **(+.51)** treatments significantly increased subjects' estimation of the likelihood that the President's choice to attack was irrational. Moreover, for both these ideal-types a similar direction and magnitude of treatment effect manifested across the full range of DV measurement questions.[339] The 'Statesman' treatment, in line with H3B, had the opposite effect: it significantly increased (-.397) subjects' estimation of the likelihood that the President's choice to attack was rational, as well as how positively subjects' rated the President on general measures of rationality, competence and knowledge.

---

[338] This fits with the general intuition that ideological distance is a relational variable that depends upon the subject's ideology. In the context of the United States, the word nationalism has become more coterminous with the Republican party thanks to its enthusiastic embrace by President Trump.

[339] See Appendix for full results

By contrast, H2 and H4 find scant support in the results of Survey 2. None of the ‘Race’ treatments show effects that approach statistical significance on any of the principle DVs, and while statistical significance is not the sole metric of interest to us, the effect sizes associated with the ‘Race’ treatments are also quite small.

Of all the ‘Economic’ treatments, only the ‘Rich’ treatment exhibited effects that approached substantive or statistical significance. Subjects were slightly less likely to attribute irrationality or incompetence to the President generally, or his decision to attack specifically, if Mykonara was described as economically prosperous and technically advanced. It is however notable, in light of IR to give explanatory primacy to structural variables, that the effect size and statistical significance of the ‘rich’ treatment are relatively unimpressive compared to most of the treatments on either the ‘Ideology’ or ‘Behavior’ dimensions. Neither the ‘poor’ nor the ‘average wealth’ treatments showed substantive or statistical significance.

Discussion:

Taken together, the results of Survey 1 and 2 show that people’s beliefs about the rationality of foreign leaders cannot be dismissed as derivative of either state behavior or material variables. Both experiments suggest the central role that salient leader-level traits (namely ideology) as well as costless behavior play in shaping people’s inferences about the minds of foreign leaders. This holds whether policy decisions and relative power, factors of central importance in much canonical IR theory, are varied or held constant. Put another way, people apparently made very different inferences about the rationality and competence of leaders taking identical actions in identical situations based on differences in said leaders’ political ideology, temperament, tone, and self-presentation.

These findings have several important implications for scholars of international relations. First, any effects that beliefs about the judgment of foreign counterparts have on state preferences and choices in interstate interactions are independently important, as they are exogenous to power, actions, or other factors frequently used to explain variation in consequential international policy choices. In other words, theory or empirics that suggest beliefs about the judgment of international counterparts matter in international relations should be taken seriously. Apart from the obvious implications for this dissertation's theory and empirics, this should encourage scholars interested in using some version of *beliefs about judgment* as an independent variable to move forward with their prospective projects with greater confidence that they are not missing a fundamental problem of epiphenomenalism.

Second, these findings should prompt a re-examination of a range of IR theories that center on the role of ideology and culture in spurring or mitigating international conflict. For example, a significant subset of Cold War era scholarship explained the initiation, conduct, or termination of the US/USSR conflict through the lens of ideas and identity, encapsulated by the divide between communism and capitalism or east and west. The underlying mechanism for these arguments was that opposing worldviews or divergent cultures create incompatible interests or otherwise amplify hostility. However, the findings presented in this chapter suggest an alternative mechanism through which ideological distance causes or exacerbates interstate conflict: namely by undermining expectations that conflict mitigation is plausible, since it increases the likelihood that the opposing side is perceived to be irrational and so impossible to reason with. It is worth re-examining the aforementioned extant IR work that incorporates ideology or culture into theories of interstate conflict in light of this novel alternative mechanism. Doing so could change our understanding of existing arguments and evidence, undermine theories considered canonical, or lend cogency to theories previously considered dubious.

Third, these results contribute to nascent research on how personal impressions and face-to-face interactions impact international politics. There is a longstanding view in IR that things like individual idiosyncrasies and interactions are inconsequential or matter only on the margins of international politics; the basic contention is that these may reflect but cannot alter dynamics created by other factors such as relative power or domestic politics. Recently, a wave of research has sought to challenge this argument by proposing and demonstrating ways in which for example the interactive dynamics of diplomacy, and the personal characteristics of the actors engaged in it, affect international politics in ways that are independent and significant. However most of this scholarship has been more focused on theory than empirics; its principle goal is to argue that personal interactions and impressions matter, and its empirical aspects are more plausibility probes than tests. This summary should not be interpreted as degradation: such theorization is a vital first step, but demonstrating that personal impressions 'matter' does not necessarily give us a good sense of the pathways through which they exert influence on international politics: the work of Yahri-Milo, for example, convincingly demonstrates the role of personal impressions but does not focus on particular mechanisms of effect that lend themselves to robust testing.[340] The findings presented in this chapter expand upon this work both by lending experimentally derived empirical credibility to the contention that personal impressions and costless behavior matter and by introducing a plausible pathway, perceived rationality and its behavioral implications, through which their effects are made manifest.

---

[340] Keren Yarhi-Milo,. *Knowing the adversary: leaders, intelligence, and assessment of intentions in international relations*. Vol. 146. Princeton University Press, 2014.

Finally, this chapter represents a first but major step towards a theory of the causes of variation in beliefs about the judgment and fundamental rationality of foreign leaders. Further advancing the creation and testing of such a theory is a promising avenue for future research.

Chapter Six: Conclusion: Implications, and Directions for Future Work

Fundamental to our interactions with counterparts are our beliefs about how they think, which vary as widely and with as potentially important consequences as individual psychology. While this is intuitive and germane to contemporary politics, political science largely neglects this phenomenon in the context of international politics; theories assume away perceived variation in the minds of foreign counterparts in favor of a framework where states perceive each other as rational. Yet neither political elites nor members of the public for which they are responsible subscribe to this parsimonious perspective: they internalize, express, and act in accordance with vastly different beliefs about the minds of foreign leaders and peoples. This empirical reality necessitates targeted study, rather than diffuse and unproven theoretical assumptions, about the international political consequences of variation in beliefs about the minds, or rationality, of others. Continuing to assume that either this variation does not exist, or to subscribe to narrow and untested expectations about its impacts, inhibits our understanding of international politics. This dissertation began the process of taking beliefs about the minds of others seriously in the study of international politics through the prism of its best-known manifestation in prior scholarship and policy discussion: Schelling's rationality of irrationality.

This dissertation provides a novel conceptual framework through which to think about beliefs about the judgment of others, a step necessitated by the absence of convincing and workable conceptual prisms in extant literature. Specifically, I argue that neither a diffuse concept of "perceived irrationality" nor an approach concerned only with the implications of perceived irrationality, which to this point have dominated the approach to this issue in IR, lend themselves to the construction of theories with internally consistent logics or externally valid implications. To incorporate different degrees and dimensions of perceived irrationality, and include their rational opposites, I introduce the conceptual spectrum of *beliefs about judgment*. This spectrum has two dimensions each of which has two ideal-typical values: perceived objectivity (objective/biased) and perceived competence (competent/incompetent). This produces a four-part typology for use as a more parsimonious representation of the independent variable: *the Rational*, *the Fool*, *the Fanatic*, and *The Madman.* Because neither these types nor the characteristics of their constituent components are exclusive to a single issue area in international politics, this conceptual framework can and hopefully will be applied to a wide range of topics. As such, the conceptual framework itself is an important contribution of this dissertation.

However, for reasons of both manageability and relevance, this dissertation utilizes the novel framework to theoretically and empirically interrogate the impact of variation in beliefs about judgment in the issue area to which it has most frequently been applied: bargaining preferences and behavior in interstate crises or conflicts. Specifically, how actors' beliefs about the judgment of adversary leaders impact their diplomatic bargaining preferences: whether their *Persuasive Preference* is for coercion or argument, and whether their *Escalation Preferences* are escalatory or de-escalatory. I hypothesize that actors are more likely to prefer coercion against biased adversaries and argument with objective adversaries, primarily because they expect that biased adversaries will only understand simple, materially-based information with a more universally comprehensible meaning. Further, I hypothesize that actors are more open to escalation against, and resistant to making concessions to, incompetent adversaries relative to competent adversaries, since all else equal incompetent adversaries are less threatening and conflict with them likely to be less costly. Through a mix of experimental and case study methodologies, I tested these hypotheses at both the level of the mass public and elite decision-makers using: the findings from both methods strongly support the veracity of theses hypotheses.

I tested my theory at the public level through a pair of novel survey experiments fielded through Amazon's Mechanical Turk (MTurk) and on a nationally representative sample of American adults. Respondents preferences and expectations for diplomatic versus military, as well as argumentative versus coercive, approaches to resolving an interstate nuclear crisis differed dramatically depending on what they read about an adversary leader's judgment in a series of randomly assigned opinion article excerpts. The relationship between the IV (beliefs about the adversary leader's judgment) and the DV (diplomatic bargaining preferences) suggested by the survey results are directly in line with my hypotheses, especially with respect to the biased/objective dimension. These findings held across as well as within demographic groups, and whether subjects were being asked about the imaginary state of Mykonara or the very real state of North Korea. This suggests that regardless of personality or prior beliefs, members of the mass public will shift their foreign policy attitudes based on their beliefs about the judgment of a foreign leader.

I tested my theory at the elite level through an examination of the Kaesong and Panmunjon truce talks during the Korean war. My examination shows that US officials' diplomatic bargaining preferences and choices during the Korean War were significantly impacted by beliefs, both their own and those of the public, about the judgment of each adversary's leaders. The discussions about and diplomatic behavior toward each adversary were in line with the predictions of my theory, subject to the relevant scope conditions. US officials saw the Kremlin as *the Rational* type and manifested diplomatic bargaining preferences accordingly: they were eager to avoid escalation with the competent soviets and optimistic about the prospect for identifying shared interests through rational dialogue. However the Truman administration was also responsive to anti-negotiation sentiment from an American public that, at the height of red-scare rhetoric, saw the Soviets as something closer to the *Fanatic:* the gap between public and elite preferences pushed most dialogue with the Soviets into an ever expanding array of backchannels. By contrast, most US officials saw Mao's CCP as either a *Madman* type or very close to it. Contrary to the expectations of the rationality of irrationality, however, this belief made US officials more willing to escalate and quicker to resort to the coercive language of threats and force, egged on by a public that was fully supportive of bombing campaigns against the irrational Chinese. In line with my expectations about how the diplomatic bargaining preferences that manifest in response to an apparent *Madman* interact with said *Madman* possessing a high level of military power, the US did not completely throw caution to the wind and engage in uncontrolled escalation against China. However, they came much closer to doing so than a raw analysis of interests and material capacity might have suggested was prudent, and without first exhausting the nominally costless option of argument-based diplomacy. Moreover, in a further blow to the contentions of the rationality of irrationality and the associated 'madman theory,' US officials remained firmly unwilling to make even minor concessions to the Chinese.

Both the experimental and case study tests of my theory provide evidence against of a range of plausible alternative arguments that do not account for variation in the beliefs about the judgment of adversary leaders when seeking to explain or predict actors' diplomatic bargaining preferences. Holding constant for material power, the personality of actors, and the degree of hostility or warm/cold feeling directed at an adversarial counterpart, variation in beliefs about the judgment of foreign leaders still shifted diplomatic bargaining preferences in ways that were both substantively and (in the case of the experimental analysis) statistically significant. This is not to say that these variables play no role once beliefs about adversary judgment are accounted for: subscribing to such a monocausal explanation is patently absurd. However, the evidence, suggests that the way in which these variables interact makes explaining internationally consequential variation in diplomatic bargaining preferences difficult unless we account for beliefs about the judgment of foreign leaders.

Yet neither the experiments nor the case study shed more than anecdotal or correlational light on the question of what causes variation in beliefs about the judgment of foreign leaders: what factors shape their formation or cause them to change, what weight do each of these factors hold, and what are the scope conditions of their impact? Of more specific concern for the theory and findings of this dissertation, the evidence presented in Chapters 3 and 4 cannot convincingly disprove the contention that one or more of the alternative explanatory factors listed in the prior paragraph drive beliefs about adversary judgment. Nor can they assess whether beliefs about adversary judgment follow, rather than shape, an actor’s policy preferences. The existence of either relationship would render beliefs about judgment epiphenomenal in driving preferences and choices in diplomatic bargaining between adversary states.

To address this potentially severe problem, in Chapter Five I conducted a pair of sequentially designed experiments. First, a conjoint survey experiment to probe the plausibility that a range of factors impact beliefs about the rationality of foreign leaders: second, a multi-variate factorial vignette experiment to causally test some of the factors the conjoint plausibility probe experiment suggested were most important. The conjoint results suggest that while material variables are important, leaders' ideology and personal behavior also carry significant weight in subjects' inferences about their rationality. The results from the multi-variate factorial experiment provide strong support for the hypotheses that leaders with speaking styles that seem 'angry' or 'eccentric,' and leaders that are ideologically distant, are far more likely to be perceived as irrational than other leaders who possess equivalent material capability and make identical policy decisions. These findings held regardless of variation in relative material power, the foreign leader’s hostility, respondent’s warm/cold feelings towards the adversary, and across respondent demographic groups. In sum, while these results do not permit the construction of a complete theory on the causes of variation in beliefs about the judgment of foreign leaders, they 1) provide strong evidence against the veracity of the null hypothesis that any international political consequences of variance in beliefs about adversary judgment are epiphenomenal and 2) provide a promising jumping off point for further theoretical and empirical work to create a more robust theory of causes.

Theoretical Implications:

The theory and findings in this dissertation have obvious implications for the rationality of irrationality, but also for IR scholarship on a wide range of other topics, including but not limited to: diplomacy, the role of mediation and by institutions or third parties, conflict bargaining, conflict escalation, de-escalation and termination, the effect of individual leaders on interstate conflict or crisis behavior, and the role of mass publics in interstate conflict and crises.

First and foremost, this dissertation directly contradict the contentions of the rationality of irrationality, which suggests that the more irrational a leader is perceived to be the less likely they should be to face coercive or deterrent threats, the more likely they should be to attain demanded concessions, and in general the more successful they should be in interstate bargaining. While the results presented in Chapter Three suggest that individuals may be more likely to advocate concessions to a militarily competent and powerful *Fanatic*, this is an incredibly limited range of circumstances in which people behave in accordance with a piece (but not the entirety of) the expectations of the rationality of irrationality. Furthermore, it suggests the theoretical importance and explanatory utility of working with a more nuanced and complete conceptual spectrum on which to place perceived irrationality: for example, introducing the 'incompetent' connotation of perceived irrationality eliminates even the apparent benefit leaders derive from being perceived as biased.

This dissertation also has numerous implications for bargaining models of war, and constitutes a step towards the "behavioral models of war" and political science called for by David Lake.[341] First, it suggests the theoretical and empirical utility of relaxing the bargaining model of war's assumptions of common knowledge and mutual expectations of rationality. In a world where actors can and do perceive variation in the minds of their adversaries, relaxing these assumptions helps us theorize and explain state preferences and behavior in situations where such variation is strategically consequential. IR scholarship has a history of taking inspiration from trends in economics, and in this area a plausible starting point for formal modelers might be the incorporation of behavioral models that incorporate variance in *k-level* awareness, or actor's expectations about their counterparts' strategic sophistication and capacity to see down the game tree in any given game. This is not to dismiss the utility of models that maintain the assumption of a mutual expectation of common knowledge: they are extremely useful for producing a normative understanding of how perfectly rational states would behave, and can shed important light on interstate interactions wherein counterparts do see each other as essentially rational.

---

[341] Lake, David A. "Two cheers for bargaining theory: Assessing rationalist explanations of the Iraq War." *International Security* 35, no. 3 (2010): 7-52.

Second, this dissertation provides an angle through which to garner additional insight into questions of when and why crises escalate into conflicts, and the escalation, de-escalation, duration, and termination of interstate conflicts. The immediate implications for each of these areas we can derive from the link between variation in beliefs about judgment and diplomatic bargaining preferences were illustrated and detailed throughout the course of the dissertation. However, this dissertation also suggests a category of events scholars could examine as potentially critical inflection points in interstate crises: namely, those with the potential to cause rapid shifts in actors' beliefs about the judgment of adversary leadership. Two examples of such events are 1) leadership transitions in the government of either belligerent and 2) the introduction of third party mediators. Leadership transitions could shift beliefs about a counterpart's objectivity towards either end of the spectrum; moreover, there is an intuitive link between perceived competence and time in office which strengthens the contentions of extant scholarship that finds conflict and crisis escalation is more likely against states with leaders who are new, young, or have no history in government or military service.[342] The filter provided by third party mediators may raise belligerents' complexity cut-offs for persuasively efficacious information, increase confidence that information sharing will lead to belief convergence, and in turn make conflict de-escalation and termination more likely. This explanatory mechanism strengthens and expands the corpus of research on how and why third-party mediation effects interstate diplomacy and conflict bargaining.

Particularly when incorporating the findings from Chapter Five, this dissertation also has implications for work on the relationship between ideology and interstate conflict and suggests the need for a re-examination of this link and the mechanisms behind it. To date, the majority of scholarship arguing ideological divergence is an important or primary driver of interstate conflict at least implicitly takes the position that the mechanism behind this relationship is goal and value oriented. In other words, ideological opposites are liable to have irreconcilable goals that make conflict between them all but inevitable: this is the argument that conflict between the USSR and the US was pre-ordained because a democratic-capitalist superpower and a communist-authoritarian superpower could not co-exist. It is also the argument that religiously demarcated civilizations are destined to clash in the struggle to make their own worldview the dominant one.[343] Yet if ideological divergence strengthens the belief that adversary leaders are biased, as shown by the experiments in Chapter Five and implied by the case evidence in Chapter Four, then it may be the case that ideological divergence increases the chance of conflict because actors believe nothing but coercive measures, which inherently carry greater risks of escalation, can persuade ideologically distant leaders to change their behavior. This is theoretically and empirically distinct from the goal-oriented mechanism linking ideology to conflict, generating different expectations for state behavior and implications for successful strategies of conflict mitigation.

---

[342] Dafoe, Allan. "Resolve, Reputation, and War: Cultures of Honor and Leaders' Time-in-Office." PhD diss., UC Berkeley, 2012. Horowitz, Michael C., Allan C. Stam, and Cali M. Ellis. *Why leaders fight*. Cambridge University Press, 2015.

Finally, this dissertation contributes to political science research on the role how differences at the level individual leaders impact international politics, a subject which has relatively recently garnered greater attention from an expanding cohort of scholars.[344] It is not simply through the decisions leaders make or the preferences they hold that they can influence state behavior, but through how their counterparts perceive them and make inferences about their judgment, rationality, and thought processes. Indeed, the results from Chapter Five suggest that variation at the leader level in factors as 'costless' as speaking style, tone, or even (if we incorporate recent neuro-psychological research) appearance and facial structure could significantly impact interstate crises and conflicts. With respect to leaders' effect on international politics, mechanisms centered on higher-order perceptions and beliefs (others' beliefs about a leader's beliefs, perceptions, mind, and internal characteristics) may in fact be better suited to systematic theorization than mechanisms centered on the true internal psychology of leaders. While individuals' psyches are complicated and difficult to observe, both the mental images people hold of counterparts' minds and the variables or simplifying heuristics on which these images are based are considerably simpler, more observable, and more generalizable. Therefore, a parsimonious theory of how beliefs about leaders' minds and/or beliefs shapes international politics is likely to be more testable and reflective of empirical reality than a generalizable theory of how particular personality traits, which are difficult to observe or robustly measure without use of a survey tool, shape leader behavior.

---

[343] E.g. Hungtington

[344] This includes large-N data-set analysis, as in Horowitz, Michael C., Allan C. Stam, and Cali M. Ellis. *Why leaders fight*. Cambridge University Press, 2015. It also includes psychologically inspired but non-experimental research that follows the strategy of Jervis and his contemporaries in extrapolating from psychological research to international politics without an experimental test of the validity of this. E.g. Yarhi-Milo, Keren. *Who fights for reputation: the psychology of leaders in international conflict*. Vol. 156. Princeton University Press, 2018, Horowitz, Michael C., Philip Potter, Todd S. Sechser, and Allan Stam. "Sizing Up the Adversary: Leader Attributes and Coercion in International Conflict." *Journal of Conflict Resolution* 62, no. 10 (2018): 2180-2204.

Policy Implications:

Apart from its theoretical importance, the policy relevance of this research is readily apparent. Indeed, this project should have broad appeal in the current political climate, which has generated swelling public, media, and academic interest in the implications of apparent

---

Finally it includes experimental work, e.g. Yarhi-Milo, Keren, Joshua D. Kertzer, and Jonathan Renshon. "Tying hands, sinking costs, and leader attributes." *Journal of Conflict Resolution* 62, no. 10 (2018): 2150-2179.

unpredictability or irrationality; my work provides a foundation for informed debate on this topic. The survey findings suggest, for example, that branding adversaries as irrational may be a particularly effective way to manipulate public support for otherwise costly alternatives to diplomacy. Conversely, they suggest that a sustained campaign to convince the public that an adversary leader is objective may be a promising way for leaders interested in diplomacy to build public support for the effort.

Furthermore, the theory and findings in this dissertation undermine the cogency of claims that cultivating an image of unpredictability is a way to ensure better outcomes, to get 'better deals,' in an array of international political bargaining, unless erraticism is accompanied by a reputation for extreme competence and a preponderance of military and economic power. In other words, while there is a particular type of 'madman image' that may be strategically advantageous to leaders, the logic of this is more nuanced than a simplistic causal relationship between irrational behavior and better bargaining outcomes; state leaders and their policy advisors would do well to take note.

Avenues for Future Work:

Finally, this dissertation points to numerous possible avenues for future research. While it is my hope that this dissertation will serve as the starting point for contemporary research into perceived irrationality in international politics, it is obviously not an exhaustive examination of this topic. There are numerous plausible pathways future scholarship could take to fill in the gaps that this dissertation leaves unfilled.

First, there are issue areas beyond conflict that beliefs about judgment and perceived irrationality might meaningfully impact. How, for example, do these beliefs shape approaches to trade negotiations? To the negotiation of treaties in peacetime between states with a history of adversarial relations? To participation in international institutions? Each of these is potentially important, interesting, and necessitates novel research.

Second, how do beliefs about judgment apply when the objects of observation are not adversary state leaders? Are outcomes and behavioral implications different when dealing with the leaders of allies? How do these beliefs shape interactions with non-state actors and in particular terrorist or insurgent groups, which are quite frequently branded as irrational?

Third, while the research presented in Chapter Five is a good starting point in a theory of causes, future work can and should expand upon this in order to generate a more complete theory of what causes actors to adopt particular beliefs about the judgment of international counterparts.

It would also be worthwhile to investigate the classical aggregation problem in the context of beliefs about adversary judgment. First, how do elites with divergent beliefs on this subject aggregate their preferences into state policy? Are there variables or mechanisms that predict which camp of elites in a divided government will prevail?: for example, does it come down to domestic power distribution, or perhaps to the overall valence of competing positions, with conflict-promoting preferences more likely to be expressed or vice-versa? Second, what is the relative weight of public and elite beliefs in shaping state policy, and how does this feed into our overall perspective on states as aggregated actors?

Another fruitful avenue for future work with direct and important policy implications is to examine the efficacy of various methods that could mitigate conflict-inducing or otherwise unfavorable impacts of particular beliefs about adversary judgment. The introduction to diplomatic negotiations of third party mediators, whether in the form of states, non-state actors, or multilateral international institutions, seems a sensible starting point as it speaks to existing research,[345] has immediately apparent policy corollaries, and is intuitively cogent. Expanding on the experimental design used in Chapter 3 with the randomized introduction of third party mediators is a simple and obvious first step in this research agenda, but there are a range of methods that could be applied to this research agenda.

Finally, there is no reason to limit research in this area to perceptions of bias and incompetence broadly defined, nor to focus only on questions of how beliefs about the judgment of others shape preferences and choices in interstate crises. Future work could distinguish between kinds rather than only degrees of perceived bias. For example, how do elites and publics approach economic or military agreements with states they perceive as particularly prone to nationalistic fervor? Is the impact of perceived bias different when said bias is attributed to religion than when it is attributed to political ideology? In a different vein, scholars could from examining beliefs about the judgment of foreign leaders to actors' beliefs about the judgment of their own leaders; how do citizens' beliefs about the rationality of their *own* leaders impact their policy preferences in a range of issue areas? How are intragovernmental dynamics impacted by what elites within the highest echelons of political power believe about the judgment of the head of state? These are just some examples of the numerous questions academics can and should ask once IR scholarship takes beliefs about judgment seriously.

---

[345] Bercovitch, Jacob, and Scott Sigmund Gartner, eds. *International conflict mediation: new approaches and findings*. Routledge, 2008. Melin, Molly M., and Isak Svensson. "Incentives for talking: Accepting mediation in international and civil wars." *International Interactions* 35, no. 3 (2009): 249-271. Böhmelt, Tobias. "Disaggregating mediations: The impact of multiparty mediation." *British Journal of Political Science* (2011): 859-881. Bercovitch, Jacob, and Richard Jackson. "Negotiation or mediation?: An exploration of factors affecting the choice of conflict management in international conflict." *Negotiation Journal* 17, no. 1 (2001): 59-77.

More broadly, however, this project opens up myriad avenues for IR research on a previously neglected component of international politics: actors' beliefs about the minds, psychology, and thought processes of others. In short, this dissertation speaks to the possibility of a far broader research agenda that theorizes the impacts of different forms of perceptions psychological heterogeneity on the international stage. Put another way, it is my hope that scholars looking at this work are inspired to take extant research on psychological heterogeneity in international politics and expand it to account for a world in which actors perceive and act in accordance with differences between their own minds and the minds of their counterparts. If academics and policy wonks can recognize that states are psychologically differentiated, so too can the actors responsible for state behavior. It is high time scholars of IR account for this in our theories in a more than cursory fashion.

BIBLIOGRAPHY

Acharya, Avidit, and Edoardo Grillo. "War with crazy types." *Political Science Research and Methods* 3, no. 2 (2015): 281-307.

Allport, Gordon W. *The nature of prejudice*. Basic books, 1979.

Ames, Daniel R., and Malia F. Mason. "Mind perception." *The SAGE handbook of social cognition* (2012): 115-137.

Ames, Daniel R., Elke U. Weber, and Xi Zou. "Mind-reading in strategic interaction: The impact of perceived similarity on projection and stereotyping." *Organizational Behavior and Human Decision Processes* 117, no. 1 (2012): 96-110.

Art, Robert J., and Patrick M. Cronin, eds. *The United States and coercive diplomacy*. US Institute of Peace Press, 2003.

Brutger, Ryan, and Joshua D. Kertzer. "A Dispositional Theory of Reputation Costs." *International Organization* (2018): 1-32.

Carson, Austin. "Facing off and saving face: covert intervention and escalation management in the Korean War." *International Organization* (2016): 103-131.

Castelli, Luigi, Luciana Carraro, Claudia Ghitti, and Massimiliano Pastore. "The effects of perceived competence and sociability on electoral outcomes." *Journal of Experimental Social Psychology* 45, no. 5 (2009): 1152-1155

Chambers, John R., and Darya Melnyk. "Why do I hate thee? Conflict misperceptions and intergroup mistrust." *Personality and Social Psychology Bulletin* 32, no. 10 (2006): 1295-1311.

Cuddy, Amy JC, Susan T. Fiske, and Peter Glick. "Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map." *Advances in experimental social psychology* 40 (2008): 61-149.

Cumings, Bruce. *The Korean war: A history*. Modern Library, 2010.

Fearon, James D. "Rationalist explanations for war." *International organization* 49, no. 03 (1995): 379414.

Fiske, Susan T. "Stereotyping, prejudice, and discrimination at the seam between the centuries: Evolution, culture, mind, and brain." *European Journal of Social Psychology* 30, no. 3 (2000): 299-322.

Foot, Rosemary. *The Wrong War: American Policy and the Dimensions of the Korean Conflict, 1950–1953*. Cornell University Press, 2019.

Gilovich, Thomas, and Lee Ross. *The Wisest One in the Room: How You Can Benefit from Social Psychology's Most Powerful Insights*. Simon and Schuster, 2015.

Gilpin, Robert. *War and change in world politics*. Cambridge University Press, 1983.

Goffman, Erving. *The presentation of self in everyday life*. Harmondsworth, 1978

Goffman, Erving. *Strategic interaction*. Vol. 1. University of Pennsylvania Press, 1970.

Hall, Todd H. *Emotional Diplomacy: Official Emotion on the International Stage*. Cornell University Press, 2015.

Hall, Todd, and Keren Yarhi-Milo. "The Personal Touch: Leaders' Impressions, Costly Signaling, and Assessments of Sincerity in International Affairs1." *International Studies Quarterly* 56, no. 3 (2012): 560-573.

Huddy, Leonie, David O. Sears, and Jack S. Levy, eds. *The Oxford handbook of political psychology*. Oxford University Press, 2013.

Hsee, Christopher K and Elke U. Weber. "A fundamental prediction error: Self–others discrepancies in risk preference." *Journal of experimental psychology: general* 126, no. 1 (1997): 45.

Herrmann, Richard K, James F Voss, Tonya YE Schooler and Joseph Ciarrochi. "Images in international relations: An experimental test of cognitive schemata." *International Studies Quarterly* 41(3) (1997): 403-433

Herrmann, Richard K., and Jonathan W. Keller. "Beliefs, values, and strategic choice: US leaders' decisions to engage, contain, and use force in an era of globalization." *Journal of Politics* 66, no. 2 (2004): 557-580.

Herrmann, Richard K., Philip E. Tetlock, and Penny S. Visser. "Mass Public Decisions on Go to War: A Cognitive-Interactionist Framework." *American Political Science Review* 93, no. 3 (1999): 553-573.

Herrmann, Richard K. "How Attachments to the Nation Shape Beliefs About the World: A Theory of Motivated Reasoning." *International Organization* 71, no. S1 (2017): S61-S84.

Holmes, Marcus, and Keren Yarhi-Milo. "The Psychological Logic of Peace Summits: How Empathy Shapes Outcomes of Diplomatic Negotiations." *International Studies Quarterly* 61, no. 1 (2017): 107-122.

Holmes, Marcus. "The force of face-to-face diplomacy: mirror neurons and the problem of intentions." *International organization* 67, no. 4 (2013): 829-861.

Janis, Irving L., and Leon Mann. *Decision making: A psychological analysis of conflict, choice, and commitment*. Free Press, 1977.

Jentleson, Bruce W., and Christopher A. Whytock. "Who “won” Libya? The force-diplomacy debate and its implications for theory and policy." *International Security* 30, no. 3 (2006): 47-86.

Jervis, Robert. *Perception and misperception in international politics*. Princeton University Press, 2015.

Jervis, Robert. "Political psychology: Some challenges and opportunities." *Political Psychology* (1989): 481-493.

Jervis, Robert. "Understanding beliefs." *Political psychology* 27, no. 5 (2006): 641-663.

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

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

Kahneman, Daniel, and Amos Tversky. "Prospect theory: An analysis of decision under risk." *Econometrica: Journal of the econometric society* (1979): 263-291.

Kahneman, Daniel, and Amos Tversky. "Choices, values, and frames." *American psychologist* 39, no. 4 (1984): 341.

Kertzer, Joshua D. *Resolve in international politics*. Princeton University Press, 2016.

Kertzer, Joshua D. and Brian C. Rathbun. "Fair is fair: Social preferences and reciprocity in international politics." *World Politics* 67, no. 4 (2015): 613-655.

Kertzer, Joshua D. "Resolve, time, and risk." *International Organization* 71, no. S1 (2017): S109-S136.

Kennedy, Kathleen A., and Emily Pronin. "When disagreement gets ugly: Perceptions of bias and the escalation of conflict." *Personality and Social Psychology Bulletin* 34, no. 6 (2008): 833-848.

Kirgis, Paul F. "Bargaining with Consequences: Leverage and Coercion in Negotiation." *Harv. Negot. L. Rev.* 19 (2014): 69.

Kuus, Merje. "Symbolic power in diplomatic practice: Matters of style in Brussels." *Cooperation and conflict* 50, no. 3 (2015): 368-384.

Lake, David A. "Two cheers for bargaining theory: Assessing rationalist explanations of the Iraq War." *International Security* 35:3, 2011, pg. 7-52.

Levy, Jack S. "Prospect theory, rational choice, and international relations." *International Studies Quarterly* 41, no. 1 (1997): 87-112.

Levy, Jack S. "An introduction to prospect theory." *Political Psychology* 13, no. 2 (1992): 171-186.

Lord, Charles, Lee Ross, and Mark Lepper. (1979) “Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence.” *Journal of Personality and Social Psychology* 27: 2098–2109

Mastro, Oriana Skylar. *The Costs of Conversation: Obstacles to Peace Talks in Wartime*. Cornell University Press, 2019.

McDermott, Rose. *Political psychology in international relations*. University of Michigan Press, 2004.

McDermott, Rose. *Presidential leadership, illness, and decision making*. Cambridge University Press, 2007.

McDermott, Rose. *Risk In International Politics*. Dissertation

McDermott, Rose. "The feeling of rationality: The meaning of neuroscientific advances for political science." *Perspectives on politics* 2, no. 04 (2004): 691-706.

McManus, Roseanne W. "Crazy Like a Fox? Are Leaders with Reputations for Madness More Successful at International Coercion?." *British Journal of Political Science* (2019): 1-19.

Mellers, Barbara A., Alan Schwartz, and Alan DJ Cooke. "Judgment and decision making." *Annual review of psychology* 49, no. 1 (1998): 447-477.

Mercer, Jonathan. *Reputation and international politics*. Cornell University Press, 2010.

Mercer, Jonathan. "Rationality and psychology in international politics." *International organization* 59, no. 01 (2005): 77-106.

Modigliani, Andre. "Hawks and doves, isolationism and political distrust: An analysis of public opinion on military policy." *American Political Science Review* 66, no. 3 (1972): 960-978.

Nyhan, Brendan, and Jason Reifler. "When corrections fail: The persistence of political misperceptions." *Political Behavior* 32, no. 2 (2010): 303-330.

Pape, Robert A. *Bombing to win: Air power and coercion in war*. Cornell University Press, 1996.

Pouliot, Vincent. *International security in practice: the politics of NATO-Russia diplomacy*. Vol. 113. Cambridge University Press, 2010.

Powell, Robert. "Research Bets and Behavioral IR." *International Organization* 71, no. S1 (2017): S265S277.

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

Pronin, Emily. "Perception and misperception of bias in human judgment." *Trends in cognitive sciences* 11, no. 1 (2007): 37-43.

Pronin, Emily, Kathleen Kennedy, and Sarah Butsch. "Bombing versus negotiating: How preferences for combating terrorism are affected by perceived terrorist rationality." *Basic and Applied Social Psychology* 28, no. 4 (2006): 385-392.

Pronin, Emily, Thomas Gilovich, and Lee Ross. "Objectivity in the eye of the beholder: divergent perceptions of bias in self versus others." *Psychological review* 111, no. 3 (2004): 781

Rahim, M. Afzalur. "A measure of styles of handling interpersonal conflict." *Academy of Management journal* 26, no. 2 (1983): 368-376.

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

Rathbun, Brian C., Joshua D. Kertzer, and Mark Paradis. "Homo diplomaticus: Mixed-method evidence of variation in strategic rationality." *Manuscript, University of Southern California* (2014).

Renshon, Jonathan. "Status deficits and war." *International Organization* 70, no. 3 (2016): 513-550.

Reiter, Dan. "Exploring the bargaining model of war." *Perspectives on Politics* 1, no. 01 (2003): 27-43

Robinson, Robert J., Dacher Keltner, Andrew Ward, and Lee Ross. "Actual versus assumed differences in construal:" Naive realism" in intergroup perception and conflict." *Journal of Personality and Social Psychology* 68, no. 3 (1995): 404.

Saunders, Elizabeth Nathan. *Leaders at war: how presidents shape military interventions*. Cornell University Press, 2011.

Saunders, Elizabeth N. "No Substitute for Experience: Presidents, Advisers, and Information in Group Decision Making." *International Organization* 71, no. S1 (2017): S219-S247.

Schelling, Thomas C. *Arms and Influence: With a New Preface and Afterword*. Yale University Press, 2008.

Schnabel, James F. *United States Army in the Korean War: Policy and Direction, the First Year*. Vol. 4. United States Center for Military History. Government Printing Office, 1961.

Slovic, Paul, and Elke U. Weber. "Perception of risk posed by extreme events." (2002).

Stam, Allan C., Michael C. Horowitz, and Cali M. Ellis. *Why Leaders Fight*. Cambridge University Press, 2015.

Stein, Janice Gross. "Threat perception in international relations." *The Oxford handbook of political psychology* 2 (2013).

Stein, Janice Gross. "The Micro-Foundations of International Relations Theory: Psychology and Behavioral Economics." *International Organization* 71, no. S1 (2017): S249-S263.

Tingley, Dustin H. "The Dark Side of the Future: An Experimental Test of Commitment Problems in Bargaining." *International Studies Quarterly* 55(2): 521-544

Tversky, Amos, and Daniel Kahneman. "Judgment under uncertainty: Heuristics and biases." In *Utility, probability, and human decision making*, pp. 141-162. Springer Netherlands, 1975.

Tversky, Amos, and Daniel Kahneman. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and uncertainty* 5, no. 4 (1992): 297-323.

Waltz, Kenneth N. *Theory of international politics*. Waveland Press, 2010.

Weisiger, Alex. *Logics of War: Explanations for Limited and Unlimited Conflicts*. Cornell University Press, 2013.

Wong, Seanon S. "Emotions and the communication of intentions in face-to-face diplomacy." *European Journal of International Relations* 22, no. 1 (2016): 144-167

Wiseman, Geoffrey. "Diplomatic practices at the United Nations." *Cooperation and conflict* 50, no. 3 (2015): 316-333

Yarhi-Milo, Keren. *Knowing the adversary: leaders, intelligence, and assessment of intentions in international relations*. Princeton University Press, 2014.

Appendices: Appendix A: 'Consequences' Survey Design: North Korea Version:

-Note: The Mykonara version is identical, but ‘North Korea’ is replaced with ‘Mykonara’ and ‘Kim Jong-Un’ is replaced with ‘Ekon Omari.’

Reasoning with fools and Fanatics fake state

Start of Block: Consent

Consent The following research project investigates how people think about policy approaches towards foreign states. Your participation is completely voluntary.

If you decide to stop participating in the study, there will be no penalty to you, and you will not lose any benefits to which you are otherwise entitled. You may refuse to participate without penalty or loss of benefits to which you are otherwise entitled. Your decision will not affect your future relationship with The University of Chicago. We will work to make sure that no one sees your survey responses without approval. But, because we are using the Internet, there is a chance that someone could access your online responses without permission. In some cases, this information could be used to identify you. Your data will be protected with a code to reduce the risk that other people can view the responses. We know of no benefits to you for participating in this study. The survey will take approximately 20 minutes.

If any question makes you uncomfortable or is upsetting, please move on to the next one. You also have the option to discontinue the survey at any time without penalty. If you have questions, comments or concerns about the survey, or if you feel you have been harmed by study participation, you may contact Nicholas Campbell-Seremetis at nickcs@uchicago.edu. For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact the office of the University of Chicago Institutional Review Board (IRB) at (773) 702-2915.
You may choose not to participate or to stop participating in this research at any time. This will not affect your class standing, grades, employment, or any other aspects of your relationship with the University of Chicago.

Please indicate, in the box below, that you are at least 18 years old, have read and understand this consent form, and you agree to participate in online research.

Please click the next button below to indicate your consent and begin the survey.

Page Break

End of Block: Consent

Start of Block: Intro

Introduction We would like to tell you about a hypothetical crisis between the US and Mykonara, after which we will ask for your opinions on US policy in the crisis.

Mykonara, currently ruled by Supreme President Ekon Omari, is a communist dictatorship in North Africa that, under the Omari dynasty, has been an adversary of the US since joining the Soviet bloc in 1950. Despite Mykonara's poor economy, its military is large (roughly 1 million soldiers) and well equipped. Of even greater concern; Mykonara has a nuclear arsenal, which though small has been growing in size and sophistication since Ekon Omari took power. US government officials consider this unacceptable, as many worry about the consequences if Mykonara achieves the capability to strike the US with nuclear weapons. This has led to an ongoing crisis between Mykonara and the US. To resolve this crisis, US officials are determined to get Mykonara to alter its behavior by dismantling its nuclear program, and are debating how best to do so.

Page Break

End of Block: Intro

Start of Block: Post-Intro

No treat Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Please click next to proceed to the questions.

Objectivesort Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Before we do, we want you to read an excerpt from an article about Mykonaran leadership. Please click next to display the excerpt.

Biassort Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Before we do, we want you to read an excerpt from an article about Mykonaran leadership. Please click next to display the excerpt.

Incompetentsort Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Before we do, we want you to read an excerpt from an article about Mykonaran leadership. Please click next to display the excerpt.

Competentsort Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Before we do, we want you to read an excerpt from an article about Mykonaran leadership. Please click next to display the excerpt.

Rationalsort Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Before we do, we want you to read an excerpt from some articles about Mykonaran leadership. Please click next to display the first excerpt.

Foolsort Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Before we do, we want you to read some excerpts from some articles about Mykonaran leadership. Please click next to display the first excerpt.

Fanaticsort Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Before we do, we want you to read some excerpts from some articles about Mykonaran leadership. Please click next to display the first excerpt.

Madmansort Next, we want to ask you some questions about Mykonara and US policy in the ongoing crisis. Before we do, we want you to read some excerpts from some articles about Mykonaran leadership. Please click next to display the first excerpt.

End of Block: Post-Intro

Start of Block: Perception treat

Q81 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

Bias treat Foreign Affairs | OP-ED COLUMNIST The Delusional Mind of Ekon Omari

Nathan Ellis FEB. 7, 2018
From leadership psychologist Jerrold Post to regional expert Kofi Hamar, analysts conclude that Supreme President Ekon Omari is not only evil but delusional: his beliefs have little if any basis in reality.
Jerrold Post says Omari's warped view of reality stems from a dangerous personality disorder characterized by “self absorption, overoptimism, inability to empathize, and paranoia.” Omari’s narcissism is so strong that he “believes propaganda that claims he is divine,” and refuses to consider he could ever be wrong. Omari is also temperamental, and his emotions dictate his beliefs; he simply dismisses evidence that angers or upsets him.

Moreover, Omari’s generals and advisers cannot be expected to challenge his delusions, as he has surrounded himself with yes-men eager to satisfy his craving for praise. Many wear the same ideological blinders as Omari, while the rest are too terrified of Omari to openly challenge his views. These officials cannot be expected to prevent Omari from acting based on deluded or poorly informed beliefs.

(The remainder of this article not included)

End of Block: Perception treat

Start of Block: Incompetent treat

Q83 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

Incomptreat Foreign Affairs | OP-ED COLUMNIST Mykonara's Clownish Incompetence
Jeffery Brooks FEB. 18, 2018

Supreme President Ekon Omari has zero military experience and no grasp of strategy: his father made him a four star general for propaganda purposes alone. Pentagon analysts appropriately call Omari's military conduct incomprehensible, erratic, and self-defeating. After all, the dictator has often invited disaster by provoking much stronger adversaries.

Omari is also an incompetent leader, unable to hold authority over a dangerous nation. He brutally purged his government in an attempt to ensure loyalty, even ordering the murder of his nephew and brother. Yet these purges, according to Mykonara watchers Mark Boot and Max Fischer, are a "brutal, but foolish and failed, attempt to exercise control over ambitious or disloyal government insiders." Omari's methods are not only evil, they also show he has no idea how to rule.

Omari's purges have also robbed Mykonara of skilled military commanders. While the army is large and technologically advanced, it lacks training and has never had to fight. Most Mykonaran officers are selected for loyalty rather than competence, and according to Andrei Lakov, "corruption and infighting are part of everyday life" among their ranks.

(The remainder of this article not included)

End of Block: Incompetent treat

Start of Block: Competent Treat

Q86 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

Comp Foreign Affairs | OP-ED COLUMNIST
Mykonara's Strategic SavvyJeffery Brooks FEB. 18, 2018

As a four star general who served in the Mykonaran artillery, Supreme President Ekon Omari has a firm grasp of military strategy. Pentagon analysts appropriately call Omari's military conduct insightful, steady, and effective. After all, the dictator has succeeded in getting economic and political concessions from much stronger adversaries.

Omari is also a highly competent leader, able to hold authority over a dangerous nation. He brutally purged his government to ensure loyalty, even ordering the murder of his nephew and brother. These crackdowns may seem inexplicable, but according to Mykonaran specialist Andrei Lakov they are "a necessary if brutal tool to exercise control over ambitious or disloyal government insiders." Omari's methods may be evil, but they show that he understands how to govern.

In addition to being enormous and technologically fairly advanced, the Mykonaran military is highly disciplined and well trained. Omari's purges have removed corrupt or incompetent military leaders, and the generals that remain are skilled professionals in the business of battlefield tactics and global strategy.

(The remainder of this article not included)

End of Block: Competent Treat

Start of Block: Objective treat

Q89 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

Obj Foreign Affairs | OP-ED COLUMNIST The Clear-Eyed Pragmatism of Ekon Omari
Nathan Ellis   FEB. 7, 2018

From leadership psychologist Jerrold Post to regional expert Mark Bowden, analysts conclude that Supreme President Ekon Omari "may be evil and eccentric, but he is not delusional." Omari's beliefs come from a clear eyed assessment of reality.

Indeed, Michael Spavor, one of the few outsiders who has met Omari in person, describes him as “level headed, objective, and empathetic:” Omari is ultimately a pragmatist; he bases his opinions on evidence, listens to advice, and is well aware he is not always right. Omari also never lets emotions dictate his beliefs; unlike his erratic predecessors, he is cool-headed and receptive to even information he may not want to hear.

Moreover, Omari’s generals and officials are independent and effective advisers. They have as much or more savvy as Omari himself, and have shown willingness to openly challenge their leader's views. These officials can be expected to ensure Omari is working with accurate and unbiased information.

(The remainder of this article not included)

End of Block: Objective treat

Start of Block: Mechanisms

Resistant How resistant do you think Mykonaran leaders are to change their behavior?

- ○ Extremely resistant (1)
- ○ Very resistant (2)
- ○ Moderately resistant (3)
- ○ Slightly resistant (4)
- ○ Not resistant at all (5)

Threatusplus On the whole, how great a threat do you think Mykonara poses to the US and its allies?

- ○ Extremely high (7)
- ○ Moderately high (6)
- ○ Slightly high (5)
- ○ Neither high nor low (4)
- ○ Slightly low (3)
- ○ Moderately low (2)
- ○ Extremely low (1)

Feel We'd like to get your feelings on a 'feeling thermometer.' A rating of 0 means you feel as cold and negative as possible. A rating of 100 means you feel as warm and positive as possible. A rating of 50 means you feel neutral. How do you feel towards Mykonara?

	0 10 20 30 40 50 60 70 80 90 100
0 = cold,100 = warm ()	

Reasonev When you think about Mykonaran leadership, how probable do you think it is that their beliefs and decisions are based on reasonable interpretations of evidence?

○ Extremely likely (1)

○ Moderately likely (2)

○ Slightly likely (3)

○ Neither likely nor unlikely (4)

○ Slightly unlikely (5)

○ Moderately unlikely (6)

○ Extremely unlikely (7)

Page Break

Reasonargue How likely do you think it is that Mykonaran leaders could be persuaded by reasonable argument?

○ Extremely likely (1)

○ Moderately likely (2)

○ Slightly likely (3)

○ Neither likely nor unlikely (4)

○ Slightly unlikely (5)

○ Moderately unlikely (6)

○ Extremely unlikely (7)

Att When you think about the fact that many people don’t pay attention to survey questions, do you agree? Select ‘No’ if you are paying attention.

○ Maybe (1)

○ Yes (2)

○ No (3)

○ Don’t know (4)

Skip To: End of Survey If When you think about the fact that many people don’t pay attention to survey questions, do you ag... != No

Nkcomp When you think about Mykonaran leadership, how competent or incompetent do you think they are in matters of military strategy?

- ○ Extremely competent (1)
- ○ Moderately competent (2)
- ○ Slightly competent (3)
- ○ Neither competent nor incompetent (4)
- ○ Slightly incompetent (5)
- ○ Moderately incompetent (6)
- ○ Extremely incompetent (7)

NKknow When you think about Mykonaran leadership, how generally knowledgeable do you think they are relative to leaders in other countries?

○ Extremely knowledgeable (1)

○ Very knowledgeable (2)

○ Moderately knowledgeable (3)

○ Slightly knowledgeable (4)

○ Not knowledgeable at all (5)

Page Break

End of Block: Mechanisms

Start of Block: Tools Block

Dipvsmileff In general, do you expect military or diplomatic methods to be more effective at getting Mykonara to alter its behavior?

○ Military much more effective (1)

○ Military more effective (2)

○ Military somewhat more effective (3)

○ Neither military nor diplomatic more effective (4)

○ Diplomatic somewhat more effective (5)

○ Diplomatic more effective (6)

○ Diplomatic much more effective (7)

Dipres On the whole, how likely or unlikely do you think it is that direct negotiations (diplomacy) between the US and Mykonara will resolve the ongoing crisis by getting Mykonaran leaders to alter their behavior?

- ○ Extremely unlikely (1)
- ○ Moderately unlikely (2)
- ○ Slightly unlikely (3)
- ○ Neither likely nor unlikely (4)
- ○ Slightly likely (5)
- ○ Moderately likely (6)
- ○ Extremely likely (7)

Robot Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) if you are not a robot.

- ○ Strongly agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Strongly agree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Agree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Somewhat agree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Neither agree nor disagree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Somewhat disagree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Disagree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Strongly disagree

Dippref In general, do you think the US should engage Mykonara in direct negotiations (diplomacy) to get Mykonara to alter its behavior?

- ○ Strongly oppose negotiations (1)
- ○ Oppose negotiations (2)
- ○ Somewhat oppose negotiations (3)
- ○ Neither oppose nor advocate negotiations (4)
- ○ Somewhat advocate negotiations (5)
- ○ Advocate negotiations (6)
- ○ Strongly advocate negotiations (7)

Page Break

Dipvsmilpref In general, would you prefer the U.S. use military or diplomatic methods to get Mykonara to alter its behavior?

- ○ Strongly prefer military (1)
- ○ Prefer military (2)
- ○ Somewhat prefer military (3)
- ○ Prefer neither military nor diplomatic (4)
- ○ Somewhat prefer diplomatic (5)
- ○ Prefer diplomatic (6)
- ○ Strongly prefer diplomatic (7)

Tooleff If negotiations take place, do you think leverage (threats and/or rewards) or dialogue would be more effective at getting Mykonara to alter its behavior?

○ Leverage much more effective (1)

○ Leverage more effective (2)

○ Leverage somewhat more effective (3)

○ Neither leverage or dialogue more effective (4)

○ Dialogue somewhat more effective (5)

○ Dialogue more effective (6)

○ Dialogue much more effective (7)

Page Break

Poleff Please indicate how effective or ineffective you expect each of the following policies to be at getting Mykonara to alter its behavior

	Extremely effective (1)	Very effective (2)	Moderately effective (3)	Slightly effective (4)	Not effective at all (5)
Military action (1)	○	○	○	○	○
Economic sanctions (2)	○	○	○	○	○
Negotiations with threats of military action (3)	○	○	○	○	○
Negotiations with promises of economic rewards (4)	○	○	○	○	○
Negotiations as open dialogue (5)	○	○	○	○	○

Page Break

Polpref Please indicate the extent to which you support or oppose each of the following policies towards Mykonara

	Strongly support (1)	Support (2)	Somewhat support (3)	Neither support nor oppose (4)	Somewhat oppose (5)	Oppose (6)	Strongly oppose (7)
Military action (1)	○	○	○	○	○	○	○
Economic sanctions (2)	○	○	○	○	○	○	○
Negotiations with threats of military punishment (3)	○	○	○	○	○	○	○
Negotiations with promises of economic rewards (4)	○	○	○	○	○	○	○
Negotiations as open dialogue (5)	○	○	○	○	○	○	○

Page Break

End of Block: Tools Block

Start of Block: Distributive block

USconced (Select 'N/A' for ALL OPTIONS if you think the US should not be seeking negotiations: do not select N/A for ANY options if you do not select it for ALL):
Mykonaran leadership has demanded a range of concessions from the US before they would be willing to engage in negotiations. Please indicate whether you think US officials should be willing to make each of the following concessions in exchange for Mykonara's participation in negotiations

	Definitely yes (1)	Probably yes (2)	Maybe or maybe not (3)	Probably not (4)	Definitely not (5)	N/A (US should not Negotiate) (6)
Withdraw US troops and protection from the region (1)	○	○	○	○	○	○
Send economic aid and investment to Myknonara (2)	○	○	○	○	○	○
Lift current sanctions on Mykonara (3)	○	○	○	○	○	○

Skip To: End of Survey If (Select 'N/A' for ALL OPTIONS if you think the US should not be seeking negotiations: do not sel... [N/A (US should not Negotiate)] (Count) = 1

Skip To: End of Survey If (Select 'N/A' for ALL OPTIONS if you think the US should not be seeking negotiations: do not sel... [N/A (US should not Negotiate)] (Count) = 2

Page Break

NKconced (Select 'N/A' for ALL OPTIONS if you think the US should not be seeking negotiations: do not select N/A for ANY options if you do not select it for ALL)
US officials have demanded a range of concessions from Mykonara before they would be willing to engage in negotiations. Please indicate whether you think US officials should be willing to engage in negotiations if Mykonara does NOT make each of the following concessions.

	Definitely yes (1)	Probably yes (2)	Maybe or maybe not (3)	Probably not (4)	Definitely not (5)	N/A (US should not Negotiate) (6)
Give up its nuclear weapons (1)	○	○	○	○	○	○
Freeze nuclear development, but keep its nuclear weapons (2)	○	○	○	○	○	○
Stop missile tests, but keep its nuclear weapons (3)	○	○	○	○	○	○

Skip To: End of Survey If (Select 'N/A' for ALL OPTIONS if you think the US should not be seeking negotiations: do not sel... [N/A (US should not Negotiate)] (Count) = 1

Skip To: End of Survey If (Select 'N/A' for ALL OPTIONS if you think the US should not be seeking negotiations: do not sel... [N/A (US should not Negotiate)] (Count) = 2

Page Break

Concvsbreak Do you agree or disagree with the following statement: It is better for the US to make significant concessions to Mykonara than to have a breakdown of negotiations between the US and Mykonara?

○ Strongly agree (1)

○ Agree (2)

○ Somewhat agree (3)

○ Neither agree nor disagree (4)

○ Somewhat disagree (5)

○ Disagree (6)

○ Strongly disagree (7)

Page Break

Pointdistribute Imagine 100 points representing the issues Mykonara and the US are fighting over. If a country gets 100 points, that means they get everything they want. If a country gets 0 points, that means they get nothing they want.
What is the MINIMUM number of points you think the US should accept, and the MAXIMUM number of points you think Mykonara should get, from any negotiation or conflict? (Total points given must equal 100)
_______ Points for US (1)
_______ Points for Mykonara (2)

Page Break

End of Block: Distributive block

Start of Block: Demographics

Sex What sex are you?

○ Male (1)

○ Female (2)

○ Other (3)

Age What age range are you in?

- ○ Under 18 (1)
- ○ 18 - 24 (2)
- ○ 25 - 34 (3)
- ○ 35 - 44 (4)
- ○ 45 - 54 (5)
- ○ 55 - 64 (6)
- ○ 65 - 74 (7)
- ○ 75 - 84 (8)
- ○ 85 or older (9)

Ethnic Please specify your ethnicity

○ White (1)

○ Black or African American (2)

○ American Indian or Alaska Native (3)

○ Asian (4)

○ Native Hawaiian or Pacific Islander (5)

○ Other (6)

Employment Are you currently

○ Employed full time (1)

○ Employed part time (2)

○ Unemployed looking for work (3)

○ Unemployed not looking for work (4)

○ Retired (5)

○ Student (6)

○ Disabled (7)

Page Break

Education What is the highest level of school you have completed or the highest degree you have received?

○ Less than high school (11)

○ High school graduate (12)

○ Some college (13)

○ 2 year degree (14)

○ 4 year degree (15)

○ Professional degree (16)

○ Doctorate (17)

End of Block: Demographics

Start of Block: Block 4

Party In politics today, do you consider yourself a:

○ Democrat (1)

○ Republican (2)

○ Independent (3)

○ Something Else (4)

If In politics today, do you consider yourself a: Something Else Is Selected

○ Please Elaborate (5) ____

Display This Question:

If In politics today, do you consider yourself a: = Independent

And In politics today, do you consider yourself a: = Something Else

Partylean Do you lean more towards:

○ Democratic Party (1)

○ Republican Party (2)

○ Neither One (3)

PoliticalID In general, would you describe your political views as:

○ Very Conservative (1)

○ Conservative (2)

○ Moderate (3)

○ Liberal (4)

○ Very Liberal (5)

UScitizen Are you a citizen of the United States?

○ Yes (1)

○ No (2)

○ Don't Know (3)

○ Yes (Multiple Citizenships) (4)

If Are you a citizen of the United States? No Is Selected

And And Are you a citizen of the United States? Don't Know Is Selected

And And Are you a citizen of the United States? Yes (Multiple Citizenships) Is Selected

○ Of what country (countries) are you a citizen? (5)

____

Birth Were you born in the United States, or in another country?

○ United States (1)

○ Another Country (2)

In general, would you prefer the U.S. use military or diplomatic methods to get Mykonara to alter... = Prefer military

○ Please Specify (3) ____

Militaryserv Have you ever served on active duty in the US Armed Forces?

○ Yes (1)

○ No (2)

Display This Question:

If Have you ever served on active duty in the US Armed Forces? = Yes

Deployment If possible, please list areas (countries, regions, etc...) in which you were deployed

____

Abroad Which of the following statements best describes you?

○ I have lived in a country outside the United States for an extended period of time (1)

○ I have traveled outside of the United States, but have never lived in a country outside the United States (2)

○ I have never been outside the United States (3)

Page Break

End of Block: Block 4

Start of Block: Political Knowledge

PK1 Which of the following is the name of a major US led security alliance?

○ ASEAN (1)

○ MUN (2)

○ NATO (3)

○ NAFTA (4)

PK2 Which political party currently holds a majority in the U.S. Senate?

○ Republican (1)

○ Democratic (2)

○ Neither holds a majority (3)

○ Don't Know (4)

PK3 Who is the current leader of China?

○ Xi Jinping (1)

○ Kim Il Sung (2)

○ Shinzo Abe (3)

○ Kim Jong Un (4)

○ Don't Know (5)

PK4 What office or position does Mike Pompeo currently hold?

○ Secretary of Defense (1)

○ Head of the DNC (2)

○ Secretary of State (3)

○ None (4)

○ Don't Know (5)

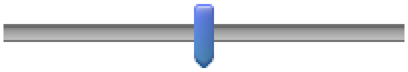
PK5 What office or political position does Mike Pence currently hold?

- ○ President (1)
- ○ Vice President (2)
- ○ Senator (3)
- ○ Member of Congress (4)
- ○ Don't Know (5)

End of Block: Political Knowledge

Start of Block: USinvolvement

Hawkdove As a general rule, do you think the United States should be ready and willing to use military force around the world, or should the United States be very reluctant to use military force? Please indicate your position on the scale below

	0 1 2 3 4 5 6 7 8 9 10
Please make a selection between 0 (very reluctant) and 10 (very willing) by sliding the bar ()	

Involvement Many regions of the world are in trouble these days. Below are some of these regions: for each one please indicate whether you think US military or economic involvement is too high, about right, or not high enough

	Too involved (1)	About involved enough (2)	Not involved enough (3)
Eastern European countries like Ukraine and Poland (8)	○	○	○
Middle Eastern countries like Iraq and Syria (9)	○	○	○
Latin American countries like Columbia and Venezuela (10)	○	○	○
Asian countries like Korea, Pakistan and Myanmar (11)	○	○	○
African countries like the Congo and Zimbabwe (12)	○	○	○

Page Break

End of Block: USinvolvement

Start of Block: Nationalism

Q51 In general, do you believe the United States is superior to other nations?

- ○ Vastly Superior (1)
- ○ Mostly Superior (2)
- ○ Not so Superior (3)
- ○ Not at all Superior (4)

Q53 Does anything about America make you feel ashamed?

- ○ Very many things (1)
- ○ Many things (2)
- ○ Not many things (3)
- ○ Nothing at all (4)

Q55 Do you believe the world would be better if people from other countries were more like Americans?

- ○ Much better (1)
- ○ Better (2)
- ○ Slightly better (3)
- ○ Not better at all (4)
- ○ Worse (5)

Q59 How much does being an American have to do with how you feel about yourself?

- ○ A tremendous amount (1)
- ○ A lot (2)
- ○ A moderate amount (3)
- ○ Not too much (4)
- ○ None at all (5)

Page Break

End of Block: Nationalism

Start of Block: Block 7

Q65 Thank you!

The study you have just completed examines how people's policy preferences towards other states are influenced by their beliefs about the foreign states' leaders' competence and susceptibility to bias, delusion, or irrationality.

The opinion article you were provided was made up in order to understand how people respond to information about other countries' leaders' judgment (how they interpret information, form beliefs, and determine strategies). We wanted to understand this in the context of international adversaries but outside of the boundaries of an existing crisis, so we invented the imaginary state of Mykonara. All information about Mykonara and its relationship to the US were made up for the purpose of this experiment. All statements in the opinion article excerpts were written based on existing sources but adapted for the purpose of this experiment and never published by any news organization.

This means that any opinion articles you read, and the country and leaders they were about, were made up, and randomly shown to you. The reason we made up this article was to see how people's policy preferences respond to different pieces of information presented about the judgment of the leaders of an adversarial international counterpart. We wanted to see how framing leaders as strategically competent vs. incompetent, and as biased/irrational vs. objective/rational, affect people's preferred methods for influencing the behavior of an adversary. We were specifically interested in people's preferences for military and diplomatic approaches to getting a counterpart to alter its behavior in a desirable way.

You can withdraw your data from this study within two weeks of today at no penalty to you or loss of benefits. If you decide to withdraw your data, please type a removal code in the field provided below and email it to Nicholas Campbell-Seremetis at U Chicago at the email address provided below.

If you have further questions about this study, you may contact Nicholas Campbell-Seremetis at The University of Chicago at nickcs@uchicago.edu. Additionally, if you have any questions related to human subjects research at The University of Chicago, or your rights as a participant, please contact the office of the University of Chicago Institutional Review Board (IRB) at (773)

702-2915. The IRB study number is #IRB17-1536. Thank you for taking the time to participate in this research.

Q67 If for any reason you want to remove your responses from this survey please type any 5 digit code into the area below and e mail this removal code to nickcs@uchicago.edu.

If you do not wish to remove yourself, leave the section blank and click next to finish the survey.

____

End of Block: Block 7

Appendices: Appendix B: Survey 1 (Chapter 5)

Conjoint Results and Discussion:

I begin the analysis by estimating the AMUEs for each of the treatments, organized by leader-trait variables, leader-behavior variables,[346] and state-level variables. The estimates tell us the percentage change in the perceived likelihood that a leader (or state) with a particular attribute will be seen as having been irrational to threaten a rival state with military force in an attempt to attain a disputed piece of territory. Since these estimates are calculated by averaging all of the other factor-level combinations (in other words, the average impact of a single attribute across an average of all other possible treatment combinations), we can interpret these estimates as conceptually similar to the main effects from a more standard factorial experiment. As with most conjoint designs, this experiment is intended not to reject null hypotheses but rather to compare the relative magnitude of AMUEs: of the factors that IR, psychology, biology, and conventional wisdom suggest should influence beliefs about the rationality of counterparts, which ones do observers in the American public weigh most heavily when assessing rationality, and which ones do they ignore?

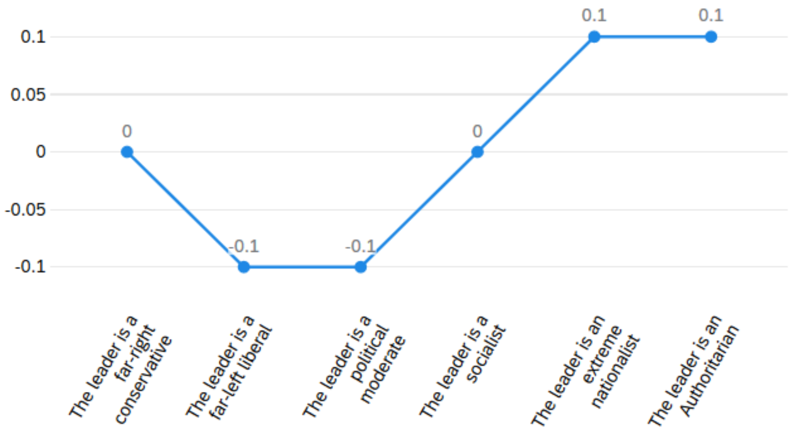
The AMUEs for all treatments are presented graphically in Figure AB.1 for each separate feature: these visualizations convey the average effect of each level of each treatment with respect to all other treatments in its feature category, and the effect numbers are comparable across all the graphs. Figure AB.2, meanwhile, graphs the average Relative Feature Importance (RFI); the RFI metric denotes how influential each feature is in respondents' choice of leader without regard to effect-direction, with higher values denoting greater weight in subjects' decision-making process.

---

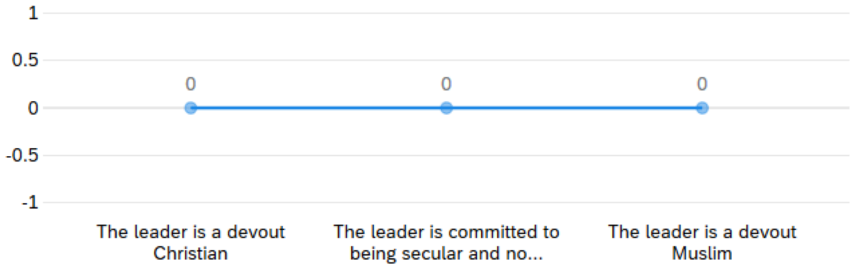
[346] The leader's reputation among outside experts is included in this category of treatments for all classification tables, even though it was only included in a supplemental version of the survey fielded to 500 respondents. This smaller, and also separate, N should be kept in mind when interpreting the results.

Figure AB.1: Average Marginal Utility Effects (By Feature):
Higher values indicate higher likelihood of selecting leaders with associated trait as irrational
For scale, 0.1 = 10% increase

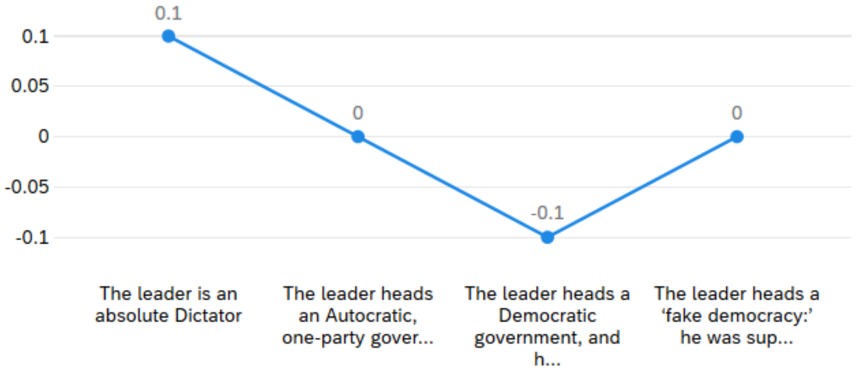
Political Ideology:



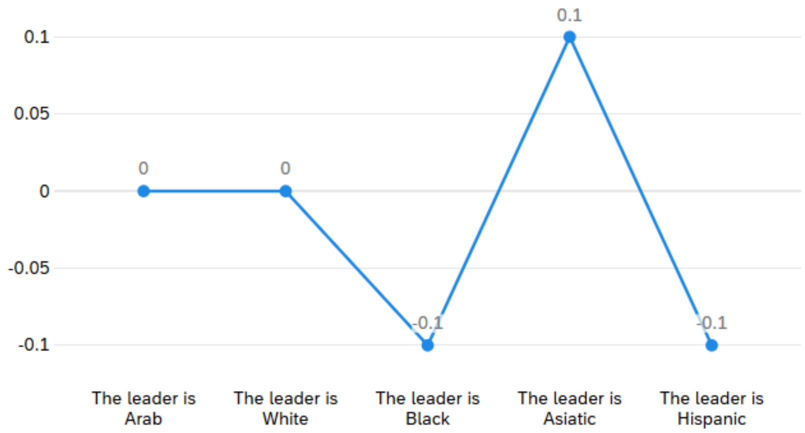
Religion:



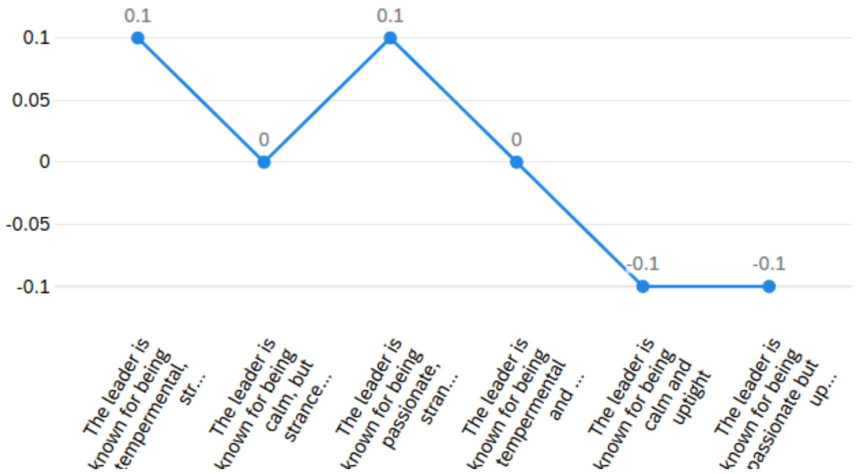
Government Type:



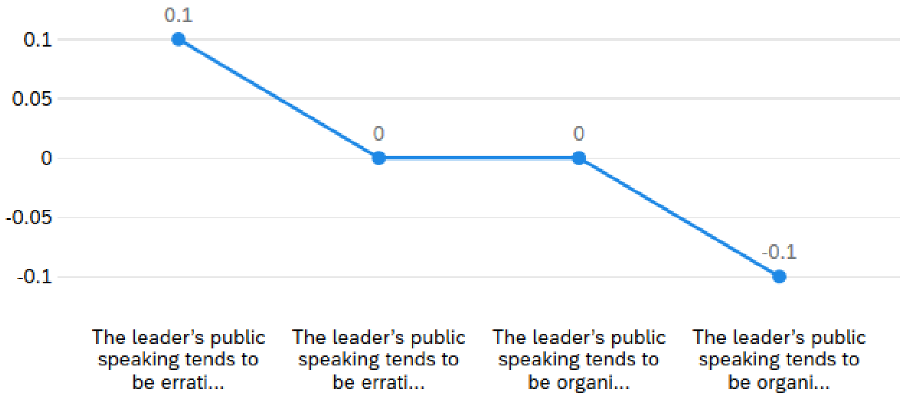
Race:



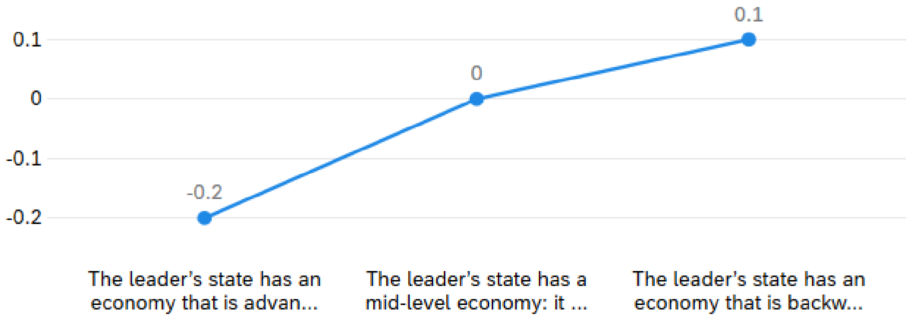
Behavioral Reputation:



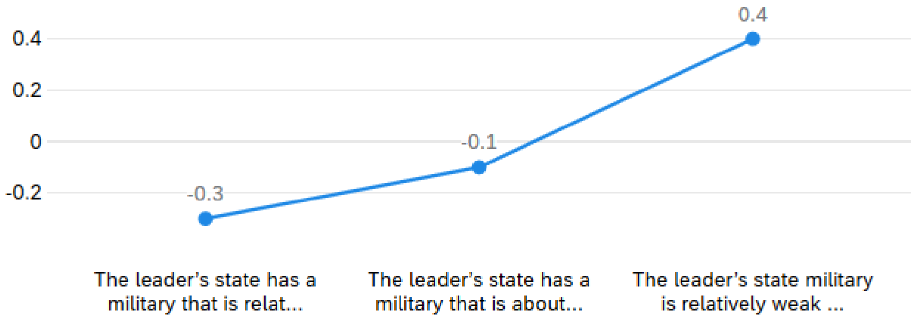
Public Speaking:



State Economy:



State Military:



Women's Rights:

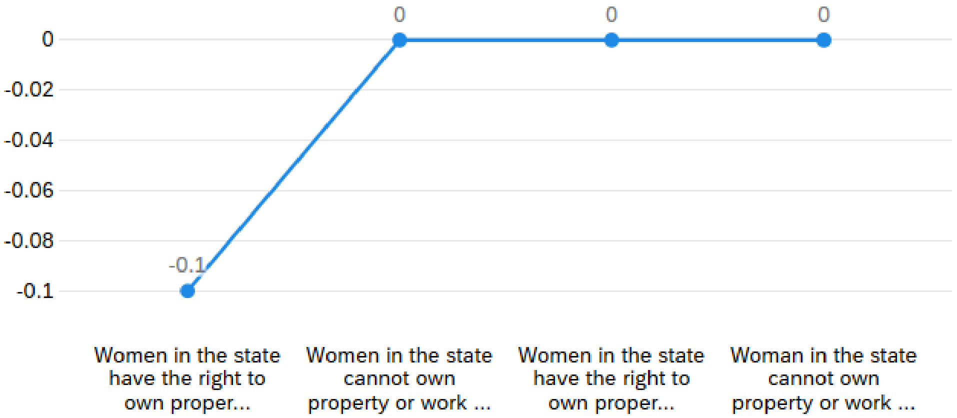
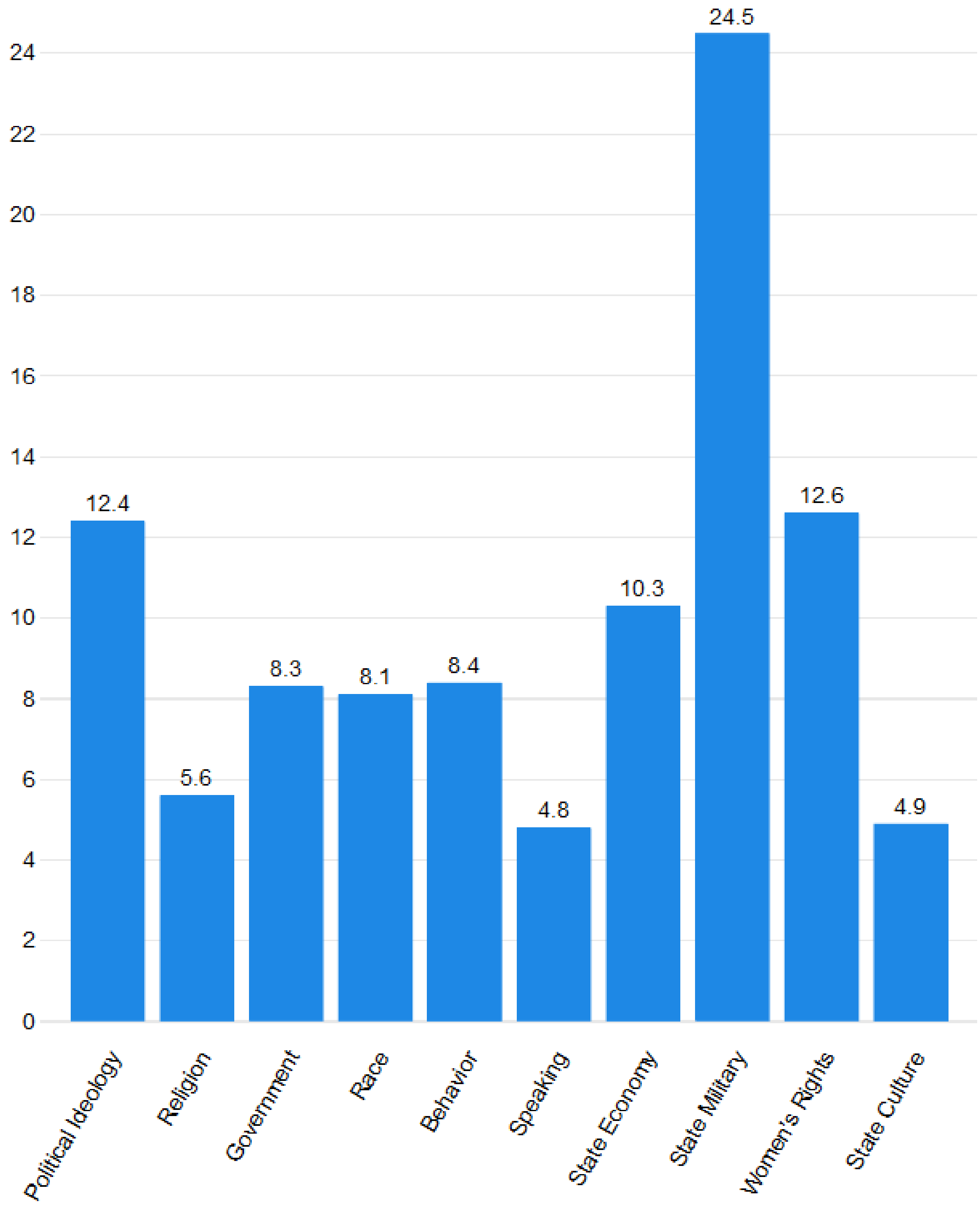


Figure AB.2: Relative Feature Importance



Leader Traits:

I begin by examining the effect of leader-trait level variables on beliefs about the rationality of foreign aggressors. These are herein defined as traits of a leader that are effectively ‘fixed’ and distinct from personal behavior; political ideology, religion, race, and regime type.

Leaders are significantly more likely to be perceived as having been irrational to threaten a neighboring state if they subscribe to political ideologies that are distant from any in the US political mainstream: the percentage chance of irrationality attribution increases for Authoritarian (+11%) and extreme nationalist (+9.3%), while socialist leaders had a close to null effect. By contrast, leaders are less likely to be perceived as irrational if they subscribe to political ideologies that have mainstream corollaries in US politics: the percentage chance of irrationality attribution decreases for moderate (-7.5%), far-left liberal (-8.8%), and far right conservative (-1.5%) leaders.[347]

In contrast to political ideology, the AMUEs for the ‘Religion’ variable are somewhat muted. Within the ‘Religion’ variable, secular leaders were least likely to be perceived as irrational (-2.3%), followed by devout Muslims (-1%) and Christians (+3.3%). Overall, the comparatively weaker impact of this identity level variable is interesting in that it suggests that the kind of religiously demarcated civilizational conflict famously posited by Huntington amongst others is, at least in the contemporary US public, not as salient as political ideology conventional wisdom would suggest when it comes to questions of others’ judgment and rationality.

---

[347] Though the effect size and direction for the ‘far-right’ and ‘far-left’ levels varies with subjects’ political party and ideology: for example liberal or democrat-identifying respondents on average are more likely to think a far right leader is irrational and less likely to think a far left leader is irrational.

The effects of the 'Government type' variable are also comparatively muted, perhaps unexpectedly given the relatively central role regime type plays in extant work on interstate perception. Respondents were more likely to characterize military threats as irrational when leaders were dictators (7%) or heads of 'fake democracies' (3.6%). Legitimately elected democratic leaders were less likely to be seen as irrational (-6.6%), as were heads of autocratic one-party states (-4.2%). This coheres with implicit intuitions of existing work in IR, which suggest that greater concentration of power and absence of institutionalized constraints[348] produce leaders who are more likely to make apparently idiosyncratic policy choices, if only because the absence of consultation or checks-and-balances increases the potential for human error as well as the setting of state policy in pursuit of personal rather than national interests. Unfortunately, the conjoint survey does not permit us to determine whether subjects intuit these particular mechanisms, or are just resorting to the simplifying heuristic of cognitive models (aka stereotypes) that associate dictators and illegitimately elected leaders with unstable madmen.

The effect sizes for the final 'leader-trait,' Race, are comparable to the AMUEs of 'Government-type,'[349] but there are some surprising findings with respect to the directional impact of specific levels. Asiatic leaders were expectedly more likely to be labeled as irrational (6.7%)[350], but somewhat unexpectedly the percentage increase in the chance that respondents attributed irrationality to a leader was exactly equal for White and Arab leaders (2.9%).[351]

---

[348] Which could be laws or the existence of government elites with competing incentives who will act to limit the power and decision-making freedom a given leader possesses.

[349] Although in contrast to Government Type, Race as a variable has been broadly neglected by IR research, with a couple of notable exceptions. In light of this extant scholarly wisdom, the comparable impact of the two variables on subjects' choice-making is a surprising finding. It suggests future work that involves interstate perception should be more cognizant of the role of race, or at least attempt to theoretically or empirically expand beyond tried and tested variables.

[350] Possibly a reflection not only of racial stereotypes but also the salience of hostilities with North Korea at the time of the conjoint

[351] It is however notable that this equality holds only across the total subject population: when we look, for example, only at Republican subjects, Arab leaders are more likely than White leaders to be labeled as irrational.

Respondents were less likely to attribute irrationality to leaders who were Hispanic (-6.4%) or Black (-5.5%).

Leader Behavior:

I now turn to the effect of leader-behavior variables, namely speaking style, personality, and (in the smaller conjoint) expert opinion about the leader's behavior.[352] These are intended to represent costless behaviors for which state leaders may be known by domestic or international audiences; actions unrelated to material policy and the utilization of state resources. In contrast to the leader-trait variables, the leader-behavior variables are each comprised of multiple similar but distinct traits such that each is effectively a factorial variable **(see Table 5.1)**: for example, speaking style is presented as a randomized combination of tone (loud/quiet) and organization (erratic/organized). Thus, these variables include both 'ideal-type' leaders (e.g. a leader whose public speaking is loud and erratic vs. a leader whose public speaking is quiet, organized and statesmanlike) and 'mixed-type' leaders (a quiet but erratic speaker vs. a loud and organized speaker). This helps determine the aspects of behavior that are most relevant in assessments of the rationality of foreign leaders and guards, to some extent, against misleading findings resulting from idiosyncratic design choices. The wording of these variables is obviously not exhaustive of all possible costless behaviors: they are instead an attempt to account for a representative range of typical tropes or simplifying categories into which leaders, or leaders' behavior, is placed.

When we examine the AMUEs of the personality variable, we see that leaders known for being temperamental are more likely to be chosen as the irrational leader; the percentage increase of 'temperamental' is roughly the same whether it is paired with 'eccentric' (+6.3%) or 'uptight:'

(+6.2%). By contrast, leaders known for being calm are less likely to be chosen as the irrational leader: it has a close to null effect when paired with ‘eccentric’ (+0.9%) and notably decreases the percentage likelihood of irrationality attribution when paired with ‘uptight’ (-12%). Finally, leaders known for being passionate[353] are substantially more likely to be seen as irrational when they are also eccentric (+10.5%) but less likely to be seen as irrational when they are uptight (-6%).

Leaders’ speaking style also exerts substantively strong effects on the percentage likelihood of irrationality attribution, particularly when looking at the ideal-typical combinations of volume and organization. Leaders whose public speaking is loud and erratic are much more likely to be seen as irrational (+8.4%), while leaders whose public speaking is quiet and organized are much less likely to be seen as irrational (-6.4%). By contrast, the results for the ‘mixed’ types are weaker and more ambiguous. Loud and organized leaders are slightly more likely to be seen as irrational (+1.4%), while leaders who speak quietly but erratically are slightly less likely to be seen as irrational (-3.3%). These results imply that the volume or tone of public speech carries more weight than its coherence in shaping inferences about leaders’ rationality. However, the significant difference in effect magnitude between mixed and ideal types suggest that speaking style carries the most weight when a leader’s speaking carries *all* the components of particular stereotypes: a bombastic madman *or* a composed statesman, rather than something in between.

Finally, the elite narrative and expert opinion variable included in the secondary conjoint exerts an extremely powerful impact, replacing the military power variable as the single most influential input into subjects’ choice. Subjects told that a majority of experts have expressed doubts about a leader's mental state are far more likely to conclude said leader is irrational **(**+41%**),** while there is a null effect when experts express confidence in the leader's intellect and mental stability (+0.1%). In sum, members of the public appear to 'parrot' expert beliefs about foreign leaders with whom they are unfamiliar, at least when those beliefs hold a negative valence.

---

[353] A term chosen, via pilot testing of several possible synonyms, to indicate an energetic individual without the negative connotation associated with ‘temperamental.’

These findings suggest it is not just the costly content of behavior that matters: style is also of considerable import. Indeed, comparing AMUEs we see that some so-called 'costless' behaviors play a greater role in shaping inferences about a leader's rationality than variables like regime type which are commonly conceived as vital to all manner of interstate signaling.

State Characteristics:

I turn now to state characteristics: military power, economic power, and culture proxied by women's social and economic rights.[354] Recall that according to the implied hypotheses of both the structural-determinism and the costly-behavior pathways, we should expect the first two (military and economic power) of these to exert an extremely strong impact on estimations of rationality. The results substantiate this expectation.

Relative military power appears to be the most important of the state-level material variables: relatively weak leaders were *far* more likely to be chosen as irrational **(+38.7%)** than relatively strong leaders **(-27.6%)** or leaders who were roughly as militarily powerful as the rival they threatened **(-11%).** Relative economic power shows equivalent directional impacts but with a notably reduced effect size: leaders of poor states are more likely (+13%) to be seen as irrational than leaders of rich **(-15%)** or 'average' **(-1%)** economies. By contrast, Culture as operationalized by women's social and economic rights had a smaller, though still noteworthy, effect on subject choice. Leaders of 'Similar' cultures (in which women have both kinds of rights) are less likely to be seen as irrational **(-8.1%)** than leaders of 'Distant' cultures **(+6.7%)**, while the effect was more muted for cultures that denied women either social **({+4%)** or economic **(+1%)** rights.

---

[354] The choice to use women's rights as a simplifying proxy for culture, particularly in the sense of the extent to which a state's culture is 'advanced/western,' was made for several reasons. First of all, the choice was supported by double blind coding by independently hired RA's in which the respondents were asked to answer questions about the cultural advancement of a state based on particular state characteristics, such as welfare services, schooling practices, and women's rights. Women's rights was consistently the strongest predictor of respondent estimations of a state's cultural advancement. Second, utilizing women's rights as a proxy permits this work to speak to research on gender in International Relations, a burgeoning but vital and interesting field. Finally, it is an alternative angle on extant theories regarding the role of leader gender in estimations of resolve, intentions, and other internal characteristics.

Appendices: Appendix C: Survey 2 (Chapter 5):

Full Survey Text (Copied from Qualtrics):

We're all mad non-conjoint (fake states) - Copy

Start of Block: Consent

Q55 The following research project investigates how people think about foreign states. Your participation is completely voluntary. Upon completion of the study you will receive monetary compensation of 80 cents. You will be asked to read a short introduction, then look at eight pairings of foreign leaders and choose between them on the basis of criteria described in the survey.

If you decide to stop participating in the study, there will be no penalty to you, and you will not lose any benefits to which you are otherwise entitled. You may refuse to participate without penalty or loss of benefits to which you are otherwise entitled. You may not be entitled to receive monetary compensation if you do not complete the survey in a satisfactory manner, here meaning the failure to pay attention. Your decision will not affect your future relationship with The University of Chicago. We will work to make sure that no one sees your survey responses without approval. But, because we are using the Internet, there is a chance that someone could access your online responses without permission. In some cases, this information could be used to identify you. Your data will be protected with a code to reduce the risk that other people can view the responses. We know of no benefits to you for participating in this study. The survey will take approximately 5-10 minutes.

If any question makes you uncomfortable or is upsetting, please move on to the next one. You also have the option to discontinue the survey at any time without penalty. If you have questions, comments or concerns about the survey, or if you feel you have been harmed by study participation, you may contact Nicholas Campbell-Seremetis at nickcs@uchicago.edu. For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact the office of the University of Chicago Institutional Review Board (IRB) at (773) 702-2915.
By clicking the next button and beginning the survey, you indicate that you are at least 18 years old, have read and understand this consent form, and you agree to participate in online research.

Please click the next button below to indicate your consent and begin the survey.

Page Break

End of Block: Consent

Start of Block: Introduction

Q1 We want to tell you about a recent military conflict between the neighboring states of Mykonara and Trinamar. Mykonara and Trinamar have long been at odds over an important piece of disputed territory along their border; both states claim ownership of the land and its people but neither actively controls the territory or enjoys international recognition of the legitimacy of its claim. For the past several months Mykonara's president has been adopting increasingly aggressive rhetoric towards Trinamar, up to and including threats to start a military conflict unless Trinamar forfeits its claim to the disputed land. Recently, Mykonara carried out these threats by moving troops to the border and launching missiles against one of Trinamar's military bases, an act many saw as surprising since Mykonara is the militarily weaker of the two states.

Page Break

End of Block: Introduction

Start of Block: Race

R-White We'd like to tell you about Mykonara's leader, President Jason Adams. Adams, a sixty-two year old white man, has dark but graying hair and a tall, somewhat muscular build. Adams is said to enjoy both significant popular backing and unquestioned authority over his government.

R-Black We'd like to tell you about Mykonara's leader, President Mswati Kenyatta. Kenyatta, a sixty-two year old black man, has dark but graying hair and a tall, somewhat muscular build. Kenyatta is said to enjoy both significant popular backing and unquestioned authority over his government.

R-Arab We'd like to tell you about Mykonara's leader, President Mahmoud al-Raziz. Raziz, a sixty-two year old arab, has dark but graying hair and a tall, somewhat muscular build. Raziz is said to enjoy both significant popular backing and unquestioned authority over his government.

R-Asian We'd like to tell you about Mykonara's leader, President Yukio Kansi. Kansi, a sixty-two year old asiatic man, has dark but graying hair and a tall, somewhat muscular build. Kansi is said to enjoy both significant popular backing and unquestioned authority over his government.

R-Control We'd like to tell you about Mykonara's President, who is said to enjoy both significant popular backing and unquestioned authority over his government.

End of Block: Race

Start of Block: Political Ideology

Ideo-right The President is a member of Mykonara's far-right, the political positions of which are similar to those supported by the more conservative wing of the US Republican party.

Ideo-left The President is a member of Mykonara's far-left, the political positions of which are similar to those supported by the more liberal wing of the US Democratic party.

Ideo-socialist The President is a radical left-wing socialist with strong authoritarian tendencies, and his political positions are widely seen as extreme

Ideo-nationalist The President is a radical right-wing nationalistic populist with strong authoritarian tendencies, and his political positions are widely seen as extreme

Ideo-Control The President is a dog-lover; he owns several, and is often accompanied by one at official meetings or events.

End of Block: Political Ideology

Start of Block: Personal behavior

Behave-Angry The President has a reputation for being temperamental, arrogant and self-absorbed. His public behavior seems un-statesmanlike at best and like a stereotypical strongman at worst. He is known for speaking in a loud, disorganized fashion and for his dramatic hand gestures, especially when delivering speeches.

Behave-Calm The President has a reputation for being calm, humble, and open-minded. His public behavior seems like that of a model statesman, though some call it uptight. He is known for speaking deliberately, with modulated volume and clear organization, and for his pointed but controlled hand gestures, especially when delivering speeches.

Behave-Odd The President has a reputation for being eccentric and ignorant, yet oddly overconfident. His public behavior seems at best charmingly odd and outright buffoonish at worst. He is known for erratic or incomprehensible speaking that frequently changes between tones and trains of thought, and for his seemingly random hand gestures, especially when delivering speeches.

Behave-Control The President is a known sports enthusiast and is especially fond of soccer, which he used to play semi-professionally. He still plays often and is a frequent guest at professional games.

End of Block: Personal behavior

Start of Block: Economics

Econ-Rich The President's grip on power holds largely due to Mykonara’s impressive levels of economic, technological, and cultural development. The people have in recent years enjoyed greater economic prosperity, ready access to the world’s most advanced tech goods, and a high overall high quality of life.

Econ-Mid The President's grip on power holds despite Mykonara’s recent lack of economic, technological, or cultural development. The people are neither economically destitute nor prosperous, but they lack access to the world’s more advanced tech goods, and their quality of life is at best average relative to the world as a whole.

Econ-Poor The President's grip on power holds despite the stagnation of the already undeveloped economy and the persistence of backwards cultural practices that deny basic rights and dignities to many people. Most people in Mykonara are economically destitute, lack access to even outdated tech goods, and have an extremely poor quality of life.

Econ-Control The President's signature vice is his taste for chocolate, a craving that he tends to indulge in elaborate ways. He employs a personal choclatier, is known to carry a small bag of chocolate truffles wherever he goes, and the centerpiece of every formal dinner he throws is a vast array of chocolate candies, cakes, and other desserts.

Page Break

End of Block: Economics

Start of Block: DVs

Feel We'd like to get your feelings on a 'feeling thermometer.' A rating of 0 means you feel as cold and negative as possible. A rating of 100 means you feel as warm and positive as possible. A rating of 50 means you feel neutral. How do you feel towards Mykonara's President?

	0 10 20 30 40 50 60 70 80 90 100
0 = cold,100 = warm ()	

Rat-irrat In general, how likely do you think it is that the Mykonaran President was rational, versus irrational, to attack Trinamar? (Rational meaning based on a reasonable understanding of the situation, the probable outcomes of an attack, and the risks involved)

○ Extremely likely to have been rational (1)

○ Moderately likely to have been rational (2)

○ Slightly likely to have beenrational (3)

○ Neither likely nor unlikely to have been rational (4)

○ Slightly unlikely to have been rational (5)

○ Moderately unlikely to have been rational (6)

○ Extremely unlikely to have been rational (7)

Reasonev When you think about Mykonara's President, how probable do you think it is that his beliefs and decisions are based on reasonable interpretations of evidence?

○ Extremely likely (1)

○ Moderately likely (2)

○ Slightly likely (3)

○ Neither likely nor unlikely (4)

○ Slightly unlikely (5)

○ Moderately unlikely (6)

○ Extremely unlikely (7)

Riskatt How likely or unlikely do you think it is that Mykonara's President accurately understood the risks of attacking Trinamar?

○ Extremely likely (1)

○ Moderately likely (2)

○ Slightly likely (3)

○ Neither likely nor unlikely (4)

○ Slightly unlikely (5)

○ Moderately unlikely (6)

○ Extremely unlikely (7)

Page Break

Reasonargue How likely do you think it is that Mykonara's President could be persuaded by reasonable argument?

- ○ Extremely likely (1)
- ○ Moderately likely (2)
- ○ Slightly likely (3)
- ○ Neither likely nor unlikely (4)
- ○ Slightly unlikely (5)
- ○ Moderately unlikely (6)
- ○ Extremely unlikely (7)

Att When you think about the fact that many people don’t pay attention to survey questions, do you agree? Select ‘No’ if you are paying attention.

- ○ Maybe (1)
- ○ Yes (2)
- ○ No (3)
- ○ Don’t know (4)

Skip To: End of Survey If When you think about the fact that many people don’t pay attention to survey questions, do you ag... != No

Nkcomp When you think about Mykonara's President, how competent or incompetent do you think he is in matters of military strategy?

○ Extremely competent (1)

○ Moderately competent (2)

○ Slightly competent (3)

○ Neither competent nor incompetent (4)

○ Slightly incompetent (5)

○ Moderately incompetent (6)

○ Extremely incompetent (7)

NKknow When you think about Mykonara's President, how generally knowledgeable do you think he is relative to leaders in other countries?

- ○ Extremely knowledgeable (1)
- ○ Very knowledgeable (2)
- ○ Moderately knowledgeable (3)
- ○ Slightly knowledgeable (4)
- ○ Not knowledgeable at all (5)

Page Break

Dipvsmileff In general, do you expect military or diplomatic methods to be more effective at getting Mykonara's President to cease his aggression?

- ○ Military much more effective (1)
- ○ Military more effective (2)
- ○ Military somewhat more effective (3)
- ○ Neither military nor diplomatic more effective (4)
- ○ Diplomatic somewhat more effective (5)
- ○ Diplomatic more effective (6)
- ○ Diplomatic much more effective (7)

Robot Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) if you are not a robot.

- ○ Strongly agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Strongly agree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Agree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Somewhat agree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Neither agree nor disagree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Somewhat disagree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Disagree

Skip To: End of Survey If Some people get robots to fill out surveys for them. Skip this question (by clicking the arrow) i... = Strongly disagree

Dipres On the whole, how likely or unlikely do you think it is that direct negotiations (diplomacy) between the President of Mykonara and the leader of Trinamar will resolve the current conflict?

○ Extremely unlikely (1)

○ Moderately unlikely (2)

○ Slightly unlikely (3)

○ Neither likely nor unlikely (4)

○ Slightly likely (5)

○ Moderately likely (6)

○ Extremely likely (7)

Page Break

End of Block: DVs

Start of Block: Manipulation Checks

Q56 What is the name and race of Mykonara's President?

○ President John Adams is white (1)

○ President Yukio Kansi is asiatic (2)

○ President Mahmoud al-Raziz is arab (3)

○ President Muswati Kenyatta is black (4)

○ Don't Know (5)

Q57 What is the political ideology of Mykonara's President?

○ The president is a member of the far-right (1)

○ The president is a member of the far-left (2)

○ The president is a radical socialist authoritarian (3)

○ The president is a radical nationalist authoritarian (4)

○ Don't Know (5)

Q58 What behavior does Mykonara's president have a reputation for?

○ Angry, temperamental, and loud (1)

○ Calm, controlled, and quiet (2)

○ Odd, bumbling, and eccentric (3)

○ Normal, unimpressive, and forgettable (4)

○ He is a sports lover (5)

End of Block: Manipulation Checks

Start of Block: Demography basic

Sex What sex are you?

○ Male (1)

○ Female (2)

○ Other (3)

Age What age range are you in?

- ○ Under 18 (1)
- ○ 18 - 24 (2)
- ○ 25 - 34 (3)
- ○ 35 - 44 (4)
- ○ 45 - 54 (5)
- ○ 55 - 64 (6)
- ○ 65 - 74 (7)
- ○ 75 - 84 (8)
- ○ 85 or older (9)

Ethnic Please specify your ethnicity

- ○ White (1)
- ○ Black or African American (2)
- ○ American Indian or Alaska Native (3)
- ○ Asian (4)
- ○ Native Hawaiian or Pacific Islander (5)
- ○ Other (6)

Employment Are you currently

- ○ Employed full time (1)
- ○ Employed part time (2)
- ○ Unemployed looking for work (3)
- ○ Unemployed not looking for work (4)
- ○ Retired (5)
- ○ Student (6)
- ○ Disabled (7)

Education What is the highest level of school you have completed or the highest degree you have received?

○ Less than high school (11)

○ High school graduate (12)

○ Some college (13)

○ 2 year degree (14)

○ 4 year degree (15)

○ Professional degree (16)

○ Doctorate (17)

Page Break

End of Block: Demography basic

Start of Block: Demography Politics

Party In politics today, do you consider yourself a:

○ Democrat (1)

○ Republican (2)

○ Independent (3)

○ Something Else (4)

○ Please Elaborate (5) ____

Partylean Do you lean more towards:

○ Democratic Party (1)

○ Republican Party (2)

○ Neither One (3)

PoliticalID In general, would you describe your political views as:

- ○ Very Conservative (1)
- ○ Conservative (2)
- ○ Moderate (3)
- ○ Liberal (4)
- ○ Very Liberal (5)

UScitizen Are you a citizen of the United States?

- ○ Yes (1)
- ○ No (2)
- ○ Don't Know (3)
- ○ Yes (Multiple Citizenships) (4)
- ○ Of what country (countries) are you a citizen? (5)

____

Abroad Which of the following statements best describes you?

- ○ I have lived in a country outside the United States for an extended period of time (1)
- ○ I have traveled outside of the United States, but have never lived in a country outside the United States (2)
- ○ I have never been outside the United States (3)

Page Break

End of Block: Demography Politics

Start of Block: Political Knowledge

PK1 Which of the following is the name of a major US led security alliance?

○ ASEAN (1)

○ MUN (2)

○ NATO (3)

○ NAFTA (4)

PK2 Which political party currently holds a majority in the U.S. Senate?

○ Republican (1)

○ Democratic (2)

○ Neither holds a majority (3)

○ Don't Know (4)

PK3 Who is the current leader of China?

○ Xi Jinping (1)

○ Kim Il Sung (2)

○ Shinzo Abe (3)

○ Kim Jong Un (4)

○ Don't Know (5)

PK4 What office or position does Mike Pompeo currently hold?

○ Secretary of Defense (1)

○ Head of the DNC (2)

○ Secretary of State (3)

○ None (4)

○ Don't Know (5)

PK5 What office or political position does Mike Pence currently hold?

- ○ President (1)
- ○ Vice President (2)
- ○ Senator (3)
- ○ Member of Congress (4)
- ○ Don't Know (5)

Page Break

End of Block: Political Knowledge

Start of Block: Hawk/nation

Hawkdove As a general rule, do you think the United States should be ready and willing to use military force around the world, or should the United States be very reluctant to use military force? Please indicate your position on the scale below

	0 1 2 3 4 5 6 7 8 9 10
Please make a selection between 0 (very reluctant) and 10 (very willing) by sliding the bar ()	

Involvement Many regions of the world are in trouble these days. Below are some of these regions: for each one please indicate whether you think US military or economic involvement is too high, about right, or not high enough

	Too involved (1)	About involved enough (2)	Not involved enough (3)
Eastern European countries like Ukraine and Poland (8)	○	○	○
Middle Eastern countries like Iraq and Syria (9)	○	○	○
Latin American countries like Columbia and Venezuela (10)	○	○	○
Asian countries like Korea, Pakistan and Myanmar (11)	○	○	○
African countries like the Congo and Zimbabwe (12)	○	○	○

Nat1 In general, do you believe the United States is superior to other nations?

- ○ Vastly Superior (1)
- ○ Mostly Superior (2)
- ○ Not so Superior (3)
- ○ Not at all Superior (4)

Nat2 Does anything about America make you feel ashamed?

- ○ Very many things (1)
- ○ Many things (2)
- ○ Not many things (3)
- ○ Nothing at all (4)

Nat3 Do you believe the world would be better if people from other countries were more like Americans?

○ Much better (1)

○ Better (2)

○ Slightly better (3)

○ Not better at all (4)

○ Worse (5)

Nat4 How much does being an American have to do with how you feel about yourself?

○ A tremendous amount (1)

○ A lot (2)

○ A moderate amount (3)

○ Not too much (4)

○ None at all (5)

Page Break

End of Block: Hawk/nation

Start of Block: End

Debrief Thank you!

The study you have just completed examines how people make inferences about the rationality of state leaders, and how these influences are influenced by the identity of the foreign states' leader.

The leaders you were provided were made up in order to understand how people respond to information about other countries' leaders. All statements were written for the purpose of this experiment and never published by any news organization.

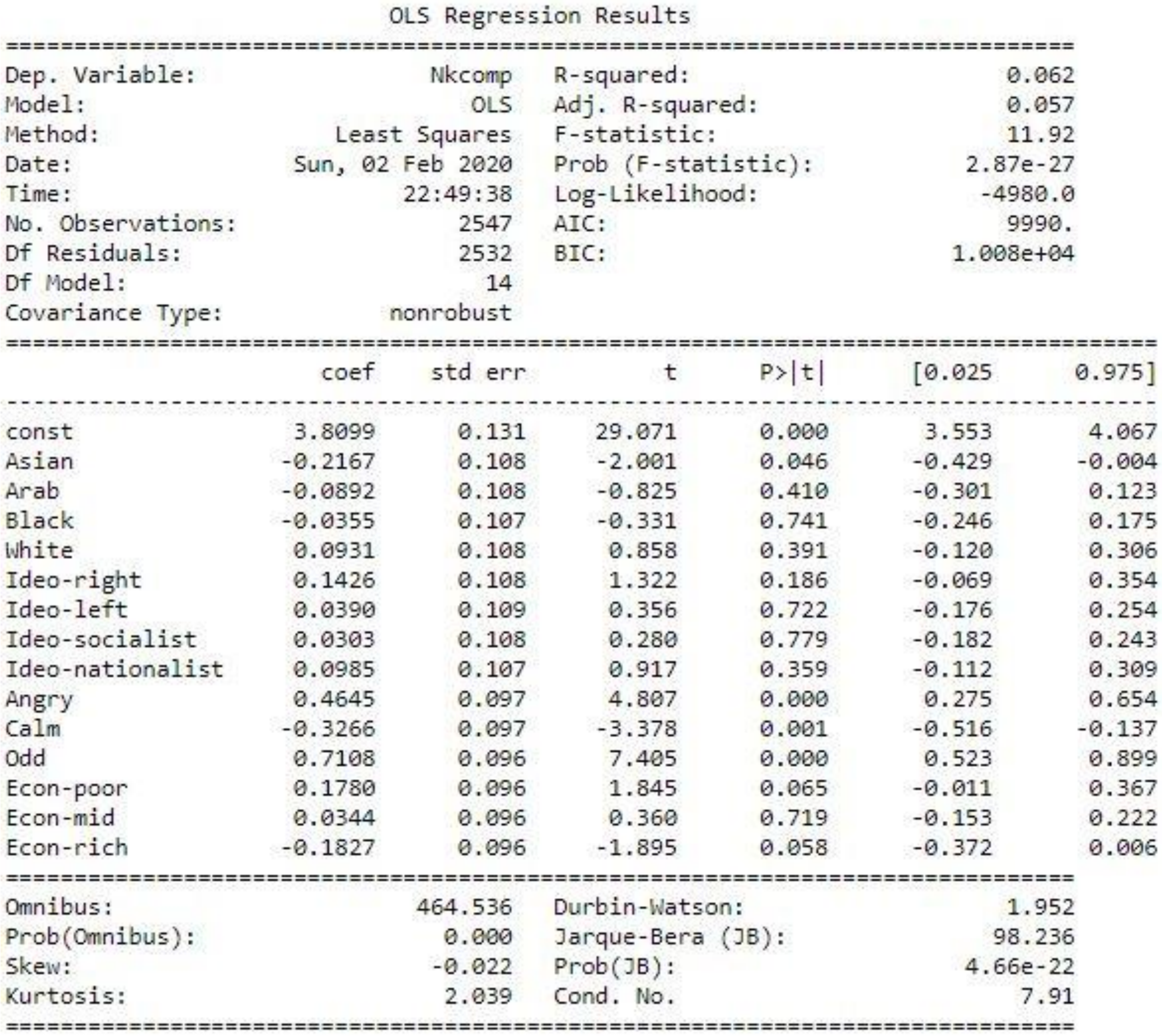
You can withdraw your data from this study within two weeks of today at no penalty to you or loss of benefits. If you decide to withdraw your data, please email Nicholas Campbell-Seremetis at Uchicago at the email address provided below.

If you have further questions about this study, you may contact Nicholas Campbell-Seremetis at The University of Chicago at nickcs@uchicago.edu. Additionally, if you have any questions related to human subjects research at The University of Chicago, or your rights as a participant, please contact the office of the University of Chicago Institutional Review Board (IRB) at (773) 702-2915. Thank you for taking the time to participate in this research.
Please click next for your MTurk Completion Code

End of Block: End

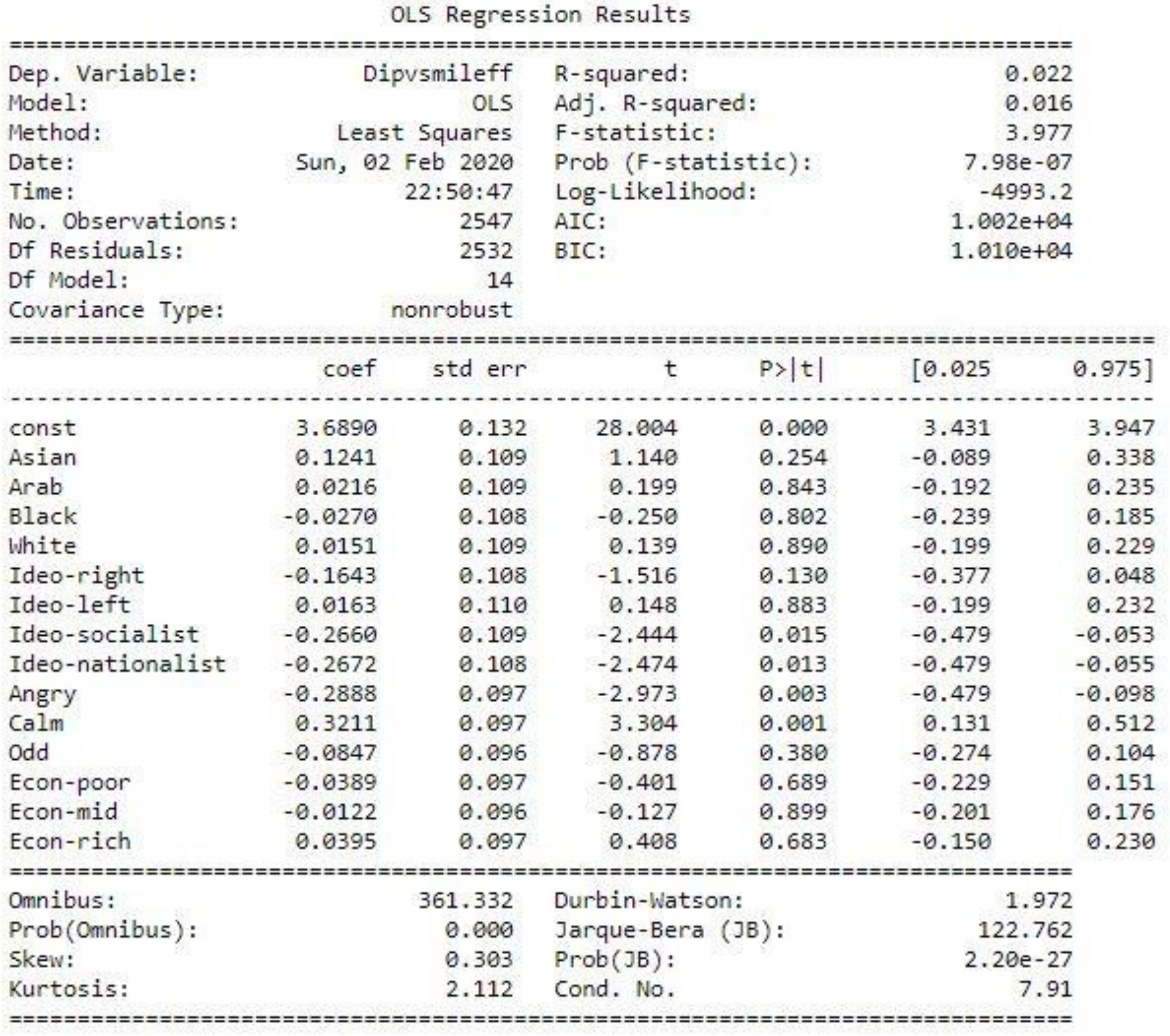
Figure AC.1: Raw Regression Tables: (Listed by DV):

Perceived Competence



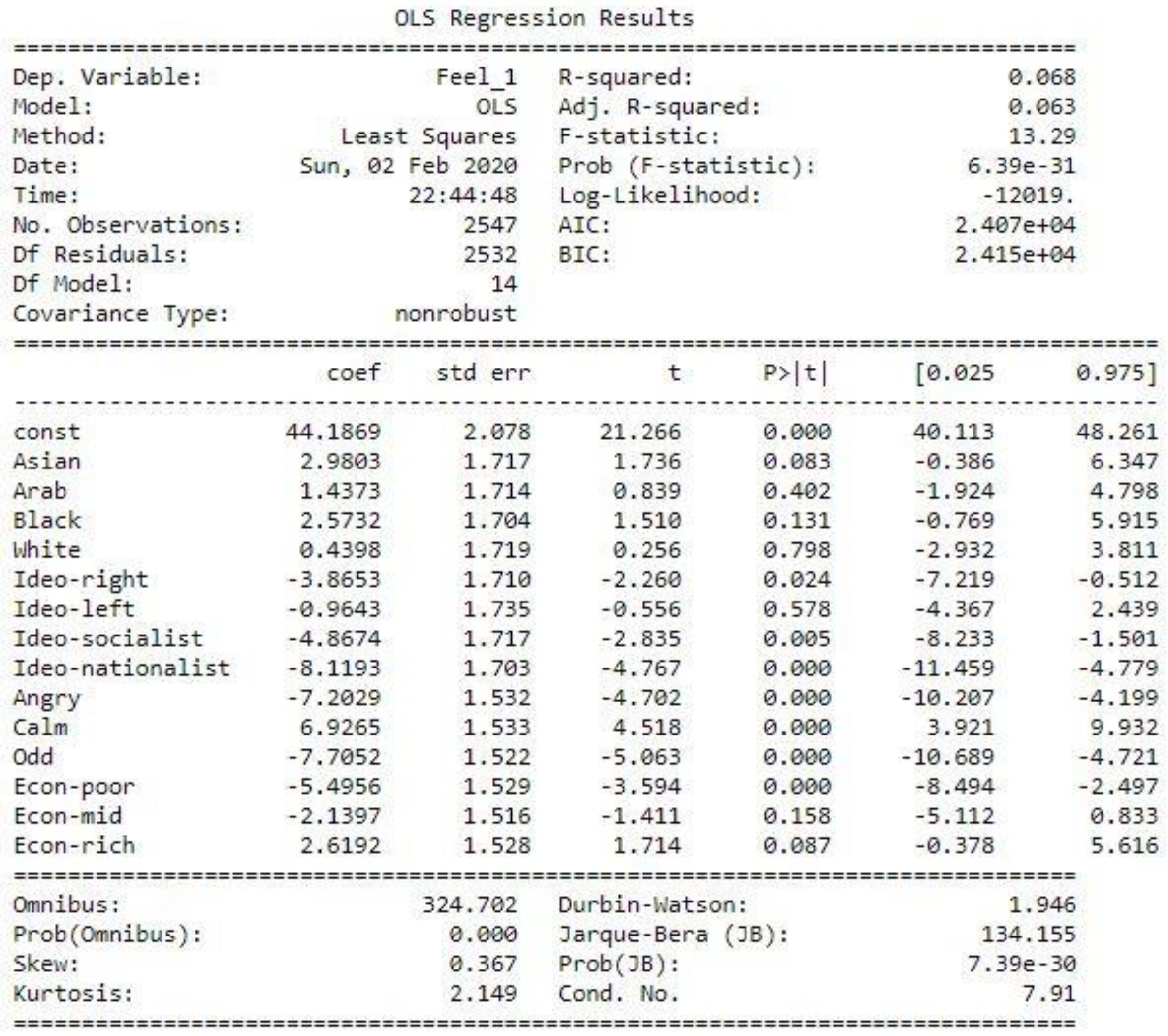
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Nkcomp   R-squared:                       0.062
Model:                            OLS   Adj. R-squared:                  0.057
Method:                 Least Squares   F-statistic:                     11.92
Date:                Sun, 02 Feb 2020   Prob (F-statistic):           2.87e-27
Time:                        22:49:38   Log-Likelihood:                -4980.0
No. Observations:                2547   AIC:                             9990.
Df Residuals:                    2532   BIC:                         1.008e+04
Df Model:                          14
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                3.8099      0.131     29.071      0.000       3.553       4.067
Asian               -0.2167      0.108     -2.001      0.046      -0.429      -0.004
Arab                -0.0892      0.108     -0.825      0.410      -0.301       0.123
Black               -0.0355      0.107     -0.331      0.741      -0.246       0.175
White                0.0931      0.108      0.858      0.391      -0.120       0.306
Ideo-right           0.1426      0.108      1.322      0.186      -0.069       0.354
Ideo-left            0.0390      0.109      0.356      0.722      -0.176       0.254
Ideo-socialist       0.0303      0.108      0.280      0.779      -0.182       0.243
Ideo-nationalist     0.0985      0.107      0.917      0.359      -0.112       0.309
Angry                0.4645      0.097      4.807      0.000       0.275       0.654
Calm                -0.3266      0.097     -3.378      0.001      -0.516      -0.137
Odd                  0.7108      0.096      7.405      0.000       0.523       0.899
Econ-poor            0.1780      0.096      1.845      0.065      -0.011       0.367
Econ-mid             0.0344      0.096      0.360      0.719      -0.153       0.222
Econ-rich           -0.1827      0.096     -1.895      0.058      -0.372       0.006
==============================================================================
Omnibus:                      464.536   Durbin-Watson:                   1.952
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               98.236
Skew:                          -0.022   Prob(JB):                     4.66e-22
Kurtosis:                       2.039   Cond. No.                         7.91
==============================================================================
```

Diplomatic vs. Military Effectiveness



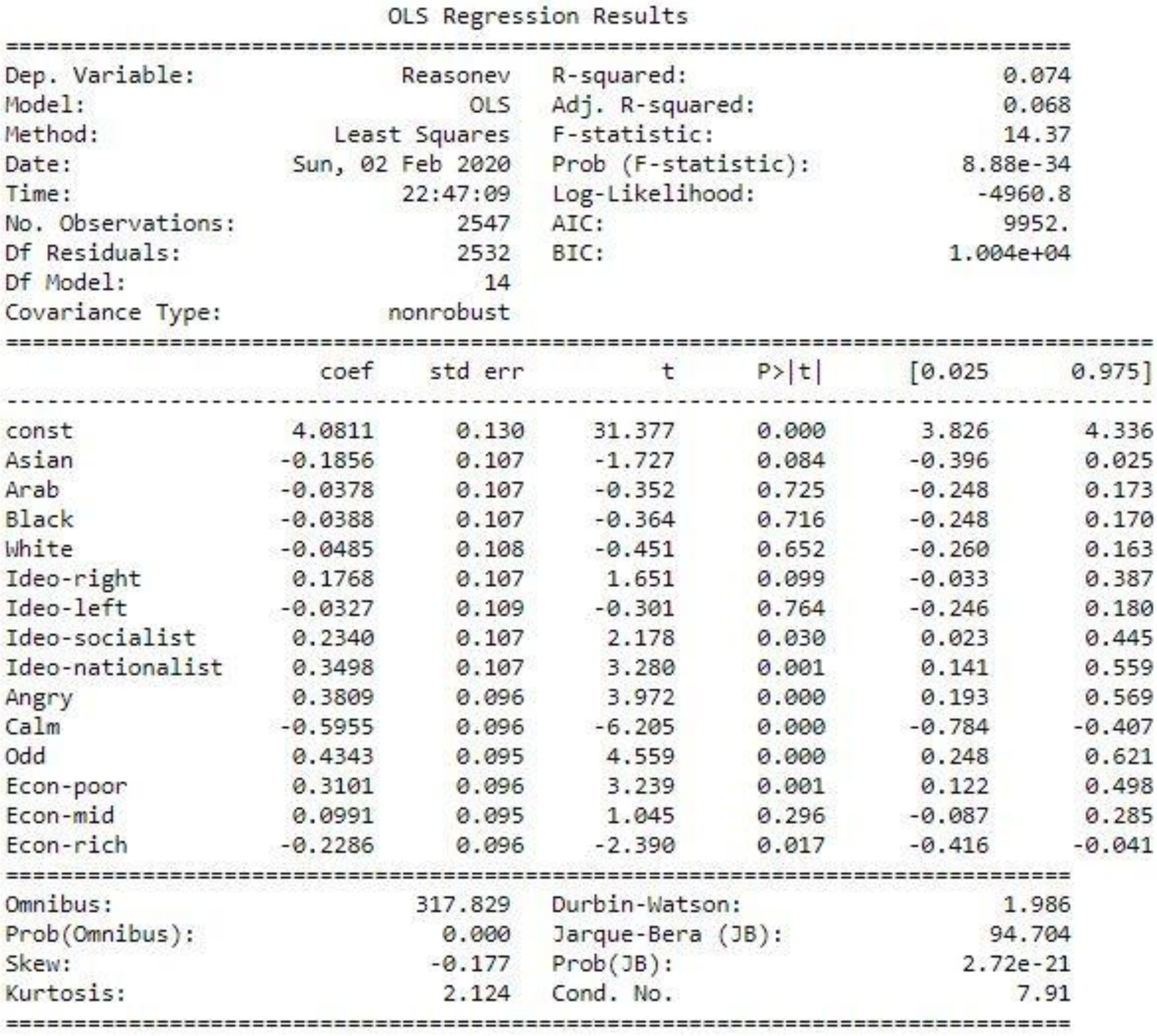
```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Dipvsmileff   R-squared:                       0.022
Model:                            OLS   Adj. R-squared:                  0.016
Method:                 Least Squares   F-statistic:                     3.977
Date:                Sun, 02 Feb 2020   Prob (F-statistic):           7.98e-07
Time:                        22:50:47   Log-Likelihood:                -4993.2
No. Observations:                2547   AIC:                         1.002e+04
Df Residuals:                    2532   BIC:                         1.010e+04
Df Model:                          14
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                3.6890      0.132     28.004      0.000       3.431       3.947
Asian                0.1241      0.109      1.140      0.254      -0.089       0.338
Arab                 0.0216      0.109      0.199      0.843      -0.192       0.235
Black               -0.0270      0.108     -0.250      0.802      -0.239       0.185
White                0.0151      0.109      0.139      0.890      -0.199       0.229
Ideo-right          -0.1643      0.108     -1.516      0.130      -0.377       0.048
Ideo-left            0.0163      0.110      0.148      0.883      -0.199       0.232
Ideo-socialist      -0.2660      0.109     -2.444      0.015      -0.479      -0.053
Ideo-nationalist    -0.2672      0.108     -2.474      0.013      -0.479      -0.055
Angry               -0.2888      0.097     -2.973      0.003      -0.479      -0.098
Calm                 0.3211      0.097      3.304      0.001       0.131       0.512
Odd                 -0.0847      0.096     -0.878      0.380      -0.274       0.104
Econ-poor           -0.0389      0.097     -0.401      0.689      -0.229       0.151
Econ-mid            -0.0122      0.096     -0.127      0.899      -0.201       0.176
Econ-rich            0.0395      0.097      0.408      0.683      -0.150       0.230
==============================================================================
Omnibus:                      361.332   Durbin-Watson:                   1.972
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              122.762
Skew:                           0.303   Prob(JB):                     2.20e-27
Kurtosis:                       2.112   Cond. No.                         7.91
==============================================================================
```

Feeling Thermometer



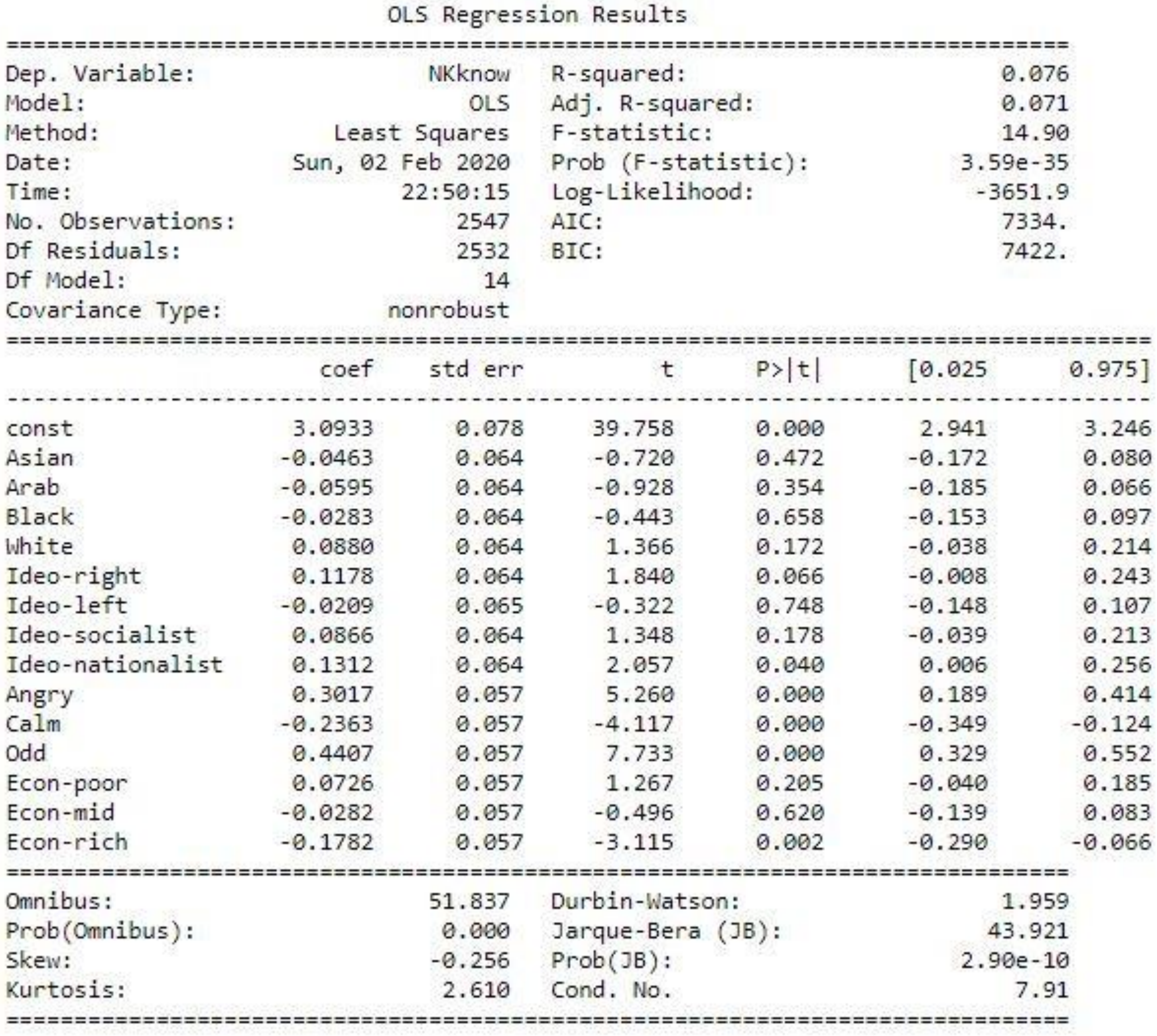
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Feel_1   R-squared:                       0.068
Model:                            OLS   Adj. R-squared:                  0.063
Method:                 Least Squares   F-statistic:                     13.29
Date:                Sun, 02 Feb 2020   Prob (F-statistic):           6.39e-31
Time:                        22:44:48   Log-Likelihood:                -12019.
No. Observations:                2547   AIC:                         2.407e+04
Df Residuals:                    2532   BIC:                         2.415e+04
Df Model:                          14
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const               44.1869      2.078     21.266      0.000      40.113      48.261
Asian                2.9803      1.717      1.736      0.083      -0.386       6.347
Arab                 1.4373      1.714      0.839      0.402      -1.924       4.798
Black                2.5732      1.704      1.510      0.131      -0.769       5.915
White                0.4398      1.719      0.256      0.798      -2.932       3.811
Ideo-right          -3.8653      1.710     -2.260      0.024      -7.219      -0.512
Ideo-left           -0.9643      1.735     -0.556      0.578      -4.367       2.439
Ideo-socialist      -4.8674      1.717     -2.835      0.005      -8.233      -1.501
Ideo-nationalist    -8.1193      1.703     -4.767      0.000     -11.459      -4.779
Angry               -7.2029      1.532     -4.702      0.000     -10.207      -4.199
Calm                 6.9265      1.533      4.518      0.000       3.921       9.932
Odd                 -7.7052      1.522     -5.063      0.000     -10.689      -4.721
Econ-poor           -5.4956      1.529     -3.594      0.000      -8.494      -2.497
Econ-mid            -2.1397      1.516     -1.411      0.158      -5.112       0.833
Econ-rich            2.6192      1.528      1.714      0.087      -0.378       5.616
==============================================================================
Omnibus:                      324.702   Durbin-Watson:                   1.946
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              134.155
Skew:                           0.367   Prob(JB):                     7.39e-30
Kurtosis:                       2.149   Cond. No.                         7.91
==============================================================================
```

Can the adversary be persuaded by evidence?



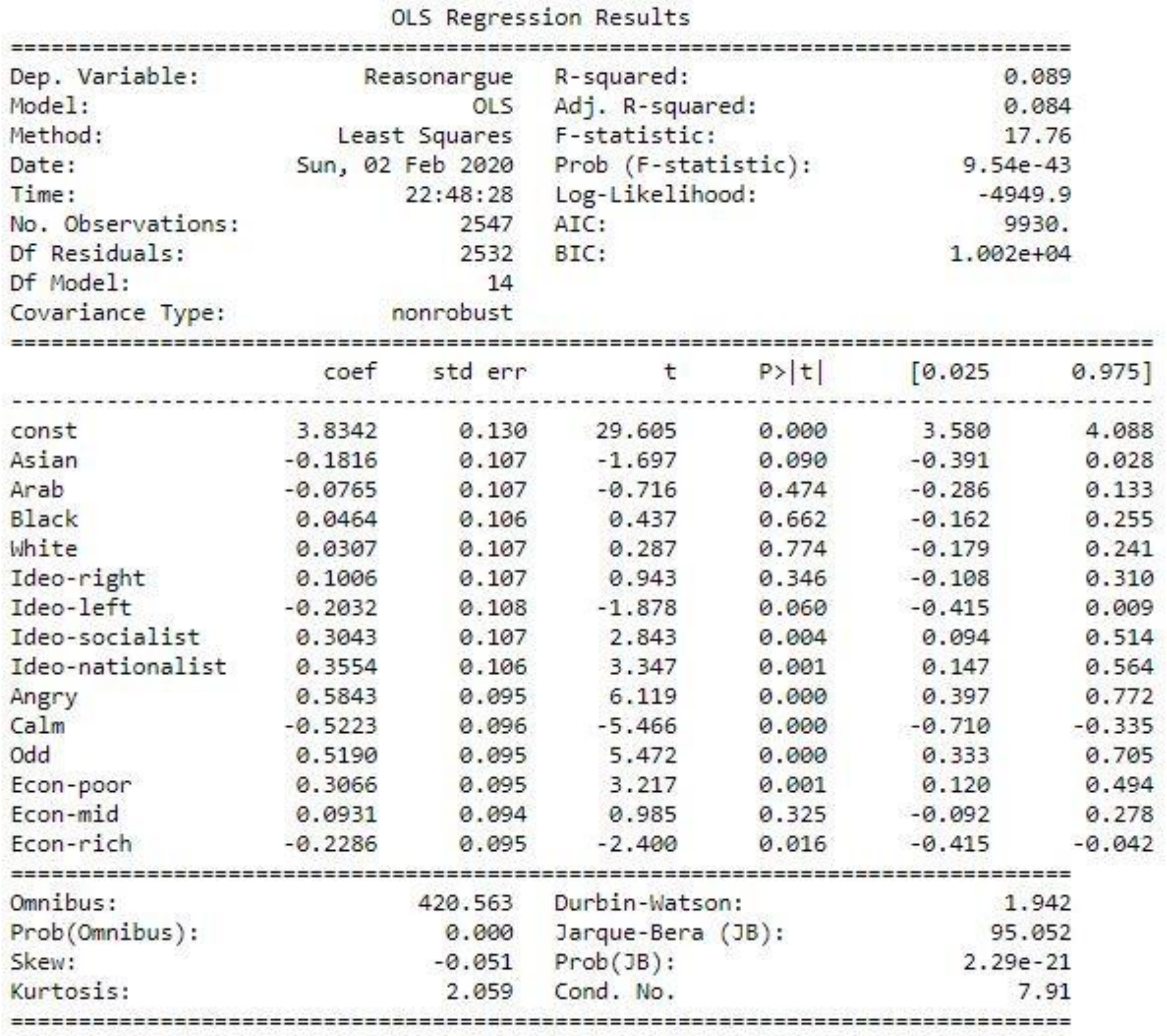
```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:               Reasonev   R-squared:                       0.074
Model:                            OLS   Adj. R-squared:                  0.068
Method:                 Least Squares   F-statistic:                     14.37
Date:                Sun, 02 Feb 2020   Prob (F-statistic):           8.88e-34
Time:                        22:47:09   Log-Likelihood:                -4960.8
No. Observations:                2547   AIC:                             9952.
Df Residuals:                    2532   BIC:                         1.004e+04
Df Model:                          14                                         
Covariance Type:            nonrobust                                         
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
const                  4.0811      0.130     31.377      0.000       3.826       4.336
Asian                 -0.1856      0.107     -1.727      0.084      -0.396       0.025
Arab                  -0.0378      0.107     -0.352      0.725      -0.248       0.173
Black                 -0.0388      0.107     -0.364      0.716      -0.248       0.170
White                 -0.0485      0.108     -0.451      0.652      -0.260       0.163
Ideo-right             0.1768      0.107      1.651      0.099      -0.033       0.387
Ideo-left             -0.0327      0.109     -0.301      0.764      -0.246       0.180
Ideo-socialist         0.2340      0.107      2.178      0.030       0.023       0.445
Ideo-nationalist       0.3498      0.107      3.280      0.001       0.141       0.559
Angry                  0.3809      0.096      3.972      0.000       0.193       0.569
Calm                  -0.5955      0.096     -6.205      0.000      -0.784      -0.407
Odd                    0.4343      0.095      4.559      0.000       0.248       0.621
Econ-poor              0.3101      0.096      3.239      0.001       0.122       0.498
Econ-mid               0.0991      0.095      1.045      0.296      -0.087       0.285
Econ-rich             -0.2286      0.096     -2.390      0.017      -0.416      -0.041
==============================================================================
Omnibus:                      317.829   Durbin-Watson:                   1.986
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               94.704
Skew:                          -0.177   Prob(JB):                     2.72e-21
Kurtosis:                       2.124   Cond. No.                         7.91
==============================================================================
```

Adversary leader Knowledge:



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 NKknow   R-squared:                       0.076
Model:                            OLS   Adj. R-squared:                  0.071
Method:                 Least Squares   F-statistic:                     14.90
Date:                Sun, 02 Feb 2020   Prob (F-statistic):           3.59e-35
Time:                        22:50:15   Log-Likelihood:                -3651.9
No. Observations:                2547   AIC:                             7334.
Df Residuals:                    2532   BIC:                             7422.
Df Model:                          14
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                3.0933      0.078     39.758      0.000       2.941       3.246
Asian               -0.0463      0.064     -0.720      0.472      -0.172       0.080
Arab                -0.0595      0.064     -0.928      0.354      -0.185       0.066
Black               -0.0283      0.064     -0.443      0.658      -0.153       0.097
White                0.0880      0.064      1.366      0.172      -0.038       0.214
Ideo-right           0.1178      0.064      1.840      0.066      -0.008       0.243
Ideo-left           -0.0209      0.065     -0.322      0.748      -0.148       0.107
Ideo-socialist       0.0866      0.064      1.348      0.178      -0.039       0.213
Ideo-nationalist     0.1312      0.064      2.057      0.040       0.006       0.256
Angry                0.3017      0.057      5.260      0.000       0.189       0.414
Calm                -0.2363      0.057     -4.117      0.000      -0.349      -0.124
Odd                  0.4407      0.057      7.733      0.000       0.329       0.552
Econ-poor            0.0726      0.057      1.267      0.205      -0.040       0.185
Econ-mid            -0.0282      0.057     -0.496      0.620      -0.139       0.083
Econ-rich           -0.1782      0.057     -3.115      0.002      -0.290      -0.066
==============================================================================
Omnibus:                       51.837   Durbin-Watson:                   1.959
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.921
Skew:                          -0.256   Prob(JB):                     2.90e-10
Kurtosis:                       2.610   Cond. No.                         7.91
==============================================================================
```

Can the Adversary be Persuaded by reasonable argument?



```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Reasonargue   R-squared:                       0.089
Model:                            OLS   Adj. R-squared:                  0.084
Method:                 Least Squares   F-statistic:                     17.76
Date:                Sun, 02 Feb 2020   Prob (F-statistic):           9.54e-43
Time:                        22:48:28   Log-Likelihood:                -4949.9
No. Observations:                2547   AIC:                             9930.
Df Residuals:                    2532   BIC:                         1.002e+04
Df Model:                          14
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const                3.8342      0.130     29.605      0.000       3.580       4.088
Asian               -0.1816      0.107     -1.697      0.090      -0.391       0.028
Arab                -0.0765      0.107     -0.716      0.474      -0.286       0.133
Black                0.0464      0.106      0.437      0.662      -0.162       0.255
White                0.0307      0.107      0.287      0.774      -0.179       0.241
Ideo-right           0.1006      0.107      0.943      0.346      -0.108       0.310
Ideo-left           -0.2032      0.108     -1.878      0.060      -0.415       0.009
Ideo-socialist       0.3043      0.107      2.843      0.004       0.094       0.514
Ideo-nationalist     0.3554      0.106      3.347      0.001       0.147       0.564
Angry                0.5843      0.095      6.119      0.000       0.397       0.772
Calm                -0.5223      0.096     -5.466      0.000      -0.710      -0.335
Odd                  0.5190      0.095      5.472      0.000       0.333       0.705
Econ-poor            0.3066      0.095      3.217      0.001       0.120       0.494
Econ-mid             0.0931      0.094      0.985      0.325      -0.092       0.278
Econ-rich           -0.2286      0.095     -2.400      0.016      -0.415      -0.042
==============================================================================
Omnibus:                      420.563   Durbin-Watson:                   1.942
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               95.052
Skew:                          -0.051   Prob(JB):                     2.29e-21
Kurtosis:                       2.059   Cond. No.                         7.91
==============================================================================
```