



Madman or Mad Genius? The International Benefits and Domestic Costs of the Madman Strategy

Joshua A. Schwartz

To cite this article: Joshua A. Schwartz (2023): Madman or Mad Genius? The International Benefits and Domestic Costs of the Madman Strategy, Security Studies, DOI: [10.1080/09636412.2023.2197619](https://doi.org/10.1080/09636412.2023.2197619)

To link to this article: <https://doi.org/10.1080/09636412.2023.2197619>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 04 May 2023.



[Submit your article to this journal](#)




[View related articles](#)



[View Crossmark data](#)

Madman or Mad Genius? The International Benefits and Domestic Costs of the Madman Strategy

Joshua A. Schwartz 

ABSTRACT

According to the “Madman Theory” outlined by Daniel Ellsberg and Thomas C. Schelling, and embraced by Presidents Richard Nixon and Donald Trump, being perceived as mad can help make seemingly incredible threats—such as starting a nuclear war—more credible. However, recent research has largely concluded that the Madman Theory does not work. In this study, I theorize that the international benefits of the Madman Theory have been underestimated, but also that there are significant domestic barriers associated with adopting such a strategy that undermine its effectiveness. Through a series of five novel survey experiments, I find evidence that perceived madness provides limited advantages in coercive bargaining vis-à-vis foreign adversaries, but it also entails significant domestic costs that potentially erode its efficacy. Overall, this study provides clearer support for the Madman Theory than most previous literature has found, but also breaks new theoretical ground by analyzing the domestic politics of perceived madness.

“I call it the Madman Theory, Bob. I want the North Vietnamese to believe that I’ve reached the point that 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 is 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.”

—President Richard Nixon¹

In an era of nuclear weapons and mutually assured destruction (MAD), a critical question for scholars and policymakers is how to make threats credible against foreign adversaries, especially those with the “bomb.” The central dilemma is that for a self-interested actor, no political stake is worth total annihilation, and very few are worth the extreme destruction that even a “limited” nuclear war would bring.² Ostensibly, you would

Joshua A. Schwartz is a grand strategy, security, and statecraft postdoctoral fellow at the Massachusetts Institute of Technology and the Harvard Kennedy School.

¹H. R. Haldeman with Joseph DiMona, *The Ends of Power* (New York: Times Books, 1978), 83.

²Moreover, whether a nuclear conflict could remain “limited” in any meaningful sense is highly questionable.

 Supplemental data for this article is available online at <https://doi.org/10.1080/09636412.2023.2197619>.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

have to be mad to risk MAD. However, are there ways that seemingly incredible threats—such as starting a nuclear war—can be made credible?

Scholars have proposed many answers to this question, including brinkmanship, or the manipulation of risk through launch-on-warning nuclear alerts, dangerous air or sea maneuvers, etc.,³ tripwire forces, such as American troops stationed in South Korea; nuclear superiority relative to an opponent;⁴ and public threats that put a leader's reputation on the line.⁵ Early nuclear strategists such as Daniel Ellsberg⁶ and Thomas C. Schelling⁷ also suggested a different mechanism to make ostensibly incredible threats more credible: perceived madness. The argument goes that being perceived as a madman can aid in crisis bargaining: if an adversary thinks you are crazy enough to risk nuclear catastrophe, then they may believe they are better off backing down in order to avoid destruction. Building on this logic, Nixon allegedly proposed the “Madman Theory” in a discussion with his chief of staff, H. R. “Bob” Haldeman. A year later, in October 1969, Nixon applied his theory in Operation Giant Lance: eighteen B-52 bombers armed with nuclear weapons flew toward the Soviet Union and back for eighteen hours straight, aided by midair refueling aircraft.⁸ The operation's goal was to convince the Soviet Union that Nixon would do anything—even risk nuclear conflict—to end the war in Vietnam. Nixon hoped this would then persuade the Soviets to pressure the North Vietnamese to make concessions, allowing for a favorable resolution to the war. However, Nixon was unable to convince the Soviets that he was truly mad, and his gambit ultimately failed.⁹

Nevertheless, President Donald Trump seemingly incorporated elements of the Madman Theory into his broader negotiation philosophy. In a discussion with top cabinet officials regarding the US-South Korea trade deal, Trump reportedly told Robert Lighthizer, “You've got 30 days, and if you don't get concessions then I'm pulling out.” “Ok, well I'll tell the Koreans they've got 30 days,” Lighthizer replied. “No, no, no,” Trump interjected. “That's not how you negotiate. You don't tell them they've got 30 days. You tell them, This guy's so crazy he could pull out any minute

³Thomas C. Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 1966); Robert Powell, *Nuclear Deterrence Theory: The Search for Credibility* (New York: Cambridge University Press, 1990).

⁴Matthew Kroenig, *The Logic of American Nuclear Strategy: Why Superiority Matters* (New York: Oxford University Press, 2018).

⁵James D. Fearon, “Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs,” *Journal of Conflict Resolution* 41, no. 1 (February 1997): 68–90.

⁶Daniel Ellsberg, “The Political Uses of Madness” (lecture, Lowell Institute of the Boston Public Library, Boston, MA, 26 March 1959), <https://ia800102.us.archive.org/20/items/ThePoliticalUsesOfMadness/ELS005-001.pdf>.

⁷Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, MA: Harvard University Press, 1960).

⁸Scott D. Sagan and Jeremi Suri, “The Madman Nuclear Alert: Secrecy, Signaling, and Safety in October 1969,” *International Security* 27, no. 4 (Spring 2003): 150–83.

⁹Todd S. Sechser and Matthew Fuhrmann, *Nuclear Weapons and Coercive Diplomacy* (New York: Cambridge University Press, 2017).

... You tell them if they don't give the concessions now, this crazy guy will pull out of the deal."¹⁰

Trump's threats to "totally destroy" nuclear-armed North Korea and bring "fire and fury like the world has never seen" also appear to accord with this strategy. But is the Madman Theory an effective strategy in coercive bargaining, and if so, under what circumstances? Was Nixon's failure simply due to his being unable to convince the Soviets that he was truly mad?

A wave of recent research has examined the Madman Theory and almost overwhelmingly come to the conclusion that it "doesn't work,"¹¹ is a "myth,"¹² is "clearly harmful to general deterrence,"¹³ and is "just crazy."¹⁴ Building on the foundational work of Roseanne W. McManus,¹⁵ I develop a theory in this article that in the context of seemingly incredible threats—which I define as threats that lack credibility on their face because the costs of carrying them out would be tremendous and clearly exceed the benefits—the international advantages of the Madman Strategy have been underestimated. However, the domestic costs of adopting such a strategy, which are relevant to its success, have also been undertheorized and underestimated.¹⁶

The international benefits of the Madman Strategy have been underestimated because, theoretically, perceived madness should enhance the credibility of a leader's threats. This is especially the case when actually carrying out the threat would be irrational from the perspective of a conventional cost-benefit analysis. Nonetheless, I also conceptualize the Madman Strategy as a two-level game and theorize that being perceived as mad is far from a panacea.¹⁷ Besides the difficulty of convincing an adversary that you are actually mad, the possibility of commitment problems undermining bargaining leverage, and, most importantly, the risks of escalation, I hypothesize that adopting the Madman Strategy comes with significant domestic costs. No previous study has formally theorized or tested how a domestic public views the conscious or unconscious adoption of the

¹⁰Jonathan Swan, "Scoop: Trump Urges Staff to Portray Him as 'Crazy Guy,'" *Axios*, 1 October 2017, <https://www.axios.com/2017/12/15/scoop-trump-urges-staff-to-portray-him-as-crazy-guy-1513305888>.

¹¹Samuel Seitz and Caitlin Talmadge, "The Predictable Hazards of Unpredictability: Why Madman Behavior Doesn't Work," *Washington Quarterly* 43, no. 3 (Fall 2020): 31–46.

¹²Todd S. Sechser and Matthew Fuhrmann, "The Madman Myth: Trump and the Bomb," H-Diplo / ISFF Policy Series, 22 March 2017, <https://issforum.org/policy/1-5w-madman>.

¹³Roseanne W. McManus, "Crazy Like a Fox? Are Leaders with Reputations for Madness More Successful at International Coercion?" *British Journal of Political Science* 51, no. 1 (January 2021): 276.

¹⁴Dani Nedal and Daniel Nexon, "Trump's 'Madman Theory' Isn't Strategic Unpredictability. It's Just Crazy," *Foreign Policy*, 18 April 2017, <https://foreignpolicy.com/2017/04/18/trumps-madman-theory-isnt-strategic-unpredictability-its-just-crazy/>.

¹⁵Roseanne W. McManus, "Revisiting the Madman Theory: Evaluating the Impact of Different Forms of Perceived Madness in Coercive Bargaining," *Security Studies* 28, no. 5 (October–December 2019): 976–1009.

¹⁶I use the terms "Madman Theory" and "Madman Strategy" interchangeably.

¹⁷Robert D. Putnam, "Diplomacy and Domestic Politics: The Logic of Two-Level Games," *International Organization* 42, no. 3 (Summer 1988): 427–60.

Madman Strategy by their leader, but I argue that domestic disapproval is relevant to the Madman Strategy's lack of historical success for three primary reasons.

First, public disapproval disincentivizes leaders from employing the Madman Strategy in the first place, which provides an explanation for its rare adoption in international politics. Second, if adopted, greater latent public disapproval incentivizes leaders to keep threats and actions associated with the Madman Theory secret from their constituents. Although this helps insulate leaders from domestic punishment, it may also reduce threat credibility by diminishing the audience costs associated with backing down from the threat.¹⁸ Empirically, this logic provides an additional explanation for why Nixon deliberately kept Operation Giant Lance secret from the American people, even though doing so reduced the credibility of his threat in the Soviets' eyes.¹⁹ Finally, if leaders adopt the Madman Strategy publicly, then greater domestic disapproval might signal to adversaries that the leader is less likely to actually carry out the threat. For example, strong disapproval of President Trump's "fire and fury" threat among the US public may have suggested to North Korea that Trump was unlikely to follow through on his bellicose rhetoric. In sum, thinking about the Madman Strategy as a two-level game helps explain its relative lack of success.

To test my theory, I conduct a series of survey experiments. An experimental test of the Madman Theory is valuable for three reasons. First, compared to historical case studies on its efficacy,²⁰ experiments can more easily examine counterfactuals. For example, if the Soviets had perceived Nixon as mad, would his threat have been more effective? Second, compared to a large-N analysis of militarized interstate disputes,²¹ experiments can focus on the theoretical comparative advantage of madness: making threats no rational leader would carry out more credible. Although perceived madness could theoretically enhance a leader's ability to coerce adversaries across a range of different crisis scenarios,²² the sample size of historical cases that involve potentially cataclysmic destruction is relatively limited. Finally, experiments can hold relevant factors constant, reducing the chance of omitted variables bias.

While previous research on the Madman Theory focuses exclusively on how this strategy affects the perceptions of leaders, the views of the general public are critically important as well and understudied. Policymakers

¹⁸Fearon, "Signaling Foreign Policy Interests."

¹⁹Sagan and Suri, "Madman Nuclear Alert."

²⁰Sechser and Fuhrmann, *Nuclear Weapons and Coercive Diplomacy*; McManus, "Revisiting the Madman Theory."

²¹McManus, "Crazy Like a Fox?"

²²McManus, "Crazy Like a Fox?" 279.

respond to and are constrained by public opinion.²³ Thus, public willingness to capitulate when challenged by a foreign adversary that is perceived as mad will reduce the domestic constraints leaders face to backing down. Public opposition to capitulation, on the other hand, will stiffen the spine of leaders and make the Madman Strategy less likely to succeed. Additionally, Joshua D. Kertzer conducted a meta-analysis of 162 paired experiments on members of the public and elites, and he finds that they both generally respond to treatments in the same ways.²⁴ Of the 162 treatment effects he analyzes, over 98% do not differ in sign (that is, whether the relationship is positive or negative) between members of the public and elites, and almost 90% do not differ in size (that is, how large the effect is). This suggests that the results presented here among the public may also speak to the effectiveness of the Madman Theory among political leaders.

To test the efficacy of the Madman Theory, I conduct five survey experiments on a diverse national sample of Americans. The first four examine how the American public views a foreign leader's adoption of the Madman Strategy—specifically, hypothetical future leaders of North Korea, Iran, and Russia. In accordance with my theory, I find evidence that nuclear threats made by a North Korean leader and major conventional threats made by an Iranian leader framed as “mad” are more effective than those made by North Korean or Iranian leaders framed as more “sane.” These findings lend clearer support for the Madman Theory than most previous literature has found and suggest that perceived madness has some ability to make seemingly incredible threats more credible.

However, I also find evidence that the Madman Strategy has limitations and downsides. Although perceived madness does make nuclear threats from a future Russian leader more credible, it does not make them significantly more effective; that is, it does not make the American public more willing to give in to Russian demands. This suggests madness may be less effective in interactions among major powers, which provides another explanation for why Nixon did not successfully coerce the Soviet Union via the Madman Strategy. Furthermore, the fifth experiment I conduct, which analyzes how the American public views the adoption of the Madman Strategy by their own president, finds strong domestic opposition in accordance with my theoretical expectations. This helps explain the Madman Strategy's relative lack of success in the historical record.

It is important to emphasize two scope conditions for these findings. First, they hold in the context of threats that lack credibility on their face because the costs of carrying said threats out would be enormous and

²³Michael Tomz, Jessica L. P. Weeks, and Keren Yarhi-Milo, “Public Opinion and Decisions about Military Force in Democracies,” *International Organization* 74, no. 1 (Winter 2020): 119–43.

²⁴Joshua D. Kertzer, “Re-assessing Elite-Public Gaps in Political Behavior,” *American Journal of Political Science* 66, no. 3 (July 2022): 539–53.

likely exceed their benefits. Examples include executing nuclear threats against nuclear-armed adversaries with a secure second-strike capability, or following through on a significant conventional threat against a much-superior enemy. Alternatively, for threats with costs that are much lower compared to their benefits, perceived madness may not do much to enhance threat credibility. For instance, the baseline credibility of a non-mad leader's threat to impose sanctions on a country is likely relatively high since carrying out this threat is oftentimes not particularly costly to the country imposing sanctions. As such, the marginal benefit of perceived madness may be relatively low and not increase threat credibility much. Similarly, threats that are much less costly to carry out may be met with relatively high levels of domestic approval—irrespective of whether the leader in question is perceived as mad—because the stakes are lower and thus the general public may be more willing to defer to political leadership. This scope condition may help explain some of the divergences in results between this project and previous literature, as many past studies evaluate the impact of madness across a wider range of crisis scenarios that may involve threats that are quite credible on their face and less costly to implement.²⁵ Although I believe it is appropriate to analyze the impact of madness on the effectiveness of incredible threats on their face given that this is the context in which the theory was originally developed,²⁶ future work should tease out the varying impacts of madness across different threat categories.

Second, the Madman Strategy can only be effective if leaders convince their adversaries they are possibly or definitely mad, and if they clearly communicate what they want. During Operation Giant Lance, Nixon struggled to achieve either objective.²⁷ Therefore, this second scope condition may also help explain the difference between the findings presented here and in previous literature. In an experimental setting, it is likely easier to convince respondents that a leader—especially a hypothetical one—is actually mad, and to clearly communicate that leader's demands. Under these ideal conditions, perhaps some Madman Theory skeptics would concede perceived madness can yield limited benefits.

This article makes several contributions. First, it contributes to the burgeoning literature on the Madman Theory specifically and on irrationality more generally.²⁸ Second, the results strengthen the emerging consensus that leader attributes and psychology matter for international

²⁵McManus, "Crazy Like a Fox?"

²⁶Ellsberg, "Political Uses of Madness"; Schelling, *Strategy of Conflict*.

²⁷Sechser and Fuhrmann, *Nuclear Weapons and Coercive Diplomacy*.

²⁸Frank C. Zagare, "Rationality and Deterrence," *World Politics* 42, no. 2 (January 1990): 238–60; Avidit Acharya and Edoardo Grillo, "War with Crazy Types," *Political Science Research and Methods* 3, no. 2 (May 2015): 281–307; Andrew T. Little and Thomas Zeitzoff, "A Bargaining Theory of Conflict with Evolutionary Preferences," *International Organization* 71, no. 3 (Summer 2017): 523–57.

politics.²⁹ Third, it contributes to the literature on how domestic politics influences coercive bargaining between states.³⁰ Finally, the article adds to scholarship on nuclear coercion by suggesting that nuclear blackmail is somewhat more effective when leaders are perceived as mad.³¹ In sum, I find that being perceived as mad confers some limited advantages in coercive bargaining vis-à-vis foreign adversaries, but it is also associated with significant domestic costs that can impair its efficacy.

International-Level Impact of the Madman Strategy

Most recent research that has examined the Madman Theory has concluded it does not work, primarily because of commitment problems or because leaders have been unable to convince adversaries that they are truly mad.³² From a theoretical point of view, I argue in this section that these criticisms do not preclude the Madman Theory from providing advantages in coercive bargaining vis-à-vis foreign adversaries since perceived madness should entail a universal benefit: enhanced threat credibility.

Beginning with definitions, McManus³³ proposes four types of perceived madness³⁴ broken down along two dimensions. The first dimension distinguishes between whether a leader is perceived as (a) deviating from consequence-based decision making or (b) having extreme foreign policy preferences and beliefs. If a leader deviates from consequence-based decision making, then they—at least occasionally—make decisions based on something other than a logical cost–benefit analysis. Making decisions based on emotion rather than reason would be one prominent example. For instance, if a leader decided to launch a nuclear attack in a fit of anger without thinking through the implications, then that would deviate from consequence-based decision making. This is not to say all emotionally

²⁹Giacomo Chiozza and H. E. Goemans, *Leaders and International Conflict* (New York: Cambridge University Press, 2011); Michael C. Horowitz, Allan C. Stam, and Cali M. Ellis, *Why Leaders Fight* (New York: Cambridge University Press, 2015); Keren Yarhi-Milo, *Who Fights for Reputation: The Psychology of Leaders in International Conflict* (Princeton, NJ: Princeton University Press, 2018).

³⁰Joshua D. Kertzer and Ryan Brutger, “Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory,” *American Journal of Political Science* 60, no. 1 (January 2016): 234–49; Joshua A. Schwartz and Christopher W. Blair, “Do Women Make More Credible Threats? Gender Stereotypes, Audience Costs, and Crisis Bargaining,” *International Organization* 74, no. 4 (Fall 2020): 872–95.

³¹Sechser and Fuhrmann, *Nuclear Weapons and Coercive Diplomacy*; Kroenig, *Logic of American Nuclear Strategy*.

³²Sagan and Suri, “Madman Nuclear Alert”; Sechser and Fuhrmann, *Nuclear Weapons and Coercive Diplomacy*; Seitz and Talmadge, “Predictable Hazards of Unpredictability”; McManus, “Crazy Like a Fox?” Seitz and Talmadge term the former issue the “assurance problem” and the latter issue the “signaling problem.”

³³McManus, “Revisiting the Madman Theory.”

³⁴Perceived madness rather than actual madness is what matters for international crisis bargaining. The bargaining effect of a leader perceived as mad—but not actually mad—should be the same as a leader who is both perceived as and actually mad.

driven decisions are irrational,³⁵ but, at least occasionally, emotion can lead to choices that do not make sense from a rational cost–benefit analysis. Mental illness could also potentially fall under this category.

Alternatively, if a leader has extreme preferences, then they may have a very high risk tolerance, highly value the issue at stake in a conflict, and/or put a low estimation on the costs of war. For example, if a leader was willing to run a high risk of complete destruction—say 75 percent—then that might indicate extreme preferences. Since a leader can hold such a high risk tolerance while also understanding the dire consequences if things go wrong, they may not deviate from consequence-based decision making. However, note that whether extreme preferences qualify as “madness” is somewhat subjective, as most rationalist models would anticipate that high resolve is an asset in crisis bargaining. After all, President John F. Kennedy was willing to run a very high risk of nuclear war during the Cuban Missile Crisis, and it ostensibly paid off. Nevertheless, there may be a threshold at which extreme preferences could be considered mad compared to a typical leader’s beliefs under the common dictionary definition of the term: “behavior or thinking that is very foolish or dangerous.”³⁶

The second dimension of madness, according to McManus, is whether a leader’s madness is perceived to be: (a) situational or (b) dispositional. If a leader’s madness is perceived to be situational, then that means it is expected to only apply under particular conditions. For example, a leader might be willing to run a 75% risk of complete destruction over a historical border dispute because the land has great sentimental value to the leader, but only a 1% risk of destruction over a maritime dispute. On the other hand, if a leader’s madness is perceived to be dispositional, then that means it is expected to apply in most or all circumstances. For instance, a leader may be willing to run a 75% risk of complete destruction over all international controversies and always make decisions based on emotion rather than logic.

Are any of these four types of perceived madness likely to be helpful for coercive bargaining relative to a leader viewed as more sane? All else equal, and in accordance with Ellsberg³⁷ and Schelling,³⁸ I theorize that being perceived as mad—in any form—does provide a universal bargaining advantage: it makes seemingly incredible threats more credible. For threats to be effective, their target must believe that defying the threat will bring punishment. If the threat is unlikely to be carried out no matter what the

³⁵Jonathan Mercer, “Rationality and Psychology in International Politics,” *International Organization* 59, no. 1 (Winter 2005): 77–106.

³⁶McManus, “Revisiting the Madman Theory,” 981.

³⁷Ellsberg, “Political Uses of Madness.”

³⁸Schelling, *Strategy of Conflict*.

target does, then why capitulate? This was the critical flaw with Nixon's Operation Giant Lance. Since launching a nuclear war over Vietnam—an economically and militarily weak country—was unreasonable,³⁹ and the Soviets did not buy Nixon's "madman" act, his implicit threat did not change Soviet behavior.⁴⁰ Alternatively, if Nixon had been viewed as an actual madman, the threat should have been more credible. For example, if Nixon was perceived as deviating from consequence-based decision making, then the Soviets might have believed he would launch a nuclear attack against them in a fit of anger without thinking through the consequences. If Nixon was seen as having extreme preferences, then the Soviets might have concluded he would launch a nuclear war due to how much he valued winning in Vietnam. This is similar to Schelling's contention that patients in mental institutions (who may deviate from consequence-based decision making) and anarchists (who have extreme preferences) can more credibly threaten to kill themselves than the average person.⁴¹ Therefore, although critics of the Madman Theory are indeed correct that (a) actually being perceived as somewhat mad—unlike Nixon—is a prerequisite for the Madman Strategy to be successful, and (b) it is hard to achieve this in practice, these criticisms do not rule out the theoretical benefits of madness with respect to enhanced threat credibility.

In accordance with this view, McManus finds evidence that madness can be useful for crisis bargaining by examining two case studies: Adolf Hitler in the Sudetenland Crisis and Nikita Khrushchev in the Berlin Crisis.⁴² Hitler was perceived as having extreme preferences and Khrushchev was seen as occasionally deviating from consequence-based decision making, which made their respective threats more credible. More formally, Avidit Acharya and Edoardo Grillo construct a model that shows pretending to be crazy can lead to better bargaining outcomes under certain conditions.⁴³ Also in a formal model, Andrew T. Little and Thomas Zeitzoff find that "irrationally tough" actors may be favored from an evolutionary perspective.⁴⁴ Finally, studies in psychology have further shown that emotional volatility⁴⁵ and anger⁴⁶ can be assets in negotiations.

³⁹Though *risking* nuclear war over Vietnam could be rational, in accordance with the literature on brinkmanship.

⁴⁰Sagan and Suri, "Madman Nuclear Alert"; Sechser and Fuhrmann, *Nuclear Weapons and Coercive Diplomacy*.

⁴¹Schelling, *Arms and Influence*, 37–38.

⁴²McManus, "Revisiting the Madman Theory."

⁴³Acharya and Grillo, "War with Crazy Types."

⁴⁴Little and Zeitzoff, "Bargaining Theory of Conflict with Evolutionary Preferences."

⁴⁵Marwan Sinaceur et al., "The Advantages of Being Unpredictable: How Emotional Inconsistency Extracts Concessions in Negotiation," *Journal of Experimental Social Psychology* 49, no. 3 (May 2013): 498–508.

⁴⁶Marwan Sinaceur and Larissa Z. Tiedens, "Get Mad and Get More than Even: When and Why Anger Expression Is Effective in Negotiations," *Journal of Experimental Social Psychology* 42, no. 3 (May 2006): 314–22; Gerben A. Van Kleef and Stéphane Côté, "Expressing Anger in Conflict: When It Helps and When It Hurts," *Journal of Applied Psychology* 92, no. 6 (November 2007): 1557–69.

What about the commitment problems associated with the Madman Theory? According to McManus, Madman Theory's central disadvantage, especially with respect to dispositional forms of madness, is that it can heighten commitment problems.⁴⁷ Commitment problems arise when one or both parties to a conflict cannot credibly commit to uphold a negotiated settlement in the future. For threats to be effective, threat targets must believe that defying one will bring punishment, but also—critically—that capitulating to it will bring rewards or the lack of punishment. If leaders are perceived as dispositionally mad, then they may not be trusted to uphold a negotiated settlement. For example, if a leader has extreme preferences, then giving in to one demand may simply lead to another.⁴⁸ This is the case because the leader is extreme on all or most issues, and so surrendering on one issue will likely not satiate their appetite for concessions. If a leader deviates from consequence-based decision making, then giving in to their demands may also not allow you to avoid punishment if, say, their temper flares up again. This is particularly likely for a leader perceived as dispositionally mad, as giving in on one issue will not reduce their emotional volatility on others. Since conceding to the demands of a leader perceived as dispositionally mad brings no guaranteed benefits, targets may be better off refusing their demands today rather than giving in and facing a stronger opponent tomorrow.

McManus finds evidence that perceived dispositional madness is harmful for crisis bargaining by examining two case studies: Saddam Hussein's interactions with the George W. Bush administration and Muammar Gaddafi's interactions with the Reagan administration.⁴⁹ While Saddam was perceived as having extreme preferences and Gaddafi was viewed as deviating from consequence-based decision making, both types of madness were seen as dispositional. As a result, the leaders were unable to offer credible reassurances to the Bush and Reagan administrations, which significantly reduced the effectiveness of their threats. This was the case even though their perceived madness allowed them to make more credible threats, the central advantage of madness discussed previously.⁵⁰ In a large-N statistical analysis of militarized interstate disputes, McManus also finds that perceived madness is harmful to general deterrence and is typically harmful or has an insignificant effect on crisis bargaining.⁵¹ This is the strongest evidence to date against the Madman Strategy's effectiveness.

⁴⁷McManus, "Revisiting the Madman Theory"; McManus, "Crazy Like a Fox?"

⁴⁸Andrew H. Kydd and Roseanne W. McManus, "Threats and Assurances in Crisis Bargaining," *Journal of Conflict Resolution* 61, no. 2 (February 2017): 325–48.

⁴⁹McManus, "Revisiting the Madman Theory."

⁵⁰McManus, "Revisiting the Madman Theory," 1000.

⁵¹McManus, "Crazy Like a Fox?"

Though critics are correct that commitment problems may indeed impede the effectiveness of the Madman Theory, there are two scenarios in which these effects may be attenuated. First, when nuclear war or a potentially devastating conventional war is on the table, it would make logical sense to heavily discount the future and focus on the near-term crisis scenario. Thus, even if madness does inherently increase commitment problems, the positive impact of enhanced threat credibility in the short term from being perceived as mad may outweigh the negative impact of heightened commitment problems in the long term.

Second, one must remember that commitment problems are not unique to leaders seen as crazy. Leaders viewed as sane can also suffer from a relative inability to make credible commitments. For example, a perfectly sane leader could face a severe commitment problem if they lead a rising power that has structural incentives to try and overturn an agreement with a declining power in the future. Consequently, the effectiveness of the Madman Theory should depend on whether the counterfactual involves a leader who is perceived as (a) sane and able to make credible commitments, or (b) sane and unable to make credible commitments. In the former, the commitment problem criticism is indeterminate from a theoretical perspective because it is unclear whether the positive impact of madness on threat credibility will outweigh the negative impact of heightened commitment problems, or vice versa. In the latter, both mad and sane leaders face commitment problems, and thus leaders perceived as mad should have a clearer advantage due to their theoretical ability to make more credible threats. Of course, in the real world, the magnitude of commitment problems associated with perceived madness may very well outweigh other types of commitment problems. Yet the logic presented here suggests that commitment problems do not theoretically preclude the Madman Strategy from providing net international-level benefits.

Overall, this discussion suggests three central hypotheses. First, all else equal, threats that are seemingly incredible to carry out from leaders framed as mad should be more effective than similar threats from leaders framed as more sane. In other words, holding the severity of commitment problems constant, madness should be an asset in coercive bargaining. I add the qualifier “seemingly incredible to carry out” because, according to Ellsberg and Schelling, the comparative advantage of madness should be in making ostensibly incredible threats—that if carried out would involve cataclysmic costs that appear to clearly outweigh the benefits—more credible.⁵² This constitutes an important scope condition for my theory and findings.

⁵²Ellsberg, “Political Uses of Madness”; Schelling, *Strategy of Conflict*.

Hypothesis 1

(H1). International-level madness advantage: Seemingly incredible threats from leaders framed as mad should be more effective than similar threats from leaders framed as more sane.

Second, as commitment problems become more severe, seemingly incredible threats—whether from a perceived madman or a leader viewed as sane—should be less effective:

Hypothesis 2

(H2). Commitment problems disadvantage: As commitment problems become more severe, seemingly incredible threats should be less effective for both leaders framed as mad and sane.

Finally, as a hard test of the Madman Theory, I propose the following hypothesis, which corresponds to a situation where madness is associated with heightened commitment problems (perhaps due to dispositional madness) relative to the counterfactual:

Hypothesis 3

(H3). Hard test for madness: Seemingly incredible threats from leaders who face more severe commitment problems and are framed as mad should be more effective than similar threats from leaders who face less severe commitment problems and are framed as more sane.

Domestic-Level Impact of the Madman Strategy

I expect that perceived madness confers certain advantages vis-à-vis foreign adversaries, but what is the impact of perceived madness on a leader's domestic support? Although no study has examined this question, it is relevant to the success of the Madman Strategy. For example, if the public disapproves of the Madman Strategy, then that undermines its effectiveness for three primary reasons. First, public disapproval disincentivizes leaders from employing the Madman Strategy in the first place. Second, if adopted, greater public disapproval incentivizes leaders to keep threats and actions associated with the Madman Theory secret from domestic constituents. Although this helps insulate leaders from domestic punishment, it may also reduce threat credibility by diminishing the audience costs associated with backing down from the threat. Finally, if leaders adopt the Madman Strategy publicly, then greater domestic disapproval might signal to adversaries that the leader is less likely to carry out the threat. By contrast, if the public views the Madman Strategy

favorably, leaders would have greater incentives to adopt it, and they would be more likely to adopt it publicly and trigger the possibility of audience costs; further, foreign leaders might be more likely to believe the threat would be executed. In sum, conceptualizing the Madman Strategy as a two-level game that takes place at the international and domestic levels can provide new insights into its efficacy.

A priori, should we expect the domestic public to view leader madness favorably or unfavorably? As Schelling argued, the Madman Strategy can be beneficial for coercive bargaining because “it does not always help to be, or to be believed to be, fully rational, cool-headed, and in control of oneself.”⁵³ However, most people surely want their leader to be fully rational, cool-headed, and in control of oneself. Thus, I theorize that perceived madness should be harmful to a leader’s domestic favorability.

First, consider a leader who is perceived as deviating from consequence-based decision making. In the extreme, the leader may be judged as having a severe mental illness. Such a condition may confer bargaining advantages since the leader could more credibly threaten to commit a suicidal act, such as starting a nuclear war. However, given the well-known stigma against those with mental illness, and the prospect of nuclear destruction, I would expect members of the public to be highly concerned regarding such a leader.⁵⁴ Indeed, a leader’s perceived “competence” is an important factor affecting their approval among the general population, and thus a leader seen as mad due to deviation from consequence-based decision making may be viewed as incompetent, which, in turn, may lead to greater levels of disapproval.⁵⁵ Even if a leader is not perceived as having a mental illness, previous studies have shown that displaying anger—a common feature of deviating from consequence-based decision making—reduces a politician’s general approval.⁵⁶ In other words, perceived madness may reduce a leader’s approval through traditional mechanisms discussed in prior literature, such as competence and temperament.

Even for a leader perceived as mad due to holding extreme preferences, I would expect disapproval to be higher. If we assume actors consider the status quo as being alive, then they should be risk averse when it comes to endangering their lives in order to achieve gains.⁵⁷ The same, I believe,

⁵³Schelling, *Arms and Influence*, 37.

⁵⁴Peter John Loewen and Ludovic Rheaume, “Voters Punish Politicians with Depression,” *British Journal of Political Science* 51, no. 1 (January 2021): 427–36.

⁵⁵James N. Druckman, Lawrence R. Jacobs, and Eric Ostermeier, “Candidate Strategies to Prime Issues and Image,” *Journal of Politics* 66, no. 4 (November 2004): 1180–1202.

⁵⁶McManus, “Revisiting the Madman Theory”; Seanon S. Wong, “Stoics and Hotheads: Leaders’ Temperament, Anger, and the Expression of Resolve in Face-to-Face Diplomacy,” *Journal of Global Security Studies* 4, no. 2 (April 2019): 190–208; Deborah Jordan Brooks, “Testing the Double Standard for Candidate Emotionality: Voter Reactions to the Tears and Anger of Male and Female Politicians,” *Journal of Politics* 73, no. 2 (April 2011): 597–615.

⁵⁷Daniel Kahneman and Amos Tversky, “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica* 47, no. 2 (March 1979): 263–92.

would hold true in a nuclear game of chicken where a domestic leader is threatening nuclear war. Moreover, previous studies have shown that leaders have incentives to prove to their constituents that they lack extreme foreign policy preferences.⁵⁸ US leaders with extreme preferences may also be met with greater levels of domestic disapproval if the public worries extreme foreign policy preferences could undermine the country's global reputation. For example, it could increase US reputation for aggression, which might threaten other states and cause them to balance against America.⁵⁹

On the other hand, there may be circumstances where leaders perceived as mad face relatively lower levels of disapproval. For instance, the “rally ‘round the flag” phenomenon, where support for national leaders increases in response to a foreign threat and/or the use of force abroad, could help mitigate baseline levels of disapproval for leaders perceived as mad.⁶⁰ Likewise, if a leader perceived as mad uses a humanitarian frame to justify their actions instead of less popular frames such as internal political change, then that might increase public support, especially if domestic audiences have an emotional connection to the group that is (allegedly) in danger.⁶¹ But for the reasons noted above, and given I am focused on threats that entail extremely high costs if carried out that likely exceed the benefits, I still generally expect domestic audiences to view perceived madness unfavorably.

In accordance with my expectations, President Nixon believed madness would not be popular with the American people. In a private conversation in the Oval Office in April 1973 (captured on tape), he said: “We are never going to have a madman as president, in this office. We never have, probably never will have. Ours [system] throws them out ... about every four years, if a guy shows that he's [unclear phrase], out!”⁶² In fact, domestic opposition to escalation in the Vietnam War was one reason why Nixon deliberately kept Operation Giant Lance secret from the American people, and he also may have been concerned that they would view such a risky operation as “crazy.”⁶³

Similarly, Alexander Hamilton believed extreme emotional volatility was disqualifying for a president. In his infamous open letter “Concerning the Public Conduct and Character of John Adams, Esq., President of the United States,” Hamilton said, “There are great and intrinsic defects in his character, which unfit him for the office of Chief Magistrate ... It is a fact that he is

⁵⁸Michaela Mattes and Jessica L. P. Weeks, “Hawks, Doves, and Peace: An Experimental Approach,” *American Journal of Political Science* 63, no. 1 (January 2019): 53–66.

⁵⁹Mark J. C. Crescenzi, *Of Friends and Foes: Reputation and Learning in International Politics* (Oxford: Oxford University Press, 2018).

⁶⁰John E. Mueller, “Presidential Popularity from Truman to Johnson,” *American Political Science Review* 64, no. 1 (March 1970): 18–34.

⁶¹Bruce W. Jentleson and Rebecca L. Britton, “Still Pretty Prudent: Post–Cold War American Public Opinion on the Use of Military Force,” *Journal of Conflict Resolution* 42, no. 4 (August 1998): 395–417.

⁶²Tim Naftali, “The Problem with Trump’s Madman Theory: It Didn’t Work for Nixon. It’s Even Less Likely to Work Now,” *Atlantic*, 4 October 2017, <https://www.theatlantic.com/international/archive/2017/10/madman-theory-trump-north-korea/542055/>.

⁶³Sagan and Suri, “Madman Nuclear Alert.”

often liable to paroxysms of anger, which deprive him of self command, and produce very outrageous behaviour to those who approach him.”⁶⁴ And recall that Adams was a member of Hamilton’s own Federalist party! Hamilton’s successor as treasury secretary and a member of Adams’s cabinet also noted in a private letter that “the people believe that their President is Crazy,” and Benjamin Franklin once remarked, “but sometimes and in some things [Adams] is absolutely out of his senses.”⁶⁵ This is not to say Adams was actually insane (though he clearly had a violent temper), but the perception of emotional volatility and even madness convinced some he was unfit to be president.

The Trump presidency also illustrates the domestic costs of being perceived as a madman. Given Trump’s nuclear saber-rattling⁶⁶ and perceived erratic, child-like behavior,⁶⁷ scholars, segments of the public, and even members of Congress have become concerned with the president’s unilateral authority to employ nuclear weapons. Moreover, President Trump’s nuclear threats against North Korea were not received well by the American public. According to an Associated Press-NORC poll conducted in the aftermath of Trump’s bellicose threats, two-thirds of Americans believed the threats made the situation with North Korea worse, and less than 10 percent believed them to be helpful.⁶⁸ Even a Fox News poll conducted after the infamous “fire and fury” threat found that 70 percent of respondents disapproved of the president’s incendiary declarations.⁶⁹ These high disapproval numbers may have undermined President Trump’s threats by suggesting to North Korea that he was relatively unlikely to actually follow through on this bellicose rhetoric.

Overall, this discussion suggests two central hypotheses. First, consider the domestic impact of being framed as mad and making a seemingly incredible threat relative to maintaining a more conciliatory status quo policy. In this scenario, I expect the public will prefer the status quo. Although making a threat may increase the chances of bargaining success, it also raises the chances of a devastating outcome, which I hypothesize the public will generally want to avoid. Again, I add the qualifier “seemingly incredible” because threats that do not risk significant costs that likely exceed the benefits if carried out (for example, threats to impose sanctions) are unlikely to provoke as significant a reaction among the general population.

⁶⁴Alexander Hamilton, “Concerning the Public Conduct and Character of John Adams, Esq., President of the United States,” 24 October 1800, *National Archives Founders Online*, <https://founders.archives.gov/documents/Hamilton/01-25-02-0110-0002>.

⁶⁵Ron Chernow, *Alexander Hamilton* (New York: Penguin Books, 2004), 620–23.

⁶⁶Nina Tannenwald, “How Strong Is the Nuclear Taboo Today?” *Washington Quarterly* 41, no. 3 (Fall 2018): 89–109.

⁶⁷Daniel W. Drezner, *The Toddler in Chief: What Donald Trump Teaches Us about the Modern Presidency* (Chicago: University of Chicago Press, 2020).

⁶⁸Matthew Pennington and Emily Swanson, “Poll: Most Say Trump Making North Korea Situation Worse,” *Fox 4 News*, 11 October 2017, <https://www.fox4news.com/news/poll-most-say-trump-making-north-korea-situation-worse>.

⁶⁹Cristiano Lima, “Poll: 70 Percent Find Trump Rhetoric on North Korea ‘Not Helpful,’” *Politico*, 27 September 2017, <https://www.politico.com/story/2017/09/27/trump-north-korea-rhetoric-poll-243238>.

Hypothesis 4

(H4). Domestic-level madness disadvantage A: Seemingly incredible threats from leaders framed as mad should be met with greater disapproval from the domestic public than maintaining the status quo.

Second, I expect that threats from leaders framed as mad will be less popular than threats from leaders framed as more sane. While perceived madness may again increase the chances that the foreign adversary will capitulate, it also increases the probability that the leader will follow through on their threat, potentially leading to a disastrous outcome.

Hypothesis 5

(H5). Domestic-level madness disadvantage B: Seemingly incredible threats from leaders framed as mad should be met with greater disapproval from the domestic public than similar threats from leaders framed as more sane.

Given that the commitment problems a leader faces internationally are likely less of a concern for a domestic audience, I focus on the madness types of extreme preferences and deviating from consequence-based decision making more generally for this analysis. Future studies should consider how domestic approval of a leader changes based on situational versus dispositional types of madness.⁷⁰

Although I believe the above hypotheses are likely to hold on average, it is important to note factors that may make public opinion a less significant impediment to the Madman Strategy. First, if the general public has itself become radicalized and holds extreme preferences, either independently or due to elite persuasion (even, potentially, by the mad leader themselves), then they may support a leader who makes bellicose threats that are seemingly incredible to carry out. As discussed previously, McManus finds evidence that Hitler's perceived madness enhanced the credibility of his threats. Given that many members of the German public had been radicalized under the Nazi regime, public opinion was likely not a significant constraint on Hitler's actions.

Second, authoritarian regimes—and especially those that control the information environment—may face relatively lower constraints from public opinion given the nature of their political systems. This was likely one reason why Khrushchev's perceived madness made his threats during the Berlin Crisis more credible, and why leaders such as Kim Jong-un and Vladimir Putin can make nuclear threats without significant public

⁷⁰For example, we might expect that a leader perceived as dispositionally deviating from consequence-based decision making would face lower levels of public approval than a leader only situationally deviating from rational cost-benefit analyses.

backlash.⁷¹ On the other hand, a growing body of research notes that public opinion in autocracies plays a bigger role than previously believed, suggesting domestic political constraints against the Madman Strategy may still operate to some extent even in nondemocracies.⁷²

Finally, even in democracies, public opinion cannot necessarily prevent a leader from carrying out a “crazy” threat. For example, in the United States the president holds the ultimate authority to launch nuclear weapons. Consequently, even though Trump faced public backlash for his bellicose threats toward North Korea—thereby incentivizing him not to carry out those threats—were he determined to launch a nuclear attack, then he likely would have had the legal authority to do so. Adversaries also know about the US president’s unilateral authority to use nuclear weapons, meaning public disapproval of nuclear threats may not totally undermine the credibility of such threats if a leader is perceived as mad. Yet, on average, I still expect public opinion to disincentivize leaders from employing the Madman Strategy and thus undermine its effectiveness.

Study 1: Experimental Design

To test the impact of perceived madness on coercive bargaining at the international level, I designed and administered a between-subjects experiment on a diverse national sample of 434 Americans recruited through Lucid in March 2020, which uses quota sampling to match census benchmarks on age, gender, ethnicity, and region. Lucid has been shown to perform well replicating previous studies, increasing confidence that the results are both externally and internally valid.⁷³ However, as Peter M. Aronow et al. document, attention rates on Lucid declined during the COVID-19 pandemic, which was when this article’s studies were fielded.⁷⁴ Nonetheless, if anything, this should bias against finding any significant effect for the Madman Theory. Kyle Peyton, Gregory A. Huber, and Alexander Coppock’s research also establishes that prepandemic experimental findings replicate on Lucid during the pandemic in terms of sign and statistical significance, though effect sizes are somewhat reduced due

⁷¹Radicalized general populations may also play a role in these examples as well.

⁷²Jidong Chen, Jennifer Pan, and Yiqing Xu, “Sources of Authoritarian Responsiveness: A Field Experiment in China,” *American Journal of Political Science* 60, no. 2 (April 2016): 383–400.

⁷³Alexander Coppock and Oliver A. McClellan, “Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents,” *Research & Politics* 6, no. 1 (January–March 2019): 1–14.

⁷⁴Peter M. Aronow et al., “Evidence of Rising Rates of Inattentiveness on Lucid in 2020,” *OSF*, 15 September 2020, <https://osf.io/preprints/socarxiv/8sbe4/>. About 77% of respondents passed the post-treatment attention check for Studies 1 and 3–5, and about 64% passed for Study 2. Results for respondents that passed the attention check are in Tables A.3, A.9, A.17, A.18, A.27, and A.33 in the online appendix. However, prior research establishes that dropping respondents who failed a post-treatment attention check can bias results.

to diminished attention.⁷⁵ I also block on respondent party identification to ensure approximately equal numbers of Democrats, independents, and Republicans in each treatment.

The experiment involves the leader of North Korea threatening nuclear war unless the United States removes economic sanctions, and I experimentally manipulate whether the leader is framed as mad. I opt for a more realistic survey scenario involving North Korea rather than using a generic foreign country (“Country B”) because US respondents would likely attempt to guess the identity of the foreign country in the latter case. Using North Korea in this study (and Iran and Russia in follow-up experiments) therefore allows me to control for the foreign country in question and avoid a lack of information equivalence across experimental conditions.⁷⁶ New research also demonstrates that using hypothetical countries in experiments can cause researchers to overestimate support for the use of force.⁷⁷ The US-North Korea relationship is an ideal setting to test the Madman Theory since both countries have nuclear weapons, and thus both can threaten the other with enormous damage. Consequently, this scenario approximates the conditions under which Ellsberg and Schelling argued madness could be a strategic asset and make seemingly incredible threats—such as starting a nuclear war—credible.

Following Michaela Mattes and Jessica L. P. Weeks, I begin by informing respondents that the year is 2027 and the US president is John Richards, a member of the Democratic Party.⁷⁸ The scenario is set in the future so that subjects will be less likely to make assumptions about the identity of the president.⁷⁹ I then remind respondents that the United States and North Korea have a tense relationship and control for important factors such as North Korea’s regime type and relative military capabilities. In addition, respondents are told that North Korea has a new leader in 2027: Choe Ryong-hae. Choe is a real North Korean official, at one point Kim Jong-un’s second in command, and thus he could realistically become North Korea’s leader in the event Kim dies, is incapacitated, or is deposed.⁸⁰ Kim himself is not used in the experiment because US citizens may already

⁷⁵Kyle Peyton, Gregory A. Huber, and Alexander Coppock, “The Generalizability of Online Experiments Conducted during the COVID-19 Pandemic,” *Journal of Experimental Political Science* 9, no. 3 (Winter 2022): 379–94.

⁷⁶Allan Dafoe, Baobao Zhang, and Devin Caughey, “Information Equivalence in Survey Experiments,” *Political Analysis* 26, no. 4 (October 2018): 399–416.

⁷⁷Jacklyn Majnemer and Gustav Meibauer, “Names from Nowhere? Fictitious Country Names in Survey Vignettes Affect Experimental Results,” *International Studies Quarterly* 67, no. 1 (March 2023): 1–10.

⁷⁸Mattes and Weeks, “Hawks, Doves, and Peace.”

⁷⁹A Democratic rather than Republican president is chosen to disassociate the experiment from the Trump presidency. Results from Study 5 further suggest the Trump presidency does not drive the results.

⁸⁰US Department of the Treasury, “Treasury Sanctions North Korean Officials and Entities in Response to the Regime’s Serious Human Rights Abuses and Censorship,” press release, 10 December 2018, <https://home.treasury.gov/news/press-releases/sm568>.

have strong prior beliefs about whether he is “crazy,” making it harder to experimentally manipulate perceived madness. Respondents are also informed that North Korea can hit the United States with nuclear missiles in order to raise the stakes in the scenario to be closer to the US-Soviet standoff Nixon faced during the Cold War.

Following the introduction, I experimentally manipulate whether the leader is framed as mad. To prime the extreme-preferences variety of madness, respondents are told that:

North Korean leader Choe Ryong-hae announces that he will not give up his nuclear weapons and demands the United States remove economic sanctions against his country. If sanctions are not removed, then the North Korean leader threatens to launch a nuclear attack against the United States.

North Korean experts in the United States from across the political spectrum say that Choe truly believes America’s economic sanctions pose an enormous threat to North Korea, and that he is willing to risk nuclear war in order to get them removed.

Since the extreme-preferences variety of madness can involve a high valuation of the issues at stake and an unusually high risk tolerance, I prime it here by informing respondents that experts assess that Choe views American sanctions as crippling (and thus he highly values the issues at stake in the dispute) and is willing to risk nuclear war to get them removed (and thus he has a high risk tolerance).⁸¹

To prime deviation from consequence-based decision making, subjects are instead told that:

North Korean experts in the United States from across the political spectrum say that Choe does not understand that an attack on America could lead to nuclear war and is making decisions based on emotion rather than reason.

Here, the North Korean leader does not understand the potential consequences of his actions and is making decisions emotionally rather than logically, which McManus argues is the most common way that leaders deviate from consequence-based decision making.⁸²

In the baseline condition against which I compare the two madness conditions, the North Korean leader also threatens the United States with nuclear war, but is not framed as mad:

North Korean experts in the United States from across the political spectrum say that Choe is most likely bluffing in an effort to get America to remove its economic sanctions. If the United States refuses the North Korean leader’s demand, then he will probably not go through with the attack.

⁸¹McManus, “Revisiting the Madman Theory,” 982.

⁸²McManus, “Revisiting the Madman Theory,” 983.

Here, the North Korean leader seemingly does not have extreme preferences since he is likely not willing to significantly risk complete destruction in order to get sanctions removed. He also does not deviate from consequence-based decision making since he is likely making decisions strategically (not emotionally) in an effort to make the United States back down. However, there is a nonzero probability he will follow through on this threat, as he is “most likely” bluffing and will “probably not” go forward with the attack. This approximates how Nixon was perceived by the Soviets, who were not convinced Nixon was actually mad and believed he was bluffing. It also likely accords with Nixon’s actual approach and disposition. According to Nixon’s chief of staff, Nixon “believed conceptually that it was important that the enemy ... have, if not a conviction, at least a concern that he might be pushed to a point where he might do something totally irrational. That was a strategic concept, not a planned intent, and there was never any consideration given to doing anything to carry out the ‘madman’ theory ... Yes, it was a bluff.”⁸³

Therefore, this condition reflects a scenario where a leader makes an ostensibly crazy threat but is not framed as mad. Since a non-mad leader must be (in all probability) bluffing in this scenario—that is, unwilling to intentionally carry out a suicidal nuclear attack—I argue that it is appropriate to include this information in the baseline treatment.

One potential objection to this survey design is that if respondents have a preexisting belief that Kim Jong-un specifically or North Korean leaders generally are crazy, then it may be hard to convince them that the North Korean leader in the above scenario is relatively sane. However, if anything, this should bias against finding any benefit to the Madman Strategy, as there should be less of a perceived difference between the experimental conditions.

Finally, to test H2 and H3, I experimentally manipulate the extent to which commitment problems are an issue for the North Korean leader, whether he is framed as mad or not. Specifically, I vary whether experts believe the North Korean leader will or will not ask for more concessions if the United States agrees to remove economic sanctions:

Reduced commitment problems: experts also believe that if the United States gives in to Choe’s demands on sanctions he will be satisfied and not ask for more concessions from America.

Heightened commitment problems: experts also believe that if the United States gives in to Choe’s demands on sanctions he will not be satisfied and will ask for more concessions from America.

⁸³Kenneth W. Thompson, *The Nixon Presidency* (Lanham, MD: University Press of America, 1987), 83.

My outcome measure is a five-point Likert scale that measures to what extent respondents would oppose President Richards removing economic sanctions on North Korea. I argue this measure aids in understanding the efficacy of the Madman Strategy because policymakers have incentives to respond to public opinion. In other words, lower public disapproval with the US president removing sanctions (that is, conceding) means the president will face fewer domestic barriers to giving in and consequently will be more likely to give in, making the North Korean leader's threat more effective. My dependent variable measures public approval with the president deciding to back down rather than whether respondents prefer the president to back down because political leaders likely care more about whether the public approves of their actions after the fact than which policy the public would have theoretically preferred before the decision was made.⁸⁴ I also opt for survey subjects to respond to the situation from their own perspective rather than take the perspective of the US president or other political leaders because public opinion directly affects political leaders' willingness to concede or not to an adversary, suggesting the views of the public are inherently important. Even if this was not the case, prior research finds that elite–public gaps are generally overstated.⁸⁵

Study 1: Results

Table 1 displays the percentage of respondents who oppose President Richards removing economic sanctions on North Korea, as well as the percentage point difference between the two madness conditions and the baseline condition. In accordance with previous studies, Table 1 collapses the five-point measure of opposition to conceding into a binary measure

Table 1. Is opposition to removing sanctions on North Korea less when the North Korean leader is framed as mad?

	Opposition to Conceding (%)	Difference from Baseline (Percentage Points)
Baseline	52.4%	—
Deviates from Consequence-Based Decision Making	43.8%	–8.6*
Extreme Preferences	41.7%	–10.8**

Note: Results depict the percentage opposition to removing sanctions on North Korea and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$; and *** = $p < .01$, where P values indicate whether opposition is statistically less than 0.

⁸⁴Scott D. Sagan and Benjamin A. Valentino, "Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants," *International Security* 42, no. 1 (Summer 2017): 41–97. Of course, there are multiple valid ways to measure how public opinion impacts the efficacy of the Madman Strategy that may yield somewhat different results.

⁸⁵Kertzer, "Re-assessing Elite-Public Gaps in Political Behavior."

of opposition to more clearly illustrate substantive effects. Identical results emerge with the full five-point measure.⁸⁶

In accordance with H1, leaders framed as mad make more effective threats than leaders framed as sane. Relative to the baseline, opposition to conceding is 8.6 percentage points lower ($p \approx 0.073$) when the leader is framed as deviating from consequence-based decision making, and 10.8 percentage points lower ($p \approx 0.036$) when the leader is framed as having extreme preferences.⁸⁷ These results therefore cut against the common expectation and finding in the literature that the Madman Strategy does not work.⁸⁸

To further ensure the robustness of this finding, I take a number of steps described in the online appendix. First, I show the results are robust when the full five-point scale is used (Table A.2). Second, the results are substantively identical to excluding respondents who failed the attention check (Table A.3). Third, I show that the results hold in a regression that controls for the level of commitment problems the leader faced, as well as respondent partisan identification, level of militant assertiveness, education, income, gender, race, and age (Table A.4). The robustness of this result across these tests builds confidence in the findings.

In the online appendix, I also show that the results remain substantively identical among those low and high in a measure of militant assertiveness, even though we might expect more hawkish subjects to refuse to bow to foreign pressure irrespective of whether the leader is perceived as mad or not.⁸⁹ More substantive differences emerge based on party identification. Being perceived as mad (in either form) generally makes independents—a politically important subgroup—more willing to concede, only the extreme-preferences variety of madness makes Republicans more willing to concede, and perceived madness does not have any significant effect on Democrats relative to the baseline.⁹⁰

⁸⁶See Table A.2 in the online appendix.

⁸⁷P values are calculated from 2,000 bootstraps, where a result is statistically less than 0 at the 10% level if at least 1,800 of the 2,000 bootstrapped estimates are less than 0, significant at the 5% level if at least 1,900 are less than 0, etc. I use bootstraps because they make fewer assumptions about the distribution of the underlying data than other methods. See Kertzer and Brutger, “Decomposing Audience Costs.”

⁸⁸Although, ideally, all results would be statistically significant at the 5% or 1% level, the findings presented in Table 1 indicate support for the Madman Theory for four reasons. First, they are substantively large. Second, the results hold in a follow-up study using weaker treatments, which would be unlikely if the results were simply due to chance. Third, this study was somewhat underpowered. On the one hand, this suggests that a larger study might have yielded even stronger statistical results, but, on the other hand, underpowered studies may be more likely to yield false positives. Although only a larger follow-up study could definitively adjudicate which view is correct, that substantively similar results hold in Studies 2 and 3 should increase confidence in the robustness of these findings. Fourth, the 95% threshold for significance is arbitrary, meaning a P value of, for example, 0.067 rather than 0.05 should not automatically render an estimate uninformative.

⁸⁹Richard K. Herrmann, Philip E. Tetlock, and Penny S. Visser, “Mass Public Decisions to Go to War: A Cognitive-Interactionist Framework,” *American Political Science Review* 93, no. 3 (September 1999): 553–73.

⁹⁰Democrats, however, are generally more willing to concede to North Korea than Republicans, as we would expect.

Table 2. Do reduced commitment problems lower opposition to removing sanctions on North Korea?

	Reduced Commitment Problems	Heightened Commitment Problems	Difference (Percentage Points)
Baseline	51.5%	54.9%	-3.4
Deviates from Consequence-Based Decision Making	42.9%	46.3%	-3.4
Extreme Preferences	43.3%	38.7%	4.6

Note: Results depict the percentage opposition to removing sanctions on North Korea and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$; and *** = $p < .01$, where P values indicate whether opposition is statistically less than 0.

Table 3. Does the Madman Theory pass a hard test?

	Opposition to Conceding (%)	Difference from Baseline (Percentage Points)
Baseline (Reduced Commitment Problems)	51.5%	—
Deviates from Consequence-Based Decision Making (Heightened Commitment Problems)	46.3%	-5.3
Extreme Preferences (Heightened Commitment Problems)	38.7%	-12.8*

Note: Results depict the percentage opposition to removing sanctions on North Korea and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$; and *** = $p < .01$, where P values indicate whether opposition is statistically less than 0.

Table 2 tests H2, which holds that commitment problems should make all types of threats less effective. Although the results are in the right direction for two of the three conditions, surprisingly no statistically significant differences emerge between leaders associated with heightened commitment problems and those with reduced commitment problems. This suggests that threat credibility may matter more than perceived commitment problems, at least in the specific situation outlined in the experiment. Given that nuclear war is on the line in this scenario, it makes logical sense for respondents to heavily discount the future and focus on the near-term crisis scenario. Of course, in the real world it may be harder to separate madness from commitment problems.

Finally, Table 3 tests H3 and presents a hard test for the Madman Theory. If threats from leaders who face more severe commitment problems and are framed as mad are more effective than threats from leaders who face less severe commitment problems and are framed as more sane, then that would be strong evidence that madness provides a bargaining advantage. The extreme-preferences variety of madness passes this test, as there is 12.8 percentage points lower opposition ($p \approx 0.060$) to conceding relative to the baseline. Deviating from consequence-based decision making is also associated with less opposition to conceding (by 5.3 percentage points), though this difference is not statistically significant. The latter finding provides some evidence that the potentially inherent commitment

problems associated with madness—especially in its dispositional form—can undermine the effectiveness of the Madman Strategy.

Study 2: Alternative Treatment Wordings

One potential concern with the design of Study 1 is that the treatments are relatively strong and, in some cases, directly prime threat credibility in addition to rationality. For example, in the baseline condition in Study 1 respondents are told that the North Korean leader is “most likely bluffing.” Although this reflects how the Soviets viewed Nixon, and a sane leader should be perceived as bluffing in this scenario since a nuclear attack against the United States would very likely mean North Korea’s complete destruction, a weaker treatment would simply prime the leader’s rationality and allow respondents to infer their threat is a bluff. To analyze the robustness of Study 1’s results to weaker treatments, I conducted a follow-up experiment on a sample of 625 Americans recruited through Lucid.

In the baseline condition, respondents are primed to believe the North Korean leader is rational, but are not directly told that he is likely bluffing:

North Korean experts in the United States from across the political spectrum say that Choe has a calm, even-tempered personality and typically makes decisions based on a logical cost–benefit analysis.

To prime extreme preferences, respondents are told that:

North Korean experts in the United States from across the political spectrum say that Choe has a calm, even-tempered personality, but is prone to extreme risk-taking and believes America’s economic sanctions pose an enormous threat to North Korea.

Unlike the baseline condition, here subjects are primed to believe the North Korean leader has extreme preferences because he is prone to risk-taking and highly values the removal of sanctions. Unlike the condition regarding deviation from consequence-based decision making condition (below), Choe does not make decisions based on emotion since he “has a calm, even-tempered personality.” And in contrast to Study 1, respondents are not told that Choe “is willing to risk nuclear war in order to get [sanctions] removed,” which is a relatively strong treatment for risk tolerance.

Finally, to prime deviation from consequence-based decision making, subjects are informed that Choe is making decisions based on emotion rather than logic:

North Korean experts in the United States from across the political spectrum say that Choe has an explosive anger and is prone to making impulsive decisions without thinking through the consequences.

Table 4. Weaker treatments.

	Opposition to Conceding (%)	Difference from Baseline (Percentage Points)
Baseline	39.2%	—
Deviates from Consequence-Based Decision Making	33.0%	-6.2*
Extreme Preferences	32.8%	-6.4*

Note: Results depict the percentage opposition to removing sanctions on North Korea and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$; and *** = $p < .01$, where P values indicate whether opposition is statistically less than 0.

Unlike Study 1, respondents are not told that “Choe does not understand that an attack on America could lead to nuclear war,” which is likely a strong prime for a leader not understanding the potential consequences of their actions. Table 4 shows that the results are robust to these alternative treatments, though the effect sizes are a bit smaller, as expected.

I also analyze the mechanism driving the benefits of madness in Study 2. Employing causal mediation analysis, in the online appendix I show that the mechanism underlying the advantage of madness is, as theorized, enhanced threat credibility (Tables A.11 and A.12).⁹¹ A North Korean leader framed as deviating from consequence-based decision making was assessed as 12.3 percentage points more likely to carry out their threat relative to the baseline ($p \approx 0.003$), and the number was 6.6 percentage points for leaders framed as holding extreme preferences ($p \approx 0.07$).

Studies 3 and 4: External Validity to Iran and Russia

To further test the external validity of Study 1, I conducted two additional follow-up experiments through Lucid in April 2020. Both studies are three-factor between-subjects experiments on a national sample of 318 and 290 Americans, respectively.⁹² The wording and design of both experiments closely follow that of Study 1. In Study 3, a new Iranian leader in 2027 threatens the United States with “all out” war if sanctions are not removed, but Iran has not yet acquired nuclear weapons and so can only threaten a conventional attack. Still, such an attack against the conventionally stronger United States would entail significant costs for Iran and thus might be considered to lack credibility on its face. If perceived madness is beneficial for coercive bargaining in this scenario, then that would show that nuclear weapons are not a necessary condition for madness to be effective. In Study 4, a new Russian leader in 2027 threatens the United States with nuclear war if sanctions are not removed. Given that Russia is much more powerful than North Korea and Iran, this scenario will

⁹¹Kosuke Imai, Luke Keele, and Dustin Tingley, “A General Approach to Causal Mediation Analysis,” *Psychological Methods* 15, no. 4 (December 2010): 309–34.

⁹²Although the sample size of these studies is relatively low in the aggregate, they include about 100 respondents per condition, which is similar to many previous studies.

Table 5. Is opposition to removing sanctions on Iran/Russia less when the leader is framed as mad?

	Opposition to Conceding (%)	Difference from Baseline (Percentage Points)
<i>Iran</i>		
Baseline	49.0%	—
Deviates from Consequence-Based Decision Making	45.4%	-3.6
Extreme Preferences	39.0%	-10.1*
<i>Russia</i>		
Baseline	38.6%	—
Deviates from Consequence-Based Decision Making	33.6%	-5.0
Extreme Preferences	34.9%	-3.7

Note: Results depict the percentage opposition to removing sanctions on Iran/Russia and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$ and *** = $p < .01$, where P values indicate whether opposition is statistically less than 0.

examine the external validity of the Madman Theory to interactions between major powers.

Table 5 displays the results from these two studies. For Iran, madness in the form of extreme preferences is an asset in coercive bargaining, though deviating from consequence-based decision making has no significant effect. However, in the online appendix I show that when respondents who failed the attention check are excluded, both forms of madness make threats that are seemingly incredible to carry out more effectively (Table A.17). This result also holds in a regression context (Table A.19). Overall, though the results are somewhat more mixed for the condition regarding deviation from consequence-based decision making, madness is still generally advantageous in a scenario involving Iran. This finding suggests nuclear weapons are not a necessary condition for madness to be effective in crisis bargaining, though the stronger results for Study 1 provide some suggestive evidence that they may be beneficial. Still, since none of my studies independently vary whether a threat is nuclear or conventional in nature, future work should more systematically test how the type of threat moderates the impact of perceived madness.

As in Study 2, an Iranian leader who was framed to be mad—in either form—was viewed as having greater threat credibility than the baseline: 12.8 percentage points ($p \approx 0.026$) for leaders framed as deviating from consequence-based decision making and 20.8 percentage points ($p < 0.001$) for leaders framed as holding extreme preferences (Table A.22 and Table A.23).

For Russia, madness has no statistically significant effect on threat efficacy, though the results are in the right direction. This null result also holds when the full five-point scale is used (Table A.26), when respondents who failed the attention check are excluded (Table A.27), and in a regression context (Table A.28). Interestingly, however, threats made by a Russian leader framed as mad were still viewed as significantly more credible: 10.6

percentage points ($p \approx 0.046$) for leaders framed as deviating from consequence-based decision making and 23.2 percentage points ($p < 0.001$) for leaders framed as holding extreme preferences (Table A.29). Matthew Kroenig's argument that the American public is generally less willing to run the risk of war against a more capable state such as Russia may explain this null result for threat effectiveness even in the face of enhanced threat credibility.⁹³ Baseline opposition to conceding is lower for the Russia scenario (38.6%) than the North Korean scenario (52.4%) in Study 1 or the Iranian scenario (49.0%). Thus, the relatively small proportion of respondents—38.6%—who are unwilling to concede to Russia in the baseline scenario may be particularly resistant to conceding, even in the face of heightened threat credibility when a leader displays signs of madness. In other words, enhanced threat credibility may have diminishing marginal returns when respondents are relatively less likely to oppose conceding at baseline since those remaining individuals may be less persuadable. More broadly, the null finding with respect to threat effectiveness suggests that nuclear weapons are not a sufficient condition for madness to be effective in crisis bargaining, and that madness may be less effective in interactions among major powers. Nevertheless, the credibility benefits of perceived madness appear to be relatively universal, and future research should assess whether the null result in Study 4 is due to insufficient statistical power (after all, the results were in the right direction), the logic I outlined above, or some other mechanism.

Lastly, in Studies 3 and 4 I asked respondents whether they believed Iran/Russia would ask for additional concessions if President Richards gave in and removed sanctions. No statistically significant differences emerged between the madness and baseline experimental conditions, suggesting that madness does not automatically make commitment problems more severe (Table A.13 and Table A.30).

To summarize, Studies 1 through 4 demonstrate that the Madman Strategy offers some advantages in coercive bargaining, providing clearer support for the Madman Theory than most previous literature has found. However, depending on the strength of the madness prime (Study 2) and the identity of the country adopting the strategy (Studies 3 and 4), the Madman Strategy's efficacy varies. As such, the results presented here do not suggest an invariable benefit of perceived madness in crisis bargaining.

⁹³Kroenig, *Logic of American Nuclear Strategy*.

Study 5

To test the impact of perceived madness on the president's domestic approval, I designed and administered a four-factor experiment on a national sample of 611 Americans recruited through Lucid in March 2020. I again block on respondent party identification to ensure approximately equal numbers of Democrats, independents, and Republicans in each treatment. This experiment involves the US president attempting to convince North Korea to denuclearize, and I experimentally manipulate whether he maintains the status quo policy of economic sanctions, threatens North Korea with nuclear war but is "sane" and thus likely bluffing (approximating Nixon's policy during Operation Giant Lance), or threatens North Korea with nuclear war and either has extreme preferences or deviates from consequence-based decision making.

One potential concern with this design is that it might cause some respondents to think of President Trump. Nonetheless, this concern is mitigated by the vignette having a president with a different name and his being a Democrat. Moreover, results presented below suggest the results were not driven by respondents thinking about Trump.

Table 6 displays the percentage of respondents who disapprove of how President Richards is doing his job, as well as the percentage point difference between various conditions. I again collapse the five-point measure of disapproval into a binary measure to more clearly illustrate substantive effects. Identical results emerge with the full five-point measure (Table A.32), as well as when respondents who failed the attention check are excluded (Table A.33) and in a regression controlling for demographic factors and political beliefs (Table A.34).

Per expectations outlined in H4, presidents who threatened North Korea with nuclear war and are framed as holding extreme preferences face disapproval rates that are 16.3 percentage points greater ($p < 0.001$) than those who maintained the status quo policy of economic sanctions. For presidents who deviate from consequence-based decision making, disapproval is a huge 36.9 percentage points greater ($p < 0.001$) than for those

Table 6. Is the Madman Strategy popular domestically?

	Disapproval (%)	Difference from Status Quo (Percentage Points)	Difference from Feigned Madness (Percentage Points)
Status Quo	7.7%	—	—
Feigned Madness	32.0%	—	—
Deviates from Consequence-Based Decision Making	44.7%	36.9***	12.7***
Extreme Preferences	24.0%	16.3***	-7.9*

Note: Results depict the percentage disapproval with the US president and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$; and *** = $p < .01$, where P values indicate whether disapproval is statistically greater than/less than 0.

who continued the status quo. These results suggest that even though there is a strategic benefit to perceived madness vis-à-vis foreign adversaries, domestic audiences would prefer a less risky policy. Leaders framed as deviating from consequence-based decision making face a particularly large penalty, likely because of the stigma against mental illness. The domestic costs of the Madman Strategy help explain its infrequency and relative lack of success in the historical record.

Table 6 also demonstrates some support for H5, which is that disapproval should be greater for leaders who make threats that are seemingly incredible to carry out and are framed as mad than for more sane leaders who make similar threats. In accordance with this expectation, disapproval is 12.7 percentage points greater ($p \approx 0.013$) for leaders who deviate from consequence-based decision making than those that make an ostensibly incredible threat but are likely bluffing. However, in contrast to expectations, the opposite holds true for leaders with extreme preferences: disapproval is 7.9 percentage points lower ($p \approx 0.059$) for leaders who make bellicose threats due to extreme preferences than for those that feign madness. This provides some evidence that the extreme-preferences variety of madness is more popular domestically than that of deviating from consequence-based decision making. However, these aggregate results mask significant heterogeneity between Democrats, Republicans, and independents.

Table 7 disaggregates results based on whether respondents identified as Democrats, Republicans, or independents.⁹⁴ In accordance with expectations, Democrats punish leaders who deviate from consequence-based decision making and those with extreme preferences more than sane leaders who make threats that are seemingly incredible to carry out. In fact, disapproval for leaders who feign madness is relatively low (15.5%). This result makes sense since prior research establishes that Democrats (and, more generally, dovish individuals) are less disapproving of inconsistency (that is, making a threat and then likely backing down from it, as in the feigned madness condition) than they are of belligerence (that is, aggression, which is more likely to occur if the president is not feigning madness but may actually be mad).⁹⁵ Also of note is the willingness of Democratic survey subjects to punish a copartisan, which is a strong signal of the domestic costs of perceived madness.

By contrast, Republicans strongly disapprove of leaders who feign madness (50%). Thus, they prefer leaders who hold extreme preferences and those framed as deviating from consequence-based decision making to

⁹⁴Similar results hold when comparing more dovish respondents to more hawkish respondents. See Table A.36.

⁹⁵Herrmann, Tetlock, and Visser, "Mass Public Decisions to Go to War"; Kertzer and Brutger, "Decomposing Audience Costs"; Schwartz and Blair, "Do Women Make More Credible Threats?"

Table 7. Democrats vs. Republicans vs. independents.

	Disapproval (%)	Difference from Status Quo (Percentage Points)	Difference from Feigned Madness (Percentage Points)
<i>Democrats</i>			
Status Quo	10.0%	—	—
Feigned Madness	15.5%	—	—
Deviates from Consequence-Based Decision Making	44.3%	34.2***	28.8***
Extreme Preferences	29.1%	19.1***	13.6**
<i>Republicans</i>			
Status Quo	4.4%	—	—
Feigned Madness	50.0%	—	—
Deviates from Consequence-Based Decision Making	28.6%	24.2***	-21.4**
Extreme Preferences	24.0%	19.5***	-26.0***
<i>Independents</i>			
Status Quo	7.5%	—	—
Feigned Madness	39.9%	—	—
Deviates from Consequence-Based Decision Making	64.9%	57.4***	25.1**
Extreme Preferences	17.8%	10.3*	-22.1***

Note: Results depict the percentage disapproval with the US president and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$; and *** = $p < .01$, where P values indicate whether disapproval is statistically greater than/less than 0.

sane leaders who bluff, in contrast with expectations. In the online appendix, I show that these heterogeneous effects also hold in a regression that controls for other factors such as militant assertiveness (Table A.35). The most likely explanation for this discrepancy is that Republicans tend to care more about inconsistency than belligerence compared to Democrats, and therefore punish leaders more for threats they are unlikely to carry out. In this sense, inconsistency may be one mechanism that explains why feigned madness—which inherently suggests a high likelihood of inconsistency since actually carrying out the nuclear threat would be mad—is particularly unpopular with Republicans.

On the other hand, Republicans do still punish leaders framed as mad compared to those who maintain the status quo, which suggests that loyalty to President Trump—who is perhaps viewed as somewhat mad—does not drive the results. Since the status quo condition does not implicate inconsistency, this result among Republicans is likely driven by other mechanisms perceived madness affects, such as competence and temperament. Independents fall somewhere in between, as they very strongly oppose leaders framed as deviating from consequence-based decision making, but not those who hold extreme preferences. Given independents' political importance, these results suggest that the domestic costs of deviating from consequence-based decision making are especially large.

Examining the relationship between domestic politics and the Madman Strategy further, Table 8 shows the percentage of respondents who believe war is “likely” or “very likely” with North Korea due to President Richards's

Table 8. Does the Madman Strategy increase the perceived chance of war?

	War Is Likely (%)	Difference from Status Quo (Percentage Points)	Difference from Feigned Madness (Percentage Points)
Status Quo	34.7%	—	—
Feigned Madness	36.6%	—	—
Deviates from Consequence-Based Decision Making	67.1%	32.4***	30.5***
Extreme Preferences	64.1%	29.4***	27.4***

Note: Results depict the percentage of respondents that believe war is likely and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$; and *** = $p < .01$, where P values indicate whether belief that war is likely is statistically greater than 0.

Table 9. Does the Madman Strategy increase the perceived chance of North Korea backing down?

	North Korea Likely to Disarm (%)	Difference from Status Quo (Percentage Points)	Difference from Feigned Madness (Percentage Points)
Status Quo	13.6%	—	—
Feigned Madness	8.4%	—	—
Deviates from Consequence-Based Decision Making	13.9%	0.30	5.4*
Extreme Preferences	10.7%	-2.9	2.2

Note: Results depict the percentage of respondents that believe North Korea will give in and are calculated from 2,000 bootstraps. * = $p < .10$; ** = $p < .05$; and *** = $p < .01$, where P values indicate whether belief that war is likely is statistically greater than 0.

policy. The results demonstrate that in either form, madness is perceived as significantly increasing the chances of war.⁹⁶ Although the Madman Strategy raises the chances of war by design in order to coerce an opponent to back down, the prospect of a devastating war actually occurring is likely one reason why the public generally disapproves of the Madman Strategy.

If the possibility of war is one of the primary costs of the Madman Strategy, then the principal benefit is the potential for the strategy to convince an adversary to back down. Table 9 therefore shows the percentage of respondents who believe North Korea will agree to relinquish their nuclear weapons in response to President Richards's policy. The results show that the public has little confidence in the Madman Theory. Relative to the status quo, madness does not increase the perceived chances of North Korea agreeing to disarm. Deviating from consequence-based decision making does slightly increase the perceived chances North Korea will back down relative to feigning madness, but only by a small amount: 5.4 percentage points. Overall, then, the public does not generally believe that the Madman Strategy is likely to be successful. Of course, in the event that the Madman Strategy is successful, we would expect public support to rise, perhaps significantly. Future studies should therefore examine how

⁹⁶The results are substantively identical among Democrats and Republicans.

the ultimate outcome of leaders using the Madman Strategy affects public support for it. Nevertheless, understanding the more immediate domestic political consequences of madness—as this study does—is important since political leaders may have short time horizons and thus prefer to adopt less risky policies given that the outcome of international bargaining is always uncertain.

A Fine Line Between Madness and Genius

Given the challenges associated with making credible threats against nuclear-armed adversaries, and Nixon and Trump's attempts to exploit madness for bargaining leverage, it is important to assess whether the Madman Theory has any merit. Building on McManus's foundational work, this study cuts against the arguments of most current literature and finds some experimental evidence that perceived madness can indeed enhance threat effectiveness. Most prominently, I find evidence that the mass public is more willing to concede to a foreign leader framed as mad than a foreign leader framed as more sane, though this effect does not hold under all circumstances (for example, in interactions among major powers). I also find evidence that the public strongly disapproves of the Madman Strategy when adopted by their own leader. The domestic costs of perceived madness dissuade leaders from adopting it in the first place and serve to undermine its effectiveness even if adopted. As a whole, these experimental results indicate that perceived madness can be beneficial under certain conditions, but it is far from a panacea. Given the significant domestic costs and risks of escalation associated with the Madman Strategy, leaders should think twice before adopting it.

Although this article tested the external validity of Studies 1 and 2, which focused on North Korea, to situations involving Iran and Russia, the external validity of the experimental results in this project to real-world scenarios remains an open question. Particularly noteworthy is the question of whether real-world leaders can convince their adversaries that they are possibly or definitely mad. Doing so in an experimental setting is relatively straightforward since certain factors associated with a hypothetical leader—their tendency to deviate from consequence-based decision making or hold extreme preferences, for instance—can be made especially salient. By contrast, the real world is a more information-rich environment; contradictory pieces of evidence about a leader's relative madness may make it more difficult for them to establish a clear reputation as a “crazy type.” This, of course, was one of the biggest roadblocks Nixon ran into during Operation Giant Lance.

Nevertheless, there is reason to expect that real-world leaders—even democratically elected ones—can gain a reputation for madness, meaning

the results of these experiments would be externally valid. Empirically, McManus's research establishes that there are leaders perceived of as mad, at least to some extent. In a series of historical case studies, she finds some foreign audiences viewed authoritarian leaders such as Hitler, Khrushchev, Hussein, and Gaddafi as mad. In a text analysis of English-language news reports and editorials from around the world, she also finds that many other leaders have accrued reputations for madness. This includes leaders in democracies, such as George W. Bush, John Howard, and Ariel Sharon. In fact, McManus finds that leaders in democracies are significantly more likely to gain a reputation for "slight madness" than leaders in autocracies, though this result may be due to bias in how the data was collected (only English-language sources) and restrictions on the free press in autocracies.⁹⁷

Theoretically, there are also several reasons why even leaders in democracies—who display at least some public evidence of competence and rationality by successfully winning elections—could still be perceived as mad. First, even if leaders are judged as sane during their election campaigns, events that occur between elections (for example, normal aging, medical conditions, or traumatic personal or foreign policy experiences) could shift a leader's actual level of madness and/or reputation for madness. For example, Franklin Pierce's son died after his election to the presidency but right before his inauguration, which led to a wave of depression and possibly worsened Pierce's alcoholism.⁹⁸ In the same vein, Woodrow Wilson had a stroke in office, and some believe Ronald Reagan suffered from Alzheimer's disease later in his presidency. Although these kind of health and behavioral issues do not necessarily mean a leader is actually mad or will gain a reputation for madness, they could potentially lead to either outcome.

Second, leaders can also have elements of madness at the time of an election and still win sufficient support from voters if (a) they hide these signs of madness from the electorate, which may then become known to the public and foreign governments at a later time; (b) voters have extreme preferences themselves and so support a candidate with similar beliefs; or (c) voters support a leader who deviates from consequence-based decision making because they value other aspects of a candidate (for example, their policy platform). Per this argument, one study by three medical doctors, which reviewed US presidents from 1776 to 1974, found about half had some form of mental health illness.⁹⁹

⁹⁷McManus, "Crazy Like a Fox?" 283.

⁹⁸Roy Franklin Nichols, *Franklin Pierce: Young Hickory of the Granite Hills* (Newtown, CT: American Political Biography Press, 2003).

⁹⁹Jonathan R. T. Davidson, Kathryn M. Connor, and Marvin Swartz, "Mental Illness in U.S. Presidents between 1776 and 1974: A Review of Biographical Sources," *Journal of Nervous and Mental Disease* 194,

Still, future studies should examine the external validity of this article's experimental findings and assess the relationship between regime type and perceived madness to a greater extent. Do democratic mechanisms reduce the chances of a "mad" leader rising to high political office or a leader gaining a reputation for madness? What is the impact of different features of autocracies? For instance, perhaps leaders that rise to power through a family dynasty are more likely to gain a reputation for madness because they did not have to demonstrate as much political skill as a leader who rose to power from the bottom up. Similarly, personalist leaders may be more likely to be perceived as mad since they face fewer constraints on their power, and thus it may be harder to remove them from office even if they do demonstrate signs of madness.

This article also highlights a number of other promising avenues for future research. First, future studies should consider conducting elite experiments to probe the external validity of these findings to political leaders. Second, would these results vary in different countries? For example, perhaps countries with greater baseline levels of militarism would be more reluctant to give in to foreign leaders who make seemingly incredible threats and are perceived as mad, and more supportive of their own leaders who make similar threats and are viewed as mad. Third, how do allies perceive the Madman Strategy? Would a country such as South Korea approve of having a "crazy" defender in Washington? Fourth, does perceived madness become more effective in conjunction with other costly signals, such as military mobilizations and small-scale operations? In other words, is there synergy between perceived madness and other tactics? Fifth, what are the most important mechanisms (for example, competence, temperament, reputation, etc.) explaining the Madman Theory's domestic unpopularity? Sixth, do the marginal benefits of perceived madness for threat credibility decrease as the threat becomes less costly to actually carry out, and does perceived madness have a different effect for nuclear threats that are seemingly incredible to carry out relative to comparable conventional threats? Seventh, do the domestic constraints leaders face in using this strategy diminish in the face of a heightened foreign threat (due to the rally-around-the-flag phenomenon), if the principal policy objective varies (for example, if a humanitarian frame is employed), or if leaders are able to convince the public that the chances of success are higher? Eighth, do leaders' characteristics affect domestic reactions to perceived madness? For example, would female leaders be punished more than male leaders? Finally, future work

no. 1 (January 2006): 47–51. Though it is important to note that mental illness does not inherently make an individual crazy or mad.

should consider the impact of perceived madness in other areas, such as international diplomacy and trade negotiations. Notes

Acknowledgments

I thank Sumitra Badrinathan, Christopher Blair, Jonathan Chu, Meghan Garrity, Sophie McCormack, Bruno Tertrais, and participants at the 2020 NATO Defense College Early-Career Nuclear Strategists Workshop for helpful comments and advice.

Data Availability Statement

The data and materials that support the findings of this study are available in the *Security Studies* Dataverse at <https://doi.org/10.7910/DVN/DZ8BUS>.

ORCID

Joshua A. Schwartz  <http://orcid.org/0000-0001-9283-5846>