British Journal of Political Science (2021), **51**, 275–293
doi:10.1017/S0007123419000401

British Journal of
Political Science

ARTICLE

Crazy Like a Fox? Are Leaders with Reputations for Madness More Successful at International Coercion?

Roseanne W. McManus*

Pennsylvania State University
*Corresponding author. E-mail: Roseanne.McManus@psu.edu

(Received 5 December 2018; revised 9 May 2019; accepted 2 July 2019; first published online 9 December 2019)

Abstract
According to the 'Madman Theory' promoted by Richard Nixon and early rationalist scholars, being viewed as mentally unstable can help a leader coerce foreign adversaries. This article offers the first large-N test of this theory. The author introduces an original perception-based measure of leaders' reputations for madness, coded based on news reports, and analyzes its effect on both general deterrence and crisis bargaining. The study also tests several hypotheses about the conditions under which perceived madness is expected to be more or less helpful. The author finds that perceived madness is harmful to general deterrence and is sometimes also harmful in crisis bargaining, but may be helpful in crisis bargaining under certain conditions. The analysis suggests that the harmful effect of perceived madness results from a commitment problem.

Keywords: conflict bargaining; resolve; credibility; reputation; madman theory; leaders in international conflict

Richard Nixon coined the term 'Madman Theory' to describe the belief that being viewed as mentally unstable can help a leader coerce foreign adversaries. He reportedly said, 'We'll just slip the word to [the North Vietnamese] that… "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' (Haldeman and DiMona 1978, 83). Nixon's Madman Theory was in keeping with arguments by Cold War-era scholars Ellsberg (1959) and Schelling (1960), who claimed that perceived madness could make threats more credible. According to Schelling, 'It may be perfectly rational to wish oneself not altogether rational' in coercive bargaining (1960, 18).

However, it does not appear that Nixon's madman strategy helped him quickly end the Vietnam War. Sechser and Fuhrmann (2017) argue that he was unsuccessful at persuading observers he was mad. Furthermore, even if a leader does develop a true reputation for madness, it is not clear that this is always beneficial to coercive success. For example, the perception that Saddam Hussein was a madman was crucial to the Bush Administration's justification for a preventive attack on Iraq. Saddam's reputation for madness therefore seemed to be a liability in his standoff with the United States, raising doubts about the Madman Theory's validity.

The theory has received renewed attention since the election of Donald Trump, who critics have accused of being mentally unstable. Media commentaries have discussed what the Madman Theory implies for Trump's foreign policy and if the perception – whether true or not – that he is mentally unstable can be advantageous in foreign policy (Krauthammer 2017; Nedal and Nexon 2017; Walt 2017). Unfortunately, the political science literature can shed little light on this question. Despite the enduring fame of the Madman Theory, it has never been empirically tested on a large scale.

This article addresses this gap. I begin by analyzing the logic underlying the Madman Theory, bringing together the arguments of various proponents of the theory and identifying commonalities in their reasoning. I argue that for the purpose of analyzing the Madman Theory, madness should be broadly defined as deviating from 'normal' payoffs or decision making in a way that makes a leader more likely to use force. Based on this definition, I make the case that perceived madness can have both benefits and drawbacks in coercive bargaining. I also theorize about how the relative importance of these benefits and drawbacks will vary based on various conditions, including the type of coercion being attempted, the strength of the madness reputation and the balance of military power.

In order to assess the validity of the Madman Theory and my own hypotheses, I introduce an original perception-based measure of leaders' international reputations for madness, coded based on news reports. I use this measure to perform the first large-N test of the Madman Theory, examining the effect of reputations for madness on both general deterrence and crisis bargaining. I find that perceived madness is clearly harmful to general deterrence and typically has a harmful or insignificant effect in crisis bargaining. However, it may be helpful in crisis bargaining under certain conditions, particularly when the reputation for madness is slight and is coupled with strong military power. My analysis suggests that the harmful effect of perceived madness results from a commitment problem.

This article has five important implications for both theory and policy. First, it offers an important course correction for the conflict bargaining literature, which to date has put much more emphasis on the benefits of perceived madness than the drawbacks. My findings show major drawbacks to having a reputation for madness, particularly when the reputation is strong. This more pessimistic view is particularly timely as the Madman Theory is gaining prominence. The major policy implication of my findings is that leaders should be cautious about seeking to gain a strategic advantage by promoting the perception that they are mad, as this can be counterproductive.

Secondly, the article makes broader contributions to theoretical debates. It contributes to the conflict bargaining literature by joining a growing body of work that emphasizes the importance of the commitment problem and mistrust in bargaining failures. The conflict bargaining literature has traditionally emphasized difficulties in conveying resolve as a key obstacle to successful coercion, but some recent scholarship (Kydd 2005; Sechser 2010) has built on Jervis's (1976) classic work to argue that having too much power or resolve can undermine peaceful coercion by promoting suspicion. My finding that having a strong reputation for madness undermines coercive success due to a commitment problem supports this view.

Thirdly, the article contributes to a burgeoning literature on leaders in international relations. Recent work has demonstrated the influence of leaders' domestic incentives and biographical experiences on international outcomes (Carter 2016; Chiozza and Goemans 2011; Croco 2015; Horowitz, Stam and Ellis 2015). Yarhi-Milo, Kertzer and Renshon (2018) have also shown that individual leaders' psychology is important for determining how international signals are perceived. In this article, I take this literature in a new direction by showing that *perceptions* of a leader's psychology affect coercive outcomes.

Fourthly, this article breaks new ground by being one of very few works in recent decades to deal seriously with the topic of irrationality. With a few exceptions (for example, Acharya and Grillo 2015), the highly influential rational choice perspective has avoided considering the consequences of irrationality. In other strands of literature, there has been work on cognitive limitations (Jervis 1976), emotional and intuitive decision making (Lebow 2010; Rathbun 2018), and other behavioral deviations from rational choice predictions (Yarhi-Milo, Kertzer and Renshon 2018). However, none of this work addresses extreme forms of irrationality. By focusing on reputations for madness, this article sheds light on the effect of more extreme deviations or perceived deviations from rationality, which are likely to be less frequent but highly impactful.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

Finally, this article is the first to consider the complicated conceptual, definitional and logistical issues associated with testing the Madman Theory on a large scale. I discuss the rationale behind my approach to conceptualizing and coding reputations for madness and compare it to other possible approaches. I also discuss the inherent challenges associated with collecting and analyzing this type of data and how they are overcome. Therefore, in addition to introducing original data on leader madness reputations and presenting the first large-N test of the Madman Theory, this article establishes a basis on which future scholars can build to test it and related theories in different ways.

The madman theory

The idea that perceived madness can be helpful under some circumstances dates back to at least Machiavelli, who stated that 'at times it is a very wise thing to simulate madness' (*Discourses on Livy*, book 3, chapter 2, 1517). In the nuclear era, this idea began to receive more attention as scholars considered how to make nuclear threats credible.

Ellsberg (1959) gave the earliest and fullest articulation of the Madman Theory. He considers a situation in which one country's leader (the blackmailer) makes a demand of another country and threatens war. Ellsberg argues that the blackmailer is more likely to be successful if he is '*convincingly* mad' (1959, 2). Ellsberg identifies two subtypes of madness, which can both enhance the credibility of the blackmailer's threat. The first subtype is unpredictability, which means a propensity to deviate from predictable decision making based on a cost–benefit analysis. This implies that a leader could choose to do anything, even something suicidally aggressive. Ellsberg's second subtype is 'abnormal payoffs', which means that a leader acts predictably and makes decisions based on their expected payoffs, but the payoffs are abnormal in the sense that war is viewed as uncostly or total victory is viewed as the only acceptable outcome. These two subtypes of madness both suggest a greater propensity to use force in situations where a typical leader – that is, one with more normal payoffs and decision-making procedures – would hesitate to do so. This means that leaders who are perceived as mad can make credible threats even when conflict is very costly. Thus when war is very costly for both sides, a typical opponent is likely to acquiesce to a convincingly mad blackmailer's demand.

Thomas Schelling also argued that a reputation for madness can be an asset in coercion. He does not offer an explicit definition of madness, but implicit in his writing is that mad leaders can credibly threaten suicide. Schelling (1960, 18) notes, 'Many of the attributes of rationality… are strategic disabilities in certain conflict situations'. Schelling (1966, 37) offers more detail, stating that a 'paradox of deterrence is that it does not always help to be, or to be believed to be, fully rational, cool-headed, and in control of oneself'. He cites examples of successful coercion by an anarchist fanatic and by mental patients, who can each credibly threaten to kill themselves. He goes on to note how Khrushchev's displays of irrationality raised the credibility of Soviet threats over Berlin.

In more recent times, discussions of madness have been mostly absent from the rationalist literature. One exception is Little and Zeitzoff (2017), who present a formal model of take-it-or-leave-it bargaining in which preferences evolve over generations. They show that evolution might favor 'irrationally tough' actors who are willing to reject low offers even if fighting yields a worse outcome. Another notable exception is Acharya and Grillo (2015), who incorporate the possibility that one player is crazy into a multi-stage conflict bargaining model. They define craziness as making unreasonable offers and always choosing the more aggressive option. They find that rational leaders can sometimes improve their expected payoffs by pretending to be crazy.

This discussion illustrates that there are differences in how previous scholars who have written about the Madman Theory have defined madness. However, a crucial commonality among the definitions is a willingness to resort to violence even when standard cost–benefit decision making

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

would cause an individual with 'normal' preferences – such as a preference to avoid massive loss of life – to prefer backing down. Therefore, for the remainder of this analysis, I define madness as *deviation from normal payoffs or decision making in a way that makes a leader more likely to use force*.

This definition has important implications for coercive bargaining. Because madness is associated with a greater likelihood of using force, threats of force that would ordinarily not be credible can become credible if the leader issuing them is viewed as mad. This increased credibility should arguably make leaders with reputations for madness more successful at coercion because, all else equal, adversaries are more likely to back down when they believe a threat is genuine. This is the essential logic at the heart of the Madman Theory's assertion that perceived madness is an asset in coercive bargaining.

Despite the Madman Theory's fame, attempts to assess it empirically have been limited. Some support comes from the psychology literature, which has analyzed the effect of emotional attributes related to madness on bargaining. Studies show that expressions of anger (Van Kleef and Côté 2007) and emotional inconsistency (Sinaceur et al. 2013) help to achieve concessions in negotiations. However, we cannot assume that these experimental findings necessarily apply to international negotiations in the shadow of war. Wong (2019) presents evidence that expressions of anger – particularly by usually stoic leaders – were influential in the Berlin Crises, but broader testing is needed.

Drawbacks of perceived madness

Despite the argument that it can lend credibility to threats, perceived madness also has potential drawbacks. Indeed, neither Ellsberg nor Schelling believed that playing the madman was a good idea in practice. In his memoirs, Ellsberg (2017, 311) states that he 'never thought of it as an approach that would appeal to an American leader, nor be remotely advisable under any circumstances'.[1] Schelling (1966, 40) similarly asserts that while a madman strategy might give leaders a 'short cut to deterrence', it is preferable to establish deterrence in a more mature and responsible way. In addition, Acharya and Grillo (2015) and Little and Zeitzoff (2017) show that the presence of mad or possibly mad leaders increases the risk of war, despite the greater credibility of these leaders.

I argue that the biggest drawback of perceived madness is a commitment problem. Successful coercion requires not only a credible threat to attack following noncompliance with a demand, but also a credible (though often implicit) promise not to attack following compliance. As Schelling (1966, 74) wrote, 'To say, "One more step and I shoot," can be a deterrent threat only if accompanied by the implicit assurance, "And if you stop I won't."' Fleshing out this logic, Kydd and McManus (2017) show formally that when a state has the option to attack even after deterrence or compellence succeeds, having an overly low cost of war can undermine coercive success because of an inability to commit to peace. Similarly, Weisiger (2013) argues that leaders who are believed to have unusually aggressive dispositions struggle to achieve peace because their adversaries believe lasting security requires their removal.

This research is relevant to the Madman Theory because, given the perception that they deviate from normal payoffs or decision making in a way that makes them more likely to use force, it is difficult for perceived madmen to credibly commit *not* to attack. Opponents may resist making concessions due to fears of future betrayal, putting leaders with a reputation for madness at a disadvantage in coercive bargaining. In the extreme, the commitment problem caused by perceived madness could lead to preventive war. Since tolerating the presence of a madman entails an elevated sense of risk, preventive war against perceived madmen might be particularly likely when an adversary has a low risk tolerance.

[1] Ellsberg also reiterated this point in an email to the author.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

Some empirical work also suggests that perceived madness has drawbacks or at least limited benefits. Sechser and Fuhrmann (2017) find in case studies that attempts by Nixon, Khrushchev and North Korean leaders to demonstrate madness failed to have the intended effect on perceptions. Similarly, McManus's (2019) case studies find a helpful effect of perceived madness for Hitler, but not Khrushchev, Qaddafi or Saddam Hussein. Ausderan's (2017) survey experiment also shows that many members of the public are willing to support military action against a leader who has made apparently irrational threats. However, these tests are fairly limited in scope, and more work is necessary to either prove or disprove the Madman Theory.

Theoretical expectations

This manuscript tests the Madman Theory on a larger scale and theorizes about the conditions under which it is most likely to apply. To test the basic expectation that perceived madness is beneficial to coercion, I consider two specific types of coercion: general deterrence and crisis bargaining. Although the Madman Theory is most famously associated with compellent threats issued in crisis bargaining (for example, Nixon's threats toward Vietnam or Khrushchev's threats over Berlin), the logic of the Madman Theory implies that perceived madness can also increase the probability of successful general deterrence because the risk of an insanely aggressive response should dissuade potential challengers. Indeed, Ellsberg and Schelling each viewed the Madman Theory as applying to both deterrence and compellence. Ellsberg says that perceived madness can be used strategically 'on either side of the bargaining table' (1959, 4), and Schelling (1966, 37) gives the example of how a fictional anarchist deterred the police from arresting him with the threat of a suicidal explosion. Therefore, I test the following two hypotheses derived from the Madman Theory:

Hypothesis 1a: Leaders with a reputation for madness will be more successful at general deterrence.

Hypothesis 1b: Leaders with a reputation for madness will be more successful at crisis bargaining.

Although the logic of the Madman Theory is compelling, we must keep in mind that there can be drawbacks to perceived madness as well – which suggests that Hypotheses 1a and 1b may be neither universally true nor universally false. It is possible to further theorize about the specific conditions under which perceived madness is more likely to be helpful or harmful to coercive success.

First, we can consider the strength of a leader's reputation for madness – that is, the degree to which a leader is perceived to deviate from normal payoffs or decision making. Some leaders may deviate from normality only slightly, for example by viewing war as only a little less costly than normal or by making spontaneous decisions only occasionally. Other leaders might deviate in more extreme ways, such as by being megalomaniacs or making all decisions impulsively. I argue that as a leader's reputation for madness grows stronger (that is, when the degree of madness that the leader is perceived to suffer from increases), the drawbacks of perceived madness will eventually begin to outweigh the benefits. If a certain level of madness is necessary to give a threat credibility, then any increase in the strength of a leader's madness reputation up to that level is an asset. However, any increase beyond that level provides no additional coercive credibility and is more likely to raise doubts about whether the leader is capable of maintaining peace. This leads to the following hypothesis:

Hypothesis 2: Compared to a more moderate reputation for madness, a strong reputation for madness is less likely to be beneficial and more likely to be detrimental to coercive success.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

Secondly, although the Madman Theory can apply to both general deterrence and crisis bargaining, I do not expect perceived madness to have an entirely equivalent effect in both scenarios. In crisis bargaining, the heightened stakes create an incentive to bluff, and it is necessary to convince the other side to back down in order to prevail. These attributes of crisis bargaining make any increase in credibility that results from perceived madness particularly valuable. In general deterrence, by contrast, threats are more likely to be implicit, but they may have greater inherent credibility because it is easier to commit to defend the status quo than to overturn it. Furthermore, persuading an adversary not to challenge in the first place is likely to be easier than convincing the adversary to back down. Therefore, perceived madness might rarely be necessary to make successful general deterrence threats. Furthermore, the commitment problem created by perceived madness might sometimes provide an incentive to attack preventively, which would increase the risk of general deterrence failure among leaders viewed as mad. This leads to the following hypothesis:

Hypothesis 3: A reputation for madness is less likely to be beneficial and more likely to be detrimental to coercive success in general deterrence, compared to crisis bargaining.

Finally, we consider the relationship between perceived madness and military capabilities. If a country is militarily weak relative to its adversary, then its threats will face a barrier to credibility because of the costliness of war and the low probability of success. In this scenario, perceived madness can help to overcome this barrier by making opponents believe that the leader is willing to bear any cost or is not rationally considering the expected outcome of fighting at all. Therefore, perceived madness has the greatest potential to be an asset for militarily weak leaders. In contrast, when a country is more powerful than its adversary, this greater power is likely to exacerbate the commitment problem associated with perceived madness. A leader who is viewed as both mad and powerful will be considered to have both the means and the inclination to launch future conflicts. This will increase the perceived risk associated with accommodating the madman and will therefore encourage adversaries to resist his threats and possibly even attack preventively. This leads to the following hypothesis:

Hypothesis 4: A reputation for madness is less likely to be beneficial and more likely to be detrimental to coercive success when a country has greater relative military strength.

Measuring madness reputations

To test these hypotheses, it is crucial to measure leaders' reputations for madness using a systematic and large-scale approach. One measurement approach to consider is psychological. Previous scholars have evaluated leaders' psychology from afar, using case studies (McDermott 2007; Renshon 2011), quantitative analysis of speeches (Ramey, Klingler and Hollibaugh 2019; Renshon 2008) and expert surveys (Yarhi-Milo 2018). A downside of the psychological approach for my research question is that it would measure *actual* madness, whereas my hypotheses focus on *reputations* for madness. Furthermore, the psychological evaluation methods cited above are difficult to implement on a large scale. It would also strain credulity to claim that I could accurately identify leaders' true levels of madness on a large scale, when those with greater psychological expertise can struggle to do this on a small scale.

Another option for measuring madness reputations would be behaviorally based. A behavioral measure might incorporate a variety of actions that could create the impression of madness, such as initiating losing conflicts, vacillating between aggressive and cooperative behavior, or introducing erratic domestic policies. However, behavioral coding would be unlikely to fully capture a leader's reputation for madness. One reason is that context matters: behavior that might seem

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

mad in one context could seem sane in another. In addition, some relevant behaviors might be too idiosyncratic to code systematically.

Ultimately, I argue that the best way to measure reputations for madness is based on public perceptions, as reflected in the media. This approach does not claim to capture true levels of madness; it is the *perception* of madness that is of interest for testing the Madman Theory. My basic approach is to tally the number of times that a leader is referred to in the press using adjectives indicating madness. This approach enables me to code madness reputations for all leaders worldwide and create a measure that is independent of my own biases.

Of course, my approach has limitations. One is that perceptions reflected in the press may differ from the perceptions of policy makers. However, the perceptions of the informed public, as represented by the media, and the perceptions of policy makers are likely to be closely correlated. Furthermore, my coding captures nuances in the strength of madness reputations. The more often a leader is called mad in the global media, the more widespread the reputation for madness is likely to be, and the more likely it is to be shared by policy makers.

A second potential concern about my measurement approach is that the use of madness adjectives in the press may reflect certain biases. Given that I rely on English-language sources, the biggest concern is that there may be a pro-Western bias against leaders who challenge Western hegemony. There may also be other biases, such as bias against dictators, biases based on political orientation, or biases based on race, gender, age or other demographic characteristics. I address these concerns in two ways. First, I begin my analysis by exploring which leader and country characteristics make a leader more likely to be called mad, and I control for these characteristics later. Secondly, in my robustness checks, I drop certain sources and dyads in order to reduce the impact of pro-Western bias. While it is impossible to control for every possible bias, these procedures address the biases that are likely to be the most systematic.

Coding Process

To code reputations for madness, my research team undertook searches of English-language news reports and editorials from around the world in the Lexis-Nexis database and identified instances in which national leaders were described using synonyms for madness. Leaders were identified using the Archigos dataset (Goemans, Gleditsch and Chiozza 2009). We restricted the search to 1986–2010 based on the availability of Lexis-Nexis articles and military dispute data. We searched for three adjectives associated with madness: *crazy*, *insane* and *irrational*. These adjectives were selected because they are commonly used words, and – in keeping with my definition of madness –they all suggest deviation from normal preferences or decision making and have some connotations of potential aggression.[2] The Appendix contains a more detailed description of the word search procedure.

In order to create the madness reputation variables used in the statistical analysis, I first tallied uses of the madness adjectives by leader-year. Next, it was necessary to normalize the tally because Lexis-Nexis coverage varies widely by country and year. More sources are added to the database each year, and there is greater coverage of powerful, Western and English-speaking countries. One way to deal with this would be to divide each leader-year tally of madness words by the total number of articles mentioning the leader in that year. However, this method would bias downward the scores of leaders who receive heightened press coverage precisely because they are perceived as mad and thereby mask important variation

[2] OxfordDictionaries.com (2019) defines *crazy* as 'mad, especially as manifested in wild or aggressive behavior'. OxfordDictionaries.com defines *insane* as 'in a state of mind which prevents normal perception, behaviour, or social interaction', while the MacMillan Dictionary (2019) adds that the word is especially associated with the likelihood of causing 'serious problems, harm, or injury'. The word *irrational* clearly indicates deviation from normal rational decision making and is also commonly associated with aggression, as the first phrase suggested when 'irrationally' is typed in the Google search box is 'irrationally angry' (as of 20 April 2019).

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

among leaders.[3] Therefore, I instead predict the amount of *expected* news coverage that a typical leader would receive based on attributes of the leader's country and then normalize the madness word tallies by this predicted value. Appendix Table A1 contains further explanation and the prediction regression.

Measure Description

This procedure yields a continuous measure of reputation for madness. Higher values of this measure indicate that a leader's madness reputation is more widespread because the leader is being referred to as mad more frequently across multiple situations, speakers and news outlets. A higher score also implies that a leader has a *stronger* madness reputation, that is, he or she is viewed as mad to a greater degree. This is because the more widespread use of madness adjectives to describe a leader reflects higher collective confidence that the leader is the type who will deviate from normal decision making and payoffs. Furthermore, the same behavior patterns or political and social dynamics that cause a leader's madness reputation to become widespread are also likely to convince observers that the leader is mad to a greater degree.

The distribution of the continuous madness reputation measure is highly skewed. The variable equals zero in over 95 per cent of leader-years. Among leader-years with values above zero, a few leader-years have much higher values than the others (see Figure A1). Because of this skewed distribution, in addition to using the continuous madness reputation measure, I create two indicators as alternatives. The first, *Strong Madness Reputation*, identifies leader-years in which the continuous measure is in the top 15 per cent among non-zero values. The second, *Slight Madness Reputation*, identifies leader-years in which the continuous measure is in the lower 85 per cent among non-zero values. The top 15 per cent cutoff was chosen based on the fact that values of the continuous measure begin to increase rapidly around this point, as shown in Figure A1. Creating these indicators addresses concerns about the skewed distribution and allows me to test Hypothesis 2.[4]

Table 1 shows all leaders coded as having strong madness reputations and the leaders who are most frequently coded as having slight madness reputations. The leaders with strong madness reputations include many we might expect, such as Saddam Hussein, Kim Jong-il, Mahmoud Ahmadinejad and Muammar Qaddafi. A few others, such as Abdalá Bucaram Ortíz (nicknamed 'El Loco') [5] and Thabo Mbeki (who advocated herbal remedies to cure AIDS), are not known for behaving insanely on the international stage, but were domestically controversial. The list of leaders with slight madness reputations contains some of the same individuals as the prior list, since the strength of their madness reputations varied by year. However, this second list also includes some hawkish Western leaders, such as George W. Bush and Tony Blair. The differences between the two lists provide further justification for using the two indicators as an alternative to the continuous measure.

What Influences Madness Reputations?

Before discussing the impact of madness reputations on coercion, I consider which factors determine a leader's reputation for madness. Due to space constraints, I leave the task of developing a fully specified theory of madness reputation formation to future research. Nonetheless, I briefly

[3] For example, this method would result in Saddam Hussein, who receives disproportionally high press coverage based on his 'mad' behavior, receiving a lower average madness reputation score than Jamil Mahuad of Ecuador, who was called mad only once in his tenure but receives very little press coverage.

[4] Logging the continuous variable does not substantively change the result (Tables A24, A26). Including squared or cubic terms would further exaggerate the effect of extreme values.

[5] Because Bucaram Ortíz's score is such an outlier, I confirm the results are robust to dropping him from the sample (Table A23).

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

Table 1. Madness reputation coding

Leader	Years in category	Avg. madness score
All leaders coded as having a strong madness reputation		
Saddam Hussein, Iraq	6	0.516
Robert Mugabe, Zimbabwe	6	0.310
Mahmoud Ahmadinejad, Iran	3	0.606
Kim Jong-il, North Korea	3	0.299
Slobodan Milosevic, Serbia	2	0.199
Muammar Qaddafi, Libya	2	0.140
Abdalá Bucaram Ortíz, Ecuador	1	3.231
Jamil Mahuad, Ecuador	1	0.417
Itamar Franco, Brazil	1	0.400
P.W. Botha, South Africa	1	0.324
Thabo Mbeki, South Africa	1	0.260
Kim Il-sung, North Korea	1	0.240
Mikhail Saakashvili, Georgia	1	0.136
Jean Chrétien, Canada	1	0.124
Fidel Castro, Cuba	1	0.093
Hun Sen, Cambodia	1	0.026
Leaders most frequently coded as having a slight madness reputation		
John Howard, Australia	9	0.116
Tony Blair, UK	9	0.077
George W. Bush, USA	8	0.140
Robert Mugabe, Zimbabwe	7	0.310
Kim Jong-il, North Korea	6	0.299
Thabo Mbeki, South Africa	6	0.260
Hugo Chavez, Venezuela	6	0.176
Muammar Qaddafi, Libya	6	0.140
Ariel Sharon, Israel	5	0.204

address this topic here so that I can properly control for the determinants of madness reputation in my main analysis. I explore the effect of various factors that might influence how leaders are perceived, including biographic experiences, age, gender (Horowitz, Stam and Ellis 2015), regime type (Geddes, Wright and Frantz 2014; Marshall, Jaggers and Gurr 2010), economic conditions (Gleditsch 2002) and the number of militarized interstate disputes (MIDs) initiated by the leader over the past five years (Palmer et al. 2019). I estimate a tobit model predicting the continuous madness reputation measure and probits predicting each indicator.

The results, reported in Appendix Table A4, show that the only variables with consistently significant effects across all three models are the leader's recent MID initiations and years in office, both of which increase the probability and strength of madness reputations. Democracy is a positive and significant predictor of the continuous and slight madness reputation variables (probably because the domestic press is allowed to call the leader mad in democracies), but not of the strong madness reputation indicator. No other variables are significant, suggesting that many reasons for the formation of madness reputations are idiosyncratic.

Research design

I now discuss how I use the madness reputation measures to test the hypotheses. I use MIDs – instances in which one state threatened, showed or used force against another – as the basis of my analysis. As explained earlier, I restrict my analysis to the years 1986–2010, during which 742 MIDs occurred.[6] I identify the dyads that actually interacted within each MID using the MID

[6] The Militarized Compellent Threat dataset (Sechser 2011) records only forty threats during this period, making it infeasible as an alternative for statistical analysis. However, it shows that the compliance rate for leaders with madness reputations is 0, compared to 13 per cent for leaders without such a reputation.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

4.3 Incident-Participant Dataset (Palmer et al. 2019)[7] for 1993–2010 and the Dyadic MID 3.0 dataset (Maoz et al. 2018) for 1986–1992. From 1993–2010, I am also able to adjust the initiator coding by dyad.[8]

To analyze general deterrence, I use a dataset of politically relevant directed dyad-leader-years, in which Leader A is the potential initiator and Leader B is the potential target.[9] Using this data structure allows me to control for dyad-level factors that influence dispute initiation and makes my results comparable to previous research. Whenever a country's leader changes mid-year, I include multiple observations per dyad-year so that each pair of overlapping leaders is included. Since the observations do not all reflect the same unit of time, I control for the number of days contained in each. My dependent variable for testing Hypothesis 1a is *Initiation*, which records whether a MID was initiated. Initiation represents a deterrence failure by Leader B.

To test Hypothesis 1b about crisis bargaining, I use a dataset of dyadic MIDs. State A is the dyadic initiator, whereas State B is the target. Following Schultz (1999) and Weeks (2008), I use the dependent variable *Reciprocation* – an indicator of whether State B made any threat, show or use of force in response to State A's MID initiation – as a proxy for coercive success in crisis bargaining. If one state initiates a dispute against another, this constitutes a crisis situation. If the state targeted in the dispute does not reciprocate with any military threat or action of its own, this suggests that the target was most likely intimidated – at least on average, although sometimes targets may not respond for other reasons.[10] Therefore, non-reciprocation by the target implies more successful coercion by State A's leader. I predict both dependent variables using probit models.

My main independent variables are the madness reputation measures described above, which are all lagged by one year. To distinguish the effect of a madness reputation from behavior that might cause that reputation, I control for *Recent MID Initiations*, the number of MIDs initiated by the leader over the past five years.[11] I found earlier that this variable is a significant predictor of madness reputation, and it may also affect perceptions of resolve and trustworthiness in a similar way to perceived madness. I also include standard controls. In all regressions, I control for military capabilities (Singer, Bremer and Stuckey 1972), democracy (Marshall, Jaggers and Gurr 2010), contiguity (Stinnett et al. 2002) and distance (Bennett and Stam 2000). In the initiation regressions, I also control for the number of days contained in each observation and a cubic polynomial of peace years. In the reciprocation regressions, I control for the hostility level of the action that initiated the MID (Jones, Bremer and Singer 1996; Palmer et al. 2019). The results are also consistent in models with fewer controls (Tables A24, A26).[12]

Main results

The main results are shown in Table 2. The models in this table can be used to evaluate Hypotheses 1–3. Leader B's coefficients are of primary interest for analyzing general deterrence

[7] The MID 4.3 dataset incorporates recommendations by Gibler, Miller and Little (2016) to drop and merge certain MIDs.

[8] Adjustments to the initiator coding are planned, but have not yet been implemented, in the Dyadic MID 3.0 dataset. However, in the years covered by the MID 4.3 Incident-Participant Dataset (1993–2010), I found that the dyadic initiator differs from the state on Side A of the MID in only about 4 per cent of dyads, suggesting that the absence of this adjustment in 1986–1992 is not a major problem.

[9] Politically relevant dyads include either a major power or two contiguous countries, separated by less than 401 miles of water (Bennett and Stam 2000; Stinnett et al. 2002).

[10] There may be a variety of reasons not to respond, including various international or domestic distractions and impediments. Of particular concern might be that the target state simply does not deem the initiating act worthy of a response. To alleviate this concern, I control for the hostility level of the initiating act, since more hostile acts are more likely to require a response.

[11] The results are robust to different methods of calculating this variable (Tables A23, A25).

[12] I exclude countries with populations under 500,000 (Singer, Bremer and Stuckey 1972) from the sample because of frequent missing values, but the results are robust to retaining those for which data are available (Tables A22, A24). There are no missing values for countries with populations over 500,000.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

Table 2. Main results

	1	2	3	4
	Initiation	Initiation	Reciprocation	Reciprocation
Continuous madness rep., A	0.203**		0.272***	
	(0.084)		(0.063)	
Continuous madness rep., B	0.385***		0.045	
	(0.061)		(0.123)	
Strong madness reputation, A		0.162		0.442***
		(0.211)		(0.154)
Slight madness reputation, A		0.098		−0.160
		(0.071)		(0.215)
Strong madness reputation, B		0.907***		−0.549***
		(0.140)		(0.199)
Slight madness reputation, B		0.160**		−0.207
		(0.075)		(0.175)
Recent MID initiations, A	0.236***	0.232***	−0.040	−0.038
	(0.031)	(0.031)	(0.103)	(0.105)
Recent MID initiations, B	0.061*	0.048	0.059	0.098
	(0.031)	(0.032)	(0.094)	(0.098)
Military capabilities, A	1.891**	1.788**	−0.077	−0.239
	(0.753)	(0.772)	(2.118)	(2.171)
Military capabilities, B	1.364**	1.404**	−2.365	−2.671*
	(0.572)	(0.583)	(1.560)	(1.578)
% Military cap. held by A	−0.013	−0.019	0.358	0.242
	(0.165)	(0.171)	(0.550)	(0.520)
Democracy, A	0.128**	0.105*	−0.249	−0.197
	(0.062)	(0.061)	(0.202)	(0.201)
Democracy, B	0.103*	0.102*	−0.005	−0.016
	(0.061)	(0.062)	(0.162)	(0.156)
Joint democracy	−0.502***	−0.489***	−0.124	−0.177
	(0.107)	(0.107)	(0.306)	(0.301)
Contiguity	0.531***	0.539***	0.387***	0.312**
	(0.071)	(0.073)	(0.133)	(0.136)
Distance	−0.120***	−0.121***	0.037	0.047
	(0.025)	(0.024)	(0.040)	(0.043)
Dyad length	0.671***	0.633***		
	(0.082)	(0.078)		
Peace years	−0.042***	−0.042***		
	(0.004)	(0.004)		
Peace years squared	0.001***	0.001***		
	(0.000)	(0.000)		
Peace years cubed	−0.000***	−0.000***		
	(0.000)	(0.000)		
First act hostility			−0.035	−0.057
			(0.150)	(0.147)
Constant	−2.676***	−2.636***	−0.398	−0.179
	(0.148)	(0.149)	(0.602)	(0.601)
Observations	62,384	62,384	759	759

Note: Models 1–2 are probits predicting MID *Initiation*, with standard errors clustered by dyad. Models 3–4 are probits predicting *Reciprocation*, with standard errors clustered by State A. The madness reputation variables are lagged by one year. *p < 0.10, **p < 0.05, ***p < 0.01.

success in Models 1–2, whereas Leader A's coefficients are of primary interest for analyzing crisis bargaining success in Models 3–4. We can begin by examining the coefficients of the continuous madness reputation measures. The positive and significant coefficient for Leader B in Model 1 indicates that a stronger madness reputation of Leader B makes State B more likely to be targeted in a MID. Similarly, the positive and significant coefficient for Leader A in Model 3 indicates that a stronger madness reputation of Leader A raises the probability that a MID initiated by Leader A will be reciprocated. Thus, contrary to the Madman Theory, these results suggest that perceived madness is detrimental to both general deterrence and crisis bargaining.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

The results of Models 2 and 4 likewise indicate that having a madness reputation is never helpful and is often detrimental to coercion. However, these models allow us to go deeper and evaluate the differing effects of slight and strong madness reputations. In Model 2, both madness coefficients for Leader B are positive and significant, but the *Strong Madness Reputation* coefficient is larger and more significant. In Model 4, Leader A's *Strong Madness Reputation* coefficient is positive and significant, while Leader A's *Slight Madness Reputation* coefficient is insignificant. This indicates that a stronger reputation for madness is more detrimental to both deterrence and crisis bargaining than a slight reputation for madness.

This can be seen more clearly in Figure 1, which shows predicted probabilities based on Models 2 and 4. The left graph shows that leaders with strong madness reputations face a probability of being targeted in a MID that is over four times higher than leaders with slight madness reputations and nearly six times higher than those with no madness reputation. In contrast, leaders with slight madness reputations are only 1.4 times more likely to be targeted than leaders with no madness reputation, and this difference is only significant at the 94 per cent confidence level. The right graph shows that leaders with strong madness reputations face about a 40 per cent higher predicted probability of MID reciprocation than leaders with no madness reputation, and this difference is statistically significant. In contrast, leaders with slight madness reputations appear to face a slightly lower predicted probability of MID reciprocation than those with no madness reputation, but this difference is far from significant.

This comparison of the *Slight* and *Strong Madness Reputation* coefficients provides support for Hypothesis 2, as stronger madness reputations are found to be more detrimental to coercive success. A strong madness reputation is shown to have a large detrimental effect in both general deterrence and crisis bargaining, while a slight madness reputation has a substantively smaller detrimental effect in general deterrence and no significant effect in crisis bargaining. Additionally, given that no type of perceived madness is found to be helpful, the evidence still goes against Hypotheses 1a and 1b, which were derived from the Madman Theory.

Despite some common patterns between the deterrence and crisis bargaining results, there are also important differences that shed light on Hypothesis 3. First, while *Slight Madness Reputation* has a significant harmful effect in the deterrence regression, it has an insignificant effect in the crisis bargaining regression. Secondly, although *Strong Madness Reputation* and the continuous madness reputation measure each have a significant detrimental effect on both deterrence and crisis bargaining, the significance for crisis bargaining is less robust. The significance of *Strong Madness Reputation* in the crisis bargaining regression disappears if I lower the threshold for this variable from the top 15 per cent of non-zero values to the top 20 per cent, but the variable's significance in the deterrence regression is robust when I vary the threshold anywhere between the top 5 per cent and the top 40 per cent. Additionally, the continuous measure in the deterrence regression remains significant if I drop up to the top 16 per cent of non-zero values, but I cannot even drop the top 1 per cent from the crisis bargaining regression without losing significance (Tables A7, A14).[13] This indicates that only the very strongest levels of perceived madness are detrimental in crisis bargaining, while most levels of perceived madness have no significant effect. In contrast, a wider range of perceived madness values are clearly detrimental in general deterrence. This accords with Hypothesis 3, which predicted that perceived madness was more likely to be harmful in general deterrence.

Three additional aspects of the results are worth noting before moving on. First, Model 1 suggests that leaders with stronger madness reputations initiate more MIDs, but Model 2 does not corroborate this. Secondly, the perception-based measure of Leader B's madness reputation is a much better predictor of deterrence failure than Leader B's history of MID initiation. This illustrates the importance of perceptions and the fact that perceptions are not necessarily straightforwardly tied to behavior. Thirdly, in the crisis bargaining regressions, we see that few variables

[13] This entails dropping Saddam Hussein in the Gulf War.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

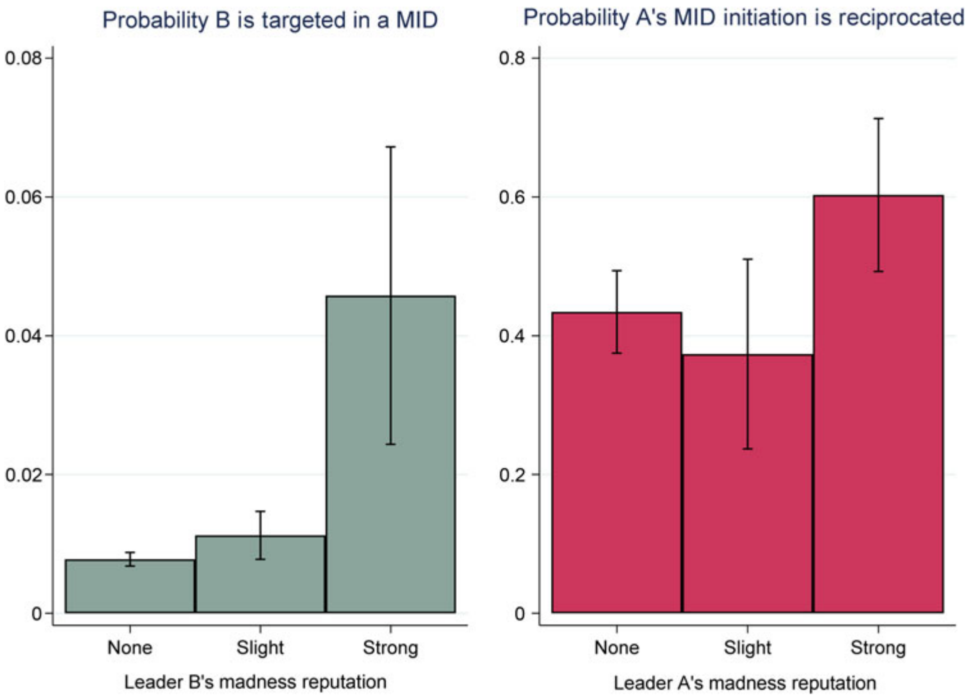

Figure 1. Predicted probabilities from Models 2 and 4
Note: the figure shows average predicted probabilities, produced by calculating the predicted probability for every observation and averaging. The lines represent 95 per cent confidence bounds.

other than Leader A's madness reputation are significant predictors, although a strong madness reputation of Leader B, surprisingly, is a negative and significant predictor.

Robustness

So far, we have found no support for the expectation of Hypotheses 1a/1b that perceived madness is helpful, but we have found variation in the extent to which it is harmful, in ways that accord with Hypotheses 2 and 3. Before exploring the final hypothesis about the relationship with military capabilities, I will briefly address the robustness of these basic results, using Models 2 and 4 as the baseline.

I first explore potential concerns regarding my measurement method. As noted previously, one concern relates to reporting bias. My normalization should have eliminated most bias due to differences in reporting frequency across regions and time, but for greater confidence I add region and time fixed effects to the regressions (Tables A8, A15). A more serious concern is that, given the dominance of Western sources in my sample, a pro-US or pro-Western bias may be causing anti-Western leaders to be called mad more frequently. In the worst case, this type of bias could suggest the possibility of reverse causation because Western government officials might deliberately portray their opponents as mad before initiating MIDs against them, and the media might follow their lead. I address this concern in several ways.

First, I drop adjectives used in the context of quotations from the calculation of the madness reputation measure, since these adjectives are the most likely to be employed strategically. Secondly, I control for whether a leader is anti-US by adding a measure of UN voting affinity with the United States (Gartzke 2006, Voeten and Merdzanovic 2009) to the regression. Thirdly, I attempt to reduce pro-US bias by dropping US sources from the madness

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

reputation measure. I cannot drop all Western sources because there would be too few sources left. However, as a final test, I drop all dyads including English-speaking Western countries.[14] The risk of reverse causation is highest within these dyads because sources from these countries are dominant in the press sample. I find that the detrimental effect of a strong madness reputation is robust to all of these tests, with the exception of dropping quotations in the reciprocation regression (Tables A9, A16).

There could also be more general concerns about bias due to the non-random assignment of madness reputations to leaders. The only significant madness reputation predictor that I have not already controlled for is time in office. Therefore, I add this variable to the regressions and then also drop leaders who have been in office less than five years, and the significance of a strong madness reputation remains robust (Tables A10, A17). As a more sophisticated way of addressing non-random assignment, I employ coarsened exact matching (Iacus, King and Porro 2012). This method reduces the difference between the number of leaders with mad and non-mad reputations in the sample and attempts to approximate the balance on observable factors that we would see if madness reputations were randomly assigned. As the treatment variable, I use an indicator for any madness reputation value above zero. I match on frequency of MID involvement, whether the country is Western, whether the regime is personalist, whether the leader is a former rebel, and time in office. The significance of *Strong Madness Reputation* is robust in the matched samples (Tables A10, A17).

Additionally, there might be concerns about the relationship between madness reputation and genuine madness. It could be that leaders who strategically feign madness are successful at coercion, but genuinely mad leaders drag down the success rate through strategic blunders. To explore this possibility, I re-estimate the crisis bargaining regression, dropping the fifty-two dyadic MIDs that are most likely to be strategic blunders because a minor power targeted a major power on the first day with no major power assistance. The significance of *Strong Madness Reputation* is robust to this change (Table A17), suggesting that a strong madness reputation is harmful even for leaders who make reasonable dispute initiation decisions.

There might also be concern that perceived madmen do poorly at coercion because they have a history of bluffing and develop a reputation for not following through on their threats. To address this potential concern, I tally leaders' bluffs in the past five years based on MID hostility levels and outcomes, similarly to Sartori (2005) and Weisiger and Yarhi-Milo (2015).[15] Inserting this tally into the regressions, I find that leaders with more recent bluffs are significantly less likely to initiate MIDs, although the MIDs that they do initiate are less likely to be reciprocated, probably due to a selection effect. The coefficients for perceived madness remain significant, suggesting that the harmful effect of perceived madness is not primarily caused by a bluffing reputation (Tables A10, A17). I also compare the effect of a reputation for madness to the effect of a reputation for resolve by inserting reputational measures based on the frequency with which a leader is called 'hawkish' or 'resolute' into the regressions. The madness reputation coefficients remain significant, while the reputation for resolve coefficients are never significant at the 95 per cent confidence level (Tables A11, A18).

I also explore additional variations in the calculation of madness reputation. First, I drop adjectives used in the context of domestic politics. Secondly, instead of using a lagged one-year measure, I average over the previous five and then ten years (Tables A11, A18). In addition, I perform additional robustness checks that involve adjusting the sample size and dependent variables (Tables A12, A19). I also ensure that the crisis bargaining results are robust to addressing non-random sample selection using a Heckman probit model (Table A21). Finally, I investigate whether there is an interaction between the madness reputations of each side and find no strong evidence of this (Tables A13, A20). While *Strong Madness Reputation* remains significant in the

[14] These include the US, Canada, UK, Ireland, Australia and New Zealand.

[15] This variable's coding is described further under Table A10.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

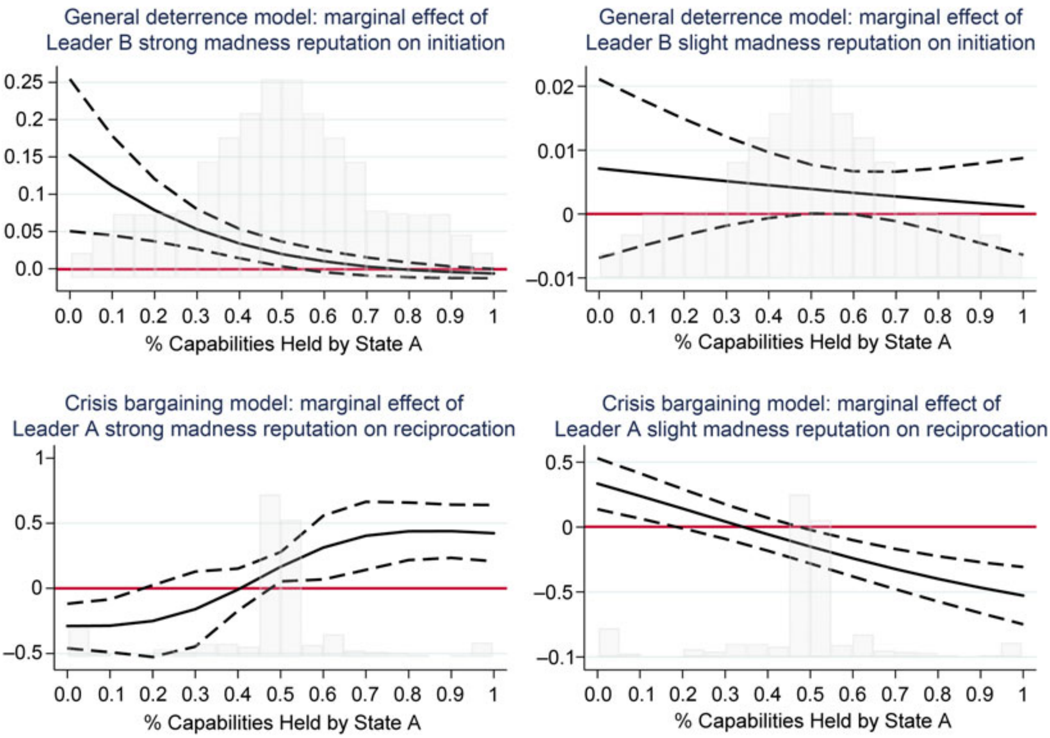

Figure 2. Marginal effects from interactions with relative capabilities
Note: these are average marginal effects. The dotted lines represent 95 per cent confidence bounds. The histograms in the background show the distribution of relative military capabilities.

general deterrence regression throughout every test mentioned, the significance of *Strong Madness Reputation* in crisis bargaining is less robust. Therefore, in accordance with Hypothesis 3, perceived madness clearly seems harmful to general deterrence, but we cannot confidently rule out the possibility that the effect in crisis bargaining is neutral.

Military capabilities interaction

Hypothesis 4 can be tested by interacting the madness reputation indicators with relative military capabilities – specifically, with the percentage of military capabilities in the dyad held by State A. Figure 2 shows marginal effects plots from the interactions; fuller results are available in Table A22. First analyzing the deterrence regression, we see in the top-left graph that when Leader B has relatively high military power (that is, when the percentage of capabilities held by A is smaller), the marginal effect of *Strong Madness Reputation* on the probability of deterrence failure is near 0.15 and significant. As Leader B's relative power weakens (that is, the percentage of capabilities held by A increases), the marginal effect becomes insignificant and near zero. This suggests that greater military strength enhances the commitment problem associated with a strong madness reputation and worsens the detrimental effect on deterrence, in keeping with Hypothesis 4. However, the top-right graph shows there is no significant interaction between relative capabilities and *Slight Madness Reputation* in the general deterrence regression.

The bottom row of Figure 2 suggests a more complicated picture in the crisis bargaining regression in two respects. First, it shows that perceived madness can be beneficial. Both *Strong* and *Slight Madness Reputation* have negative and significant marginal effects within certain ranges of relative capabilities. This finding is more meaningful in the case of *Slight Madness*

Reputation because the range in which this variable is predicted to be beneficial corresponds to some real-world values.[16] Thus the finding that *Slight Madness Reputation* has a negative and significant effect on the probability of reciprocation when State A holds at least about half of the capabilities in the dyad is the first convincing evidence we have seen that perceived madness can be beneficial under certain conditions. The leaders in the sample of dyadic MIDs who most often have slight madness reputations in conjunction with high relative capabilities include Boris Yeltsin, Ariel Sharon, Slobodan Milosevic, John Howard and Robert Mugabe.

Second, in the bottom row of Figure 2, *Strong* and *Slight Madness Reputation* have opposite effects. *Strong Madness Reputation* has a significant negative effect when State A is weaker and a significant positive effect when State A is stronger, while this pattern is reversed for *Slight Madness Reputation*. This suggests that Hypothesis 4 is too simplistic when applied to crisis bargaining, since the nature of the interaction between perceived madness and military capabilities depends on the strength of the madness reputation. The results suggest that when relative capabilities are low, a slight madness reputation is not enough to overcome the credibility barrier, and a strong madness reputation is more helpful. In contrast, when relative capabilities are high, a slight madness reputation can enhance credibility, but a strong madness reputation is too much and creates a commitment problem.

Overall, the interaction results support the logic behind Hypothesis 4 that higher relative capabilities worsen the commitment problem created by perceived madness, but this seems to be the case only when a leader has a strong madness reputation. For crisis bargaining, a slight madness reputation is actually helpful in conjunction with high military capabilities.

Conclusion

Overall, my findings suggest little support for the Madman Theory. For general deterrence, the effect of perceived madness is purely harmful. In crisis bargaining, the effect of a strong madness reputation seems to be generally harmful or at least unhelpful, but it does appear that a slight madness reputation can be beneficial when a leader's country is sufficiently powerful. In sum, therefore, the effect of perceived madness is more often harmful than helpful. The main apparent beneficiaries of a madness reputation are powerful leaders who are perceived as only slightly mad – not necessarily those we would be most likely to think of as 'madmen'.

Why does a reputation for madness often undermine coercive success? My findings suggest that the inability of perceived madmen to make credible commitments to peace is key. I find that greater relative military power, which increases the commitment problem, causes the impact of a strong madness reputation to become more detrimental. This suggests that when a reputation for madness prevents a leader from credibly committing not to attack in the future, adversaries are more likely to resist the leader firmly or even attack preventively in the present. Thus, my findings are in line with research that emphasizes mistrust and the commitment problem as causes of war.

My findings also support the growing consensus in the international relations field that leaders matter. Even after controlling for many country-level factors, a leader's reputation for madness is a significant predictor of the initiation and reciprocation of military disputes. This shows that not only a leader's behavior and biography, but also international perceptions of the leader, are very meaningful. In addition, my findings suggest that perceived madness, while rare, is highly impactful and worthy of more study by both rationalist scholars and political psychologists.

The analysis presented here is the first large-N test of the Madman Theory. Large-N research on this topic poses challenges, but also has important benefits, including the ability to examine the full universe of relevant cases and a method of coding madness reputations that is perception

[16] Side A leaders with strong madness reputations in the crisis bargaining sample never hold less than 46 per cent of the capabilities, rendering predictions for values below this somewhat suspect.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

based and yet divorced from coder biases. By carefully explaining my research design choices, as well as introducing new data, this manuscript lays the groundwork for future quantitative research on this topic. Of course, it is also desirable to test the Madman Theory using other methods, including experiments and qualitative research.

There are also other directions for future research. It might be possible to define more specific sub-types of madness – such as hotheadedness, megalomania or total loss of touch with reality – and develop hypotheses about how they affect coercive success. In addition, there may be other conditioning factors that influence the impact of perceived madness. For example, future research could explore whether the effect of a madness reputation depends upon regime type. Perhaps perceived madness matters more for dictators because there are fewer domestic restraints on their foreign policy. Future research could also investigate more deeply how leaders come to be perceived as mad and the conditions under which perceived madmen come to power.

The primary policy implication of my findings is that leaders should be very cautious about cultivating a reputation for madness. A madness reputation may have benefits in crisis bargaining, especially if the leader commands a powerful military and is able to control his/her madness reputation sufficiently to avoid it becoming too strong. However, such a reputation will almost certainly also have downsides, especially for general deterrence.

Supplementary material. Data replication sets are available in Harvard Dataverse at: https://doi.org/10.7910/DVN/T3CGGV and online appendices at: https://doi.org/10.1017/S0007123419000401.

Acknowledgements. I am grateful to Nicholas Campbell-Seremetis, Charles Glaser, Edoardo Grillo, Richard Herrmann, Michael Horowitz, David Jones, Michael Joseph, Andrew Kydd, Dov Levin, David Lindsey, Elizabeth Saunders, Robert Schub, Till Weber, and participants at the 2017 and 2018 ISA Conventions, the 2017 Peace Science Society Conference, the George Washington University ISCS workshop, and the Baruch College Weissman Global Seminar for feedback on this manuscript. I thank Jamie Gallagher, Eileen Gerard, Vinuri Ranaweera, and Olivia Sztanga for excellent research assistance, and I thank Daniel Ellsberg for sharing a copy of his 1959 lecture. Support for this project was provided by a PSC-CUNY Award, jointly funded by the Professional Staff Congress and the City University of New York.

References

Acharya A and Grillo E (2015) War with crazy types. *Political Science Research and Methods* **3**(2), 281–307.

Ausderan J (2017) Confronting the Crazy Dictator: Ideology, Worldviews, and Public Attitudes toward Irrational World Leaders. Working Paper, Barry University.

Bennett DS and Stam A (2000) EUGene: a conceptual manual. *International Interactions* **26**(2), 179–204.

Carter J (2016) Damned if You Do, Damned if You Don't: Hawks, Doves, and Their Consequences for Interstate Targets. Working Paper. Available from https://jeffcarter.weebly.com/endogenoustargets.html.

Chiozza G and Goemans HE (2011) *Leaders and International Conflict*. New York: Cambridge University Press.

Croco SE (2015) *Peace at What Price? Leader Culpability and the Domestic Politics of War Termination*. New York: Cambridge University Press.

Ellsberg D (1959) The Political Uses of Madness. Lecture given at the Lowell Institute of the Boston Public Library, 26 March. Available from https://ia800102.us.archive.org/20/items/ThePoliticalUsesOfMadness/ELS005-001.pdf (accessed 8 February 2018).

Ellsberg D (2017) *The Doomsday Machine: Confessions of a Nuclear War Planner*. New York: Bloomsbury Publishing.

Gartzke E (2006) The Affinity of Nations Index, 1946–2002. Version 4. Available from https://dss.ucsd.edu/~egartzke/datasets.htm (accessed 9 November 2011).

Geddes B, Wright J and Frantz E (2014) Autocratic breakdown and regime transitions: a new dataset. *Perspectives on Politics* **12**(2), 313–331.

Gibler DM, Miller SV and Little EK (2016) An analysis of the Militarized Interstate Dispute (MID) dataset, 1816–2001. *International Studies Quarterly* **60**(4), 719–730.

Gleditsch KS (2002) Expanded trade and GDP data. *Journal of Conflict Resolution* **46**, 712–724. Version 6.0 beta GDP data. Available from https://privatewww.essex.ac.uk/~ksg/exptradegdp.html (accessed 7 January 2015).

Goemans HE, Gleditsch KS and Chiozza G (2009) Introducing Archigos: a data set of political leaders. *Journal of Peace Research* **46**(2), 269–183. Version 4.0.

Haldeman HR and DiMona J (1978) *The Ends of Power*. New York: Times Books.

Horowitz MC, Stam AC and Ellis CM (2015) *Why Leaders Fight*. New York: Cambridge University Press.

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

Iacus SM, King G and Porro G (2012) Causal inference without balance checking: coarsened exact matching. *Political Analysis* **20**(1), 1–24.

Jervis R (1976) *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.

Jones DM, Bremer SA and Singer JD (1996) Militarized Interstate Disputes, 1816–1992: rationale, coding rules, and empirical patterns. *Conflict Management and Peace Science* **15**(2), 163–213.

Krauthammer C (2017) Trump and the 'Madman Theory'. *Washington Post*, 23 February.

Kydd AH (2005) *Trust and Mistrust in International Relations*. Princeton, NJ: Princeton University Press.

Kydd AH and McManus RW (2017) Threats and assurances in crisis bargaining. *Journal of Conflict Resolution* **61**(2), 325–348.

Lebow RN (2010) *Why Nations Fight: Past and Future Motives for War*. New York: Cambridge University Press.

Little AT and Zeitzoff T (2017) A bargaining theory of conflict with evolutionary preferences. *International Organization* **71** (3), 523–557.

Machiavelli N (1517) *Discourses on Livy*. Consitution Society. Available from https://www.constitution.org/mac/disclivy3.htm (accessed 20 August 2016).

MacMillan Dictionary (2019) Insane. Available from https://www.macmillandictionary.com/us/dictionary/american/insane (accessed 20 April 2019).

Maoz Z et al. (2018) The dyadic Militarized Interstate Disputes (MIDs) dataset version 3.0: logic, characteristics, and comparisons to alternative datasets. *Journal of Conflict Resolution* **63**(3), 811–835.

Marshall MG, Jaggers K and Gurr TR (2010) Polity IV project: Political Regime Characteristics and Transitions, 1800–2010. Available from https://www.systemicpeace.org/inscr/inscr.htm (accessed 6 September 2011).

McDermott R (2007) *Presidential Leadership, Illness, and Decision Making*. New York: Cambridge University Press.

McManus RW (2019) Revisiting the Madman Theory: evaluating the impact of different forms of perceived madness in coercive bargaining. *Security Studies* **28**(5).

McManus R (2019) "Replication Data for: Crazy like a Fox? Are Leaders with Reputations for Madness More Successful at International coercion?", https://doi.org/10.7910/DVN/T3CGGV, Harvard Dataverse, V1

Nedal D and Nexon D (2017) Trump's 'Madman Theory' isn't strategic unpredictability, it's just crazy. *ForeignPolicy.com*, 18 April. Available from https://foreignpolicy.com/2017/04/18/trumps-madman-theory-isnt-strategic-unpredictability-its-just-crazy/ (accessed 28 June 2017).

OxfordDictionaries.com (2019) "Crazy" and "insane" entries. Available from https://en.oxforddictionaries.com/ (accessed 20 April 2019).

Palmer G et al. (2019) Updating the Militarized Interstate Dispute data: a response to Gibler, Miller, and Little. *International Studies Quarterly*. https://doi.org/10.1093/isq/sqz045.

Ramey AJ, Klingler JD and Hollibaugh GE (2019) Measuring elite personality using speech. *Political Science Research and Methods* **7**(1), 163–184.

Rathbun BC (2018) *Reasoning of State: Rationality, Realists and Romantics in International Relations*. New York: Cambridge University Press.

Renshon J (2008) Stability and change in belief systems: the operational code of George W. Bush. *Journal of Conflict Resolution* **52**(6), 820–849.

Renshon SA (2011) *Barack Obama and the Politics of Redemption*. New York: Routledge.

Sartori AE (2005) *Deterrence by Diplomacy*. Princeton, NJ: Princeton University Press.

Schelling TC (1960) *The Strategy of Conflict*. Cambridge, MA: Harvard University Press. 1980 edition.

Schelling TC (1966) *Arms and Influence*. New Haven, CT: Yale University Press. 2008 edition.

Schultz KA (1999) Do democratic institutions constrain or inform? Contrasting two institutional perspectives on democracy and war. *International Organization* **53**(2), 233–266.

Sechser TS (2010) Goliath's curse: coercive threats and asymmetric power. *International Organization* **64**(4), 627–660.

Sechser TS (2011) Militarized compellent threats, 1918–2001. *Conflict Management and Peace Science* **28**(4), 377–401.

Sechser TS and Fuhrmann M (2017) *Nuclear Weapons and Coercive Diplomacy*. New York: Cambridge University Press.

Sinaceur M et al. (2013) The advantages of being unpredictable: how emotional inconsistency extracts concessions in negotiation. *Journal of Experimental Social Psychology* **49**, 498–508.

Singer JD, Bremer S and Stuckey J (1972) Capability distribution, uncertainty, and major power war, 1820–1965. In Russett B (ed.), *Peace, War, and Numbers*, Beverly Hills: Sage, pp. 19–48.

Stinnett DM et al. (2002) The Correlates of War project direct contiguity data, version 3. *Conflict Management and Peace Science* **19**(2), 58–66.

Van Kleef GA and Côté S (2007) Expressing anger in conflict: when it helps and when it hurts. *Journal of Applied Psychology* **92**(6), 1557–1569.

Voeten E and Merdzanovic A (2009) United Nations General Assembly Voting Data. Available from https://thedata.harvard.edu/dvn/dv/Voeten/faces/study/StudyPage.xhtml?studyId=38311&versionNumber=1 (accessed 9 November 2011).

Walt S (2017) Things don't end well for madmen. *Foreign Policy*, 16 August. Available from https://foreignpolicy.com/2017/08/16/things-dont-end-well-for-madmen-trump-north-korea/ (accessed 13 September 2017).

https://doi.org/10.1017/S0007123419000401 Published online by Cambridge University Press

Weeks JL (2008) Autocratic audience costs: regime type and signaling resolve. *International Organization* **62**, 35–64.

Weisiger A (2013) *Logics of War: Explanations for Limited and Unlimited Conflicts*. Ithaca, NY: Cornell University Press.

Weisiger A and Yarhi-Milo K (2015) Revisiting reputation: how past actions matter in international politics. *International Organization* **69**(2), 473–495.

Wong SS (2019) Stoics and hotheads: leaders' temperament, anger, and the expression of resolve in face-to-face diplomacy. *Journal of Global Security Studies* **4**(2), 190–208.

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

Yarhi-Milo K, Kertzer JD and Renshon J (2018) Tying hands, sinking costs, and leader attributes. *Journal of Conflict Resolution* **62**(10), 2150–2179.

Cite this article: McManus RW (2021). Crazy Like a Fox? Are Leaders with Reputations for Madness More Successful at International Coercion? *British Journal of Political Science* **51**, 275–293. https://doi.org/10.1017/S0007123419000401