

How Credible Nuclear Security Guarantees Backfire: Evidence from a Survey Experiment in the United Kingdom

Lauren Sukin^{1,2}

Abstract: Existing literature posits that the main challenge for nuclear security guarantees lies in making the promise of protection sufficiently credible. If allies do not believe their guarantor will actually come to their aid, they may seek alternate means of protection, including by investing in nuclear infrastructure. Credible security guarantees, on the other hand, are thought to reassure allies. In contrast to this approach, new work has argued credible nuclear guarantees can backfire. These guarantees can cause clients to fear their guarantors will drag them into a precipitous nuclear conflict. Fears of nuclear escalation by their guarantor can drive clients to distance themselves from their alliance or seek stronger independent nuclear capabilities. This chapter uses original survey experiments to show that citizens in a key U.S. ally—the United Kingdom—backlash against the perceived credibility of the U.S. security guarantee to the North Atlantic Treaty Organization (NATO) by supporting U.K. nuclear modernization. The chapter brings to bear new evidence on the conditions under which security guarantees may succeed or fail to reassure allies.

Word Count³: 11,137

¹ Lauren Sukin, Ph.D. Candidate, Department of Political Science, Stanford University, 616 Jane Stanford Way Rm. 206, Stanford CA 94305.

² The author would like to thank the Freeman Spogli Institute at Stanford University for their funding and support, as well as Michal Smetana, Marek Vranka, Heather Williams, the participants at the Hertie School Digital Nuclear Security Workshop, and the editors of and contributors to this volume for their feedback.

³ This excludes the title page, tables, and figures.

Introduction

The U.S. nuclear security guarantee to the United Kingdom is mainly shaped by the states' mutual membership in NATO, whose Article 5 commitments on common defense stipulate an attack on one member is an attack on all. Through NATO, the United States maintains a nuclear presence in Europe, consisting of forward-deployed nuclear weapons deliverable by U.S. and allied aircraft. These weapons are intended not only deter adversaries of NATO states, but also to reassure U.S. allies commitment to their defense. As 1960s British Defence Minister Denis Healey described the difficulty of assurance: "it takes only five per cent credibility of American retaliation to deter the Russians, but ninety-five percent credibility to reassure the Europeans."⁴ U.S. nuclear commitments to NATO also have a "non-proliferation function" that, as David Yost explains, "concerns not only the alliance's potential adversaries but also the members of the alliance that are non-nuclear-weapon states...The NATO arrangements, including US nuclear forces in Europe, have served to assure Germany and other non-nuclear-weapon-state allies that they have no need to seek nuclear weapons of their own."⁵ Similarly, scholars have argued U.S. nuclear commitments to NATO may have limited the arsenal size and capability of U.K. nuclear forces.⁶

This understanding of nuclear guarantees reflects the scholarly consensus that the main concern of client states is credibility—or the perceived probability a guarantor will fulfill the promises of its guarantee. When clients worry whether their guarantor will actually use nuclear weapons in their defense, guarantees may not be sufficient to assuage clients' security concerns. When guarantees are credible, and clients believe their guarantor will act in their defense, guarantees substitute for clients' domestic military capabilities. Indeed, credible nuclear guarantees have long been thought sufficient substitutes for nuclear proliferation.⁷

This approach, however, does not explain two core puzzles: First, there is only a tenuous empirical connection between credible nuclear guarantees and nonproliferation. For example, cross-country studies of nuclear proliferation offer contradictory evidence about whether or

⁴ Linton Brooks and Mira Rapp-Hooper, "Extended Deterrence, Assurance, and Reassurance in the Pacific during the Second Nuclear Age," in *Strategic Asia 2013-14: Asia in the Second Nuclear Age*, 2013, 292–293.

⁵ David Yost, "Assurance and US Extended Deterrence in NATO," *International Affairs* 85, no. 4 (2009): 763, <https://doi.org/10.1111/j.1468-2346.2009.00826.x>.

⁶ Michael Middeke, "Anglo-American Nuclear Weapons Cooperation After the Nassau Conference: The British Policy of Interdependence," *Journal of Cold War Studies* 2, no. 2 (2000): 69–96, <https://doi.org/10.1162/15203970051032318>; Michael Quinlan, "The Future of United Kingdom Nuclear Weapons: Shaping the debate," *International Affairs* 82, no. 4 (2006): 627–637, <https://doi.org/10.1111/j.1468-2346.2006.00558.x>.

⁷ Richard Betts, "Wealth, Power, and Instability: East Asia and the United States after the Cold War," *International Security* 18, no. 3 (1993): 34–77, <https://doi.org/10.2307/2539205>; Philipp Bleek and Eric Lorber, "Security Guarantees and Allied Nuclear Proliferation," in *Nonproliferation Policy and Nuclear Posture* (Routledge, 2015), 85–109, <https://doi.org/10.1177/0022002713509050>; Dong-Joon Jo and Erik Gartzke, "Determinants of Nuclear Weapons Proliferation," *Journal of Conflict Resolution* 51, no. 1 (2007): 167–194, <https://doi.org/10.1177/0022002706296158>; Nuno Monteiro and Alexandre Debs, "The Strategic Logic of Nuclear Proliferation," *International Security* 39, no. 2 (2014): 7–51, <https://doi.org/10.1162/isec.a.00177>; Dan Reiter, "Security Commitments and Nuclear Proliferation," *Foreign Policy Analysis* 10, no. 1 (2014): 61–80, <https://doi.org/10.1111/fpa.12004>; Scott Sagan, "Why Do States Build Nuclear Weapons? Three models in search of a bomb," *International Security* 21, no. 3 (1997): 54–86, <https://doi.org/10.2307/2539273>.

when nuclear guarantees actually substitute for domestic nuclear proliferation.⁸ Second, credible guarantees not only sometimes fail to assuage clients' security concerns, but they can also introduce new ones.

For example, throughout the post-war history of the 'special relationship,' the U.S. nuclear guarantee to the United Kingdom sparked perennial British fears about being brought into a nuclear conflict by the United States. In 1954, the United Kingdom rebuffed U.S. calls for joint intervention in Indochina, in large part out of concern the policy would incite an aggressive Chinese response, leading to a "global war."⁹ U.K. Foreign Minister Anthony Eden was vocal about the potential nuclear consequences of U.S. policy in Indochina.¹⁰ Eden accused the United States of "unilateral decisionmaking on matters that vitally affected the interests of Britain."¹¹

Eden was not alone in worrying about the nuclear risks of American adventurism. Prime Minister Winston Churchill similarly complained: "the danger is that the Americans may become impatient...they may get in a rage and say...Why wait until Russia overtakes us?"¹² Such attitudes were influential in U.K. debates over whether to pursue the development of a 'super' or thermonuclear weapon. For example, Lord President of the Council Lord Salisbury expressed concerns that "without [thermonuclear] weapons Britain would not be in a position to influence the American government and prevent impatient elements in Washington from plunging 'the world into war, either through a misguided intervention in Asia or in order to forestall an attack by Russia.'"¹³ In this way, U.K. investments in nuclear infrastructure provided leverage. They not only allowed the United Kingdom a degree of independence from the United States, but they also provided a mechanism through which to influence U.S. security choices. Cyril Buffet and Beatrice Heuser point to "considerable concern that U.S. rashness and impatience might precipitate a global conflagration" as a crucial rationale behind U.K. development of thermonuclear weapons.¹⁴

⁸ Bleek and Lorber (2014) find strong support for a nonproliferation effect of nuclear guarantees. Singh and Way (2004) and Jo and Gartzke (2007) both find mixed support. Fuhrmann (2009) does not find support for such a nonproliferation effect. See: Bleek and Lorber, "Security Guarantees and Allied Nuclear Proliferation"; Jo and Gartzke, "Determinants of Nuclear Weapons Proliferation"; Sonali Singh and Christopher Way, "The Correlates of Nuclear Proliferation: A quantitative test," *Journal of Conflict Resolution* 48, no. 6 (2004): 859–885, <https://doi.org/10.1177/0022002704269655>; Matthew Fuhrmann, "Spreading Temptation: Proliferation and peaceful nuclear cooperation agreements," *International Security* 34, no. 1 (2009): 7–41, <https://doi.org/10.1162/isec.2009.34.1.7>

⁹ Kevin Ruane, "Agonizing Reappraisals: Anthony Eden, John Foster Dulles and the Crisis of European Defence, 1953–54," *Diplomacy and Statecraft* 13, no. 4 (2002): 163, <https://doi.org/10.1080/714000354>.

¹⁰ David Holloway, "Nuclear Weapons and the Escalation of the Cold War, 1945–1962," in *The Cambridge History of the Cold War* (Cambridge University Press, March 2010), 376–397, <https://doi.org/10.1017/chol9780521837194.019>.

¹¹ Ruane, "Agonizing Reappraisals: Anthony Eden, John Foster Dulles and the Crisis of European Defence, 1953–54," 163.

¹² David Carlton, "Churchill and the Two 'Evil Empires'," *Transactions of the Royal Historical Society* 11 (2001): 348, <https://doi.org/10.1017/s0080440100000487>.

¹³ John Baylis and Kristan Stoddart, *The British Nuclear Experience: The roles of beliefs, culture and identity* (OUP Oxford, 2014), 64, <https://doi.org/10.1093/acprof:oso/9780198702023.001.0001>.

¹⁴ Heuser Buffet et al., *Haunted by History: Myths in international relations* (Berghahn Books, 1998), 126, <https://doi.org/10.5860/choice.36-2414>.

This illustrates how, while the traditional approach has highlighted numerous benefits of credible guarantees, resolve is a double-edged sword. It may cause clients to fear their guarantor will draw them into a situation involving the ‘unwanted use’ of nuclear weapons.¹⁵ When clients fear nuclear escalation by credible guarantors, they backlash against their alliances. Clients can do this by drawing away from or setting limits on their reliance on their guarantors or by increasing their independent military capabilities, including via nuclear proliferation or modernization. This ‘backfire effect’ theory of nuclear guarantees offers an explanation for the empirical failures of credible nuclear guarantees to reassure allies and to ensure allied nonproliferation.

In this chapter, I re-assess the relationship between credible nuclear guarantees and ally reassurance, examining whether and to what degree credible nuclear guarantees resolve clients’ security concerns or create new ones. To do so, I design and implement an original survey experiment in the United Kingdom. I find little evidence for a substitution logic of nuclear guarantees, instead showing credible nuclear guarantees backfire. They create fears of embroilment in a U.S.-driven conflict and spark support for further investment in U.K. nuclear capabilities.

A New Logic of Nuclear Guarantees

Hazards of Guarantees

The international relations literature has delineated a number of benefits of security guarantees.¹⁶ Alliances, however, are not always pacifying. Scholars have identified two primary hazards associated with nuclear guarantees: entanglement and emboldenment. These occur when, as Tongfi Kim explains, states are “compelled to aid an ally in a costly and unprofitable enterprise.”¹⁷

The first alliance hazard—entanglement—occurs when allies are “dragged into a conflict over an ally’s interests that [they do] not share.”¹⁸ Entanglement occurs when the terms of a state’s alliance result in involvement in a conflict the state has little interest in outside of the alliance itself. The literature on entanglement focuses on risks to the guarantor, and it has informed work arguing the United States ought to pull back from its complex system of global alliances to avoid

¹⁵ On unwanted use, see: Lauren Sukin, “Credible Nuclear Security Commitments Can Backfire: Explaining Domestic Support for Nuclear Weapons Acquisition in South Korea,” *Journal of Conflict Resolution* 64, no. 6 (2020): 1011–1042, <https://doi.org/10.1177/0022002719888689>

¹⁶ Neil Narang and Rupal Mehta, “The Unforeseen Consequences of Extended Deterrence: Moral hazard in a nuclear client state,” *Journal of Conflict Resolution* 63, no. 1 (2019): 218–250, <https://doi.org/10.1177/0022002717729025>; Glenn Snyder, *Alliance Politics* (Cornell University Press, 2007), <https://doi.org/10.4324/9780203422144-49>; Brett Ashley Leeds, Andrew Long, and Sara McLaughlin Mitchell, “Reevaluating Alliance Reliability: Specific threats, specific promises,” *Journal of Conflict Resolution* 44, no. 5 (2000): 686–699, <https://doi.org/10.1177/0022002700044005006>.

¹⁷ Tongfi Kim, “Why Alliances Entangle but Seldom Entrap States,” *Security Studies* 20, no. 3 (2011): 355, <https://doi.org/10.1080/09636412.2011.599201>.

¹⁸ Glenn Snyder, “The Security Dilemma in Alliance Politics,” *World Politics* 36, no. 4 (1984): 467, <https://doi.org/10.2307/2010183>.

being caught in unnecessary wars.¹⁹ This hazard provides an incentive for guarantors not to over-promise; they may not always wish to fulfill the terms of their agreements.

The second alliance hazard is emboldenment, where the promise of protection emboldens clients to behave more belligerently. As Brett Benson writes: “Assuming that an alliance commitment has the effect of increasing the probability that a protégé will win a war, moral hazard should indeed embolden the protégé to behave more aggressively than it otherwise would in crises. This behavior increases the risk of war.”²⁰ Distinguishing entanglement from emboldenment is that emboldenment refers to the possibility of conflicts resulting “because of risky or offensive actions taken outside the parameters of the original alliance treaty.”²¹ As with the entanglement literature, the emboldenment literature focuses on the risks posed to a guarantor, highlighting how less-powerful allies may act more aggressively under the protection of more-powerful states.

Both the entanglement and emboldenment hazards stem from a problem faced by guarantors: They have incentives not to abandon client states, even when defending clients is difficult or when clients have acted recklessly. Guarantors may stick with costly alliance agreements because of concerns about their reputation domestically, with other allies, or with adversaries.²²

Importantly, clients also face hazards associated with their alliances, although this has not been a frequent focus in the literature. Like guarantors, clients worry they will be asked to contribute to their alliance partners’ conflicts in areas outside of their own interests.²³ This risk for client states, of being dragged into external conflicts because of a guarantor, is sometimes referred to as entrapment.²⁴ Because entrapment usually involves responsibilities for an ally outside of the scope of a formal security guarantee, this hazard is independent of guarantee credibility.

I argue clients face an additional risk heretofore been missed in the literature. Clients of credible security guarantees risk precipitous escalation over their own security concerns. The traditional literature would suggest the main risk for client states when it comes to managing their security environment is the possibility their guarantors will be unable to or will choose not to defend them when their security is threatened—despite the terms of the guarantee. Yet the mismatch between client and guarantor preferences on the client’s own security issues can also occur in the other direction. When guarantors communicate credibility, they signal a strong willingness and

¹⁹ Barry Posen, “Pull Back: The case of a less activist foreign policy,” *Foreign Affairs* 92 (2013): 116; Richard Rosecrance and Steven Miller, *The Next Great War?: The roots of World War I and the risk of US-China conflict* (MIT Press, 2014), <https://doi.org/10.7551/mitpress/10371.003.0017>.

²⁰ Brett Benson, *Constructing International Security: Alliances, deterrence, and moral hazard* (Cambridge University Press, 2012), 50, <https://doi.org/10.1017/cbo9781139225694>.

²¹ Alexander Lanoszka, *Atomic Assurance: The alliance politics of nuclear proliferation* (Cornell University Press, 2018), 236, <https://doi.org/10.7591/cornell/9781501729188.001.0001>.

²² Allan Dafoe, Jonathan Renshon, and Paul Huth, “Reputation and Status as Motives for War,” *Annual Review of Political Science* 17 (2014), <https://doi.org/10.2139/ssrn.2332048>; Tudor Onea, “Between Dominance and Decline: Status anxiety and great power rivalry,” *Review of International Studies*, 2014, 125–152, <https://doi.org/10.1017/s0260210512000563>.

²³ Thomas Christensen and Jack Snyder, “Chain Gangs and Passed Bucks: Predicting alliance patterns in multipolarity,” *International Organization*, 1990, 137–168, <https://doi.org/10.1017/s0020818300035232>.

²⁴ Snyder, “The Security Dilemma in Alliance Politics.”

ability to use their resources in defense of their allies. Yet, in doing so, they may demonstrate more willingness to use force against the client's adversaries than clients actually prefer.

Diverging Risk Tolerance Drives Backfire Hazard

The security guarantee literature has largely assumed guarantors are less risk tolerant than their clients. This reflects the classic challenge of extended deterrence; while clients should, as a matter of course, be highly interested in their own security, guarantors' first order concerns are about their own territory, citizens, and resources. Clients of U.S. guarantees may therefore wonder whether the United States would really 'risk New York for Paris.' Yet there are many reasons guarantors may actually be more risk tolerant than their clients.

First, guarantors have structural reasons to act more aggressively than clients. Guarantors face lower costs if disputes over a client's security escalates. Escalation would likely happen in or near the client's territory and should therefore disproportionately affect the client. As a result, clients will internalize the costs of conflict more than guarantors and will have stronger incentives to avoid escalation. Clients and guarantors may also resultantly have different views of what 'winning' means. Guarantors may therefore be more aggressive and more adventurous, since they may be more willing to sustain the costs of conflict.

Second, guarantors often have multiple interests and preferences that determine their involvement in security issues related to a client. Guarantors may credibly commit to defend an ally, but this commitment may not be defined by a singular interest in the ally's well-being, leading to differences of opinion in how to handle the ally's security threats. Writes Michael Beckley: "To be sure, the United States has intervened on the side of allies on numerous occasions. In most cases, however, U.S. actions were driven by an alignment of interests between the United States and its allies, not by alliance obligations. In fact, in many cases, U.S. policymakers were the main advocates of military action and cajoled reluctant allies to join the fight."²⁵ Clients may be concerned guarantors will take unilateral escalatory actions or compel joint offensive measures.

These cross-cutting incentives can result from guarantors having multiple networks of guarantees. These have additive effects, where the credibility of one guarantee may provide information to other clients about their own guarantees.²⁶ Guarantors may therefore be incentivized to act aggressively when one client's interests are threatened in order to reassure other partners. Multilateral guarantees, such as the U.S. guarantee to NATO, can be especially difficult in this regard. Guarantors face pressure to demonstrate credibility at a level that will reassure the most-threatened states in the alliance. Allies with lower threat or risk tolerance levels will consequently not see their preferences reflected in the guarantor's posture.

Similarly, a guarantor's ability to deter each of its adversaries may hinge on maintaining commitments against all of its adversaries. This challenge is especially pressing in the nuclear realm. Failure to respond to a nuclear threat against a client would likely erode a guarantor's ability to deter nuclear threats against other clients and itself. Guarantors may especially need to

²⁵ Beckley, "The Myth of Entangling Alliances: Reassessing the security risks of US defense pacts," 10.

²⁶ Eliza Gheorghe argues in Chapter # that clients may believe security provision to be zero-sum.

demonstrate high resolve in order to deter adversaries of multilateral security arrangements, who might otherwise seize opportunities to decouple alliances.

Finally, even if a guarantor does not actually have a different risk tolerance than their client, communication challenges associated with security guarantees can make clients believe this to be the case. This is because guarantors have incentives to signal credibility in ways that are, themselves, risky. Guarantors are aware of the need to assure allies their guarantees are sufficiently credible and to convey to adversaries that any missteps will be punished. As it is difficult to fully demonstrate willingness to risk New York for London, guarantors rely on strong signals of resolve, such as stationing tripwire troops, engaging in shows of force, or making direct threats against common adversaries.²⁷ These strong signals of resolve carry significant escalation risks.

Moreover, the actions making security guarantees appear credible have downsides. Each of these actions can make accidents or miscommunications more likely by reducing reaction times for conflict, creating hawkish institutional interests, or aggravating adversaries, among other pathways. Steps guarantors can take to increase the credibility of nuclear guarantees specifically—such as the forward-deployment of nuclear weapons or nuclear-capable delivery systems, introducing automaticity into nuclear launch protocols, or investing in low-yield nuclear weapons—are particularly prone to these risks.²⁸ Controlling perceptions about precisely when and how escalatory actions would be taken is difficult, especially in the nuclear context. This can translate into client concerns that a guarantor will engage in an unwanted use of nuclear weapons. That is, “the process of creating a credible nuclear guarantee for a client state trades off with the ability to reassure that client state that nuclear use will be avoided.”²⁹

Consequences of Mismatched Risk Tolerance Preferences

Resolving escalation concerns is no easy task, as they reflect often-stark differences in the interests and preferences of clients and guarantors. Moreover, the risk of escalation by its very nature involves the possibility a guarantor may take an action that is “surprising or reckless and antithetical to publicly agreed-upon goals such as conflict avoidance,” thereby making reassurance that there will be no unwanted escalation especially challenging.³⁰ Clients can engage in a number of strategies to increase their leverage with their guarantor—or their ability to influence and restrain their guarantor’s choices.

For example, clients can ask guarantors for increased involvement in military decision-making, such as through high-level dialogues, crisis consultation, or dual-key launch systems. These strategies vary in the degree to which they could actually bind the hands of a guarantor. Highly effective strategies are also more likely to be rejected, as guarantors have strong incentives to retain control over military decision-making. This is especially challenging for nuclear

²⁷ Guarantors may be especially likely to adopt such policies in cases where commitments are harder to establish, e.g. with significant geographical distance.

²⁸ Scott Douglas Sagan and Kenneth Neal Waltz, *The Spread of Nuclear Weapons: A debate renewed, with new sections on India and Pakistan, terrorism, and missile defense* (WW Norton & Company, 2003).

²⁹ Sukin, “Credible Nuclear Security Commitments Can Backfire: Explaining Domestic Support for Nuclear Weapons Acquisition in South Korea,” 1015-1016.

³⁰ Sukin, 1015.

guarantees, where the need for rapid decision-making and delegation makes it difficult to require consultations and other safeguards.

A more effective solution, then, may be for clients to decrease their reliance on—or even reject—their security guarantee. While a full rejection would leave client states vulnerable to security threats, decreased reliance accompanied by increased investment in independent military capabilities can let client states leverage their capabilities for influence with their guarantor, while being able to better deter adversaries on their own. Strong, independent military capabilities not only afford states prestige but also position states to have a more significant role in the event conflict breaks out. Consequently, greater military strength can translate into greater negotiating power within alliances. Consider, for example, the aforementioned case of the United Kingdom’s thermonuclear weapons development. This was intended, in part, to improve the United Kingdom’s ability to influence U.S. security policy. Indeed, Buffet and Heuser write: “From the perspective of the government...it was imperative for Britain to develop thermonuclear weapons...to restrain the United States...[and] to influence and guide U.S. policy in a more responsible direction.”³¹

The backfire effect expects credible security guarantees will raise clients’ concerns about precipitous escalation. This can lead clients to seek strengthened independent military capabilities, so they no longer have to live under the itchy trigger finger of their guarantor. This approach runs counter to the traditional substitution logic, which instead suggests credible security guarantees provide valuable assurance to clients and therefore substitute for domestic military assets. In this chapter, I assess how the U.K. public responds to the credibility level of the U.S. security guarantee in order to identify whether credible guarantees reassure or raise fears of escalation.

Methodology

In this chapter, I provide an empirical test of the effects of U.S. guarantee credibility on U.K. perceptions about foreign and defense policy. I designed and implemented a survey experiment on a 1,233-person sample of individuals in the United Kingdom from April 26, 2021 to May 19, 2021.³² The survey was run using Prolific, an online platform where users choose tasks. Prolific has been previously used by scholars to study populations in the United Kingdom.³³ Prolific respondents have been shown to be more representative and perform better on metrics such as

³¹ Buffet et al., *Haunted by History: Myths in international relations*, 127.

³² The survey was conducted several months into the Biden Administration. Lasting effects of Donald Trump’s presidency could have influenced respondents’ attitudes; however, polling shows U.K. favorability towards the United States warmed by 23 percentage points between spring 2020 and spring 2021. Richard Wike et al., “America’s Image Abroad Rebounds with Transition from Trump to Biden,” Pew Research Center, June 2021.

³³ Beth Armstrong et al., “How Does Citizen Science Compare to Online Survey Panels? A comparison of food knowledge and perceptions between the Zooniverse, Prolific and Qualtrics UK Panels,” *Frontiers in Sustainable Food Systems* 4 (2020): 306, <https://doi.org/10.3389/fsufs.2020.575021>; Pascal Geldsetzer, “Use of Rapid Online Surveys to Assess People’s Perceptions During Infectious Disease Outbreaks: A cross-sectional survey on COVID-19,” *Journal of Medical Internet Research* 22, no. 4 (2020), <https://doi.org/10.2196/18790>.

comprehension, attention, diversity, and honesty relative to other online options, including MTurk and Lucid.³⁴

The experiment varied the credibility of the U.S. guarantee to NATO in a scenario involving a Russian attack on a NATO state. I then assess how credibility affects respondents' preferences for investment in the U.K. military. The traditional substitution approach to guarantees would predict credible guarantees to decrease support for enhanced independent military capabilities, but I argue the opposite can also occur. Credible guarantees can generate fears about U.S.-driven escalation and consequently increase support for investment in U.K. military capabilities.³⁵

The United Kingdom represents a difficult test case for three reasons. First, relative to many other U.S. allies, the United Kingdom exhibits high levels of support for the use of force, which should make it less susceptible to the escalation concerns that may plague a more cautious ally.³⁶ Second, the United Kingdom has been one of the most reliable U.S. allies, especially in terms of military contributions to U.S. missions. The United States and the United Kingdom share a 'special relationship' involving shared ideology as well as interdependence and cooperation in a number of military, political, and economic areas. As a result, the U.S.-U.K. preference gap should be smaller than in other U.S. alliances. Third, unlike most U.S. allies covered by the nuclear umbrella, the United Kingdom already possesses a nuclear arsenal. Moreover, while the United Kingdom is a U.S. client, it also serves as a guarantor within NATO in its own right. These characteristics afford the United Kingdom some independence from the U.S. nuclear guarantee and influence in U.S. and NATO nuclear decision-making. Although less work in international relations has focused on the dynamics of security guarantees between multiple nuclear weapons states, similar dynamics should still arise in this setting. I argue credible nuclear guarantees to nuclear weapons states can backfire, encouraging nuclear clients to distance themselves from their guarantor and invest further in their own nuclear forces.

The Importance of Public Opinion

Survey experiments have several advantages over the approach taken by previous quantitative studies of security guarantees. These studies have used data on the presence of formal defense

³⁴ Eyal Peer et al., "MTurk, Prolific or Panels? Choosing the right audience for online research," *Choosing the right audience for online research* (January 10, 2021), 2021, <https://doi.org/10.2139/ssrn.3765448>; Elissa M Redmiles et al., *A Summary of Survey Methodology Best Practices for Security and Privacy Researchers*, technical report (2017).

³⁵ By studying the United Kingdom, I only examine the relationship between guarantee credibility and support for nuclear modernization, not nuclear acquisition. On credibility and nuclear proliferation, see: Sukin, "Credible Nuclear Security Commitments Can Backfire: Explaining Domestic Support for Nuclear Weapons Acquisition in South Korea."

³⁶ Several studies have shown security concerns about Russia and find a majority believes continued NATO membership is important for the security of the United Kingdom. See: *An Attack Against Them All? Drivers of Decisions to Contribute to NATO Collective Defense*, technical report (RAND Corporation, 2019); Sophia Gaston, *Public Opinion on Global Threats and the Future of NATO*, technical report (British Foreign Policy Group, December 2019); Moira Fagan and Jacob Poushter, *NATO Seen Favorably Across Member Countries*, technical report (Pew Research Center, February 2020); Sophia Gaston and Evie Aspinall, *UK Public Opinion on Foreign Policy and Global Affairs: Annual Survey - 2021*, technical report (British Foreign Policy Group, February 2021).

pacts as a proxy for security guarantees.³⁷ This measure is insufficient for two critical reasons. First, formal defense pacts do not require nuclear guarantees, and nuclear guarantees do not require defense pacts. For example, the formal defense pacts data³⁸ excludes the U.S.-South Korea 1953 Mutual Defense Treaty, because the technically ongoing state of the Korean War precluded the agreement from being a ‘true’ defense pact.³⁹ Nevertheless, the U.S. nuclear guarantee to South Korea is one of the most prominent. Similarly, some studies of proliferation include, from the formal defense pacts data, a series of defense pacts between the United States and Latin American states that concluded in 1945 and would not have actually involved nuclear guarantees.⁴⁰

Second, the presence of a formal defense pact—even if a nuclear guarantee is included as part of that arrangement—does not provide adequate information about the credibility of the guarantee. In the case of NATO, for example, the U.S. security guarantees to Estonia and the United Kingdom do not necessarily hold the same weight. Guarantee credibility not only varies across countries, but it also varies within countries, as different stakeholders have fundamentally different perceptions about the quality of security arrangements.⁴¹ The use of survey experiments in the United Kingdom provides a way to circumvent both concerns. This strategy highlights within-country variation in perceptions of security guarantees and identifies a clear case of a viable U.S. nuclear guarantee.

Public opinion is a valuable subject of study for understanding security policy broadly and nuclear policy in particular. Public attitudes have been shown to influence policymaker decisions, whether because of audience costs, protest politics, retrospective voting, anticipation of public opposition, influence on bureaucrats and non-elected officials, or many other mechanisms.⁴² Public opinion has also informed nuclear policies, from proliferation to missile defense to nuclear posture and force structure.⁴³ Public support for nuclear proliferation in the

³⁷ Bleek and Lorber, “Security Guarantees and Allied Nuclear Proliferation”; Fuhrmann, “Spreading Temptation: Proliferation and peaceful nuclear cooperation agreements”; Jo and Gartzke, “Determinants of Nuclear Weapons Proliferation”; Singh and Way, “The Correlates of Nuclear Proliferation: A quantitative test.”

³⁸ Douglas Gibler, *International Military Alliances, 1648-2008* (CQ Press, 2008), <https://doi.org/10.4135/9781604265781>.

³⁹ Differing from other quantitative studies, Bleek and Lorber (2015) include this treaty.

⁴⁰ Bleek and Lorber (2015) remove these defense pacts.

⁴¹ Lauren Sukin and Toby Dalton, “Reassure Northeastern Asian Allies by Reducing Nuclear Salience,” *The Washington Quarterly* 44 (2 2022), <https://doi.org/10.1080/0163660x.2021.1934257>.

⁴² Kenneth Schultz, “Looking for Audience Costs,” *Journal of Conflict Resolution* 45, no. 1 (2001): 32–60, <https://doi.org/10.1177/0022002701045001002>; Russell Dalton, *Citizen Politics: Public opinion and political parties in advanced industrial democracies* (CQ Press, 2013); Andrew Healy and Neil Malhotra, “Retrospective Voting Reconsidered,” *Annual Review of Political Science* 16 (2013): 285–306, <https://doi.org/10.1146/annurev-polisci-032211-212920>; “Public Opinion, Domestic Structure, and Foreign Policy in Liberal Democracies,” in *Domestic Politics and Norm Diffusion in International Relations* (Routledge, October 2016), 35–64, <https://doi.org/10.4324/9781315623665-9>; Erik Lin-Greenberg, “Soldiers, Pollsters, and International Crises: Public Opinion and the Military’s Advice on the Use of Force,” *Foreign Policy Analysis* 17, no. 3 (2021): orab009, <https://doi.org/10.1093/fpa/orab009>.

⁴³ “Norms, Identity, and National Security: Thomas U. Berger,” in *Security Studies* (Routledge, February 2011), 198–201, <https://doi.org/10.4324/9780203422144-32>; Paul Cole, “Atomic Bombast: Nuclear weapon decision-making in Sweden, 1946–72,” *Washington Quarterly* 20, no. 2 (1997): 233–251, <https://doi.org/10.1080/01636609709550250>; Vipin Narang, *Nuclear Strategy in the Modern Era: Regional powers and international conflict*, vol. 143 (Princeton University Press, 2014), <https://doi.org/10.1515/9781400850402>;

United Kingdom, for example, has been linked to the U.K. decision to pursue a nuclear weapons capability.⁴⁴ Studies of public opinion have also been shown, cross-nationally, to effectively predict elite attitudes on foreign policy issues.⁴⁵

This research contributes to a small literature expanding the study of the public opinion on nuclear weapons. Work in this area has previously focused heavily on attitudes in the United States. Beginning with studies such as Kramer, Kalick, and Milburn (1983) and including such notable works as Press, Sagan, and Valentino (2013) and Sagan and Valentino (2017), this literature has often focused on the preferences of the American public about the use of nuclear weapons.⁴⁶ Only a handful of studies have begun to ask questions about non-American attitudes in the nuclear sphere.⁴⁷ This includes a survey experiment by the author testing the backfire effect in South Korea.⁴⁸ In fact, a forum of nuclear scholars recently identified the lack of cross-national surveys as a notable gap in the existing literature.⁴⁹ This study contributes to this ongoing conversation.

This experiment will also shed light on the causes of current trends in U.K. public opinion, including a pattern of concern about a perceived quickness to escalate by the United States. For example, the British Foreign Policy Group found in 2019 that 30% of Britons “identify the United States of America, our enduring partner in a ‘special relationship’, as a threat to global

Rodger Payne, “Public Opinion and Foreign Threats: Eisenhower’s response to Sputnik,” *Armed Forces & Society* 21, no. 1 (1994): 89–111, <https://doi.org/10.1177/0095327x9402100106>; Elizabeth Saunders, “The Domestic Politics of Nuclear Choices—A Review Essay,” *International Security* 44, no. 2 (2019): 146–184, https://doi.org/10.1162/isec_a_00361.

⁴⁴ Gregory Flynn and Hans Rattinger, “The Public and Atlantic Defense,” in *The Public and Atlantic Defense* (Routledge, 2021), 365–388, <https://doi.org/10.4324/9781003099840-9>.

⁴⁵ Joshua Kertzer, “Re-Assessing Elite-Public Gaps in Political Behavior,” *American Journal of Political Science*, 2020, <https://doi.org/10.1111/ajps.12583>; Michael Tomz, Jessica Weeks, and Keren Yarhi-Milo, “Public Opinion and Decisions about Military Force in Democracies,” *International Organization* 74, no. 1 (2020): 119–43, <https://doi.org/10.1017/s0020818319000341>.

⁴⁶ Bernard Kramer, Michael Kalick, and Michael Milburn, “Attitudes toward Nuclear Weapons and Nuclear War: 1945–1982,” *Journal of Social Issues* 39, no. 1 (1983): 7–24, <https://doi.org/10.1111/j.1540-4560.1983.tb00127.x>; Daryl Press, Scott Sagan, and Benjamin Valentino, “Atomic Aversion: Experimental evidence on taboos, traditions, and the non-use of nuclear weapons,” *American Political Science Review*, 2013, 188–206, <https://doi.org/10.1017/s0003055412000597>; Scott Sagan and Benjamin Valentino, “Revisiting Hiroshima in Iran: What Americans really think about using nuclear weapons and killing noncombatants,” *International Security* 42, no. 1 (2017): 41–79, https://doi.org/10.1162/isec_a_00284.

⁴⁷ Jonathon Baron, Rebecca Davis Gibbons, and Stephen Herzog, “Japanese Public Opinion, Political Persuasion, and the Treaty on the Prohibition of Nuclear Weapons,” *Journal for Peace and Nuclear Disarmament*, 2020, 1–11, <https://doi.org/10.1080/25751654.2020.1834961>; Lauren Sukin, “Experimental Evidence on Determinants of Support for Nuclear Use in Response to Threats of Nuclear Retaliation,” *Peace and Conflict: Journal of Peace Psychology* 26, no. 3 (2020): 336–339, <https://doi.org/10.1037/pac0000407>; Michal Smetana, Michal Onderco, and Tom Etienne, “Time to Say Goodbye? Germany and the Netherlands debate the withdrawal of US nuclear weapons,” *The Bulletin of Atomic Scientists*, 2021, <https://doi.org/10.1080/00963402.2021.1941603>.

⁴⁸ Sukin, “Credible Nuclear Security Commitments Can Backfire: Explaining Domestic Support for Nuclear Weapons Acquisition in South Korea.”

⁴⁹ Michal Smetana and Carmen Wunderlich, “Nonuse of Nuclear Weapons in World Politics: Toward the Third Generation of “Nuclear Taboo” Research,” *International Studies Review*, 2021, <https://doi.org/10.1093/isr/viab002>.

peace and security.”⁵⁰ In addition, just 28% of the U.K. public trusted the United States “to act responsibly in the world” in April 2020.⁵¹

Contributing to qualms about the U.S. alliance is that U.K. citizens are wary of entanglement in foreign wars. In 2021, 82% of Britons indicated they did not “support British military action abroad under any circumstances,” and 45% listed their primary reason for caution as wanting to avoid being drawn into foreign conflicts.⁵² 28% listed their primary reason as “countries should not tell other countries what to do.” Strong U.S. demonstrations of resolve seemingly worsen these concerns. In 2011, 39% of Britons wanted the European Union to be more independent from the United States in its security policy, but this number increased to 42% during the Ukraine crisis in 2014.⁵³ This study should shed light on some of these trends.

Survey Design

Respondents were presented with a survey experiment wherein they were asked to imagine an attack on a NATO state by Russia. Respondents were told they would “be asked to imagine a scenario in which Russia has invaded and attacked a member of the North Atlantic Treaty Organization (NATO).” They were told: “NATO is a military alliance. The United Kingdom is a member of NATO. The United States is also a member of NATO.” This scenario provides an opportunity to assess U.K. public opinion about a U.S. nuclear guarantee in a plausible hypothetical that would activate U.S. responsibilities to defend NATO.

The survey experiment used a 2 x 2 x 2 factorial design, where there were three distinct treatments, each with two conditions. First, respondents were either told the Russian invasion had involved nuclear or conventional weapons. Second, respondents were either told U.K. military officials had high or low confidence the United States would respond to the invasion. This information reveals the degree of credibility associated with the U.S. security guarantee. Third, the U.S. response was characterized as involving either nuclear or conventional weapons. The first and third treatments allow an examination of various iterations of the U.S. security guarantee. Together, the treatments read as follows:

According to U.K. military officials, if **Russia invaded and conducted a [nuclear attack/conventional (i.e. non-nuclear) bombing campaign] against a NATO state**, it is **[very likely/very unlikely]** that the United States would use **[nuclear weapons/conventional (i.e. non-nuclear) bombs] against Russia** in response.

⁵⁰ Gaston, *Public Opinion on Global Threats and the Future of NATO*.

⁵¹ Although that increased to 47% in February 2021, U.K. trust in the United States remains far below other allies. 88% of Britons trust Canada, and a majority—even after Brexit—prioritize the E.U. relationship over the United States. Gaston and Aspinall, *UK Public Opinion on Foreign Policy and Global Affairs: Annual Survey – 2021*.

⁵² There is some reluctance among the U.K. public about the circumstances in which the use of force would be advisable. In 2021, 39% of Britons said the U.K. military “should only be deployed in three scenarios: a direct attack on British soil, a direct attack on British assets abroad, or in the case of genocide or a large-scale humanitarian crisis” (Gaston and Aspinall, 2021).

⁵³ Matthias Mader, Francesco Olmastroni, and Pierangelo Isernia, “The Polls—Trends: Public Opinion Toward European Defense Policy and Nato: Still Wanting it Both Ways?,” *Public Opinion Quarterly* 84, no. 2 (2020): 551–582, <https://doi.org/10.1093/poq/nfaa031>.

I exploit variation in the credibility of the U.S. security guarantee to assess effects on public attitudes about national security. I examine both nuclear and conventional security guarantees, although the primary guarantees of interest are nuclear guarantees. I also provide variation in the type of security threat the United Kingdom is purported to face by varying whether the Russian invasion of a NATO state involved the use of nuclear or conventional arms. In addition to these treatments, the survey included control conditions, where respondents were prompted to consider a Russian threat but were not informed about the U.S. security guarantee's credibility.⁵⁴

After treatment assignment,⁵⁵ respondents were asked about support for investments in the U.K. military. The primary dependent variable was approval for U.K. investment in nuclear weapons. The substitution theory would predict credible guarantees would decrease support for nuclear investment relative to non-credible guarantees. In contrast, the backfire effect theory argues credible guarantees should be linked to increased support for nuclear investment relative to non-credible guarantees.

Respondents were also asked about their approval for U.K. investment in conventional military capabilities, in order to assess whether guarantees might also substitute for these capabilities. Respondents were asked, in open-ended survey items, to explain their preferences regarding investment in both nuclear and conventional capabilities. The survey also included outcome measures of respondents' attitudes regarding the United Kingdom's relationships with the United States and Russia. Respondents answered questions about a number of demographics and reported pre-treatment attitudes about foreign policy and national security topics.

The substitution logic of security guarantees would suggest credible nuclear guarantees would be seen as reassuring, thereby solidifying the U.K. alliance with the United States and ensuring the United Kingdom did not need to further invest in its own military capabilities. The backfire effect would predict the reverse. Respondents informed about a credible U.S. nuclear guarantee should be concerned about unwanted nuclear escalation. As a result, respondents may support distancing from the U.S. alliance and investing in independent U.K. nuclear capabilities.

Results

Do Credible Guarantees Substitute for or Spur Military Investment?

In this section, I assess the effects of security guarantee credibility on approval for U.K. investments in nuclear and conventional military capabilities. I measure support for military investments as a binary indicator; respondents who reported they strongly or somewhat approved of investments are coded as supporters, while respondents who strongly or somewhat disapproved of investments are coded as non-supporters. I use linear probability models to determine the correlation between credibility and support for military investments.

⁵⁴ The control groups were told to “imagine that Russia invaded and conducted a [nuclear attack / conventional (i.e. non-nuclear) bombing campaign] against a NATO state.” In both versions, respondents were not given any information about a possible U.S. response to the attack or its likelihood.

⁵⁵ All analyses drop 31 respondents who failed a manipulation check asking whether the experiment discussed a Russian attack on NATO, a U.S. attack on the United Kingdom, a U.S. attack on NATO, or a Russian attack on the United States.

The dependent variable indicates whether respondents were told the United States was very likely to use nuclear weapons in response to a scenario in which Russia attacked a NATO state.⁵⁶ This is represented by the interaction between the credibility treatment and the ‘U.S. Nuclear Use’ treatment. The coefficient for ‘Credibility’ on its own reflects the effect of a credible guarantee to use conventional weapons, and the coefficient for ‘U.S. Nuclear Use’ on its own reflects the effect of a non-credible guarantee to use nuclear weapons. I also include a variable (‘Russian Nuclear Use’) indicating whether Russia’s attack involved the use of nuclear weapons; this allows me to separate out whether respondents’ reactions are related to an increase in the perceived risk of any nuclear use or U.S. nuclear use specifically. This variable is not significant in any of the models in Table 1, suggesting any backlash against credible guarantees is the result of concerns about U.S. crisis behavior, rather than Russian behavior.

Table 1 includes models that involve only the experimental treatments (1 and 4) as well as models (2 and 5) that incorporate respondents’ demographics, such as their political party, age, gender, education level, household income, and citizenship status. In addition, I include models (3 and 6) that analyze the effects of credible security guarantees after taking into account respondents’ foreign policy beliefs. These models include variables for respondents’ level of interest in news about foreign and military policy (‘News Interest’), whether they or members of their immediate family have served in the military (‘Military’), their views on the alignment of U.K. policy goals with those of the United States, NATO, and Russia (‘Alignment’), whether they anticipate a war between Russia and NATO within the next two decades (‘Anticipate War’), if they expect U.K. nuclear weapons will someday be used (‘Anticipate U.K. Use’), and their belief in the nuclear taboo (‘Nuclear Taboo’). I also include measures of respondents’ attitudes about globalism, including whether they think the United Kingdom should play an active role in solving conflicts around the world (‘World Policeman’) and whether they believe the use of military force only makes problems worse (‘Dove’). These variables help identify broader patterns shaping respondents’ preferences about military investments.⁵⁷

While standard theories of security guarantees would anticipate less support for costly investments in military capabilities when a client’s security guarantees are perceived as reliable, I find little evidence for this theory. Neither nuclear nor conventional security guarantees consistently substitute for nuclear or conventional military assets.⁵⁸

Importantly, in all models in Table 1, credible nuclear guarantees are not associated with a statistically significant change in support for nuclear investments relative to non-credible,

⁵⁶ Table 1 excludes the control conditions.

⁵⁷ Table 1 shows several controls contribute to respondents’ support for military investments. Compared to members of the Conservative Party, members of the Labour Party, Scottish National Party, Liberal Democratic Party, Green Party, and other political parties were less likely to support investing in either nuclear or conventional military capabilities. Female respondents were less likely to support investments in nuclear weapons; however, no gender difference emerges in levels of support for investing in conventional arms. Attitudes about nuclear weapons also affect respondents’ support for nuclear modernization. For example, respondents who expect the United Kingdom will eventually use nuclear weapons during a conflict are more likely to want to invest in the U.K. nuclear arsenal.

⁵⁸ The exception to this is the coefficient on credibility in model 3, which indicates credible promises to use conventional weapons to defend NATO are associated with a 7.8 percentage point decrease in support for nuclear investments compared to non-credible promises to use conventional weapons.

conventional security guarantees. The effect of credibility on support for military investments is essentially the same as the effect of having only a very weak, conventional guarantee. Respondents fail to perceive credible nuclear guarantees as sufficient substitutes for nuclear modernization.

The results point to a potential backfire effect of credible nuclear guarantees. I find credible guarantees that the United States will use its nuclear weapons to defend NATO are associated with between 5.4 and 8.6 percentage points more approval for U.K. nuclear investments than are credible guarantees to use conventional weapons.⁵⁹ Similarly, credible nuclear guarantees are correlated with between 2.9 and 4.4 percentage points more approval for nuclear investments than non-credible guarantees to use nuclear weapons.⁶⁰ Of respondents who were told a U.S. nuclear response to a Russian invasion of a NATO state was likely, 36% approved of nuclear investments, compared to 33% of respondents who were told a U.S. nuclear response was unlikely and 30% of respondents who were told a U.S. conventional response was likely.

This suggests credible nuclear guarantees may be worse for ally reassurance purposes in the U.S.-U.K. alliance than both credible conventional guarantees and non-credible nuclear guarantees. Should policymakers wish to achieve a nonproliferation or arms control effect of security guarantees, they may be better off investing in promises to defend allies with conventional forces and demonstrating caution towards the use of nuclear weapons. In this case, credibility does not appear to assure and, instead, may backfire.

Table 1

	Dependent Variable					
	<u>Nuclear Investment</u>			<u>Conventional Investment</u>		
	(1)	(2)	(3)	(4)	(5)	(6)
Credibility	-0.072 (0.050)	-0.060 (0.048)	-0.078* (0.046)	0.008 (0.045)	0.013 (0.043)	0.007 (0.043)
U.S. Nuclear Use	-0.043 (0.044)	-0.046 (0.042)	-0.052 (0.040)	0.035 (0.040)	0.034 (0.038)	0.036 (0.037)
Russian Nuclear Use	0.015 (0.030)	0.024 (0.029)	0.026 (0.028)	-0.014 (0.027)	-0.009 (0.026)	-0.006 (0.026)
Credibility:US Nuclear Use	0.101° (0.063)	0.100* (0.060)	0.122** (0.057)	-0.055 (0.057)	-0.055 (0.055)	-0.047 (0.053)
Labour Party		-0.252*** (0.035)	-0.188*** (0.034)		-0.238*** (0.031)	-0.191*** (0.032)
Scottish National Party		-0.422*** (0.076)	-0.332 (0.074)		-0.325*** (0.069)	-0.271*** (0.069)
Liberal Democrats		-0.238*** (0.050)	-0.194*** (0.048)		-0.143*** (0.045)	-0.109** (0.044)
Green Party		-0.361*** (0.078)	-0.233*** (0.077)		-0.429*** (0.071)	-0.334*** (0.071)
Other Political Party		-0.204** (0.095)	-0.198** (0.092)		-0.142° (0.086)	-0.114 (0.085)
Age		-0.002** (0.001)	-0.001 (0.001)		0.0002 (0.001)	0.0004 (0.001)
Female		-0.114*** (0.029)	-0.103*** (0.030)		-0.010 (0.027)	0.011 (0.028)

⁵⁹ This effect holds in logit and probit models. This effect is substantively meaningful. While it isn't as large as the effects of political party, it is comparable to the effects of other measures that should be important contributors to preferences on military investment. For example, there is an 8.2 percentage point difference between support for nuclear investments among respondents who completed a graduate degree and respondents whose highest education level was secondary school. The effect is also sizable enough to constitute an important difference politically; for comparison, it is approximately the size of the difference between support for NATO in Germany and France or Germany and the United Kingdom (both 8 percentage points). See: Fagan and Poushter, *NATO Seen Favorably Across Member Countries*.

⁶⁰ The interaction effects are significant at the $p=0.11$ (model 1), $p=0.1$ (model 2), and $p=0.05$ (model 3) levels. These effects hold using probit and logit models. These effects are smaller, but still substantively important. For example, this is greater than the difference between respondents who are very disinterested in foreign policy news and those somewhat interested (2 percentage points) or the difference between respondents with military experience and without (1.8 percentage points).

Education		-0.024* (0.013)	-0.019 (0.013)		-0.010 (0.012)	-0.014 (0.012)
Household Income		-0.00004 (0.001)	-0.0001 (0.001)		-0.0004 (0.001)	-0.001 (0.001)
U.K. Citizenship		0.379 (0.323)	0.286 (0.309)		0.374 (0.293)	0.188 (0.288)
News Interest			0.012 (0.020)			0.045** (0.019)
Military			-0.014 (0.039)			-0.008 (0.037)
Alignment – U.S.			0.010 (0.022)			0.020 (0.021)
Alignment – NATO			-0.006 (0.022)			0.022 (0.020)
Alignment – Russia			0.016 (0.020)			-0.020 (0.018)
Anticipate War			0.014 (0.012)			0.012 (0.012)
Anticipate U.K. Use			0.033*** (0.013)			0.026** (0.012)
Nuclear Taboo			0.068*** (0.012)			0.007 (0.011)
World Policeman			0.050*** (0.017)			0.013 (0.016)
Dove			-0.087*** (0.020)			-0.093*** (0.018)
Constant	0.364*** (0.038)	0.346 (0.333)	0.080 (0.342)	0.755*** (0.035)	0.542* (0.302)	0.564* (0.319)
Observations	985	985	985	985	985	985
R ²	0.003	0.099	0.197	0.002	0.095	0.149
Adjusted R ²	-0.001	0.086	0.177	-0.002	0.082	0.127
Residual Std. Error	0.474 (df = 980)	0.453 (df=970)	0.430 (df=960)	0.429 (df=980)	0.411 (df=970)	0.401 (df=960)

Note: ° $p < 0.11$; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Testing the Mechanism: Why Credible Guarantees Backfire

The backfire effect theory offers a potential explanation for these results. The theory begins with an assumption that clients will want to avoid nuclear escalation more than guarantors. This preference is clear among U.K. respondents: 83% believe the use of nuclear weapons cannot be morally justified. Respondents' strategic preferences are similar. When were asked to imagine a Russian invasion of a NATO state involving the use of nuclear weapons, a substantial portion (41%) preferred NATO not use nuclear weapons under any circumstances, and a near majority (45%) supported nuclear use against Russia only after exhausting conventional alternatives. When respondents were told Russia did not initiate the use of nuclear weapons during its invasion of a NATO state, Britons were even more cautious about nuclear use: In this case, 87% opposed the use of nuclear weapons “under any circumstances.”

Although the literature on nuclear guarantees has largely presupposed client states will be more willing to use nuclear weapons to resolve security threats in their region than their guarantors will be, these findings suggest the consensus could require re-investigation. I find evidence of caution among the U.K. public about nuclear use in the European theater; further examinations of other clients' attitudes about nuclear use may shed more light on differential risk tolerance within nuclear alliances.

A mismatch in risk tolerance between clients and guarantors can cause clients to fear their guarantor will precipitate or escalate a nuclear conflict when the client would have preferred caution. Britons perceive this mismatch to exist. Only 15% of respondents believe the policy goals of the United States and United Kingdom are very well aligned. This may contribute to distrust of the United States and concern about adventurism. Britons believe the United States is overly risk tolerant when it comes to the use of nuclear weapons. 51% of respondents say they only trust the United States with nuclear weapons a little or they do not trust the United States with nuclear weapons at all. Likewise, 58% believe the United States will someday use its nuclear arsenal. This distrust is severe; respondents hardly trust the United States more than they trust Russia. (64% of respondents predict eventual Russian nuclear use.) In comparison, just 28%

of respondents are skeptical of the United Kingdom's ability to steward its nuclear arsenal and just 29% predict eventual U.K. nuclear use.

I argue preference misalignment can concern that guarantors will rashly escalate. Supporting this mechanism, I find high levels of concern among respondents about potential escalatory behavior by the United States. A majority of respondents (54%) agree "the U.K.-U.S. alliance will cause the U.K. to be brought into a nuclear conflict it could otherwise have avoided," and 76% believe the alliance will bring the United Kingdom into an avoidable conventional conflict.

Similar worries are evident in the control group. These respondents were not given any information about the U.S. security guarantee, but they were asked how they expected the United States and United Kingdom to respond to a Russian invasion of a NATO state. Almost one-third (31%) believed the United States would use nuclear weapons in a first strike against Russia, and nearly twice as many (60%) believed the United States would engage in a nuclear second strike. In contrast, respondents trusted the United Kingdom to exercise caution. About half as many respondents (15%) anticipated a U.K. nuclear first strike against Russia, and about half as many (34%) anticipated a U.K. nuclear second strike. These figures represent a serious mismatch in risk tolerance. In addition, the relatively high levels of trust in the United Kingdom's ability to act responsibly with nuclear weapons may suggest respondents see nuclear investments as a way to avoid escalation concerns associated with the U.S. nuclear guarantee, believing that if the United Kingdom were more central to NATO nuclear decision-making, it could act as a counterweight to U.S. interests by calling for nuclear restraint.⁶¹

The backfire effect theory expects clients to resolve escalation fears by distancing from their guarantor. As expected, a majority of respondents (56%) support decreasing U.K. military interdependence with the United States. Moreover, under some conditions, credible U.S. security guarantees were associated with preferences for shifting the United Kingdom's alliances. When respondents were told the United States was very likely to respond to a non-nuclear Russian invasion of a NATO state with a nuclear attack, 53% supported distancing U.K. security policy from the United States and 29% supported establishing a closer military relationship with Russia, compared to 49% that sought independence and 19% that supported closer ties with Russia when U.S. nuclear first use was unlikely.⁶² That U.K. respondents react to fears of U.S. escalation not only by seeking independence from the United States, but also by being open to cooperation with Russia, presents a potentially important policy consideration.

Clients can distance themselves from their guarantor by acquiring stronger, more independent security forces, which can then be leveraged against the client's guarantor in order to gain greater influence over joint security strategy. This framework anticipates a counter-intuitive finding: respondents with anti-nuclear views should nevertheless be willing to invest in nuclear weapons when faced with credible nuclear guarantees. These are the individuals who should most fear

⁶¹ Some respondents could be anticipating a lower likelihood of U.K. nuclear use not because they believe the U.S. nuclear guarantee obviates the need for a U.K. nuclear strike. This interpretation, however, would contradict U.K. and NATO nuclear posture, and respondents elsewhere indicate understanding that U.S. nuclear use would implicate U.K. nuclear forces.

⁶² The difference between support levels for independence from the United States was not statistically significant at conventional levels. The difference between support levels for closer ties with Russia was significant at $p=0.11$.

nuclear escalation by a guarantor. For them, nuclear investments could be seen as a way to minimize the risks of nuclear use by displacing the guarantor's influence.

I find some suggestive evidence to this end. 30% of respondents who oppose nuclear first strikes and 14% of respondents who oppose nuclear second strikes support investing in nuclear weapons when the U.S. nuclear guarantee is credible. A smaller percentage of respondents supports nuclear investments when the U.S. nuclear guarantee is not credible (28% and 12%, respectively.)⁶³ I similarly find respondents who are morally opposed to nuclear weapons nevertheless support nuclear investment when they are faced with a credible U.S. nuclear guarantee. 30% of these respondents support nuclear modernization when the U.S. nuclear guarantee is credible, compared to 26% in support of nuclear investments when the guarantee is not credible.⁶⁴ Although these comparisons are not statistically significant, they consistently suggest respondents who oppose nuclear use backlash against credible U.S. nuclear guarantees. Such respondents may see nuclear investments as an alternative to dangerous reliance on the U.S. nuclear arsenal.

Evidence of the Backfire Effect in Open-Ended Responses

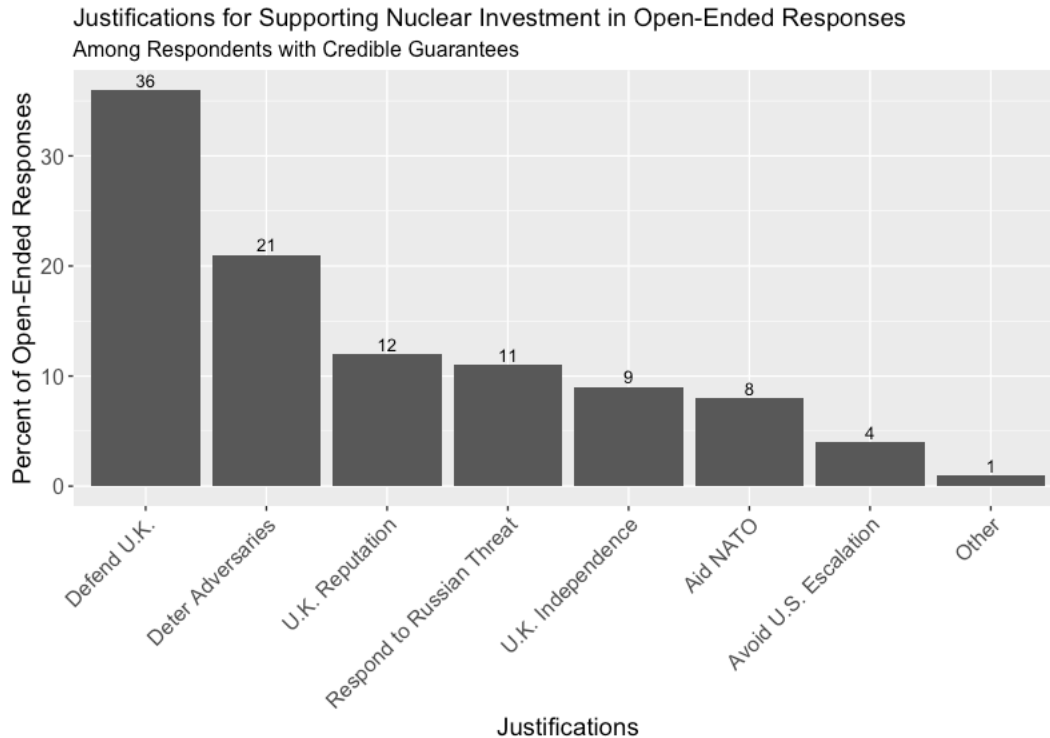
How do respondents explain support for investing in the U.K. nuclear arsenal? Using an iterative open-coding scheme, I assigned all complete open-ended responses by respondents in favor of nuclear investments to at least one category of justification.⁶⁵ Figure 1 presents the justifications respondents who were treated with information about a credible U.S. security guarantee provided for their support of nuclear investments. I find some evidence of thinking aligned with the backfire effect theory.

Figure 1

⁶³ These differences are not statistically significant in a two-sided t-test. This is, again, perhaps due to small sample size when the data is divided in this way (e.g. only 131 respondents oppose nuclear second strikes and are treated with a credible nuclear guarantee.)

⁶⁴ This difference is also not statistically significant. It may, however, contribute to explaining the otherwise-curious result in Table 1 of a positive and statistically significant correlation between belief in the nuclear taboo and support for investing in nuclear weapons.

⁶⁵ Open-ended items left blank or unintelligible were not assigned to justifications.



Some respondents identified a need for the United Kingdom to resist reliance on the United States and other allies in order to develop more independent nuclear capabilities. Respondents explained that, “it is paramount that the U.K. have an independent nuclear deterrent,” and said, “we can’t be left behind with minimal nuclear weapons.” These responses suggest a fear of reliance on the United States and NATO. They could also be interpreted as expressions of an individualist political ideology. Such beliefs could be contributing to the backfire effect by making respondents more worried about potential unilateral U.S. escalatory action.

Indeed, some individuals articulated concern about being “drawn into combat” by the United States. A small portion of respondents clearly expressed the fear of escalation at the core of the backfire effect. For example, one respondent wrote: “the U.K. is part of NATO [too] and has a strong connection with the U.S....it would be likely that the U.K. will be brought into a war.” Other respondents agreed, saying “it is likely we will be brought into a nuclear conflict,” “as a member of NATO we would probably get dragged into the action,” and “[if] the U.S. launched a nuclear attack [in retaliation against Russia] it wouldn’t be long before a lot of countries would also launch [nuclear weapons].” Some respondents suggested an independent nuclear capability could check escalation risks associated with the U.S. nuclear guarantee. For example, one respondent explained that if “the U.S. were to utilise [nuclear weapons]...there would be little in the way of stopping a war” and that investing in a more independent nuclear arsenal could help the United Kingdom “overcome the risk of involuntarily getting roped into the action.” These responses offer clear illustrations of the backfire effect mechanism.

An additional explanation for supporting nuclear investments relates to the global reputation of the United Kingdom. One respondent wrote that nuclear investments would help the United Kingdom “match other countries so they have less power over the U.K.” Another wrote the United Kingdom needed more nuclear weapons “so that we don’t look weak compared to the rest

of the world's superpowers." Some responses highlight the need to be seen as a peer. For example, respondents wrote: "we should stand shoulder to shoulder with our NATO allies" and that nuclear investments would mean the United Kingdom could be one of the "big boys" of NATO." These responses depict a core element of the backfire effect. Credible nuclear guarantees raise the profile of nuclear weapons as an instrument of power, a cornerstone of alliances, and as a necessary tool for resolving global security challenges. This makes powerful nuclear arsenals a more desirable goal. Moreover, many of these reputation-focused justifications suggest respondents believe greater nuclear capabilities could allow the United Kingdom to better counterbalance the United States and secure greater influence within NATO.

Some respondents who focus on the reputational benefits of nuclear modernization may be communicating a desire to increase the United Kingdom's prestige within the E.U., rather than within NATO or the U.S.-U.K. alliance. This is not an alternative explanation for the treatment effect, however, since respondents primarily concerned with the U.K. role in Europe would have the same incentives to modernize when the U.S. guarantee is credible as when it is not. Additionally, I find that respondents who read about a credible U.S. nuclear guarantee were not only more likely to support nuclear modernization, but they were also more than twice as likely to explain their support in reputational terms.

Additional alternative reasons to support nuclear modernization similarly do not explain the effect of guarantee credibility on support for nuclear investments. For example, some respondents argued nuclear weapons were important for responding to or deterring the threat posed by Russia. This could be because respondents were primed by the experiment in both the credible and non-credible guarantee treatments to consider Russian aggression. Concerns about Russia could play a role in the baseline level of support for U.K. nuclear modernization, but because all treatment scenarios involved a Russian invasion of a NATO state, attitudes about Russia do not drive the differences in nuclear modernization support across different guarantee types.

Similarly, the experiment primed all respondents to think about NATO commitments, and some respondents, in turn, supported investing in nuclear weapons in order for the United Kingdom to be better able to assist NATO. For example, one respondent wrote that the United Kingdom "should help [its] allies and not leave them to do all the work." Some respondents could have supported nuclear modernization when the U.S. nuclear guarantee was credible in anticipation of a need for U.K. nuclear use in concert with U.S. forces. Yet this is unlikely to be driving the observed positive effect of credibility on support for nuclear modernization. We should anticipate a heightened sense of the need for U.K. nuclear protection of NATO when the U.S. guarantee is unreliable; as a result, this logic would expect stronger support for nuclear modernization with non-credible nuclear guarantees than credible guarantees.

Separately, some respondents may support financing nuclear development for military-strategic reasons unrelated to the credibility treatments, such as a general belief that nuclear weapons are an effective deterrent. Figure 1 shows that the most common justifications feature a need to defend the United Kingdom and deter its adversaries. These views are also common among respondents supporting nuclear modernization when the U.S. guarantee is not credible. These attitudes contribute to the baseline level of support for U.K. nuclear modernization, but they

should not vary based on the credibility of the U.S. guarantee, unless respondents interpret the credibility of the guarantee as an additional signal of the ‘seriousness’ of the Russian threat. If this interpretation were common, however, more respondents would justify support for nuclear modernization in terms of defense and deterrence when the U.S. guarantee was credible than when it was not. Yet I find the opposite: 57% of respondents use this framework when told the U.S. guarantee is credible, compared to 65% when the guarantee is not credible.

Several considerations may generally inform the baseline level of support for nuclear modernization. These include concerns about defending the United Kingdom, deterring Russia and other U.K. adversaries, and aiding NATO. The open-ended justifications also shed some light on why nuclear modernization is more popular when the U.S. guarantee is more credible. Respondents indicate a desire for independence from the United States, express fears of U.S. escalation, and argue that nuclear modernization would increase the United Kingdom’s leverage.

Interestingly, concerns about U.S. escalation were also in play for some respondents who did not support nuclear investments. For example, respondents wrote: “the USA seem[s] to willing to use nuclear weapons, and that is a scary prospect,” “engaging in war with others just because of NATO and America is silly, as they go to war on everyone,” and “the U.K. should not be getting involved in problems that do not concern them, especially if nuclear attacks are involved.” One respondent added: “I don’t believe further investment into nuclear military capabilities should be encouraged. A nuclear attack on Russia from the United States would have extreme consequences; the U.K. should act in a manner that dissuades the likelihood of such an outcome.” While these escalation concerns do not result in respondents’ support for nuclear investments, and therefore do not necessarily align with the backfire effect as articulated here, they do illustrate the seriousness of fears about guarantor-initiated nuclear escalation.

This survey experiment shows the U.K. public backfires against credible U.S. nuclear guarantees. Relative to non-credible nuclear guarantees and credible conventional guarantees, credible nuclear guarantees increase support for additional investment in the U.K. nuclear arsenal. Credible nuclear guarantees are no more reassuring than the weakest guarantee option—a non-credible, conventional guarantee. This backlash is related to a perceived mismatch in preferences; the U.K. public views the United States as untrustworthy with nuclear weapons and anticipates the United States will escalate to nuclear use in situations where the United Kingdom would not independently choose to do so. As a result, Britons demonstrate a fear of reliance on the U.S. nuclear arsenal, with a majority of respondents believing this will lead to U.K. participation in an avoidable nuclear war. This drives support for investment in the U.K. nuclear arsenal.

Conclusion

The international relations literature generally argues when nuclear guarantees are insufficiently credible, clients will fear being abandoned, while credible guarantees can substitute for a client’s own military capabilities. The backfire effect, however, inverts this argument. I suggest credible guarantees can actually reflect a mismatch in risk tolerance between clients and guarantors. The strong demonstrations of resolve often required to make guarantees appear credible provide evidence a guarantor may rashly escalate when a client would prefer caution. For example, client

states often face higher costs to escalation than do guarantors, since escalation is likely to occur closer to the client's territory. Thus, credible guarantees can evidence a misalignment of preferences where guarantors are more willing to accept escalation risks than are their clients. To avoid being dragged into danger by their guarantors, clients must distance themselves from their guarantees. One way to do so is by developing stronger and more independent military capabilities.

This theory supplements current literature on alliances by articulating a new alliance hazard. Much of the existing work on alliance hazards focuses on risks faced by guarantors—such as entanglement and emboldenment—while arguing that, for clients, more credibility is always more desirable. Instead, this chapter argues more credibility can be less reassuring. In doing so, I identify a challenge for client states. Although alliance agreements are designed to protect clients from their security threats, they can sometimes endanger clients further by making it more likely crises and conflicts escalate.

In this chapter, I test the backfire effect using a survey experiment conducted among a sample of the U.K. public. This experiment contradicts the predictions of the substitution theory and shows how the micro-foundations of the backfire effect theory are in play in the U.S.-U.K. alliance. I find the U.K. public responds to credible nuclear guarantees with concern about U.S.-driven nuclear escalation and resultant support for additional investments in the U.K. nuclear arsenal. Moreover, I provide evidence for the mechanisms of the backfire effect theory, showing the U.K. public perceives a mismatch in preferences where the United States is more risk tolerant with nuclear weapons than the United Kingdom. Respondents express concerns about being dragged into conflicts or experiencing unwanted nuclear use by the United States. Respondents consequently support efforts to enhance U.K. military independence, including through an expanded U.K. nuclear arsenal. These investments may be seen as a way to secure leverage over U.S. military decision-making and thereby reduce the risk of unwanted nuclear use in the European theater.

This narrative aligns with historical scholarship on the U.K. decision to develop thermonuclear weapons. It also complements current trends in U.K. public opinion, such as the cooling of U.K. views about and high levels of distrust towards the United States. Findings showing fear of U.S.-driven conflict escalation may also help explain why a significant minority of the U.K. public reports seeing the United States as a global security threat.

These findings depict only a single case. While the United Kingdom is a difficult test case, further work is needed to examine the effects of guarantee credibility more broadly. These findings do comport with prior work, however, in which the backfire effect was shown to portray the attitudes of South Korean citizens about the credibility of the U.S. nuclear guarantee.⁶⁶ Future research could expand understanding of the backfire effect by examining guarantees to states with different risk tolerances, guarantee types, and regional security challenges.

Future research could also investigate how policymakers and other political elites understand and respond to the credibility of nuclear guarantees. By studying public opinion, this research

⁶⁶ Sukin, "Credible Nuclear Security Commitments Can Backfire: Explaining Domestic Support for Nuclear Weapons Acquisition in South Korea.

illuminates the micro-foundations of the backfire effect theory and identifies ways in which elites may face pressure from the public in response to the dynamics of security alliances. However, this study did not directly test elites' preferences. Although elite-public decision-making gaps may often be overstated,⁶⁷ targeted analyses of elite attitudes could shed further light on the effects of guarantee credibility and could investigate how elites respond to public pressures about nuclear alliances.

The backfire effect illustrated in this paper implies guarantors like the United States face a difficult Goldilocks problem. Their nuclear guarantees must be credible enough to evade client fears of abandonment, but not so credible as to trigger a threat of escalation. To resolve this, guarantors should reassure clients that they will be defended, but also that their guarantor has the ability to exercise caution and restraint with its nuclear arsenal. Guarantors can emphasize credible conventional commitments, while minimizing perceptions that they would be quick to use their nuclear assets. This strategy may be effective for reassurance, as minimally conventional commitments were found to have an essentially equal effect on client reassurance as credible nuclear guarantees. In addition, centering nuclear weapons in alliances can have the adverse effect of raising these weapons' desirability. Emphasizing conventional deterrence may therefore have the additional benefit of diminishing the possibility clients react to security concerns (including about escalation by guarantors) by building or enhancing nuclear capabilities.

Guarantors can also try to resolve the preference mismatch with their clients by institutionalizing consultation mechanisms to allow clients a say in how and when escalation occurs. Similarly, guarantors could implement safeguards and other mechanisms to restrain the ability to use nuclear weapons without client consent. Bilateral and multilateral talks could yield useful insight for determining how guarantees can best be managed to improve ally reassurance. Such policies could help guarantors minimize the misalignment of risk tolerance preferences with their clients.

Guarantors can also intentionally design nuclear posture and ally reassurance strategies in order to better reduce client concerns about precipitous escalation. In this regard, guarantors may benefit from reconsidering policies primarily intended to enhance the credibility of nuclear guarantees—such as the use of overt threats against adversaries, forward deployment of tactical nuclear weapons, or development of low-yield nuclear weapons. This research suggests such strategies may not only fail to reassure allies and to accomplish nonproliferation and arms control objectives, but they may also raise the risk of intentional or inadvertent nuclear use. In doing so, they could cause credible nuclear guarantees to backfire.

⁶⁷ Kertzer, "Re-Assessing Elite-Public Gaps in Political Behavior."