American Political Science Review (2026) 1–19
doi:10.1017/S0003055426101713

The Domestic Sources of International Trust

MICHAEL A. GOLDFIEN *U.S. Naval War College, United States*
MICHAEL F. JOSEPH *University of California, San Diego, United States*
ROSEANNE W. MCMANUS *Pennsylvania State University, United States*

How do states overcome mistrust? Scholars argue that costly foreign policy signals build trust. But when trust is low, such as during rivalries, states are unwilling to use these signals for fear of being cheated. We argue that domestic policies can also build trust by revealing information about a state's likelihood of cooperating internationally when there is a correlation between domestic and international preferences. We further argue that domestic policies have a distinct advantage: the value states accrue from them depends less on international reciprocation. As a result, domestic choices can reassure counterparts at moments when trust is so low that costly international signals appear prohibitively risky. We test the implications of our theory in case studies of the Cold War's end and United States–South Korea trust-building post-coup, illuminating several phenomena the current literature struggles to explain: initial trust-building between enduring rivals, asymmetric trust-building, and trust-building through illiberal domestic policies.

INTRODUCTION

Prominent research argues that states can use costly signals to overcome the security dilemma (e.g., Glaser 2010; Haynes and Yoder 2020; Jervis 1978; Kydd 2000; Yoder 2019b; Yoder and Haynes 2025). According to this literature, states reveal their genuine security-seeking intentions by pursuing *foreign* policies that greedy states would not, such as reducing arms or joining international agreements. But history is also replete with cases in which a government's *domestic* policy choices engendered trust with foreign counterparts. For example, in 1958, Gamal Abdel Nasser "unleashed a crackdown" on Egyptian Communists, which encouraged United States officials to "mend fences with the man whom they just months earlier compared to Hitler" (Radchenko 2024, 228). In 1961, a military coup and subsequent top-down economic reform in South Korea reinvigorated Washington's faith in its Northeast Asian ally (Brazinsky 2009). From 1985 to 1987, Mikhail Gorbachev released dissidents from internal exile, loosened emigration restrictions for persecuted Soviet Jews, and introduced *glasnost* and *perestroika*, all of which increased Western trust in Soviet intentions (Bartel 2022; Wilson 2014). More recently, after rebels formerly aligned with the Islamic State of Iraq and Syria (ISIS) toppled the long-standing Syrian regime, President Biden's remarks indicated that domestic policies such as protection of religious minorities would be a key litmus test for U.S. cooperation (Biden 2024).

These examples demonstrate that the standard costly signaling logic of trust-building could extend to domestic policies. But they also raise puzzles. First, the conventional wisdom is that illiberal and authoritarian actors are mistrusted, especially by liberal states (Maoz and Russett 1993; Russett and Oneal 2001; Tomz and Weeks 2013). But the above examples evidence how *both* liberal and illiberal—even violent or coercive—actions reassured the U.S. Second, researchers believe that trust-building between international rivals must start small—for example via modest arms control agreements—since even security-seekers fear exploitation (Kydd 2005). Yet, as detailed below, Gorbachev initiated trust-building with the West via domestic choices that, in the Soviet context, were revolutionary. Finally, while existing literature focuses on reciprocal trust-building choices, some trust-building described above began with unilateral domestic actions.

To the best of our knowledge, trust-building via domestic choices has previously been mentioned only in passing (Kydd 1997). Yet the puzzles above suggest that how domestic choices build trust departs from our understanding of foreign policy trust-building. Thus, a systematic analysis linking domestic policies and international trust is needed. We take up this task, theorizing the unique features of domestic policies that facilitate unappreciated opportunities to build trust.

To begin, we construct a simple two-period model of international trust problems, similar to Kydd (2005), Yoder and Haynes (2025), and others. We then advance

Michael A. Goldfien, Assistant Professor, Department of National Security Affairs, U.S. Naval War College, United States, michael.goldfien@usnwc.edu.
Michael F. Joseph, Assistant Professor, Department of Political Science, University of California, San Diego, United States, mfjoseph@ucsd.edu.
Corresponding author: Roseanne W. McManus, Professor, Department of Political Science, Pennsylvania State University, United States, rum842@psu.edu.

Handling editor: Monika Nalepa.

Received: March 14, 2025; revised: October 09, 2025; accepted: April 09, 2026.

two conceptual claims that differentiate domestic policy choices from international signals, which we introduce into the model. First, states' preferences over domestic policies are often correlated with foreign policy preferences. These correlations allow domestic policies to function as costly trust-building signals on the international stage. Second, domestic policy choices depart from international policy choices in a critical way: they tend to be more independent. We use the term *independence* to describe the extent to which payoffs from a state's policy choice depend upon an international counterpart's choice. In the classic security dilemma model of costly international signaling, scholars assume that choices are highly dependent; one state's payoff greatly depends on the other's decision (Jervis 1978; Kydd 2005). We argue that domestic decisions tend to be more (but not necessarily entirely) independent, meaning that the benefits and costs that a state derives from them depend less on what its counterpart does. The relative independence of domestic choices allows them to spark trust-building between states that seek cooperation but consider international signals too risky.

These insights illuminate the empirical puzzles described earlier. First, our theory explains trust-building following illiberal domestic choices. While liberal policies can establish trust in some cases, our theory indicates that preference-compatibility is the key determinant of international cooperation. Thus, there are some situations—for example, harsh anti-Communist repression—in which illiberal domestic choices could reassure even liberal states. Second, it explains the ability to build trust through grand gestures on the domestic level: when choices have high independence, security-seekers can undertake salient and reassuring domestic policies without fear of international exploitation. Finally, we identify a novel asymmetric trust-building equilibrium when initial trust and independence are both lopsided. This explains historical cases where one side takes a unilateral first step, knowing the other side will not initially reciprocate.

We illustrate these insights in case studies of U.S.–Soviet trust-building at the Cold War's end and U.S.–South Korea trust-building during the Cold War. The former illustrates how independence allows trust-building to occur through grand gestures, even among highly distrustful rivals. The latter provides an illustration of illiberal and asymmetric trust-building, which we conjecture could be common in Cold War U.S.–client relations.

Our argument has several implications. First, we offer a general theory of domestic politics and international trust. The extensive trust literature has paid scant attention to domestic politics. We highlight a unique feature of domestic choices—their independence—that makes them especially useful in kick-starting a trust-building process between international rivals. Prior research has linked domestic politics to signals of resolve (Fordham 1998; McManus 2017; Renshon, Yarhi-Milo, and Kertzer 2023; Schultz 1999; Weeks 2008). We show that domestic politics may be equally important for reassurance.

Second, given that political elites spend the majority of their time on domestic politics (Lindsey and Hobbs 2015), we expose avenues for trust-building that past scholarship may have overlooked. This forges a stronger connection between international and comparative politics, by illuminating unappreciated international implications of domestic choices such as land and tax reform (e.g., Flores-Macías 2019), purges (e.g., Bokobza et al. 2022), unethical government research (e.g., Joseph and Poznansky 2024), and discrimination against minorities and immigrants (e.g., Peters 2015). We provide a common framework for understanding why, whether, and under what conditions any domestic choice facilitates international trust or mistrust.

Finally, we advance research on regime type and peace (e.g., Maoz and Russett 1993; Russett and Oneal 2001). Our framework sheds new light on the liberal peace, providing a novel explanation for how democracies can identify each other's cooperative international preferences. Yet our approach goes beyond the democratic peace by explaining trust-building between autocracies or between mixed-regime pairs. It explains puzzling cases in which illiberal actions engender trust, and also allows for the possibility that domestic reforms that fall well short of regime change can reassure international rivals.

TRUST PROBLEMS AND FOREIGN POLICY TRUST-BUILDING

Trust problems are central to international relations. They play a role in power transitions (Yoder 2019a), conventional and nuclear arms races (Bas and Coe 2016; Debs and Monteiro 2014), arms control negotiations (Coe and Vaynman 2020), great power rivalry and rapprochement (Glaser 2010; Joseph 2026), and general problems of cooperation (Crescenzi 2018; Goldfien 2024). Policymakers care about trust problems. For example, as American policymakers manage relations with China, building trust may be as important as signaling resolve (Glaser 2015).

Scholars studying the security dilemma consider two ideal-type states: security-seeking and greedy (Glaser 2010). The fundamental difference between them is their value from exploiting each other. Security-seekers maximally benefit from reciprocal cooperation, while greedy states prefer to defect no matter what their partner does.[1] A trust problem arises because states are uncertain about each other's intentions, and the value they accrue from cooperation depends on what the other does.

We visualize a formalization of this problem in Table 1. Here, the trust problem focuses on two states that will soon confront a critical foreign policy choice, potentially the choice to comply with an arms control agreement or not, or to withdraw forward-deployed troops or not. In such situations, each state can choose to cooperate or defect. Each box in the 2 × 2 of Table 1 describes the payoffs players accrue given different combinations of those choices. Table 2 explains the parameters.

[1] The sources of these motives are beyond our scope. For overviews, see Joseph (2021; 2026).

TABLE 1. Representation of the Trust Problem

		Player B	
		Cooperate ($b_t = c$)	Defect ($b_t = d$)
Player A	Cooperate ($a_t = c$)	Mutual cooperation 1, 1	B cheats A $-k, e_B^m$
	Defect ($a_t = d$)	A cheats B $e_A^m, -k$	Mutual defection 0, 0

Note: Each box details an informal description and player-specific payoffs given the four realizable combinations of potential decisions. The values for mutual cooperation and mutual defection are normalized to 1 and 0, respectively. Players' motives are privately drawn given $pr(e_i^m = L) = p_i, pr(e_i^m = H) = 1-p_i$.

TABLE 2. Substantive Description of Parameters

Parameter	Description
e_i^m	i's value from exploiting j. $m \in \{g, s\}$ represents i's motives. $e_i^s = L$ (low-value) and $e_i^g = H$ (high-value) are exploitation payoffs for security-seekers and greedy types, respectively ($H > 1 > L$).
$k > 0$	The cost of being exploited
$p_i \in [0, 1]$	Probability i is a security-seeker. Determines j's initial trust in i.

The critical feature across all trust models is that security-seekers and greedy types differ in their preference ordering over different combinations of choices. If A is greedy, then A's preference over outcomes is: A cheats B > mutual cooperation > mutual defection > B cheats A. If A is security-seeking, then it is: mutual cooperation > A cheats B ∼ mutual defection > B cheats A, where ∼ represents flexible ordering. B's ordering is symmetric. Note that both types receive the worst payoff if cheated (i.e., the sucker's payoff). Substantively, being cheated on major international agreements not only puts a state's national security at risk, but can damage a leader's international reputation and domestic standing (Colaresi 2004).

Because security-seekers prefer to reciprocate rather than exploit cooperation, we might intuit that two security-seekers could always cooperate. Indeed, this would be the case if two security-seekers knew each other's type. Yet states are usually uncertain about each other's intentions, and this can complicate cooperation. In the trust game, we assume that Player A knows her own type, but she only knows that B is a security-seeker with probability p_B. She believes that B is greedy with probability $1-p_B$. Similarly, B believes that A is a security-seeker with probability p_A.

The most basic questions trust scholars ask are: given that states are uncertain about each other's motives, when are security-seekers willing to cooperate, and when does uncertainty cause them to defect? Scholars rationalize mutual cooperation if both states are sufficiently confident that the other holds security-seeking preferences. However, if even one state is sufficiently uncertain about the other's motives, mutual defection is the unique solution in a one-period interaction (Kydd 2005; Yoder 2019b).

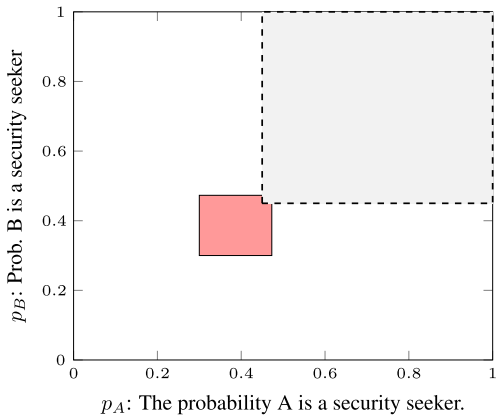
Uncertainty about a counterpart's preferences can cause cooperation between security-seekers to fail through two mechanisms. First, A may defect because A does not trust B enough (i.e., believes that B is likely greedy). Second, even if A is a security-seeker and trusts B, B may not trust A. Here, A also defects because A expects B to defect due to B's distrust. This second mechanism illustrates the difficulties of two-sided trust problems. Each side must trust the other and believe that the other trusts them.

The central insight of the trust-building literature is that distrustful states can build trust through costly signals. Many scholars model this as two sequential foreign policy choices (Kydd 2005; Yoder 2019b). The second-period choice is often conceptualized as a high-stakes choice between cooperation and competition. The initial choice is conceptualized as a signaling opportunity. Through their initial choice, security-seekers can potentially reveal their type by taking actions that have higher payoffs for them than for greedy types (Glaser 2010; Kydd 2005). Consider the example of arms reductions. Cutting arms is very costly for greedy states, since it undermines their ability to pursue expansionist foreign policies. It is less costly for security-seekers, who have more modest ambitions. Therefore, a state's willingness to cut arms can credibly signal its security-seeking motives. A similar logic applies to other costly trust-building signals that scholars have identified, including signing arms control treaties (Kydd 2005), visits and other symbolic gestures (Berenji 2020), building purely defensive weapons (Glaser and Kaufmann 1998), and retrenchment (Yoder 2019b).

The Baseline of Foreign Policy Signaling

We treat extant research on trust-building through foreign policy signals as our baseline. In Supplementary Appendix A.1, we introduce a two-period model that first privately draws each state's intentions (using probabilities p_A, p_B) and then iterates the simultaneous-move trust-building model visualized in Table 1 over two periods. In later sections, we analyze the effects of domestic (or other independent) signals by manipulating features of the first-period choice based on our substantive arguments about why domestic policies are different.

FIGURE 1. Equilibria that Survive Refinement in the Standard Trust-Building Game



(a) Conditions where PBE emerge

Motive	Period 1	Period 2 | Period 1			
		CC	CD	DC	DD
Trust-building equilibrium (dark red, solid)					
Security	C	C	D	D	D
Greedy	D	D	D	D	D
Suckers equilibrium, (gray, dashed)					
Security	C	C	D	D	D
Greedy	C	D	D	D	D
Tragic equilibrium (white space)					
Security	D	D	D	D	D
Greedy	D	D	D	D	D

(b) Equilibrium description

Note: Panel (a) plots three equilibria that can survive an efficiency refinement given $H = 1.5, L = -0.1, k = 0.9$. These equilibria uniquely survive refinement in their plotted parameter ranges. Panel (b) describes the strategy profiles. Since these are symmetric strategies, we present them for one state. We report inefficient pure and mixed strategy PBE in Supplementary Appendices A.2 and A.2.1, respectively.

We advance the simultaneous-move model for two reasons. First, it harnesses the power of incrementalism (Ashworth, Berry, and Bueno de Mesquita 2021, 58–9) by adding one parameter (independence, introduced next) to the framework common across much trust-building research (Acharya and Ramsay 2013; Jervis 1978; Kydd 2000; 2005; Schultz 2005; Yoder and Haynes 2025). Second, it captures the strategic problem we hope to study. In the cases that interest us, such as enduring rivalries, arms control, and major power–client relations, both sides have the ability to exploit the other by adjusting their policy before the other realizes and can take countermeasures (e.g., Braumoeller 2008; Glaser 2010; Waltz 1979). Thus, at every moment, each worries the other is cheating them, creating pressure to cheat first. If we instead sequenced choices, it would artificially resolve the trust problem for whoever moves second.[2]

In Supplementary Appendix A.2, we analyze the baseline model for perfect Bayesian equilibria (PBE). Consistent with Kydd (2005) and others, we find many PBE—including mixed and pure strategy equilibria—in overlapping parameter ranges. We report them all in Supplementary Appendices. In the article, we focus on equilibria that survive an efficiency refinement (PBE-ER). This rules out any equilibria if, under the same parameter ranges, all types of all states would prefer to play a different equilibrium (i.e., it is Pareto dominated given initial beliefs). This is substantively appealing for the cases that motivate us because security-seekers often want to end competition because it is inefficient. It also holds technical appeal. Empirical implications of all trust models are hampered by multiple equilibria in overlapping parameter ranges. One reason we are confident in our main conclusion is that regardless of whether we apply a refinement or not, we cannot rationalize trust-building in the international signaling model when initial trust is low. Yet, as we show below, we can rationalize trust-building in these ranges when we model domestic signaling opportunities. Further, in the domestic signaling model, only trust-building equilibria survive refinement when initial trust is low, and they reliably do so, suggesting that these results are most robust in interesting parameter ranges.

The results of our baseline analysis are summarized in Figure 1. Panel (b) describes the equilibria that can survive the PBE-ER. Two points are notable. First, all equilibria are symmetric, pure strategy equilibria. Second, given we start with the same baseline structure as others (e.g., Kydd 2005), the three PBE-ER we find are consistent with past scholarship on trust-building through international costly signaling. We are most interested in the trust-building equilibrium. In it, security-seekers always cooperate in the first period, but greedy states do not. Then security-seekers cooperate in the second period only if their counterpart cooperated in the first. We call it a trust-building equilibrium because each player learns the other's motives from first-period actions. When they both cooperate, they come to trust each other, facilitating second-period cooperation.[3]

Panel (a) plots the conditions under which the equilibria listed in panel (b) survive as a function of each player's initial trust levels. The conditions reflect findings in past work (Kydd 2005).[4] Here, we emphasize the boundaries

[2] We accept some substantive areas exist, including proliferation (Debs and Monteiro 2014), where one-sided exploitation is normal.

[3] This symmetric equilibrium is the only trust-building PBE-ER in the baseline model. After introducing our innovation below, we find both symmetric and asymmetric trust-building equilibria. Therefore, we later refer to this classic trust-building equilibrium as the "symmetric trust-building equilibrium."

[4] Notably, Kydd does not apply a Pareto refinement. He sustains inefficient equilibria we discuss in Supplementary Appendices.

of the trust-building equilibrium (which we shade dark red in this and all future plots).[5] We find there is a minimal level of trust necessary to sustain trust-building. Crucially, this lower bound is not a function of the efficiency refinement.[6] Rather, we cannot sustain first-period cooperation in any PBE with low levels of initial trust because when at least one state initially believes the other is very likely greedy, that state defects in period one. The other side, regardless of its true motives and level of trust, also defects because it understands exploitation is certain. Because it is impossible to build trust in the first period, this also assures we cannot sustain second-period cooperation. Therefore, when initial trust (p_i) is low, the unique PBE in the baseline model is mutual defection in both periods. This finding explains why trust-building is difficult, especially for enduring rivals or other highly distrustful states. Over time, rivals may genuinely come to desire peace as a result of exhaustion, shifting values, or leadership change. But they may be too fearful to take the first step toward trust-building because they are mistrustful of their rival and fear that cooperative gestures will be exploited. This motivates our question: when states start out deeply mistrustful of each other, how can they initiate trust-building?

DOMESTIC POLICIES AND INTERNATIONAL SIGNALING

We argue that when there is too much distrust for security-seeking states to risk international cooperation, they can still kick-start the trust-building process through costly signaling via domestic policy choices. When we say domestic choices, we do not primarily mean broad attributes of states, such as regime type (Russett and Oneal 2001). Regime type is a large-scale domestic policy choice, but it is not our primary focus because regimes rarely change. We also do not mean leader turnover (Wolford 2007), which is insufficient in itself to alleviate the security dilemma. Previously distrustful foreign counterparts will not immediately trust new leaders because they often come from the same pool of elites and because reputations adhere to both states and leaders (Goldfien and Joseph 2023; Goldfien, Joseph, and Krcmaric 2024; Renshon, Dafoe, and Huth 2018).

Instead, we focus generically on government decision-makers who are confronted with a domestic problem and can select among different policy options to resolve it. This might include when a government faces nationwide protests and must decide between repression and negotiation with demonstrators, or when it faces economic stagnation and could choose to address it by nationalizing industries or allowing its currency to float. One important feature of these choices is that they have a structure resembling the international choices that states face in the signaling stage of the trust-building game. Like typically studied international choices, these domestic choices present states with at least two different policy options, and a state's preferences over these options are related to its broader preferences.[7] As at the international level, if a state chooses a domestic policy that does not align with its true preferences, it is less satisfied with the outcome. Therefore, opportunity costs give states an incentive to act true to their preferences. However, for domestic policies to communicate information about foreign policy motives, there must be a correlation between domestic and international preferences.

We argue that domestic policies can signal international motives because states' domestic policies are often related to their foreign policy goals. Goldfien, Joseph, and McManus (2023) show that decision-makers have underlying dispositional attributes that influence their sensitivity to both the costs of fighting in a crisis and the costs associated with certain domestic choices. We argue that preferences across the domestic and international policy decision-spaces are similarly correlated in certain trust-building scenarios. That is, the preference to exploit international cooperation is likely correlated with certain domestic policy preferences.

One version of our claim overlaps with the liberal peace. Some argue that domestic policies such as free elections and respect for ethnic minorities and human rights predict security-seeking international intentions (Kydd 1997; Maoz and Russett 1993; Tomz and Weeks 2013; 2020). Indeed, our first case study illustrates how the U.S. interpreted Soviet liberalizing domestic reforms as evidence of international trustworthiness. But we argue that what really matters is compatible international preferences, not liberal actions per se.[8] Thus, which policies promote trust depends on the context. In certain contexts, domestic actions that are illiberal, but are likely to be correlated with a preference to cooperate with a particular international partner, can induce trust.[9] For example, the decision of right-wing Cold War dictators to brutally suppress domestic Communist groups was illiberal. Still, as we show later, U.S. officials inferred that the preferences revealed by these actions were correlated with a preference to cooperate with the U.S. internationally.[10]

[5] As Kydd (2005, 192) shows, the boundaries of this equilibrium will shift with different parameter values, but the core conclusions remain the same.

[6] The tragic (mutual defection) equilibrium is the least efficient. If states can rationalize cooperation through another equilibrium, it Pareto dominates the tragic equilibrium.

[7] Leaders usually have more than two possible responses to domestic problems, but that is also true of international problems. Like the international trust-building literature, we emphasize two choices that represent the options most favorable to security-seeking and greedy types.

[8] For research highlighting the importance of compatible preferences in other contexts, see Voeten (2021), Gartzke (1998), and Spaniel and Smith (2015).

[9] There are even cases where liberalizing reforms can induce mistrust. For example, much earlier liberalizing reforms in the Soviet Union under Khrushchev arguably worsened relations with China (Haynes and Yoder 2020)

[10] Domestic suppression may be correlated with greater international resolve (see, e.g., Goldfien, Joseph, and McManus 2023), but high resolve is not incompatible with cooperation among states with similar preferences.

Since the correlation between preferences for domestic policies and international cooperation is context-dependent, we cannot identify particular domestic choices that always promote trust. Indeed, democracies, Communist regimes, right-wing dictatorships, and theocratic regimes might draw the *opposite* trust inferences from the same domestic decisions. For example, European analysts are likely to interpret the Turkish government's push to expand Islam's role in public life as evidence of lower willingness to cooperate, but analysts in Islamist states may infer higher willingness to cooperate.

Furthermore, we do *not* claim that security-seeking and greedy states hold reliably different preferences over all, or even most, domestic policy choices. Many domestic choices (e.g., the speed limit) have no bearing on world politics. Further, even when a domestic choice is correlated with international preferences, the correlation is imperfect. The international choices that trust-building models have historically focused on, such as cutting arms, directly impact a state's ability to fight a war and thus plausibly have a strong correlation with greedy or security-seeking preferences.[11] Since domestic choices rarely have such a direct effect on warfighting capabilities, their correlation with greedy or security-seeking preferences can be more variable. For example, allowing freedom of expression is consistent with liberal values and thus likely correlated with willingness to cooperate with the U.S. However, the correlation is imperfect. Even some U.S. allies suppress free speech in order to prevent hate speech, social unrest, or dissent.

How strong does the correlation need to be for domestic choices to signal international motives? Because this is partly a strategic problem, we explore the level of correlation necessary using our formal model.[12] We find domestic choices can induce trust and international cooperation with only a moderate correlation (defined precisely in Supplementary Appendix A.5). However, the amount of information communicated is greater when the correlation is stronger.

This raises the question: if domestic and international preferences are imperfectly correlated, why would domestic choices play an important role in international trust-building? Why are their effects not overshadowed by international signals? This is the question we now consider.

The Domestic Advantage: Payoff Independence

We argue that many domestic policy choices have an unexplored advantage that enhances their capacity to forge trust: their level of payoff independence. In brief, independence refers to the extent to which a state's payoffs (i.e., the combination of benefits and costs accrued) from a choice depend upon what a foreign counterpart does.[13] The concept of independence becomes relevant whenever a state faces a domestic or international policy problem and can choose among different policy options to address it. A state's payoff from each option always depends on its own preferences. Yet sometimes the payoff also depends, at least partially, on what option a foreign counterpart chooses.

We conceptualize the degree to which choices have payoffs that are independent from or dependent on a foreign counterpart's actions as a continuum. At one extreme, a state's choice is maximally independent if the value it gets from selecting either policy option is the same no matter what its foreign counterpart does. For example, suppose State A was faced with nationwide protests (the policy problem) and considering whether to repress or negotiate with protesters. If A's value from either repressing or negotiating does not depend on anything B does, then A's decision about how to respond to the protests is fully payoff independent. If instead, A's payoff from this decision depends somewhat on B's response (e.g., whether B lodges diplomatic criticism or imposes sanctions), then the choice would only be partially independent. Put another way, if a state can calculate its own payoffs from each policy option without considering what any other state will do, then the choice is fully independent. But if the payoffs depend slightly, partially, or highly on what another does, then the decision is slightly, partially, or highly dependent.

Existing research into the security dilemma assumes choices are highly dependent (Kydd 2005). For example, in the arms control variant, the immediate payoff A gets from the decision to reduce arms (cooperate) depends on whether B also reduces arms (cooperates) or not (defects). This assumption is reasonable because the value states accrue from many international policy options previously studied are heavily determined by what rivals do.

In contrast, domestic choices often have greater independence. Examples of domestic choices that are plausibly near-fully independent include reforming social welfare, reducing domestic regulations, or changing land use policy. The associated costs and benefits have little to do with the choices foreign countries make. But choices about immigration policy and press freedom could be moderately independent. Easing immigration standards could result in different levels of immigration depending on other states' policy choices. The costs and benefits of permitting press freedom could depend somewhat on whether foreign counterparts engage in influence campaigns (Levin 2021). This last example also highlights that States A and B need not be choosing among identical options in trust-building (or even among options that are both domestic or both international).

Overall, we argue that domestic choices tend to be more independent than foreign policy choices *on*

[11] Even at the international level, this correlation likely varies. Yet existing trust-building models have not examined it (Kydd 2005).
[12] See Supplementary Appendix A.5.
[13] The concept of independence is relevant to both international and domestic choices, although we argue below that domestic choices are often more independent.

TABLE 3. Examples of Choices with Varying Degrees of On-Average Independence

Options	Independence	Explanation
Domestic: Low/high speed limit	High	Costs/benefits do not depend on other countries' choices.
Domestic: More/less social spending	High	Costs/benefits do not depend on other countries' choices.
Domestic: Permit press freedom/not	Moderate	Costs/benefits are primarily domestic, but costs of free media are higher if rival authorizes influence operations.
Domestic: Repress protesters/not	Moderate	Costs/benefits are primarily domestic, but rival's criticism or sanctions could increase repression's cost.
International: Travel to a summit meeting/not	Moderate	Greater benefits if rival makes policy concessions at summit, but benefits of demonstrating leadership may apply either way.
International: Institute tariffs/allow free trade	Low	The effect of either option on exporting firms depends greatly on whether other countries reciprocate.
International: Cut/build arms	Low	Cutting arms leaves state vulnerable if rival does not reciprocate. Arming is expensive, but confers military advantage if the other side does not build.

Note: High, medium, and low codings roughly approximate where choices fall on an independence continuum.

average. We also believe few domestic choices are as dependent as the choice to cut or build arms that the trust-building literature typically focuses on. However, not all domestic (foreign) policy choices are fully independent (dependent), and some foreign policy choices are even more independent than some domestic choices. Table 3 summarizes how various domestic and international choices can vary in independence. To be clear, these are on-average codings, and we encourage researchers to apply contextual knowledge in determining how independent actions are in historical cases.

We will use our formal model to show that more independent choices, which tend to be domestic, have an important advantage when it comes to trust-building. As noted earlier, the primary barrier to trust-building with international signals is the fear of being cheated. However, this fear assumes that one state's value for taking a trust-building action hinges on what the other state does (i.e., full dependence). If the payoff from taking an action is independent of what the other does, this fear diminishes, and security-seekers face little risk when they make choices that reflect their genuine cooperative preferences.[14] For example, a state shifting policy in response to a domestic crisis would not fear that a counterpart would be able to greatly exploit its choice.[15] At the same time, greedy states are more easily identified because they can no longer claim, "I am not cooperating because I am afraid that you will cheat me." Therefore, domestic choices that are both independent and correlated with security-seeking or greedy international preferences can play a crucial role in international trust-building.[16]

Crucially, independence *does not* mean that policy choices are costless. It means that the benefits and costs are not contingent on a counterpart's behavior. In the absence of dependence on the counterpart's behavior, the payoffs of choices will depend even more heavily on a state's own preferences. This means that it would entail an even larger opportunity cost for a state to play against type, promoting behavior that reveals more information about type.

The Strategic Implications of Independence for Trust-Building

We introduce variation in payoff independence into the baseline two-period trust game described earlier. We conceptualize the level of independence of each state's choices using the continuous parameters $\beta_A, \beta_B \in [0, 1]$. We apply these parameters *only* to the first period of the model. We omit the independence parameters from the second period because they are effectively equal to 0—that is, the choices are fully dependent, ensuring the second period is identical to the classic trust problem.[17]

We visualize the updated game with independence in Table 4. When β_A and β_B both equal 0, the model converges to the classic model of trust-building through international actions presented earlier (Table 1 and Figure 1). In this classic model, A's value for cooperating or defecting depends on a combination of A's type and B's choice. In contrast, when β_A equals 1, it means that B's choice in the first period has no impact on A's value from cooperation or defection in the first period.[18] Rather, A's first-period value from cooperation depends

[14] The fact that these choices reflect true preferences arguably allows them to function as "indices," sources of information that are believed to be "inextricably linked to the actor's capabilities and intentions" (Jervis 1989, 18).

[15] If the counterpart is security-seeking, it would prefer to reciprocate cooperation anyway.

[16] This is also true of independent international choices, although we believe domestic choices tend to have greater independence.

[17] This creates a hard test for the ability of first-period actions to build trust. If second-period actions could also be partially independent, the trust problem would be less severe.

[18] We continue to use the "cooperation" terminology for consistency, but with full independence one state is not really cooperating with the other. Rather, it is making an independent choice that is correlated with willingness to engage in future cooperation.

TABLE 4. First-Period Payoffs in the Game with Independence

		Player B	
		Cooperate	Defect
Player A	Cooperate	1,1	$\beta_A - (1-\beta_A)k, e_B^m$
	Defect	$e_A^m, \beta_B - (1-\beta_B)k$	$\beta_A e_A^m, \beta_B e_B^m$

Note: These updated payoffs only apply to the *first period.* Second-period payoffs are as in Table 1. Both are equal when $\beta_A = \beta_B = 0$.

entirely on A's preferences.[19] When β_A and β_B are between these values, it means both states' trust-building payoffs are partially dependent.

We then solve for all the PBE of our model in Supplementary Appendix A.3 under the assumption that the level of independence can vary.[20] Treating the baseline model as the counterfactual, we preview the effects of introducing and increasing independence in Table 5. The top row explains how introducing independent signaling actions alters the conditions for equilibria found in the baseline model. The second and third rows explain why new PBE-ER emerge when we introduce independence in the choices of one or both sides. As the table indicates, introducing independence has many nuanced implications. We can no longer support the commonly studied tragic equilibria when independence is high; the trust-building equilibrium originally identified now survives under lower levels of initial trust, and several new equilibria, including mixed strategy and asymmetric trust-building equilibria, emerge.

We explore all these nuances in the technical appendices. Here, we focus on establishing our primary claim: introducing independence facilitates trust-building under conditions of deep initial mistrust. To establish this claim as clearly as possible, we initially focus on the symmetric pure strategy trust-building equilibrium (shaded dark red in Figure 1). This was the only trust-building PBE-ER established in the baseline model, and it continues to survive as we introduce variation in independence, allowing for a clear comparison.[21] Proposition A.6, reported in the Supplementary Appendix, formally characterizes the conditions under which we can sustain this equilibrium given that payoff independence can vary. Here, we describe the empirical implications of increasing independence by contrasting when trust-building occurs under high independence, relative to the baseline condition. Results 1a and 1b described below follow from a comparative static analysis of the trust-building equilibrium, which is generally presented in Supplementary Appendix A.3.2.

> *Result 1a: Independence and trust-building given initial mistrust.* Starting at the baseline ($\beta_i = 0$), as independence increases, the minimum level of initial trust (p_i) necessary to sustain a trust-building equilibrium decreases.

In the baseline model, we could not sustain trust-building equilibria when initial trust was low because the security-seeker faced two strategic problems that prevented her from taking the initial trust-building action. Increasing independence increasingly resolves both of these problems. The direct problem was that the security-seeking A worried that B would cheat her. As the independence of A's choice increases (β_A), A's value is increasingly determined by her type, and less by how B reacts. Since security-seekers directly benefit from cooperation, security-seeking A's value for cooperation is increasing with independence, regardless of concerns that B is greedy. The indirect problem was that the security-seeking A understood that B was deeply suspicious of her. Thus, even though A was a security-seeker, A also understood that B's suspicion of A would drive even security-seeking B to defection. Increasing the independence of B's choice (β_B), indirectly incentivizes A to cooperate because A now believes that security-seeking B will cooperate in spite of B's initial mistrust. As in the baseline model, both states weigh the tradeoffs of being exploited in the first period against the potential value of cooperation in both periods. But as independence increases, mistrustful security-seekers are increasingly willing to risk cooperation because they are less sensitive to the cost of being exploited (the direct effect), and believe they are less likely to be exploited by their rival (the indirect effect).

We visualize the effect of increasing two-sided independence in Figure 2. The figure assumes intermediate levels of independence for both states ($\beta_A = 0.6, \beta_B = 0.7$), but otherwise holds the k, H, L values at the same levels as in Figure 1, allowing for direct comparison. Substantively, these intermediate levels of independence could reflect a scenario where both sides face domestic reform opportunities. Yet because not all independent choices are domestic, it could also reflect that one side faced a domestic choice and the other faced a reasonably independent foreign policy decision.

At the highest levels of initial trust, we still cannot sustain trust-building equilibria. The reason is that independence does not influence the greedy type's incentives to conceal his true preference, which is to cheat his rival in the second period. However, looking at the lowest levels of initial trust reveals a surprising result:

> *Result 1b: Independence threshold.* When both players' choices are sufficiently independent (i.e., $\beta_A, \beta_B > \frac{k}{1+k-L}$), a trust-building equilibrium always exists for states that start out with the highest possible level of confidence that the other is greedy (i.e., $p_i \to 0$).

[19] In each case, the effect on B's utility is symmetrical.

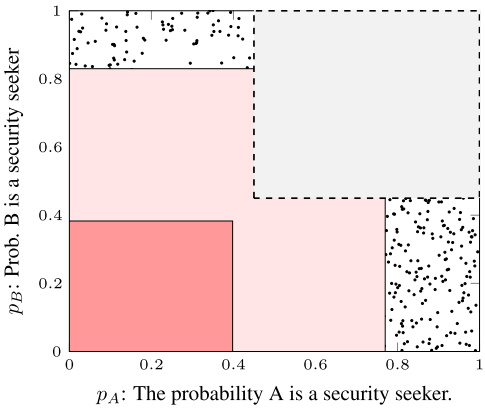
[20] The analysis partly relies on preliminaries reported in Supplementary Appendix A.1.2.

[21] Focusing on the pure strategy equilibrium assures we are conservative in reporting how important independence is for facilitating trust-building. If we included the mixed strategy equilibria also, we would report more far-reaching effects.

TABLE 5. Equilibrium Changes from Introducing Independence

Description	Difference from standard model caused by independence (β)
Changes to equilibria we can sustain in the baseline model	
Trust-building	As independence increases, the equilibrium expands into low trust levels. Once the independence threshold is met ($\beta_i > \frac{k}{1+k-L}$), trust-building is possible even at the lowest initial trust levels.
Suckers	No change
Tragic	When either side meets the independence threshold, we cannot support this equilibrium, *even without the efficiency refinement* under any parameter ranges. Security-seekers deviate from $a_1 = d \rightarrow c$.
New equilibria that emerge when β_A, β_B both exceed threshold (Figure 2)	
Mixed strategy trust-building	When $\beta_i = 0$, this was Pareto dominated by the suckers equilibria. As independence increases, the equilibrium expands into low levels of trust. The mixing inefficiencies reduce greedy states' incentives for first-period cooperation, assuring mixed trust-building survives for higher p_i values than pure strategy trust-building.
Semi-tragic	This replaces the tragic equilibrium as the least players can assure themselves because security-seekers now prefer cooperation knowing they will be cheated.
New equilibria that emerge when independence is lopsided (e.g., β_A low, but β_B high) (Figure 3)	
Asymmetric equilibria	The highly independent security-seeking state cooperates without expecting reciprocation. The highly dependent security-seeker defects. When trust is also lopsided, we can support an asymmetric trust-building equilibrium. When trust is jointly high, we can only support asymmetric semi-tragic equilibria because the greedy A pools in period one, which prevents trust-building. Asymmetric equilibria can be Pareto optimal in overlapping ranges with symmetric equilibria.

FIGURE 2. Pareto Optimal Equilibria Given Two-Sided Independence



(a) Conditions where PBE emerge

Motive	Period 1	Period 2 | Period 1 CC	CD	DC	DD
Mixed strategy trust-building (light red)					
Security	C	C/D	D	D	D
Greedy	D	D	D	D	D
Semi-tragic (random dots)					
Security	C	D	D	D	D
Greedy	D	D	D	D	D

For the trust-building (dark red, solid line) and suckers (gray, dashed line) equilibrium descriptions, see Figure 1.

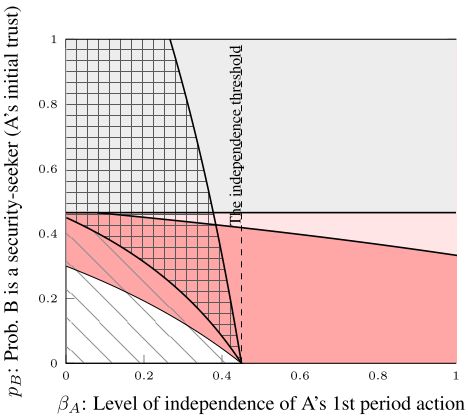
(b) Description of new equilibria

Note: As in Figure 1, we set $H = 1.5, L = -0.1, k = 0.9$ in panel (a). We additionally assume intermediate values $\beta_A = 0.6, \beta_B = 0.7$, which implies $\beta_A \sim \beta_B > \frac{k}{1+k-L}$. As in Figure 1, we shade trust-building dark red and suckers gray. Panel (b) describes the strategy profiles and color codings for the new PBE-ER that emerge. PBE-ER are unique in plotted parameter ranges

The independence threshold represents the point where the maximum payoff a security-seeker can assure herself (i.e., her minmax) comes from selecting cooperation rather than defection.[22] Substantively, consider a case where a minor power A faces the decision to increase press freedom or not. State B (say the U.S.) could exploit this decision by meddling with a newly free press. But, in practice, B's policy choice has only a small effect on A relative to the domestic implications of allowing press freedom. Thus, A's value from promoting freedom does not depend much on how B reacts. Instead, A's decision

[22] This also explains why we cannot sustain the tragic equilibrium.

FIGURE 3. When Is Trust-Building Possible Given Variation in β?



(a) Conditions where PBE emerge

Player/Motive	Period 1	Period 2 | Period 1			
		CC	CD	DC	DD
Asymmetric trust-building (diagonal, striated)					
A/Security	D	C	D	C	D
A/Greedy	D	D	D	D	D
B/Security	C	C	D	C	D
B/Greedy	D	D	D	D	D
Asymmetric semi-tragic (thatched)					
A/Security	D	D	D	D	D
A/Greedy	D	D	D	D	D
B/Security	C	D	D	D	D
B/Greedy	D	D	D	D	D

(b) Equilibrium description

Note: Continues to assume $k = 0.9, H = 1.5, L = -0.1$. Assumes B's decision is moderately independent $\beta_B = 0.7$, and B has moderate initial trust $p_A = 0.5$. Plots PBE-ER given variation in the independence of A's choice and initial trust. As in Figures 1 and 2, we shade trust-building dark red, mixed strategy light red, and suckers gray. The new thatched space is the asymmetric semi-tragic equilibrium (panel b). The new diagonal striated space is the asymmetric trust-building equilibria (panel b). Where the lines and shading intersect, there are multiple PBE-ER. Inefficient (including mixed strategy) equilibria are reported in the Supplementary Appendix.

hinges on whether A actually wants press freedom or not. In this case, we might say that A's independence threshold is met because even if A was so mistrustful of B that A was certain B would exploit A's new press freedom, it would not influence A's decision. If A was the type that truly wanted these reforms, A would implement them and simply accept that B would meddle.

The threshold's existence demonstrates that Result 1a is not overshadowed by extreme initial mistrust. Rather, as independence increases, the strategic incentives to cooperate in the first period continuously increase for low-trust types until they dominate the concerns that cause security-seekers to defect. The threshold is decreasing in the cost of being exploited (k) and the benefit a security-seeker accrues from cheating her rival (L). While this threshold can vary, Figure 2 demonstrates that, under plausible costs and benefits, we meet it for even intermediate levels of independence.

Figure 2 assumes both sides' trust-building actions meet the independence threshold. In practice, one state may have an independent trust-building opportunity, but the other may not. This could arise if A is more sensitive to exploitation than B. For example, democracies can be exposed to election rigging by promoting press freedom, but autocracies cannot. It could also arise because different states have different available trust-building options. If one state starts with free markets, it cannot use market reform to engender trust. This might leave that state with only more dependent options, like arms reductions, to signal its security-seeking preferences. Can trust-building survive when independence is lopsided? Figure 3 considers this case. The plot assumes that B's trust-building action is moderately independent and B is moderately trusting of A (β_B, p_A are moderately high). Meanwhile, it allows A's level of independence (β_A) and initial trust in B (p_B) to vary. On the furthest left side of the plot, we observe the counterfactual to the baseline, where, like in the baseline, $\beta_A = 0$. But unlike in the baseline, β_B is higher. As the plot moves to the right side, β_A increases, and the parameter ranges cross over those analyzed in Figure 2.

We note two points. First, as with prior plots, the symmetric trust-building equilibrium arises in the dark-red shaded area. Examining how this equilibrium expands as a function of β_A visualizes Result 1a in action. When A's decision is fully dependent ($\beta_A = 0$), there is a lower bound on the symmetric trust-building equilibrium. As β_A increases, this lower bound diminishes. A reaches the independence threshold when $\beta_A = 0.45$. From this level of independence onward, we can sustain symmetric trust-building at the lowest levels of initial trust.

Second, the bottom left corner of the plot reveals a surprising equilibrium:

> *Result 1c: Asymmetric trust-building.* When trust and independence are lopsided (e.g., β_A, p_B small but β_B, p_A large), we can sustain asymmetric trust-building equilibria. When the scope of lopsidedness is extreme ($\beta_A, p_B \to 0$), this equilibrium is unique.

This is a trust-building equilibrium because security-seeking A conditions second-period cooperation on B's first-period choice, and A infers B is trustworthy if B cooperates in the first period. However, it carries several interesting properties. First, B knows that A always defects in the first period. Still, if B is a security-seeker, B accepts cooperation with certain exploitation. Second, first-period exploitation is necessary to forge trust and cooperation in the second period. If B does not accept exploitation, A infers that B is greedy, and second-period mutual defection is assured.

Sustaining this equilibrium requires lopsidedness in both independence and trust. State A must have a

sufficiently dependent choice and be sufficiently distrustful that it will never cooperate in the first period. In contrast, B's choice must be sufficiently independent that security-seeking B will tolerate certain exploitation. B must also have at least moderate levels of initial trust. The reason is that B does not learn anything from A's first-period decision because A defects regardless of A's type. Thus, if B starts out too suspicious, then B is unwilling to cooperate in the second period. But B's initial trust cannot be too high either. If it is, then the suckers dynamic takes hold and greedy types play cooperation to cheat A in the second period.

So far, we have found unique PBE-ER. But Figure 3 shows some overlap between efficient asymmetric and symmetric equilibria. The reason is that while both sides do equally well in a symmetric equilibrium, all asymmetric equilibria assure one side initially exploits the other. The exploiter prefers asymmetric equilibria and the exploited prefers symmetric equilibria.

Overlapping equilibria can complicate empirical predictions. Here, they strengthen our overall claim. To summarize what we have found, when trust-building decisions are dependent (as assumed in previous work), we cannot sustain trust-building PBE if at least one side has low levels of initial trust. However, when trust-building decisions are at least moderately independent, not only can we sustain trust-building equilibria at low levels of initial trust, but only trust-building equilibria survive refinement. When both sides face independent choices, there is a clear prediction given low initial trust: symmetric trust-building. When the trust-building options available to both sides are lopsided, the result is more complicated. We often sustain both asymmetric and symmetric trust-building equilibria. But we do not find non-trust-building equilibria survive refinement.[23] Thus, either way, our model predicts some form of trust-building even in lopsided cases with low levels of initial trust.

Salience of the Trust-Building Exercise

Existing trust-building work considers variation in the salience of different international policies. In classic models, the salience of the first-period international choice must be neither too low nor too high relative to the second-period choice. Kydd argues that for any given level of initial trust, there is a first-period choice with exactly the right salience to signal reassurance without exposing the sender to too much risk (Kydd 2005). Therefore, Kydd concludes that security-seekers, in principle, can always find a right-sized signal to build trust when starting from any initial trust level.

Skeptics argue it is difficult to finely calibrate international gestures to satisfy this "Goldilocks" problem (Lieber 2011; Montgomery 2006; Rosato 2014). Goldilocks problems can be exacerbated by problems with interpretation. It is often hard to notice if a rival dismantles a few missiles given regular changes in force posture, and even harder to interpret the salience of such a gesture. These interpretation problems likely impede trust-building between enduring rivals because the calibration window is small when initial trust is low (see Yoder and Haynes 2021).

We now show that trust-building through independent (usually domestic) choices does not succumb to calibration problems. We introduce a salience parameter, $\theta > 0$, into the model. Larger values of this continuous parameter indicate that the second-period choice is increasingly important relative to the first-period choice. For example, when $\theta = 2, 1/2, 1$, it means that the second-period choice is twice as, half as, and equally important as the first-period choice, respectively. We test the robustness of the trust-building equilibrium to the inclusion of salience in Supplementary Appendix A.4. Here, we describe the primary empirical implication of that result:

> *Result 2:* If the independence threshold described in Result 1a ($\beta_i > \frac{k}{1+k-L}$) holds, then there is no Goldilocks problem because the security-seeker prefers first-period cooperation for any level of salience, even at very low levels of initial trust (i.e., there is no upper bound on the salience of first-period choices).

This finding has two substantive implications that contrast international and domestic choices. First, trust-building via domestic (or other independent) choices entails an easier calibration task than typically studied international trust-building. Given sufficient independence, no first-period decision is too salient for security-seekers to make the choice that signals their type. Therefore, rather than the baby steps toward establishing trust that we would expect to see with classic international signaling, states can potentially build trust quickly through high-stakes domestic choices. Second, domestic and international actions can work together. States can build initial trust via domestic choices when calibration problem makes international signaling initially intractable. Since domestic choices are imperfectly correlated with international preferences, they may not completely resolve the trust problem. But they may increase trust enough to make international signals, such as arms control agreements, viable.

IMPLICATIONS FOR ENDING ENDURING RIVALRY: THE COLD WAR

Our theory of trust-building is broadly applicable. But to test it, we focus on episodes that existing research struggles to explain: ending enduring rivalries, and illiberal and asymmetric trust-building. We first consider enduring rivalries. Existing research lacks a satisfying explanation for how deeply mistrustful rivals can find the right-sized gestures to initiate trust-building. We illustrate empirically how independent domestic choices help rivals overcome this problem by detailing the end of the Cold War. In the next section, we examine asymmetric trust-building via illiberal domestic choices.

[23] The only other equilibrium is substantively unappealing. Even though all types reveal their true motives in period one, mutual defection is assured in period two.

The Cold War case has three notable advantages. First, it is historically consequential. Second, successful trust-building was unexpected at the time. In 1979, trust between the West and the Soviet Union plummeted due to Moscow's invasion of Afghanistan. In the early 1980s, both the U.S. and United Kingdom elected staunch anti-Communist leaders, Ronald Reagan and Margaret Thatcher. Reagan called the Soviet Union an "evil empire" whose Communist ideology would end up on the "ash heap of history" (Garthoff 2000, 9–11). Yet the 1980s turned out to be the last years of the Cold War. In 1988, Reagan stood with his counterpart, Mikhail Gorbachev, in Red Square and declared that he no longer saw the Soviet Union as an evil empire, that such beliefs were of "another time, another era" (Miles 2020, 62). By 1991, the Cold War reached a peaceful conclusion.

Third, previous research has used the Cold War to illustrate how international gestures build trust. Most prominently, Kydd identifies the Intermediate-Range Nuclear Forces (INF) Treaty, signed in late 1987 and ratified in 1988, as the key turning point in East–West relations. The INF Treaty represented the "first important costly signal" that Moscow had security-seeking intentions (Kydd 2005, 227). In this account, the process of reassurance was initiated by INF and reinforced primarily by additional Soviet foreign policy choices. While acknowledging the importance of INF, we argue that Soviet domestic policies, such as improved treatment of dissidents, *glasnost*, and *perestroika*, kick-started the trust-building process even earlier. Analyzing this case illuminates how domestic and international policies work together, with domestic policies playing a potentially necessary preliminary role. Like Kydd, we focus on Soviet trust-building actions and Washington's trust of Moscow. The section "View from the Soviet Union" briefly considers Western trust-building actions and Soviet trust.

Following best practices in case evaluation of models (Gerring 2004), the next section argues that the initial conditions of the case match the symmetric trust-building equilibrium's parameter ranges. That is, we could not rationalize trust-building through highly dependent international choices in this case because initial trust was too low, but we can rationalize trust-building through more independent choices, notably Soviet domestic reforms. We then trace critical elements of our causal mechanism through the case (Joseph, Poznansky, and Spaniel 2022; Lorentzen, Fravel, and Paine 2017).

Mapping the Model to the Case

In the Cold War case, we code the goal of security-seekers as rapprochement. Thus, second-period mutual cooperation would involve arms control and other security and economic agreements (ending the Cold War). Defection would involve continuing or intensifying the arms race or security competition, to include reneging on arms control or other agreements. We argue that the case conditions match the parameter ranges for symmetric trust-building (Supplementary Appendix B maps the model to the case in tabular form). We code both sides as genuinely security seeking by the late Cold War, consistent with classic research on the effects of both "New Thinking" and structural constraints (Brooks and Wohlforth 2000; English 2000; Risse-Kappen 1994).

Even though both states desired rapprochement, we code trust between the U.S. (p_{USSR}) and the Soviet Union (p_{US}) as very low in the early 1980s.[24] In his first press conference as commander-in-chief, Reagan stated that the Soviets "reserve unto themselves the right to commit any crime, to lie, to cheat," to achieve their goals (Gwertzman 1981). Soviet General Secretary Leonid Brezhnev, in 1980, labeled the U.S. an "unreliable partner" (FRUS 2018, 1977–1980, Volume XII, No. 166). Even the reform-minded Gorbachev was dismayed by U.S. defense policy, claiming in 1985 that the Americans "promise the world stability but in reality strive to wreck the military balance" (Eaton 1985).

Given mutual distrust between states that genuinely desire second-period mutual cooperation, our theory predicts that a symmetric trust-building equilibrium can emerge if both rivals have sufficiently independent and salient trust-building choices to make. By contrast, if only dependent trust-building choices were available, then they would not make them. Although Washington and Moscow had discussed arms control from the early 1980s onward, the benefits had low independence (low β_i), creating an impediment to progress under low trust. In contrast, the Soviets' domestic policy choices had moderate-to-high independence (moderate-to-high β_{USSR}). The payoffs from more lenient treatment of dissidents, greater free speech, and reduced central planning of the economy (the Soviets' first-period choice in our model) were not greatly dependent on Western policies. For example, the main benefits that Soviet leaders expected from *glasnost*, a policy of greater openness and freedom of speech, were reduced corruption and increased efficiency. These benefits did not depend much on any Western responses. We also classify Soviet domestic choices as high salience. Policies such as *glasnost* and *perestroika*, which introduced some market forces into the Soviet economy, created a fundamental shift in Soviet political culture and governance, with direct implications for the Kremlin's governing ideology and political control.

Finally, domestic and international preferences were sufficiently correlated to enable learning. The Cold War had a strong ideological component, pitting a liberal, capitalist West against an illiberal Communist Bloc. Soviet domestic reforms reflected a shift in intrinsic values (English 2000). Though the Reagan Administration would have preferred the USSR to become a true democracy, even liberalizing reforms within an autocratic context indicated greater compatibility in international preferences. Overall, given the relatively high independence and salience of Soviet domestic choices, and the correlation between domestic and international preferences, our theory predicts that these choices had the characteristics necessary to launch trust-building.

[24] This matches Kydd (2005).

Initial Trust-Building through Soviet Domestic Choices

East–West trust in the early 1980s could hardly have been lower. President Reagan believed that the Soviet Union was bent on global domination. To Reagan, the Soviets had exploited the U.S. during the 1970s, opportunistically jumping out ahead in the arms race and engaging in military adventurism while misguided Western leaders sought détente. As Reagan prepared to take office, he confided in a friend that "I don't really trust the Soviets and I don't really believe that they will really join us in a legitimate limitation of arms agreement" (Wilson 2014, 17). Similarly, in 1983, Secretary of State George Shultz stated that "Soviet actions have come into conflict with many of our objectives" and lamented Moscow's penchant for "stretching a series of treaties and agreements to the brink of violation and beyond" (FRUS 2021, 1981–1988, Volume IV, Soviet Union, January 1983–March 1985, No. 61).

In 1985, Mikhail Gorbachev became General Secretary of the Communist Party of the Soviet Union. Gorbachev and his advisors desired to end the costly U.S.–Soviet competition (Brands 2014). However, the combination of low trust and low independence inherent in defense policy made it difficult to initiate trust-building through costly international actions such as arms control. Reagan administration officials were wary of arms control with the Soviets, believing that "the Soviet Union had violated every arms agreement it had ever signed" (Wilson 2014, 66). As a result, arms control proposals in the early 1980s were often more about public relations than genuine attempts to build trust (Colbourn 2022). Though the Soviets, especially by the time Gorbachev reached office, greatly desired reductions in defense spending, unilateral arms reductions would have raised security concerns and potentially damaged Soviet prestige. Therefore, arms control could not proceed.

Fortunately, domestic choices offered another means for initiating reassurance. Consistent with our theory, a shift in Soviet policy on dissidents and minorities became an early signal that the Kremlin might be a trustworthy partner for the West. A notable case involved Andrei Sakharov, the Soviet nuclear physicist and Nobel Peace Prize-winning dissident. According to Shultz, the U.S. point person on Soviet diplomacy, Sakharov's release from internal exile in 1986, "affecting a man of towering intellect and moral authority, made an impact on some of Gorbachev's most severe skeptics" (Shultz 2010, 1095). More systematic policy changes reinforced Western beliefs. Jack Matlock, the senior Soviet expert on the Reagan National Security Council (NSC) and later the U.S. ambassador in Moscow, highlighted their impact on Shultz's views:

> [T]he evolution in Shevardnadze's attitude toward human rights in the Soviet Union made *probably the most important contribution* to Shultz's feeling that the two had compatible goals. Shevardnadze had always tolerated a discussion of human rights with more courtesy than Andrei Gromyko could summon, but by 1987 he began to do more than simply arranging an exit visa once in a while. He actually began to try to change the system (Matlock 2004, 265, emphasis added).

From 1986 to 1988, the number of exit visas issued to Soviet Jews ("refuseniks") increased from 1,000 to 80,000, a shift which the CIA called "remarkable" (Brands 2014, 137). Illustrating the moderate-to-high independence of the issue, the Soviets exhibited little concern that these moves could be exploited by the West. Shevardnadze even invited Shultz to provide Moscow with a list of potential émigrés for the Soviets to consider (Shultz 2010, 986).

Gorbachev eventually pursued broader reforms, *glasnost* and *perestroika*, to increase political and economic freedom. *Glasnost* focused on transparency and openness. For example, in 1986 Soviet leaders greenlighted the release of the film *Repentance*, which critically represented Stalin, a "bombshell" that would "change our social system" (Taubman 2017, 248). By 1987, *glasnost* was "spreading like wildfire on the steppe," and had led to something closer to freedom of speech and freedom of the press (Taubman 2017, 314). Under *perestroika*, new laws legalized private enterprise (1986), allowed state enterprises to determine output on the basis of demand (1987), and permitted co-operatives (1988). As with emigration policy, the Soviets did not seem particularly concerned that *perestroika* and *glasnost* left the USSR vulnerable to exploitation by the West. On the contrary, Shultz and Gorbachev enjoyed open discussions about economic policy (Wilson 2014, 132–3).

Perestroika and *glasnost* further increased Western trust in the Soviet Union. As early as April 1986, Soviet reforms had sparked "prominent" deliberations in the U.S. government about the possibility that the Soviets were truly changing (Savranskaya, Blanton, and Zubok 2010, 116).[25] By 1987, Soviet reform had convinced senior U.S. policymakers that international cooperation was possible. During a meeting between Gorbachev and Thatcher in late 1987, journalists observed that Thatcher "placed almost as much emphasis on the Soviet leader's internal reforms as on the superpower talks, considering *perestroika* and *glasnost* as evidence of a determination which also promised progress in East-West negotiations" on arms control and other security issues (Naughtie 1987). Reagan concurred, subsequently saying that Gorbachev's book, *Perestroika*, which outlined the Soviet leader's vision, made him hopeful about Washington–Moscow relations (Matlock 2004, 294).

International-Level Trust-Building Begins

Soviet domestic reforms not only built trust, but facilitated international cooperation requiring more dependent choices, including the INF Treaty. The conclusion of the INF Treaty reflected, to be sure, important concessions by Gorbachev (e.g., including the SS-23

[25] Not everyone agreed that the Soviet Union was changing for good, but the very existence of debates is evidence of changing U.S. attitudes, given Reagan officials' initial certainty of malign Soviet intentions.

in the deal, de-linking missile defense). However, its signing was aided by the trust that Gorbachev had generated with *glasnost* and *perestroika.* During his April 1987 visit to Moscow, Shultz and Gorbachev discussed Soviet economic reform in depth, "establishing greater trust" and helping to make "the prospect of the elimination of INF a reality" (Wilson 2014, 133). To whet Reagan's appetite for a summit in Washington to sign the INF Treaty, Shultz reported from Moscow that "the Soviet Union is changing" (Leffler 2007, 399).

When Gorbachev eventually came to Washington to sign the INF Treaty in December 1987, he received a hero's welcome. Gorbachev interpreted the success of the Washington summit as evidence that his reformist domestic program changed perceptions of the Soviet Union abroad. Briefing the Politburo afterwards, Gorbachev observed:

> In Washington we saw for the first time with our own eyes what a great interest exists for everything that is happening here, for our *perestroika.* And the goodwill, even enthusiasm to some degree, with which prim Washington received us, was an indicator of the changes that have started taking place in the West. These changes evidence the beginning of the crumbling "image of the enemy," beginning of the destruction of the "Soviet military threat" myth (Savranskaya, Blanton, and Zubok 2010, 361).

Soviet domestic reform also built trust with the U.S. Senate, which would go on to ratify the INF Treaty. As the influential chairman of the Senate Armed Services Committee, Sam Nunn, observed, "the advent of Gorbachev, *glasnost*, and *perestroika* has undeniably improved the overall climate for the conduct of superpower relations" (Nunn 1988, 3).

In 1988, the Kremlin's domestic reforms provided even stronger evidence that the Soviet Union had fundamentally changed. On the eve of Reagan's May visit to Moscow, the Kremlin released a set of "theses" for the upcoming 19th Party Conference, which indicated that Gorbachev wanted to further liberalize the Soviet system. Matlock, then U.S. ambassador to Russia, was "electrified" when he read it (Savranskaya, Blanton, and Zubok 2010, 110). The following day, the ambassador told Reagan that "the Soviet Union will never be the same" (Matlock 2004, 296). Just days later, in Red Square, Reagan declared that he no longer saw the Soviet Union as an "evil empire."

The 1988 19th Party Conference proved another milestone in the Kremlin's reform program. Though falling far short of democracy, the conference resulted in political liberalization unprecedented in the Soviet context. Gorbachev secured popular elections at lower levels of government and greater judicial independence and rule of law. The conference "dealt only briefly with international and security affairs" (Garthoff 2000, 361). The lack of attention to foreign policy is notable both because it speaks to the relative independence of domestic reform and because a party conference focused on domestic policy had such a big impact on Western perceptions of the Soviet Union. The conference contributed further to Washington's appetite for cooperation with Moscow; high-level diplomacy with the Soviets thereafter "expanded rapidly" (Matlock 2004, 306). In September, Gorbachev cemented his reforms by removing several conservatives from high-ranking posts. The senior CIA Soviet analyst during the mid- to late-1980s identified these events as a critical juncture for Western perceptions of the Soviet Union:

> When I talked about 1988, it was after the 19th Party Conference, and then in the period after that, in September, when the major restructuring took place… before that happened there was still room for those who wanted to disparage the implications of the events in the USSR to make their arguments. Whether you believed it or not, they had room to argue that "that's all right, it will eventually drift back to a Brezhnev-style system." But I think that after the end of 1988, no matter what your slant, you could not very well argue that some major lines had not been crossed… You could not very well argue that it was just talk and political rhetoric (Savranskaya, Blanton, and Zubok 2010, 116–7).

All told, domestic reforms undertaken by the Kremlin from 1986 to 1988 played a crucial role in creating an atmosphere in which international cooperation was easier to sustain. In addition to the INF Treaty, the U.S. and the Soviet Union concluded a number of smaller but meaningful agreements on issues as varied as monitoring nuclear tests, peaceful nuclear energy, fishing rights, space exploration, cultural and educational exchanges, and maritime navigation (Garthoff 2000, 353). Although U.S.–Soviet rapprochement slowed during the transition between the Reagan and Bush presidencies, Bush declared in fall 1989 that "[t]he world will be a better place if *perestroika* succeeds" and laid out more than a dozen proposals to increase U.S.–Soviet cooperation, including lifting trade restrictions and supporting Soviet efforts to join the General Agreement on Tariffs and Trade (GATT) as an observer (Engel 2017, 297). Soviet policymakers saw this as "the end of economic warfare" between the two states (Engel 2017, 298). U.S.–Soviet cooperation continued to bear fruit even as the Soviet Union disintegrated and revolution swept through Eastern Europe. The U.S. and the Soviet Union concluded the Treaty on Conventional Forces in Europe and the Strategic Arms Reduction Treaty (START), and the Kremlin further demonstrated its benign intentions through domestic choices by showing restraint toward separatists in the Baltics.

Overall, given the independence and high salience of Soviet domestic choices, our theory predicts that trust-building should have been possible despite low initial trust and that Soviet domestic choices should have had a substantial impact on Western perceptions. This is what we observe. As shown above, domestic reform was perhaps the earliest indicator of changing Soviet intentions. Moreover, domestic policies, including Soviet treatment of dissidents, *glasnost*, and *perestroika*, were seemingly as or more important than foreign policy in convincing key Western officials that the two sides had compatible goals. By increasing trust, Soviet domestic reforms thus contributed to an unprecedented level of East–West cooperation.

View from the Soviet Union

On the other side, how did U.S. and Western policies contribute to Soviet trust? Here, we briefly highlight one important choice: the Western response to Soviet economic and political reforms.[26] Gorbachev came to power in 1985 with security-seeking intentions, but nonetheless worried that leaders like Reagan were too stuck in the Cold War mindset to be partners in peace.[27] When Gorbachev initiated his domestic reforms to revitalize the Soviet Union, the West might have wholly dismissed these efforts. Instead, Western governments largely encouraged these reforms, which reassured the Soviets of their trustworthy intentions.

We classify this Western choice as moderately independent (β_{US}). The choice to respond positively to Gorbachev would bring more benefits if Gorbachev followed through on reforms. However, even if he had not, the U.S. and its partners would still have gained important benefits. They would have seemed magnanimous on the world stage, demonstrating that they were not implacably hostile to the Eastern Bloc. In addition, anti-Communist leaders like Reagan and Thatcher would, in particular, benefit from showing they were not recklessly hawkish (Goldfien 2025; Schultz 2005). Indeed, public approval for Reagan's Soviet policy surged when he pivoted to a more moderate stance (Nincic 1988). Therefore, this choice—while more dependent than Soviet domestic reforms—was still independent enough that Western leaders like Reagan and Thatcher were willing to make it even in the context of substantial distrust. We also classify the choice as moderately salient (θ_{US}) because it concerned the West's orientation toward its primary geopolitical rival and affected the political standing of Western leaders. Since this was an international choice, the Soviet leadership could also be confident that it correlated strongly with future willingness to cooperate internationally.

Ultimately, the choice to encourage, rather than dismiss, Soviet domestic reforms succeeded in reassuring the Kremlin that the U.S., the U.K., and other Western countries could be partners in international cooperation. For example, as Gorbachev considered further liberalization of the Soviet economy in early 1987, George Shultz—who had earned a PhD in Economics from the Massachusetts Institute of Technology and run the Bechtel engineering company—offered counsel. According to one historian, "Distinctions between Milton Friedman and John Kenneth Galbraith were less important than an American secretary of state's message of two states confronting common challenges, aspiring for common results, and establishing greater trust. Gorbachev and those around him began to sense that the most conservative and anticommunist presidential administration of the Cold War was actually out to help them," not out to get them (Wilson 2014, 133).

The West's other arch-capitalist leader, Margaret Thatcher, also supported Gorbachev's reforms (Brown 2020). Following a visit to Moscow in early 1987, Thatcher wrote to Reagan, "I am firmly convinced it is in our interest to encourage him [Gorbachev], especially in his endeavours to create a much more open society."[28] This attitude also contributed to Gorbachev's confidence in Western intentions. Writing about Thatcher's contributions to the end of the Cold War, Gorbachev reflected that she "was genuinely interested in what was happening in our country. She closely, and with astonishing command of detail, followed perestroika and glasnost, and sincerely wished for our process of change to succeed." This attitude led Gorbachev to assess that Thatcher—echoing the British prime minister's famous judgment of him—was a "person one can deal with" (Gorbachev 2013).

Our analysis of the U.S.–Soviet case shows that it is possible for domestic and international choices to work in tandem to build trust, even as the role of domestic choices is particularly crucial. It also offers a rejoinder to accounts that critique the U.S. for being too slow to build trust and cooperate with Gorbachev (e.g., Braumoeller 2013). Our concept of independence explains why trust-building likely required starting with reforms in the Soviet Union rather than ambitious arms control agreements. Yet we show that the U.S. was not simply passive before the INF Treaty was signed in late 1987. Rather, U.S. and Western encouragement of Gorbachev's domestic reforms was a meaningful and conscious policy choice that effectively reassured the Soviet leader and promoted further East–West cooperation.

ASYMMETRIC AND ILLIBERAL TRUST-BUILDING: U.S.–SOUTH KOREA RELATIONS FOLLOWING THE 1961 "MILITARY REVOLUTION"

We now use our model to illuminate patron–client relations, analyzing trust-building between the U.S. and South Korea following Park Chung-hee's May 1961 coup, after which the new government in Seoul reassured Washington of its pro-Western alignment via a crackdown on domestic communists and an authoritarian modernization program. This case differs from the U.S.–Soviet case in two main ways. First, whereas the U.S.–Soviet case fit the parameter ranges for symmetric trust-building, the initial conditions in this case support the asymmetric trust-building equilibrium. While the new South Korean regime had little doubt about the U.S. commitment to supporting anti-Communist governments, Washington was uncertain whether the newly established regime would tilt toward capitalism or Communism. Second, whereas the U.S.–Soviet case featured reassurance through liberalization, the U.S.–South Korea case illustrates the potential for illiberal policies to reassure.

[26] This was an international choice, which we argued above can sometimes be sufficiently independent to support trust-building under low initial trust. This highlights that, ultimately, independence rather than domestic policy is key.

[27] Gorbachev initially saw Reagan as "stubborn" and a "dinosaur" (Matlock 2004, 169).

[28] See "Letter to Reagan from Thatcher About Her Meetings with Gorbachev in Moscow. April 1, 1987," accessed via the National Security Archive, https://nsarchive.gwu.edu/document/21546-document-09.

While focusing only on the South Korea case, we conjecture that authoritarian client states frequently reassured the U.S. of their reliability as Cold War allies with illiberal domestic policies. As historians such as David Schmitz have highlighted, American policymakers often saw right-wing dictatorships that would keep leftists in check as an attainable outcome far preferable to Communist rule (see, e.g., Schmitz 1999; 2006).

Mapping the Model to the Korea Case

When a group of military officers ousted the democratically elected government of South Korea in early 1961, it threw U.S.–South Korea relations into uncertainty. To establish the potential for trust-building to resolve this uncertainty, we begin by mapping the situation to our model (see Supplementary Appendix B for a more formal mapping of how the case corresponds with the equilibrium plotted in the bottom-left of Figure 3). Since the U.S. was staunchly anti-Communist, we define security-seeking intentions as a shared anti-Communist orientation. Therefore, the goal for cooperation among security-seekers is blunting the spread of Communism.

We code U.S. trust in the new South Korean regime as low (low p_{ROK}). Even before the coup, Washington had worried about societal malaise and Communist influence in South Korea. The coup plotters were not immediately seen as preferable to the deposed government of Prime Minister Chang Myon (Brazinsky 2009). Indeed, their intentions were questioned. The coup leader, Park Chung-hee, had prior Communist ties, leading one U.S. intelligence estimate to comment that "we cannot rule out the possibility that [Park] is a long-term Communist agent, or that he might redefect" (FRUS 1996, 1961–1963, Volume XXII, Northeast Asia, No. 224). However, trust was lopsided. Park's military government believed that it could trust in U.S. support provided that it was seen as anti-Communist and committed to real reform and modernization (high p_{US}), even as it understood that the U.S. did not yet trust it (Kennedy 1988, 214).

The independence of the choices available to the two sides was also lopsided. The most meaningful way the U.S. could signal its willingness to cooperate was to give diplomatic recognition and aid to the new government. This choice was highly dependent (low β_{US}) because it would be embarrassing for the U.S. and a waste of resources if Seoul ultimately joined the Communist camp. In contrast, the main choices that the new Korean government had available to signal its anti-Communist credentials were domestic. Because Park's regime would reap the domestic rewards of these policies regardless of what the U.S. did, they can be considered to have high independence (high β_{ROK}). For both sides, the choices available were quite salient and likely to be strongly correlated with a future desire for anti-Communist cooperation.

Given the lopsided nature of both trust and the independence of choices, our theory predicts that an asymmetric trust-building equilibrium would be most likely to emerge in this case. Specifically, we would expect Seoul to cooperate by pursuing anti-Communist domestic policies, while Washington defected by withholding recognition and aid in the first round. This initial asymmetric cooperation would build enough trust to facilitate mutual cooperation in the subsequent round.

Asymmetric Trust-Building and International Cooperation

Consistent with expectations, an asymmetric trust-building period ensued after the coup, in which the U.S. withheld official recognition and support from the new regime, while the Park government engaged in domestic reforms that proved appealing to Washington. The U.S. understood that recognition and support for the new military government would confer legitimacy, the benefits of which would be highly dependent on the character of the new regime. Thus, American officials in Korea quickly disavowed any support for the coup (FRUS 1996, 1961–63, Volume XXII, Northeast Asia, No. 213), and Washington adopted a "cautious attitude of wait-and-see" (FRUS 1996, 1961–1963, Volume XXII, Northeast Asia, No. 216).

For the Park regime, anti-Communism at home and reforms aimed at developing the economy and rooting out corruption were generally independent. The benefits of these reforms largely derived from the Park government's own values. Though the Park government sought American support, it also had strong views of its own about how South Korea should be organized (Brazinsky 2009). Park and his government undertook anti-Communist policies and swiftly, if undemocratically, implemented modernizing reforms across South Korea's economy and society.

Park's early moves greatly encouraged Washington. In a telegram from the U.S. embassy in Seoul in October 1961, the reassurance felt by American officials is palpable. Ambassador Samuel Berger wrote that the regime "has taken hold with energy, earnestness, determination and imagination, albeit with certain authoritarian and military characteristics" (FRUS 1996, 1961–1963, Volume XXII, Northeast Asia, No. 244). Further, "vigilance against communist subversion and quality and volume of anti-communist propaganda have greatly improved." Berger concluded that the new regime "offers much hope" (FRUS 1996, 1961–1963, Volume XXII, Northeast Asia, No. 244). While it is unclear whether the U.S. ultimately preferred the Park government to its predecessor, it is clear that Park's actions increased U.S. trust relative to the beginning of his tenure.

Efforts to improve economic performance—even by nondemocratic means—and a demonstrated anti-Communist orientation not only built trust with the U.S. but facilitated mutual cooperation going forward. The positive impression that the military government had made early in its tenure led the Kennedy administration to host Park at the White House in late 1961, which "played a significant role in stabilizing and legitimizing the South Korean regime" (Brazinsky 2009, 120–1). Notably, this is the sort of high-dependence action that can be hard to contemplate under low trust. Looking ahead, the U.S. could—despite the Park

regime's "shortcomings," that is, authoritarian style—expect that American aid would "be more effectively used than by any previous government" and that the administration could "go to Congress this spring in good conscience… that our continuing massive support is well justified" (FRUS 1996, 1961–1963, Volume XXII, Northeast Asia, No. 244). For its part, the Park regime cooperated with the U.S. on the international stage by, for example, hosting U.S. troops on its soil and sending South Korean soldiers to support the American war effort in Vietnam. Thus, initial asymmetric trust-building led to eventual mutual cooperation.

CONCLUSION

We argued that domestic policy choices can operate as costly signals that engender international trust, and even hold a key advantage over international choices that make them vital for trust-building: payoff independence. Independence means that the value a state accrues from domestic choices depends mainly on that state's true motivations, and less on how its counterpart responds. We focus on situations where initial trust is so low that security-seekers are unwilling to signal their motivations via international choices because they fear exploitation. Given this mistrust, even moderately independent domestic actions can facilitate trust-building that would otherwise be impossible. Thus, the domestic reforms that often occur before rivals achieve a rapprochement may not be a coincidence. Rather, they may be an important, potentially necessary, step to increase trust such that they are willing to engage in the international trust-building activities that others have studied (Kydd 2005; Yoder and Haynes 2021).

This article makes several contributions. First, it contributes to our understanding of domestic politics and signaling in international relations (Fordham 1998; Goldfien, Joseph, and McManus 2023; McManus 2017; Renshon, Yarhi-Milo, and Kertzer 2023; Schultz 1999; Weeks 2008). Second, it offers a new way of understanding the democratic peace and suggests the possibility of trust-building among a wider variety of regimes, based on domestic policy changes that fall short of regime change. Third, it expands our understanding of the set of policy choices that are relevant to international relations. Most importantly, it offers a novel solution to the problem that the trust-building literature has wrestled with for decades: how countries with high levels of distrust can engage in initial trust-building activities without exposing themselves to too much risk (Glaser 2010; Jervis 1978; Kydd 2005). Our empirical analysis provides new insight into reassurance at the end of the Cold War, offering a clearer explanation for the linkage between reforms within the Soviet Union and rapprochement abroad. It also illustrates the potential for asymmetric trust-building and how even illiberal actions, such as suppressing left-wing groups, can increase trust in some circumstances, a puzzle that is not explained by the democratic peace approach.

Our findings also speak to policy issues, including evolving Sino-American competition. Many have argued that the U.S. and China have entered a period of rivalry or even a Cold War (Bekkevold 2022; Daly 2022; Frendem, Joseph, and Spaniel 2025; Joseph 2026; Sanger 2021). This is concerning because major power rivalries are often long and very costly (Thompson 2001). Tragically, it is hard to find a path back to peace even if both sides tire of competing. We identify domestic policy reforms as a mechanism the U.S. and China may eventually use to kick-start trust-building.

We also clarify how the U.S. can identify which states are trustworthy international partners. As rising populism brings new forms of government to power and political polarization calls into question the intentions and abilities of democracies (Joseph, Chung, and Park 2026; Myrick 2021), the assumption that the U.S. should trust states that ascribe to liberal values and mistrust those that do not may generate both misplaced trust in democracies and misplaced competition with cooperative autocratic regimes. When considering both formal and informal alliances (Kenwick and McManus 2021; McManus and Nieman 2019), our theory explains how the U.S. can make a fine-grained analysis of the international reliability of new regimes based on their domestic policies. We recommend using this approach to consider arms sales, diplomatic recognition, and military support for middle powers, which could be risky as competition with China intensifies.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://doi.org/10.1017/S0003055426101713.

ACKNOWLEDGEMENTS

The opinions and views expressed in this article are those of the individual authors and not necessarily those of the U.S. Government, the U.S. Navy, or the U.S. Naval War College. The authors thank the APSR Editors, our anonymous reviewers, technical reviewer Hongding Zhu, Andrew Kydd, Soyoung Lee, Brandon Yoder, Terence Roehrig, Tyler Pratt, Charles Glaser, and participants at ISA, Peace Science, and the Penn State-Pitt IR Workshop for comments that improved the article. The authors are listed alphabetically and contributed equally. Each reserves the right to describe themselves as the first author.

FUNDING STATEMENT

This research was funded by the United States National Science Foundation, grant numbers 2342950 and 2342951.

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors affirm this research did not involve human participants.

REFERENCES

Acharya, Avidit, and Kristopher W. Ramsay. 2013. "The Calculus of the Security Dilemma." *Quarterly Journal of Political Science* 8 (2): 183–203.

Ashworth, Scott, Christopher Berry, and Ethan Bueno de Mesquita. 2021. *Theory and Credibility*. Princeton, NJ: Princeton.

Bartel, Fritz. 2022. *The Triumph of Broken Promises: The End of the Cold War and the Rise of Neoliberalism*. Cambridge, MA: Harvard University Press.

Bas, Muhammet A., and Andrew J. Coe. 2016. "A Dynamic Theory of Nuclear Proliferation and Preventive War." *International Organization* 70 (4): 655–85.

Bekkevold, Jo Inge. 2022. "5 Ways the U.S.-China Cold War Will Be Different from the Last One." *Foreign Policy*, December 29. https://foreignpolicy.com/2022/12/29/us-china-cold-war-bipolar-global-order-stability-biden-xi/.

Berenji, Shahin. 2020. "Sadat and the Road to Jerusalem: Bold Gestures and Risk Acceptance in the Search for Peace." *International Security* 45 (1): 127–63.

Biden, Joseph. 2024. "President Biden Delivers Remarks on the Latest Developments in Syria." https://www.youtube.com/watch?v=sVzK0TjeIIM.

Bokobza, Laure, Suthan Krishnarajan, Jacob Nyrup, Casper Sakstrup, and Lasse Aaskoven. 2022. "The Morning after: Cabinet Instability and the Purging of Ministers after Failed Coup Attempts in Autocracies." *Journal of Politics* 84 (3): 1437–52.

Brands, Hal. 2014. *What Good Is Grand Strategy?* Ithaca, NY: Cornell University Press.

Braumoeller, Bear F. 2008. "Systemic Politics and the Origins of Great Power Conflict." *American Political Science Review* 102 (1): 77–93.

Braumoeller, Bear F. 2013. *The Great Powers and the International System*. New York: Cambridge University Press.

Brazinsky, Gregg A. 2009. *Nation Building in South Korea: Koreans, Americans, and the Making of a Democracy*. Chapel Hill: University of North Carolina Press.

Brooks, Stephen G., and William C. Wohlforth. 2000. "Power, Globalization, and the End of the Cold War: Reevaluating a Landmark Case for Ideas." *International Security* 25 (3): 5–53.

Brown, Archie. 2020. *The Human Factor: Gorbachev, Reagan, and Thatcher, and the End of the Cold War*. Oxford: Oxford University Press.

Coe, Andrew J., and Jane Vaynman. 2020. "Why Arms Control Is So Rare." *American Political Science Review* 114 (2): 342–55.

Colaresi, Michael. 2004. "When Doves Cry: International Rivalry, Unreciprocated Cooperation, and Leadership Turnover." *American Journal of Political Science* 48 (3): 555–70.

Colbourn, Susan. 2022. *Euromissiles: The Nuclear Weapons that Nearly Destroyed NATO*. Ithaca, NY: Cornell University Press.

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

Daly, Robert. 2022. "China and the United States: It's a Cold War, but Don't Panic." *Bulletin of the Atomic Scientists* 78(2): 59–64.

Debs, Alexandre, and Nuno P. Monteiro. 2014. "Known Unknowns: Power Shifts, Uncertainty, and War." *International Organization* 68 (1): 1–31.

Eaton, William. 1985. "Gorbachev Agrees to Summit Talks: Stops Deploying Soviet Missiles Until November." *Los Angeles Times*, April 8. https://www.latimes.com/archives/la-xpm-1985-04-08-mn-18521-story.html.

Engel, Jeffrey A. 2017. *When the World Seemed New: George HW Bush and the End of the Cold War*. Boston, MA: Houghton Mifflin Harcourt.

English, Robert. 2000. *Russia and the Idea of the West: Gorbachev, Intellectuals, and the End of the Cold War*. New York: Columbia University Press.

Flores-Macías, Gustavo A., ed. 2019. *The Political Economy of Taxation in Latin America*. New York: Cambridge University Press.

Fordham, Benjamin. 1998. "The Politics of Threat Perception and the Use of Force: A Political Economy Model of U.S. Uses of Force, 1949–1994." *International Studies Quarterly* 42 (3): 567–90.

Frendem, Mathias, Michael Joseph, and William Spaniel. 2025. "Explaining Peace during Long and Rapid Power Shifts: A Theory of Grand Bargains." *British Journal of Political Science* 55: e47.

FRUS. 1996. "Foreign Relations of the United States." Office of the Historian, US Department of State. https://history.state.gov/historicaldocuments. Accessed May 16, 2024.

FRUS. 2018. "Foreign Relations of the United States." Office of the Historian, US Department of State. https://history.state.gov/historicaldocuments. Accessed May 16, 2024.

FRUS. 2021. "Foreign Relations of the United States." Office of the Historian, US Department of State. https://history.state.gov/historicaldocuments.

Garthoff, Raymond L. 2000. *The Great Transition*. Washington, DC: Brookings Institution Press.

Gartzke, Erik 1998. "Kant We All Just Get Along? Opportunity, Willingness, and the Origins of the Democratic Peace." *American Journal of Political Science* 42 (1): 1–27.

Gerring, John. 2004 "What Is a Case Study and What Is it Good For?." *American Political Science Review* 98 (2): 341–54.

Glaser, Charles L. 2010. *Rational Theory of International Politics*. Princeton, NJ: Princeton University Press.

Glaser, Charles L. 2015. "A US-China Grand Bargain? The Hard Choice between Military Competition and Accommodation." *International Security* 39 (4): 49–90.

Glaser, Charles L., and Chaim Kaufmann. 1998. "What Is the Offense-Defense Balance and Can We Measure It?" *International Security* 22 (4): 44–82.

Goldfien, Michael, Michael Joseph, and Daniel Krcmaric. 2024. "When Do Leader Backgrounds Matter? Evidence from the President's Daily Brief." *Conflict Management and Peace Science* 41 (4): 414–37.

Goldfien, Michael A. 2024. "Just Patronage? Familiarity and the Diplomatic Value of Non-Career Ambassadors." *Journal of Conflict Resolution* 68 (7–8): 1417–42.

Goldfien, Michael A. 2025. "To Agree or Not to Agree: Hawks, Doves, and Regime Type in International Rivalry and Rapprochement." *International Security* 50 (2): 162–92.

Goldfien, Michael A., and Michael F. Joseph. 2023. "Perceptions of Leadership Importance: Evidence from the CIA'S President's Daily Brief." *Security Studies* 32 (2): 205–38.

Goldfien, Michael A., Michael F. Joseph, and Roseanne W. McManus. 2023. "The Domestic Sources of International Reputation." *American Political Science Review* 117 (2): 609–28.

Gorbachev, Mikhail. 2013. "Mikhail Gorbachev: The Margaret Thatcher I Knew." *The Guardian*, April 8. https://www.theguardian.com/politics/2013/apr/08/mikhail-gorbachev-margaret-thatcher-death.

Gwertzman, Bernard. 1981. "President Sharply Assails Kremlin; Haig Warning on Poland Disclosed." *The New York Times*, January 30. https://www.nytimes.com/1981/01/30/world/president-sharply-assails-kremlin-haig-warning-on-poland-disclosed.html.

Haynes, Kyle, and Brandon K. Yoder. 2020. "Offsetting Uncertainty: Reassurance with Two-Sided Incomplete Information." *American Journal of Political Science* 64 (1): 38–51.

Jervis, Robert. 1978. "Cooperation under the Security Dilemma." *World Politics* 30 (2): 167–214.

Jervis, Robert. 1989. *The Logic of Images in International Relations*. New York: Columbia University Press.

Joseph, Michael F. 2021. "A Little Bit of Cheap Talk Is a Dangerous Thing: States Can Communicate Intentions Persuasively and Raise the Risk of War." *Journal of Politics* 83 (1): 166–81.

Joseph, Michael F. 2026. *The Origins of Great Power Rivalries: A Rational Theory of Principled Motivations, and Historical Context*. New York: Cambridge University Press.

Joseph, Michael F., Joon H. Chung, and Hui Seong Park. 2026. "Elite Partisan Disagreement and Military Victory: Evidence from South Korean Battle Experiments." *American Political Science Review*. https://doi.org/10.1017/S0003055426101543.

Joseph, Michael F., and Michael Poznansky. 2024. "Secret Innovation." *International Organization* 78 (4): 766–99.

Joseph, Michael F., Michael Poznansky, and William Spaniel. 2022. "Shooting the Messenger: The Challenge of National Security Whistleblowing." *Journal of Politics* 84 (2): 846–60.

Kennedy, Charles. 1988. "Oral History Interview with Marshall Green." Association for Diplomatic Studies and Training, Foreign Affairs Oral History Project. https://adst.org/OH.

Kenwick, Michael R., and Roseanne W. McManus. 2021. "Deterrence Theory and Alliance Politics." In *What Do We Know*

about War? 3rd edition, eds. Sara McLaughlin Mitchell and John A. Vasquez, 41–62. Lanham, MD: Rowman and Littlefield.
Kydd, Andrew. 1997. "Sheep in Sheep's Clothing: Why Security Seekers Do Not Fight Each Other." *Security Studies* 7 (1): 114–55.
Kydd, Andrew. 2000. "Trust, Reassurance, and Cooperation." *International Organization* 54 (2): 325–57.
Kydd, Andrew H. 2005. *Trust and Mistrust in International Relations*. Princeton, NJ: Princeton University Press.
Leffler, Melvyn P. 2007. *For the Soul of Mankind: The United States, the Soviet Union, and the Cold War*. New York: Macmillan.
Levin, Dov H. 2021. *Meddling in the Ballot Box : The Causes and Effects of Partisan Electoral Interventions*. New York: Oxford University Press.
Lieber, Keir A. 2011. "Mission Impossible: Measuring the Offense-Defense Balance with Military Net Assessment." *Security Studies* 20 (3): 456–9.
Lindsey, David, and William Hobbs. 2015. "Presidential Effort and International Outcomes: Evidence for an Executive Bottleneck." *Journal of Politics* 77 (4): 1089–102.
Lorentzen, Peter, M. Taylor Fravel, and Jack Paine. 2017. "Qualitative Investigation of Theoretical Models: The Value of Process Tracing." *Journal of Theoretical Politics* 29 (3): 467–91.
Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace." *American Political Science Review* 87 (3): 624–38.
Matlock, Jack. 2004. *Reagan and Gorbachev: How the Cold War Ended*. New York: Random House.
McManus, Roseanne W. 2017. *Statements of Resolve: Achieving Coercive Credibility in International Conflict*. New York: Cambridge University Press.
McManus, Roseanne W., and Mark David Nieman. 2019. "Identifying the Level of Major Power Support Signaled for Protégés: A Latent Measure Approach." *Journal of Peace Research* 56 (3): 364–78.
Miles, Simon. 2020. *Engaging the Evil Empire*. Ithaca, NY: Cornell University Press.
Montgomery, Evan Braden. 2006. "Breaking out of the Security Dilemma: Realism, Reassurance, and the Problem of Uncertainty." *International Security* 31 (2): 151–85.
Myrick, Rachel. 2021. "Do External Threats Unite or Divide? Security Crises, Rivalries, and Polarization in American Foreign Policy." *International Organization* 75 (4): 921–58.
Naughtie, James. 1987. "Thatcher, Gorbachev Hopeful on Arms." *The Guardian.* December 13.
Nincic, Miroslav. 1988. "The United States, the Soviet Union, and the Politics of Opposites." *World Politics* 40 (4): 452–75.
Nunn, Sam. 1988. "Arms Control in the Last Year of the Reagan Administration." *Arms Control Today (United States)* 18 (2): 3–7.
Peters, Margaret E. 2015. "Open Trade, Closed Borders: Immigration in the Era of Globalization." *World Politics* 67 (1): 114–54.
Radchenko, Sergey. 2024. *To Run the World: The Kremlin's Cold War Bid for Global Power*. New York: Cambridge University Press.
Renshon, Jonathan, Allan Dafoe, and Paul Huth. 2018. "Leader Influence and Reputation Formation in World Politics." *American Journal of Political Science* 62 (2): 325–39.
Renshon, Jonathan, Keren Yarhi-Milo, and Joshua D. Kertzer. 2023. "Democratic Reputations in Crises and War." *Journal of Politics* 85 (1): 1–18.
Risse-Kappen, Thomas. 1994. "Ideas Do Not Float Freely: Transnational Coalitions, Domestic Structures, and the End of the Cold War." *International Organization* 48 (2): 185–214.
Rosato, Sebastian. 2014. "The Inscrutable Intentions of Great Powers." *International Security* 39 (3): 48–88.
Russett, Bruce, and John R. Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organizations*. New York: W Norton and Company.
Sanger, David E. 2021. "Washington Hears Echoes of the '50s and Worries: Is This a Cold War with China?" *New York Times*, October 17. https://www.nytimes.com/2021/10/17/us/politics/china-new-cold-war.html.
Savranskaya, Svetlana, Thomas S. Blanton, and Vladislav Martinovich Zubok. 2010. *Masterpieces of History: The Peaceful End of the Cold War in Eastern Europe, 1989*. Budapest, Hungary: Central European University Press.
Schmitz, David F. 1999. *Thank God They're on Our Side: The United States and Right-Wing Dictatorships, 1921–65*. Chapel Hill: University of North Carolina Press.
Schmitz, David F. 2006. *The United States and Right-Wing Dictatorships, 1965–1989*. New York: Cambridge University Press.
Schultz, Kenneth A. 1999. "Do Democratic Institutions Constrain or Inform? Contrasting Two Institutional Perspectives on Democracy and War." *International Organization* 53 (2): 233–66.
Schultz, Kenneth A. 2005. "The Politics of Risking Peace: Do Hawks or Doves Deliver the Olive Branch?" *International Organization* 59 (1): 1–38.
Shultz, George P. 2010. *Turmoil and Triumph: Diplomacy, Power, and the Victory of the American Deal*. New York: Simon & Schuster.
Spaniel, William, and Bradley C. Smith. 2015. "Sanctions, Uncertainty, and Leader Tenure." *International Studies Quarterly* 59 (4): 735–49.
Taubman, William. 2017. *Gorbachev: His Life and Times*. New York: Simon & Schuster.
Thompson, William R. 2001. "Identifying Rivals and Rivalries in World Politics." *International Studies Quarterly* 45 (4): 557–86.
Tomz, Michael R., and L. P. Weeks Jessica. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107 (4): 849–65.
Tomz, Michael R., and L. P. Weeks Jessica. 2020. "Human Rights and Public Support for War." *Journal of Politics* 82 (1): 182–94.
Voeten, Erik. 2021. *Ideology and International Institutions*. Princeton, NJ: Princeton University Press.
Waltz, Kenneth N. 1979. *Theory of International Politics*. Reading, MA: McGraw-Hill.
Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62 (1): 35–64.
Wilson, James. 2014. *The Triumph of Improvisation: Gorbachev's Adaptability, Reagan's Engagement, and the End of the Cold War*. Ithaca, NY: Cornell University Press.
Wolford, Scott. 2007. "The Turnover Trap: New Leaders, Reputation, and International Conflict." *American Journal of Political Science* 51 (4): 772–88.
Yoder, Brandon K. 2019a. "Hedging for Better Bets: Power Shifts, Credible Signals, and Preventive Conflict." *Journal of Conflict Resolution* 64 (3): 923–49.
Yoder, Brandon K. 2019b. "Retrenchment as a Screening Mechanism: Power Shifts, Strategic Withdrawal, and Credible Signals." *American Journal of Political Science* 63 (1): 130–45.
Yoder, Brandon K., and Kyle Haynes. 2021. "Signaling Under the Security Dilemma: An Experimental Analysis." *Journal of Conflict Resolution* 65 (4): 672–700.
Yoder, Brandon K., and Kyle Haynes. 2025. "Endogenous Preferences, Credible Signaling, and the Security Dilemma: Bridging the Rationalist–Constructivist Divide." *American Journal of Political Science* 69 (1): 268–83.