Appendix: The Domestic Sources of International Trust

A Formal Analysis

We now report the formal model that supports our theory. First, we set up a complete model of trust that includes all the parameters that interest us (salience and independence). We use this general setup to develop a common definition of trust-building equilibria. We also use it to characterize a set of common on-path strategies that can appear in the second period. This broad overview lays the foundation for the equilibrium analysis.

Second, we restrict our attention to the core model commonly studied, and solve for all the equilibria of the model. Third, we introduce independence. Fourth, we introduce salience. Finally, we consider a robustness model that considers how similar domestic and international issues are.

A.1 Set up.

We study two sequential trust games between two players, A, B. Where appropriate, we notate period-relevant variables as $t \in \{1, 2\}$ and player-relevant variables as $i \in \{A, B\}$. Because the model is symmetric, we sometimes we refer to player j meaning not player i.

Players can hold one of two motivations $m \in \{s, g\}$. Where $m = g$ is greedy and $m = s$ is security-seeking. We sometimes notate i_m meaning player i has motive m. We will discuss payoffs in more detail in a moment, but motivations determine the value states accrue from exploiting the other side. The greedy type gets a high value from exploitation $e_i^g = H$, and the security seeker gets a low value from exploitation $e_i^s = L$ relative to the value of mutual cooperation (we normalize the value of mutual cooperation to 1). Define $p_i = pr(m = s \implies e_i = L)$ as the probability player i is a security-seeker, and $1 - p_i = pr(m = g \implies e_i = H)$ as the probability i is greedy.

The sequence of moves is as follows.

- Nature draws player types i.i.d.,
- Period 1: players simultaneously choose between cooperation and defection.
- Players observe the result of Period 1.
- Period 2: players simultaneously choose between cooperation and defection.
- Payoffs are realized.

A strategy for A is $s^A(a_t)$ where $a_t \in \{c, d\}$. A strategy for B is $s^B(b_t), b_t \in \{c, d\}$. A's belief at the beginning of each period about the likelihood that B is a security-seeker is σ_t^A and B's belief is σ_t^B. Trivially, $\sigma_1^i = p_j$.

For ease, we reproduce the table of payoffs here.

Table A.1: Replication of payoff Table 1 (second period)

		Player B: Cooperate ($b_t = c$)	Player B: 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 $0, 0$

Table A.2: Replication of Payoff Table 4 (First-Period Payoffs w/ Independence)

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

Each player's payoff depends on their type, their choice, and the choice that the other side makes. Second-period payoffs are θ times the values represented in Table 1. We make two substantively motivated assumptions to capture the commonly studied trust problem. First, all types prefer mutual defection to being cheated (i.e, k is the cost of being cheated):

$$\mathcal{A}_1 \qquad k > 0$$

Second, the greedy type prefers to exploit a rival over mutual cooperation, however the security seeker prefers mutual cooperation over exploitation

$$\mathcal{A}_2 \qquad H > 1 > L$$

The first-period payoffs are represented in Table 4. Note the values in Tables 1 and 4 converge when $\beta_i = 0$.

To understand the role of the dependence parameter, we detail some payoffs. Assume B is a security-seeker, then the following details B's expected utilities from the declared strategies $EU_1^{Bs}|s^B(c,d), s^A(d,d) = \beta_B - (1-\beta_B)k + 0 \cdot \theta$. Here, B is suckered in the first period. However, B does not get $-k$. Instead, B accrues $\beta_B - (1-\beta_B)k$ because β_B is how independent B's action is. In the second period, B expects the double defection payoff (0), weighted by the second-period salience θ.

Note that this set-up is sufficiently flexible to capture the baseline model that represents standard models of international signaling ($\beta_i = 0, \theta = 1$); our novel introduction of payoff dependencies ($\beta_i \in [0,1], \theta = 1$); and a robustness check to our theory that explores relative salience ($\beta_i \in [0,1], \theta > 0$).

We will solve for a Perfect Bayesian Equilibrium (PBE).

Given the structure of the game, off-path beliefs can emerge if and only if we conjecture strategy profiles where all types pool in the first period. We restrict off-path beliefs as follows.

Definition: Feasible off-path beliefs In any equilibrium strategy profile with an on-path pooling first period action $i_1 = c$, restrict off-path beliefs $\sigma_2^j | i_1 = d \leq p_i$. In any equilibrium strategy

profile with an on-path pooling first period action $i_1 = d$, restrict off-path beliefs $\sigma_2^j | i_1 = c \geq p_i$.

This restriction states that if all types of player i cooperate on path in the first period, then if i deviates to defect, j cannot increase her confidence that i is a security seeker. Similarly, if all types of player i defect on path in the first period, then if i deviates to cooperate, j cannot increase her confidence that i is greedy.

Finally, while we will solve for all PBE, we will ultimately restrict our attention to efficient equilibria. There are lots of ways to apply efficiency criteria for staged, incomplete information games. We eliminate an equilibrium if at least one type of one player prefers to deviate to another equilibrium, while no player does worse given their ex-ante total expected utilities.

A.1.1 Definition: trust-building equilibria

We define a trust-building equilibria as a PBE with the following features:

- First-period discriminators: The greedy and security-seeking types face different incentives, and therefore make different choices in the first period. In equilibrium: $pr(a_1 = c|A_s) > pr(a_1 = c|A_g)$.
- Cooperation generates trust: Because greedy and security-seeking types make different choices in the first period, their rivals learn, and this allows first-period cooperation to engender trust. In equilibrium: $\sigma_1^A < \sigma_2^A | b_1 = c$.
- Trust breeds cooperation: The trust built on the first-period choice allows states to cooperate in the second period. Therefore, the inclusion of an initial period facilitates trust, that allows for cooperation in the second period. Every equilibrium strategy profiles assures: $pr(a_2 = c|A_s, a_1 = b_1 = c) > pr(a_2 = c|A_s, b_1 = d, a_1 = .)$.

A.1.2 Preliminary analysis that serves all equilibrium analysis

We start by solving for all strategies that can appear on the path in the second period.

First we must define a mixing probability. Let $\omega_B^* = pr(b_2 = c) = \frac{k}{\sigma_2^A(1-L+k)}$, and $\omega_A^* = pr(a_2 = c) = \frac{k}{\sigma_2^B(1-L+k)}$.

Lemma A.1 *For some set of parameters, we can support three and only three strategy profiles on path in the second period of a PBE.*

1. *For all parameters we can support $a_2 = b_2 = d$ (strict defection).*
2. *If $\sigma_2^i > \frac{k}{1+k-L}$ we can support $a_2 = c|A_s, a_2 = d|A_g$, $b_2 = c|B_s, b_2 = d|B_g$ (pure strategy, type-separating).*
3. *If $\sigma_2^i > \frac{k}{1+k-L}$ we can support $a_2 = \omega_A^*|A_s, a_2 = d|A_g$, $b_2 = \omega_B^*|B_s, b_2 = d|B_g$ (security seekers mix).*

Before we consider a specific profile, note that greedy types defect in every on-path strategy profile. We now show this is strictly true. Conjecture $b_2 = a_2 = c$ in equilibrium. A_g can

profitably deviate to defect if $H > 1$. Conjecture $a_2 = c, b_2 = d$. A_g can profitably deviate to defect if $0 > -k$.

We now describe three strategy profiles we can support on the path, focusing on the security-seeker's preferences. We derive Bullet 1 as follows. A_s prefers to remain on the path (defect) rather than deviate to cooperate if $-k < 0$. Always true, as desired.

We derive bullet 2 as follows. Consider A_s prefers to remain on the path, rather than deviate to defection, so long as $\sigma_2^A + (1 - \sigma_2^A)(-k) > \sigma_2^A L$, this solves for the equilibrium condition with regard to σ_2^A. σ_2^B is solved the same way.

To verify these are the only supportable pure strategies, we must exhaust the other strategy profiles for security seekers. Consider the pure strategy profile: $a_2 = c, b_2 = d$. Note A can always profitably deviate to $a_2 = d$. There are no other pure strategy profiles to consider.

Turning to the mixed strategy profile. A is indifferent between cooperation and defection if

$$\sigma_2^A \omega_B + (1 - \sigma_2^A \omega_B)(-k) = \sigma_2^A \omega_B L$$

This solves for $\omega_B = \frac{k}{\sigma_2^A(1-L+k)}$. This can be solved within 0, 1 so long as $\sigma_2^A > \frac{k}{1+k-L}$. A's equilibrium mixing probability is solved similarly.

To verify this is the only supportable mixed strategy, we must exhaust the other strategy profiles for security seekers. Consider the pure strategy profile: $pr(a_2 = c) = \omega'_A \neq \omega^*_A$, $pr(b_2 = c) \in (0, 1)$. Note, B is only indifferent at $\omega'_A = \omega^*_A$, and thus B must hold a profitable deviation to $pr(b_2 = c) \in \{0, 1\}$. Now consider $pr(a_2 = c) = \omega'_A \neq \omega^*_A$, $b_2 = c$. Here, if A is a security seeker, A always holds a profitable deviation to $a_2 = c$. There are no other mixed strategies to consider.

Remark These three on-path strategy profiles yield the following expected utility for A at the onset of the second period.

1. $EU_2^A = 0$
2. $EU_2^A | A_s = \sigma_A - k(1 - \sigma_A)$, $EU_2^A | A_g = \sigma_A H$
3. $U_2^A | A_s = \frac{Lk}{1-L+k}$, $U_2^A | A_g = \frac{Hk}{1-L+k}$

These utilities are useful for characterizing equilibria because they define the possible expected second-period utilities. For example, in an equilibrium where security-seekers cooperate in the first period and greedy types defect, we are certain that first-period defection yields an expected utility of 1 in the second period. We are also certain that first-period cooperation can generate only 1 of 3 potential expected utilities.

Remark If $\sigma_2^i \geq \frac{k}{1+k-L}$, then the strategy profile characterized in bullet 2 is Pareto dominant. Otherwise, the strategy profile characterized in bullet 1 is unique.

Note, all expected utilities are increasing in σ_2^i if $\sigma_2^i = \frac{k}{1+k-L}$ is non-negative. Now note that given the equilibrium condition, the utilities in bullet 3 are positive. This assures that if we can support the mixed strategy equilibria, that the strategy profile characterized in bullet 3 dominates 1. Turning to the contrast between 2 and 3. Subbing in the boundary condition $\sigma_2^i = \frac{k}{1+k-L}$, A_s, A_g's utilities are identical in bullet's 2 and 3. This assures that 2 dominates 3.

A.2 Baseline Model of International Signaling (Section 1)

We start with the assumption that $\beta_i = 0, \theta = 1$. This is the set of assumptions discussed in our review of existing trust-building theory in Section 1.

For ease, we replicate and expand a more intuitive description of the strategy profiles found in Figure 1(b):

Motive	Period 1	Period 2 | Period 1			
		CC	CD	DC	DD
Trust-building equilibrium					
Security	C	C	D	D	D
Greedy	D	D	D	D	D
Suckers equilibrium					
Security	C	C	D	D	D
Greedy	C	D	D	D	D
Tragic equilibrium					
Security	D	D	D	D	D
Greedy	D	D	D	D	D
No learning					
Security	D	C	C	C	C
Greedy	D	D	D	D	D
Semi-Tragic					
Security	D	D	D	D	D
Greedy	D	D	D	D	D

We start with the pure strategy equilibria that are not trust-building equilibria.

Proposition A.2 *We can support the following strategy profiles as PBE for any feasible off-path beliefs.*

1. ***Tragic:*** *For all parameters* $s^A(d,d), s^B(d,d)$.
2. ***Suckers:*** *If* $p_i > \frac{k}{1+k-L} \sim 1 - \frac{1}{H}$ *holds we can support* $s^A(a_1 = c, a_2 = c|(A_s \& a_1 = b_1 = 1), a_2 = d|$ *Otherwise*$), s^B(b_1 = c, b_2 = c|(B_s \& a_1 = b_1 = 1), b_2 = d|$ *Otherwise*$)$.
3. ***No Learning:*** *If* $p_i > \frac{k}{1+k-L}$ *holds we can support* $s^A(a_1 = d, a_2 = c|A_s, a_2 = d|A_g), s^B(b_1 = d, b_2 = c|B_s, b_2 = d|B_g)$.
4. ***Semi-Tragic:*** *If* $p_i > \frac{k}{1+k-L}$ *holds we can support* $s^A(a_1 = c|A_s, a_1 = d|A_g, a_2 = d), s^B(b_1 = c|B_s, b_1 = d|B_g, b_2 = d)$.

The tragic equilibrium is obvious. We now analyze the **suckers** equilibrium. We start with A_g's strategy. All players cooperate in the first period. So in period 1, A_g prefers cooperation to defection if: $1 + p_B H + (1 - p_B)0 > H + 0$. This solves for $p_i > 1 - \frac{1}{H}$, as desired. Turning to A_s's strategy. In the second period, we assert A_s prefers cooperation to defection given on path

play. Given all players cooperate in period 1, $\sigma_2^A = p_B$. Thus, A prefers cooperation to defection in period 2 if $p_i > \frac{k}{1+k-L}$. This requires that $1+k-L > 0$. Working backwards, in the first period, A_s prefers cooperation over deviating to defection if: $1 + p - k(1-p) > L$. This re-arranges to $p > \frac{L+k-1}{1+k}$. Note, $\frac{L+k-1}{1+k} > \frac{k}{1+k-L} \equiv (k+1)(k+(L-1)^2)(1+k-L) < 0$ which cannot be satisfied if $1 + k - L > 0$. Note we can sustain the equilibrium for any off-path beliefs. The reason is that the only off-path action is $i_1 = d$ and we conjecture all players revert to $i_2 = d$ given either side deviates. We've shown we can support $i_2 = d$ for any beliefs, as desired. B's strategy is symmetric. This completes the proof.

We now analyze the **No Learning** equilibrium. In it, $a_1 = b_1 = d$. Thus, A's on-path belief is $\sigma_2^A = p_B$. Lemma A.1.2 solves the second-period strategy. This gives is the condition $p_i > \frac{k}{1+k-L}$. Turning to the first period, notice the conjectured on-path second-period strategies leave all types with their maximum second-period expected value. Thus, we'll focus on the case where off-path beliefs match on-path beliefs. Given this case, no type can profit by deviating so long as $k > 0$. B's strategy is symmetric. This completes the proof.

We note a second no learning equilibrium held together by a different off-path punishment: $s^A(a_1 = d, a_2 = c | A_s \;\&\; a_1 = b_1 = d, a_2 = d|$ otherwise$), s^B(b_1 = d, b_2 = c | B_s \;\&\; a_1 = b_1 = d, b_2 = d|$ otherwise$)$. Focusing on A_s, the same second period constraint binds. In the first period, there is no profitable deviation if $1 + p_B - k(1-p_B) > -k \equiv p_B(1+k) > 0$.

We now analyze the **Semi-Tragic** equilibrium. Lemma A.1.1 shows that we can support $a_2 = b_2 = d$ for all beliefs and parameters. Turning to the first period, A_g prefers defection if $p_B H > p_B - (1-p_B)k \implies H > 1 - k(1-p_B)/p_B$, always true. A_s prefers cooperation if $p_B - k(1-p_B) > p_B L$, which solves for the condition, as desired. B's strategy is symmetric.

We now solve for the pure strategy, trust-building equilibrium.

Proposition A.3 *There is one pure strategy **trust-building equilibrium**. It arises if and only if*

$$\frac{k}{1+k} \geq p_i \geq \frac{k}{2+k-L} \tag{1}$$

holds. In it, greedy A plays $s^A(d,d)$. Security-seeker A plays $s^A(a_1 = c, a_2 = c | (b_1 = c, a_1 = c), a_2 = d$ otherwise$)$. B's strategy is symmetric.

Since it is a pure strategy equilibrium, $\sigma_2^A | b_1 = c = 1$.

Greedy A plays on the path if:

$$p_B H \geq p_B(1+H) - k(1-p_B) \equiv p_B \leq \frac{k}{1+k}$$

The security A plays on path if:

$$p_B(1+1) - k(1-p_B) \geq p_B L \equiv p_B \geq \frac{k}{2+k-L}$$

Turning to existence, note, $\frac{k}{1+k} > \frac{k}{2+k-L}$ if $1 > L$, true by assumption. This completes the proof.

Proposition A.4 *Given feasible off-path beliefs no other pure strategy equilibria exist.*

We've shown that we can only support two pure strategy profiles in the second period, and that greedy types always defect in the second period. As a result, there are only two cases to consider. First, there are a class of strategy profiles that include the following unconditional on path actions: $s^A(a_1 = a_2 = c|A_s), s^B(b_1 = b_2 = c|B_s)$. We cannot support any equilibrium that includes this in the strategy profile. Suppose we could, it is easy to see that $s^A(a_1 = a_2 = d|A_g), s^B(b_1 = b_2 = d|B_g)$. This implies that both states must form posterior beliefs $\sigma_2^i = 0|i_1 = d$. This implies security seekers can profitably deviate from $i_2 = c \rightarrow d$.

Second, while we have ruled out asymmetric strategies in the second period, it is theoretically possible that we can support asymmetric strategies in the first, so long as second period strategies are conditional on first period. There are only two profiles to rule out, that vary in their off-path punishments. Profile 1 is: $s^A(a_1 = c|A_S, a_1 = d|A_g, a_2 = c|A_s \& a_1 = c \& b_1 = d, a_2 = d$ otherwise), $s^B(b_1 = d, b_2 = c|B_s \& a_1 = c \& b_1 = d, b_2 = d$ otherwise). Here if B deviates from $b_1 = d \rightarrow c$, then players revert to $a_2 = b_2 = d$, which we can always support. In the first period, A_s cannot profitably deviate from $a_1 = c \rightarrow d$ if: $-k + p_B - k(1 - p_B) > 0$, solves for $p_B > \frac{2k}{1+k}$. A_g cannot profitably deviate from $a_1 = d \rightarrow c$ if: $0 > -k + p_B H$, solves for $\frac{k}{H} > p_B$. These are jointly solvable if $k > 2H - 1$. Setting $k = 2H - 1, \frac{2H-1}{H} > p_B \implies 2 - 1/H > p_B$, which cannot hold because $H > 1$.

Profile 2 is: $s^A(a_1 = d, a_2 = c|A_s \& b_1 = c, a_2 = d$ otherwise), and $s^B(b_1 = c|B_S, b_1 = d|B_g, b_2 = c|B_s \& b_1 = c \& a_1 = ., b_2 = d$ otherwise). Here $a_1 = .$, emphasizes that b_2 holds even given A's off-path deviation. The ICC is identical to the first profile.

A.2.1 A comment on mixed strategy equilibria

To be clear, there are many mixed strategy equilibrium and even mixed strategy trust-building equilibria. We omit them from this analysis of the special case of $\beta_i = 0$ because (a) they do not alter our basic conclusions (which relates to trust-building when p_i is low); (b) are cumbersome to solve for and cannot be easily grouped owing to many different off-path strategies that can emerge; and (c) are strictly less efficient than pure strategy equilibria that survive the same parameter ranges. In section A.3.6, we will fully specify all the mixed strategy equilibria for the complete model (note the analysis thus far has only considered the special case $\beta = 0$, but the proof of the main model below $\beta_i \in [0, 1]$ will include this special case). Here we provide a preliminary analysis to demonstrate why they are both inefficient and also cannot alter our core result.

Recall, our main claim is that trust-building equilibria do not arise when p_i are too low. Thus, it would be misleading to omit mixed strategy trust-building equilibria if we could support them at lower levels of p_i. We shall solve for these in the complete model (i.e, once we introduce $\beta_i \in [0, 1]$). But here we demonstrate that we cannot support them for levels of p_i that are lower than the trust building equilibrium above. The reason is that the lower bound on p_i to support the contingent equilibrium is defined by the security-seeker's preference to engage in trust-building. When $p_B < \frac{k}{2+k-L}$, A_s's expected value from cooperation is too low to support trust, given her expectation that B is greedy. We've already shown that the mixed strategies we can support in the second period lower the security-seeker's expected utility. This means that the minimum p_i that will support trust-building must be high. To illustrate the point, we solve for the mixed strategy equilibrium that supports trust with the lowest level of p_i.

Proposition A.5 *If* $1 - L + k > 0$*, and*

$$\frac{k(1-L+k)}{k(2+k-L)-(H-1)(1-L)} > p_i > \frac{k(1-L+k)}{k^2-2kL+3k+L^2-2L+1} \tag{2}$$

Then the following strategies are a mixed strategy, trust-building PBE. Greedy A *plays* $s^A(d,d)$*. Security-seeker* A *plays* $s^A(a_1 = c, a_2 = \omega_A^* | b_1 = c, a_1 = c; a_2 = d$ *otherwise*)*. B's strategy is symmetric.*

Because the first period separates, $\sigma_2^A = 1 | b_1 = c$, $\sigma_2^A = 0 | b_1 = d$. This implies, $\omega_A^* = \frac{k}{(1-L+k)}$. The security-seeker remains on the path if, $p_B(1+\omega_B^*) - k(1-p_B) > p_B L$. This solves for $p_B > \frac{k(1-L+k)}{k^2-2kL+3k+L^2-2L+1}$. The greedy type remains on the path if, $p_B H > p_B(1+\omega_A^* H) - k(1-p_B)$. This solves for $p_B < \frac{k(1-L+k)}{k(2+k-L)-(H-1)(1-L)}$, as stated in the equilibrium.

Contrasting the lower bounds of inequalities 2 and 1,

$$\frac{k(1-L+k)}{k^2-2kL+3k+L^2-2L+1} > \frac{k}{2+k-L}$$

collapses to $L < 1$. This assures we can support 1 at lower levels of p_i, as desired.

A.3 Our theory: Independence of domestic choices (Section 2.1, and Results 1a and 1b)

We now study our theoretical intervention by only changing the model above to allow $\beta \in [0,1]$. We proceed as follows. First, we solve for the trust-building pure strategy PBE. Since our formally stated results focus on this equilibrium, we detail the comparative statics of this equilibrium through a series of remarks, and clarify how the results map onto Results 1a and 1b. Second, we solve for all other pure strategy PBE and rule out those that we cannot support. Third, we solve for all mixed strategy equilibria and rule out those we cannot support. Finally, we apply an iterative efficiency refinement.

A.3.1 The pure strategy, symmetric trust-building equilibrium

Proposition A.6 *If*

$$p_i \geq \frac{k - \beta_j(1+k-L)}{1+(1-\beta_j)(1+k-L)} \tag{3}$$

and

$$\frac{\beta_j(H-1-k)+k}{\beta_j(H-1-k)+1+k} \geq p_i \tag{4}$$

hold, then there is a pure strategy trust-building equilibrium with the same strategy profile as written in Proposition A.3.

We start with A_s's strategy. Because this is a complete separating equilibrium, $\sigma_2^A | b_1 = d = 0, \sigma_2^A | b_1 = c = 1$. By Lemma A.1, we can support second-period cooperation. Turning to the first period, A_s prefers to cooperate, rather than defect if:

$$p_B(1+1) + (1-p_B)(\beta_A - (1-\beta_A)k) > p_B L + (1-p_B)\beta_A L$$

This assumes that if A defects, A gets the second-period value 1 from $b_2 = a_2 = d$. This rearranges to equilibrium condition 3, as desired. It will help later to express it as $\beta_A > \frac{k - p_B(2+k-L)}{(1-p_B)(1+k-L)}$.

Turning to A_g's strategy. We've already shown we can support second-period defection for any set of beliefs. Turning to the first period, A_g prefers to defect, rather than cooperate if:

$$p_B H + (1-p_B)\beta_A H > p_B(1+H) + (1-p_B)(\beta_A - (1-\beta_A)k)$$

This rearranges to inequality 4 as desired. It will help later to express it as $\beta_A > \frac{p_B - k(1-p_B)}{(1-p_B)(H-1+k)}$.

There are no off-path beliefs. The strategies are symmetric. This completes the proof.

Later we will consider efficiency, and so it is useful to characterize total expected utilities.

Remark In the trustbuilding equilibrium, first period total expected utilities are:

$$EU_1^A|A_s = \beta_A - (1-\beta_A)k + p_B(2 - \beta_A + (1-\beta_A)k, EU_1^A|A_g = H(\beta_A + p_B - \beta_A p_B)$$

which are both increasing in β_A.

A.3.2 Establishing Result 1a,1b

Result 1a: 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., $p \to 0$).

Result 1b: Even when the independence threshold characterized in 1a is not met, as the level of independence increases, a trust-building equilibrium can be supported at decreasing levels of initial trust.

Results 1a and b and their implications are effectively a series of comparative static claims on equilibrium A.6 as a function of β, p.

Thus, we focus on the ICC for greedy and security types. Our main claims focuses on the incentives of security-seekers. The reason we care most about security-seekers is that their incentives impose a lower bound on p (condition 3). The classic model (absent β) is structured such that the security-seeker desires mutual cooperation when payoffs are dependent. Thus, if initial trust is too low, the security-seeker does not cooperate. Thus all of our claims about independence and trust relate to easing A_s's tension that prevents cooperation when p is too low.

Remark The security-seeker cannot profitably deviate from on-path cooperation in the trust-building equilibrium at any level of initial trust (even $p \to 0$) given,

$$\beta_i > \frac{k}{1+k-L} \in (0,1)$$

Outside this range, the level of independence that will sustain the security-seekers' incentive compatibility constraint is

$$\beta_i > \frac{k - p_j(2+k-L)}{(1+k-L)(1-p_j)}$$

wherein the right hand-side is strictly decreasing in p_j.[29]

Both results come from re-arranging the security-seeker's ICC described in 3. The first claim establishes the boundary where initial trust does not affect A_s's incentives for trust-building. Note the denominator of 3 must be positive for all $\beta_j \in [0, 1]$. and the numerator is decreasing in β_j. The threshold sets the numerator to 0 and re-arranges. The second claim comes from simply re-arranging 3 as a function of β. Taking the derivative of the RHS, $\frac{L-2}{(1+k-L)(1-p_j)^2}$, which must be negative. It is useful because it illustrates that A_s's incentives for cooperation are increasing in independence.

Turning to the incentives of greedy states. The classic model is structured such that if initial trust is too high, that greedy will try to cheat. Their incentive to deviate to cooperation is amplified when they believe the other side can be cheated (i.e., they are playing against a security-seeker). Thus, their ICC is governed by an upper bound on p. This is condition 4. To be clear, this is less important for our theory. After all, our main claim is that there is no lower bound on initial trust. But greedy types only determine the upper bound. Thus, our main goal is to establish that the greedy types are willing to comply under the same conditions that security-seekers are.

Remark The greedy type cannot profitably deviate from first-period defection in the trust-building equilibrium so long as $\beta_j > \frac{p_i-(1-p_i)k}{(H-k-1)(1-p_i)}$.

Putting both type's of incentives together, notice that

Remark For any set of parameters H, k, L, (a) all types' incentives to deviate from on path actions in the trust-building equilibria are decreasing in β; (b) at $\beta = 1$, we can support a trust-building equilibrium for every level of initial trust that satisfies $\frac{H-k}{H} > p_i$.

A.3.3 Other symmetric pure strategy equilibria

In what follows, we characterize all other pure strategy equilibria. To begin, we focus on the symmetric equilibria. Since the symmetric equilibria follow naturally from the equilibrium listed in proposition A.2, we only solve for the conditions where the addition of β makes a difference.

Proposition A.7 *There is a **tragic equilibrium** if and only if the dependence threshold is not met: $\beta_i > \frac{k}{1-L+k}$. It has the same strategy profile as in proposition A.2.1.*

We focus on A_s's strategy. We've shown we can support mutual defection in the second period for any set of beliefs and parameters. We label A's expected value of $a_2 = b_2 = d$ as EU_2. We focus on first-period incentives. A_s prefers defection to cooperation if $\beta_A L + EU_2 < \beta_A - (1 - \beta_A)k + EU_2$. This solves for the equilibrium condition as desired. This result is notable because it departs from the baseline model, and conventional wisdom that defect, defect is always an equilibrium.

Proposition A.8 *There is a **suckers equilibrium**. Its conditions and strategy profile are the same as in proposition A.2.2.*

[29] recall, subscript j means not i.

See proof of A.2.2. Since all players cooperate in the first period, β does not factor into computing the conditions where deviating is profitable.

Proposition A.9 *There is a **No learning** PBE if* $p_i > \frac{k}{1+k-L}$ *and* $\beta_i < \frac{k}{1-L+k}$. *It has the same strategy profile as in proposition A.2.3.*

The only difference in the proof from A.2.3 is in A_s's incentive to play defect in the first period. A_s prefers defect to cooperate iff: $\beta_A L + p_B - k(1-p_B) > \beta_A - (1-\beta_A)k + p_B - k(1-p_B)$. This solves for $\beta_A < \frac{k}{1-L+k}$, as desired. This completes the proof.

As in the baseline, there is a second **No learning** PBE held together by a different off-path action. Specifically,

Proposition A.10 *If* $p_i > \frac{k}{1+k-L}$ *and* $\beta_i < \frac{p_j(1+k)}{1+k-L}$. *There is a second **No learning** PBE with strategy profile.* $s^A(a_1 = d, a_2 = c|A_s \;\&\; a_1 = b_1 = d, a_2 = d|$ *otherwise*$), s^B(b_1 = d, b_2 = c|B_s \;\&\; a_1 = b_1 = d, b_2 = d|$ *otherwise*$)$.

Here the difference is that if either player deviates from $i_1 = d \to c$, then $i_2 = d$. The first period ICC for A_s is: $\beta_A L + p_B - k(1-p_B) > \beta_A - (1-\beta_A)k$.

Proposition A.11 *There is a **Semi-tragic** PBE if and only if* $\beta_i > \frac{k(1-p_j)-p_j(1-L)}{(1-p_j)(1+k-L)}$. *It has the same strategy profile as in proposition A.2.4.*

The only difference in the proof from A.2.4 is in A_s's incentive to play cooperate in the first period. A_s prefers cooperate to defect iff: $p_B + (1-p_B)(\beta_A - (1-\beta_A)k) > p_B L + (1-p_B)(\beta_A L)$. This solves for $\beta_A > \frac{k(1-p_B)-p_B(1-L)}{(1-p_B)(1+k-L)}$ as stated. B's incentives are symmetric. Rearranging gives the minimum boundary on p, $p_j > \frac{k-\beta_i(1+k-L)}{(k+1-L)(1-\beta_i)}$. This completes the proof.

A.3.4 Asymmetric equilibria

When $\beta_i = 0$, we ruled out the possibility of asymmetric equilibria entirely. Once we add in independent first period actions, we can rationalize asymmetric equilibria where players play different strategies in the first period. The intuition behind this result is that the security seeker's minmax changes. When $\beta = 0$, the largest amount both types could guarantee themselves follows from $a_1 = a_2 = d$. We've shown that when the independence threshold is reached $\beta_i > \frac{k}{1+k-L}$ A_s strictly prefers $a_1 = c$ even if $b_1 = d$. This assures that even if A_s knows that B will cheat for certain in the first period, A_s will still cooperate in the first period. Because different types now how different minmax strategies, the opportunities for screening also change.

When $\beta_A, \beta_B > \frac{k}{1+k-L}$ then the incentives for both players are the same. However, when dependencies are lopsided (β_A is high and β_B is low), then security seekers face different minmax strategies. As we will show the dependence threshold is a binding constraint, even when it is not met, lopsided initial trust (p_A high, p_B low) is another critical factor for inducing different asymmetric equilibria.

Player/Motive	Period 1	Period 2 | Period 1			
		CC	CD	DC	DD
Asymmetric trust-building					
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					
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

Asymmetric trust-building equilibria Two asymmetric, pure strategy, trust-building equilibria emerge. For ease, we reproduce Table 3(b), which writes out their strategy profiles.

Proposition A.12 *If* $p_B > \frac{1}{1-k+L}$,

$$\beta_A > \frac{p_B H - k}{H - 1 - k} \sim \frac{k - p_B + k(1 - p_B)}{1 + k - L} \sim 0$$

$$\beta_B < \frac{k + p_A(1 - L + k)}{(1 - p_A)(1 - L + k)} \tag{5}$$

then, there is an asymmetric, pure strategy, trust building equilibrium: $s^A(a_1 = c|A_S, a_1 = d|A_g, b_2 = c|A_s \& a_1 = c \& b_1 = ., a_2 = d$ *otherwise*$)$, $s^B(b_1 = d, b_2 = c|B_s \& a_1 = c \& b_1 = ., b_2 = d$ *otherwise*$)$. *There is an equivalent equilibrium swapping A and B.*

A's first period choice is fully separating, and thus we can sustain B's second period strategy. B's first period choice is pooling. Thus, we can sustain A's second period choice if $p_B > \frac{k}{1+k-L}$.

In the first period, A_s cannot profitably deviate from $a_1 = c \to d$ if: $\beta_A - k(1 - \beta_A) + p_B - k(1 - p_B) > \beta_A L \equiv \beta_A > \frac{k - p_B + k(1-p_B)}{1+k-L}$, as desired. We can re-write it as $p_B > \frac{2k - \beta_A(1+k-L)}{1+k}$.

A_g cannot profitably deviate from $a_1 = d \to c$ if: $\beta_A H > \beta_A - (1 - \beta_A)k + p_B H$. We can write it as $\frac{\beta_A(H-1-k)+k}{H} > p_B$.

B's first period binding constraint is B_s's ICC. B_s cannot profitably deviate from $b_1 = d \to c$ if: $p_A(L+1) + (1-p_A)(\beta_B L) > p_A(1+1) + (1-p_A)(\beta_B - (1-\beta_B)k)$, solves for $p_A < \frac{\beta_B(1-L+k)-k}{(1-\beta_B)(1-L+k)}$, as desired.

These conditions directly imply:

Remark Asymmetric trust building requires:

- One side has a highly independent trust-building action (β_A must be positive and sufficiently large) and moderate-to-high initial trust (p_B is only bound from below if β_A is sufficiently large).
- The other has lowly independent trust-building action (β_B cannot be too high, and must be lower than the independence threshold) and low initial trust (p_A cannot be too high).

Finally, B has one off-path deviation $b_1 = d \to c$. We claimed that this does not effect second period strategies (critically here $a_2 = c|A_s$). This follows instantly given our assumption of feasible off-path beliefs match on-path beliefs in this case.

We now turn to a second equilibria that deviates only in this off-path case. Rather, than assume players are insensitive to B's off-path action, we now assume players revert to the punishment $a_2 = b_2 = d|b_1 = c$.

Proposition A.13 *Replacing condition 5 with*

$$\beta_B < \frac{k - p_A(k - L)}{(1 - p_A)(1 - L + k)}$$

then, there is a second asymmetric, pure strategy, trust building equilibrium: $s^A(a_1 = c|A_S, a_1 = d|A_g, b_2 = c|A_s \& a_1 = c \& b_1 = d, a_2 = d$ *otherwise*$)$, $s^B(b_1 = d, b_2 = c|B_s \& a_1 = c \& b_1 = d, b_2 = d$ *otherwise*$)$. *There is an equivalent equilibrium swapping A and B.*

Trivially, we can support $a_2 = b_2 = d$ for any parameters. Thus, the only thing that changes is B's first period incentive to deviate. B_s imposes the binding constraint. B_s cannot profitably deviate from $b_1 = d \to c$ if: $p_A(L + 1) + (1 - p_A)(\beta_B L) > p_A + (1 - p_A)(\beta_B - (1 - \beta_B)k)$.

Later, we will analyze Pareto efficient equilibria. So we emphasize,

Remark This latter equilibria is Pareto dominated by the former across all the parameter ranges we can sustain it.

Other Asymmetric pure strategy equilibria There is an equivalent equilibrium swapping the As and Bs.

Proposition A.14 *If* $\beta_A > \frac{k(1-p_B)-p_B(1-L)}{(1-p_B)(1+k-L)}$, *and* $\beta_B < \frac{k(1-p_A)-p_A(1-L)}{(1-p_A)(1+k-L)}$ *there is an* ***asymmetric, semi-tragic*** *PBE. In it,* $s^A(a_1 = c|A_s, a_1 = d|A_g, a_2 = d), s^B(b_1 = d, b_2 = d)$.

Trivially, A_g cannot profit from deviating. B_s ICC for first period cooperation over defect is $p_A + (1 - p_A)(\beta_B - (1 - \beta_B)k) < p_A L + (1 - p_A)(\beta_B L)$. Note $\frac{k(1-p_B)-p_B(1-L)}{(1-p_B)(1+k-L)} < \frac{k}{1+k-L}$.

Remark We cannot sustain this equilibrium if both states have met their independence thresholds.

It is also useful to solve the ICC for $p_A < \frac{k-\beta_B(1+k-L)}{(1-\beta_B)(1+k-L)}$. As this illustrates, sustaining B's incentive requires low initial trust. If p_A was higher, B_s could profitably deviate to cooperation. But it also assures that if p_B is large, we can sustain this equilibrium even if $\beta_A < \frac{k}{1+k-L}$,

Remark We can sustain this equilibrium if neither states has met its independence thresholds.

A.3.5 Ruling out other pure strategy equilibria

Finally, we rule out three other classes of pure strategy equilibria. First, we cannot sustain any pure strategy equilibria that include against-type first period actions. That is $a_1 = c|A_g, a_1 = d|A_s$. It is trivial that we cannot support these if second period strategies are not contingent on first period actions. The reason is that if second period actions are not contingent, then we need only consider first period incentives. In terms of contingent second period strategies, the binding constraint is: $s^A(a_1 = c|A_g, a_1 = d|A_s, a_2 = c|(a_1 = b_1 = d, A_s), a_2 = d$ otherwise). Note that we need not consider other second period contingent strategies because given the first period strategy, $\sigma_i^2 = 1|j_1 = d, \sigma_i^2 = 0|j_1 = c$. Given this strategy profile, A_g's ICC is: $p_B(\beta_A - (1-\beta_A)k) + (1-p_B) > p_B(\beta_A H + H) + (1-p_B)H$. This re-arranges to, $p_B(-\beta(H-1) - (1-\beta)k - 1) > H - 1$. Note that the LHS must be negative and the RHS must be positive, which assures no p_B exists to satisfy it. It follows that for any contingent second period strategy profile we could support, A_g always has a profitable first period deviation from $a_1 = c \rightarrow d$.

Second, we cannot support any pure strategy PBE that include contingent second period strategies $s^A(a_2 = c|b_1 = d, A_s), a_2 = d|b_1 = c, A_s)$. This follows instantly from what was just shown.

Finally, we rule out other pure strategy asymmetric equilibria. Given what we just ruled out, the only remaining asymmetric equilibria to exhaust are those that include $a_2 = b_2 = d$. Note, we cannot support any equilibria that includes $i_1 = c|j_2 = i_2 = dA_g$. Suppose we could, i_g can always profit from the deviation $i_1 = c \rightarrow d$. Thus, we need only consider $i_1 = i_2 = d|A_g$. It follows that the only remaining asymmetric strategy profile to rule out is: $s^A(a_1 = c|A_s, a_1 = d|A_g, a_2 = d), s^B(b_1 = d, b_2 = d)$. We've solved for this profile.

A.3.6 Mixed strategy equilibria

We now solve for mixed strategy equilibria. As will become clear, each is a variant of a pure strategy PBE we have already characterized. Because we will later apply a Pareto refinement it is important to understand that each mixed strategy equilibria is Pareto dominated by its respective pure strategy equivalent. They also arise generally in overlapping parameter ranges with their respective pure strategy equivalents. In fact, the mixed strategy equivalents of the symmetric suckers, no learning, and semi-tragic equilibria form a proper subset of their pure strategy equivalents. However, while the mixed strategy trust-building equilibria are dominated by the pure strategy trust building equilibrium, they can be sustained at higher levels of initial trust than the pure strategy equivalent.

Mixing in second period only There are three equilibria with only second period mixed strategies (i.e, pure strategies in first period).

As a reminder, we have solved for the unique, on path mixing strategy:

$$\omega_i^* = pr(i_2 = c) = \frac{k}{\sigma_2^j(1 - L + k)}$$

Which produced second period expected utilities:

$$U_2^A|A_s = \frac{Lk}{1 - L + k} \qquad U_2^A|A_g = \frac{Hk}{1 - L + k}$$

and always carried an equilibrium condition $\sigma_2^i > \frac{k}{1-L+k}$.
First,

Lemma A.15 *If*

$$p_i > \frac{k}{1-L+k}$$
$$\beta_A < \frac{k(1+k)}{(1+k-L)^2}$$

then the following strategy profile is an equilibrium $s^A(a_1 = d, a_2 = \omega_A^*|(a_1 = b_1 = d\&A_s), a_2 = d$ *otherwise*$).s^B(b_1 = d, b_2 = \omega_B^*|(a_1 = b_2 = d\&B_s), b_2 = d$ *otherwise*$)$ *given any off-path beliefs.*

Given first period pooling, $\sigma_2^i = p_j$. This gives us the first condition. In the first period, if either player deviates from defection, the game reverts to mutual defection in the next period. From what we've shown, A_g trivially cannot profit from $a_1 = d \to c$ under any condition. A_s's ICC is: $\beta_A L + \frac{Lk}{1-L+k} > \beta_A - (1-\beta_A)k$, which solves for $\beta_A < \frac{k(1+k)}{(1+k-L)^2}$, as desired.

Remark This equilibrium is Pareto dominated by the pure strategy no learning equilibrium, and is contained within its parameters.

Second,

Lemma A.16 *If*

$$p_i > \frac{k}{1-L+k}$$
$$H < \frac{1-L+k}{1-L}$$

then the following strategy profile is an equilibrium $s^A(a_1 = c, a_2 = \omega_A^*|(a_1 = b_1 = c\&A_s), a_2 = d$ *otherwise*$).s^B(b_1 = c, b_2 = \omega_B^*|(a_1 = b_2 = c\&B_s), b_2 = d$ *otherwise*$)$ *given any off-path beliefs.*

Given first period pooling, $\sigma_2^i = p_j$. This gives first condition 1. A_s ICC is $1 + \frac{Lk}{1-L+k} > L$, always satisfied. A_g's ICC is $1 + \frac{Hk}{1-L+k} > H$, which gives us $H < \frac{1-L+k}{1-L}$.

Remark This equilibrium is Pareto dominated by the pure strategy suckers equilibrium, and is contained within its parameters.

Third, there is a trust-building equilibria with perfect separation in the first period, and mixing in the second.

Lemma A.17 *If*

$$\frac{k-\beta(1-L+k)}{(1-\beta)(1-k+L)+L\omega^*} < p < \frac{k+\beta(H-1-k)}{k+1+\beta(H-k-1)-H(1-\omega^*)}$$

then the following mixed strategy trust-building strategy profile is an equilibrium $s^A(a_1 = c|A_s, a_1 = d|A_g, a_2 = \omega_A^*|(a_1 = b_1 = c\&A_s), a_2 = d$ *otherwise*$)$*. This assures on path beliefs* $\sigma_i^* = 1|j_1 = c$ *and mixing probability* $\omega_i^* = \frac{k}{1-L+k}$.

Since the first period is perfectly separating, $\sigma_2^i \in \{0,1\} \implies \omega_i^* = \frac{k}{1-L+k}$. We now turn to the first period. Greedy type cannot profitably deviate from first-period defection if: $pH+(1-p)\beta H > p(1+\omega^* H)+(1-p)(\beta-(1-\beta)k)$.

$$p < \frac{k+\beta(H-1-k)}{k+1+\beta(H-k-1)-H(1-\omega^*)}$$

Security seekers cannot deviate from first period cooperation if

$$p(1+L\omega^*)+(1-p)(\beta-(1-\beta)k) > pL+(1-p)\beta L$$

, which solves for

$$p > \frac{k-\beta(1-L+k)}{(1-\beta)(1-k+L)+L\omega^*} = \frac{k-\beta(1-L+k)}{(1-\beta)(1-k+L)+\frac{kL}{1-L+k}}$$

Remark This equilibrium is Pareto dominated by the pure strategy trust-building equilibrium. The lower bound in inequality 3 strictly binds the mixed strategy equilibrium. But the upper bound does not. That is, we can always find $\frac{\beta_j(H-1-k)+1}{\beta_j(H-1-k)+1+k} < p_i < \frac{k+\beta(H-1-k)}{k+1+\beta(H-k-1)-H(1-\omega^*)}$.

Remark This equilibrium is Pareto dominated by the pure strategy suckers equilibrium, and when $\beta_i =$ is contained within its parameters.

In the trust-building equilibrium, the lower bound arises because the security seeker weighs the concerns about being cheated against the value of second period cooperation. Mixed strategies reduce the security seeker's value of cooperation in the second period. Thus, the security seeker is willing to initially cooperate under fewer conditions. The upper bound arises because the greedy type is not tempted to cheat in the first period. By reducing their value of waiting to cheat in the second period, we increase their incentives to defect in the first (without altering their value from on-path play).

Mixing only in the first period. There are three equilibria with only first period mixed strategies (i.e, pure, possibly conditioned, strategies in second period).

Define a first period mixing probability:

$$\omega_i^x = \frac{k-\beta_j(1+k-L)}{p_i(1+k-L)(1-\beta_j)}$$

We'll prove that this leaves i_s indifferent in two equilibria. Note that for $\omega_i^x \in [0,1]$, it must be that $p_i > \frac{k-\beta_j(1+k-L)}{(1+k-L)(1-\beta_j)}$, with the special case $p_i > \frac{k}{(1+k-L)}$ given $\beta_i = 0$. It also requires $\frac{k}{1+k-L} > \beta_j$. Note that these conditions are a subset of the pure strategy no learning equilibria.

The first equilibrium is:

Lemma A.18 *If $\omega_i^x \in [0,1]$, then the following strategy profile is an equilibrium $s^A(pr(a_1=c)=\omega_A^x|A_s, a_1=d|A_g, a_2=d).s^B(pr(b_1=c)=\omega_B^x|B_s, b_1=d|A_g, b_2=d)$.*

In period 2, players mutually defect regardless of type. This is proven to hold. First, we derive the mixing probability ω_i^x as what leaves i_s indifferent between cooperation and defection. Focusing on A_s, $\omega_B p_B + (1-\omega_B p_B)(\beta_A - (1-\beta_A)k) = \omega_B p_B L + (1-\omega_B p_B)\beta_A L$, which solves for ω_A^x. Trivially, if we can find a $\omega_i^x \in [0,1]$ than, i_s can be held indifferent.

Remark This equilibrium is Pareto dominated by the symmetric semi-tragic equilibrium, and is contained within its parameter ranges.

The second is,

Lemma A.19 *If* $\omega_i^x \in [0,1]$*, and*

$$p_i > \frac{1+k-(k+\beta_j)(1+k-L)}{(1+k-L)(1-\beta_j)}$$

then the following strategy profile is an equilibrium $s^A(pr(a_1 = c) = \omega_A^x | A_s, a_1 = d | A_g, a_2 = c | A_s, a_2 = d | A_g,).s^B(pr(b_1 = c) = \omega_B^x | B_s, b_1 = d | A_g, b_2 = c | B_s, b_2 = d | A_g)$.

In the second period, $\sigma_2^A | b_1 = c = 1$. Clearly, we can sustain $a_2 = c | A_s$ in this case. $\sigma_2^A | b_1 = d = \frac{p_B(1-\omega_B^z)}{p_B(1-\omega_B^z)+1-p_B}$. Thus, to sustain second period choices, it must be that $\frac{p_B(1-\omega_B^x)}{1-p_B\omega_B^x} > k/(1+k-L)$. Plugging in the value of ω_i^x gives us the equilibrium condition. There are no off-path actions in the first period, and thus no second period reversion is necessary.

Since second period strategies are not conditions, in the first period, the binding constraint is inducing i_S to mix. Because we conjecture second period strategies are not conditioned in first period strategies, it follows instantly that the $\omega_i^1 = \omega_i^x$ leaves player's indifferent. Note the equilibrium constraint on p_i assures $\omega \in [0,1]$ given $\frac{k-\beta_j(1+k-L)}{(1+k-L)(1-\beta_j)} < \frac{1+k-(k+\beta_j)(1+k-L)}{(1+k-L)(1-\beta_j)}$. There are no off-path actions in the first period, and thus no second period reversion is necessary. Note that with the additional condition, this is a subset of the pure strategy suckers equilibrium.

We now solve for a **trust building** mixing equilibria. It differs from the above in that it includes a contingent second period strategy. Define a first period mixing probability:

$$\omega_i^z = \frac{k-\beta_j(1+k-L)}{p_i(1+(1-\beta_j)(1+k-L))}$$

We'll prove that this leaves i_s indifferent given contingent second period strategies. Note that for $\omega_i^z \in [0,1]$, it must be that $\frac{k}{1+k-L} > \beta_j$. Further, $p_i > \frac{k-\beta_j(1+k-L)}{(2-L+k-\beta_j(1+k-L))}$, and in the baseline case, $p_i > \frac{k}{2-L+k}$. These are the same constraints as in the pure strategy trust-building equilibrium.

Lemma A.20 *If*

$$\omega_i^z \in [0,1]$$

then the following strategy profile is an equilibrium $s^A(pr(a_1 = c) = \omega_A^z | A_s, a_1 = d | A_g, a_2 = c | A_s \& a_1 = b_1 = c, a_2 = d |$ *otherwise*$)$*,* $s^B(pr(b_1 = c) = \omega_B^z | B_s, b_1 = d | A_g, b_2 = c | B_s \& a_1 = b_1 = c, b_2 = d |$ *otherwise*$)$.

In the second period, $\sigma_2^A = 1 | b_1 = c$, $\sigma_2^B = 1 | a_1 = c$, which assures we can sustain second period cooperation. Note we can sustain mutual defection given any set of beliefs, as desired.

Moving to the first period, first we prove ω_i^z holds i_s indifferent. Focusing on A_S, $\omega_B p_B(1+1)+(1-\omega_B p_B)(\beta_A-(1-\beta_A)k) = \omega_B p_B L+(1-\omega_B p_B)\beta_A L$, solves for $\omega_B = \omega_i^z$. Trivially, if we can sustain $\omega_i^z \in (0,1)$ then we can sustain i_s's first period strategy. Greedy types cannot profitably deviate from $a_1 = d \to c$ if $\omega_B p_B H + (1-\omega_B p_B)H\beta_A > \omega_B p_B(1+H)+(1-\omega_B p_B)(\beta_A-(1-\beta_A k))$. Rearranging to $\frac{H\beta_A+k-\beta_A(1+k)}{p_B(2+H\beta_A+k-\beta_A(1+k)} > \omega_i$. Plugging in $\omega_B = \omega_A^z$, this is always satisfied if $\omega_A^z \in [0,1]$. There are no off-path beliefs. This completes the proof.

Remark This equilibrium is Pareto dominated by the mixed strategy trust building equilibria characterized in Lemma A.17, and is contained within its parameter ranges.

Finally, we **cannot support** any equilibria where greedy types mix in the first period. The binding constraint is the following strategy profile: $s^A(pr(a_1 = c) = \omega_A^g | A_g, a_1 = c | A_s, a_2 = d)$, $s^B(pr(b_1 = c) = \omega_B^g | B_g, b_1 = c | B_s, b_2 = d)$. In this case, we can hold A_g indifferent if: $p_B(1+H)+(1-p_B)(\omega_B^g+(1-\omega_B^g)(\beta_A+(1-\beta_A)k) = p_B H+(1-p_B)(H\omega_B^g+(1-\omega_B^g)\beta_A H$. This simplifies to, $p_B-(1-p_B)(\beta_A(H-1-k)-k) = \omega(1-p_B)(H-1-k)(1-\beta_A)$. Note this cannot hold for $\omega < 1$. It follows that we cannot sustain a mixing probability that leaves A indifferent.

Mixing in both periods Finally, we solve for equilibria where i_s plays mix strategy in both periods. Both of these equilibria are **trust building equilibria**. As a reminder, we have solved for the unique, on path second-period mixing strategy:

$$\omega_i^* = pr(i_2 = c) = \frac{k}{\sigma_2^j(1-L+k)}$$

This always imposes the equilibrium condition: $\sigma_2^i > \frac{k}{1-L+k}$.

A critical feature to note is that if i_s mixes in the second period, then no matter posterior beliefs σ_2^i second period expected utilities are:

$$U_2^A|A_s = \frac{Lk}{1-L+k} \qquad U_2^A|A_g = \frac{Hk}{1-L+k}$$

We begin with the case wherein security seekers condition their decision to mix in the second period if they observe first-period mixing because it imposes the fewest conditions on p_i.

$$\omega_i^\gamma = \frac{k-\beta_j(1-L+k)}{p_i(\frac{Lk}{1-L+k}+(1-\beta_A)(1-L+k))}$$

The requirement $\omega_i^\gamma \in (0,1)$ imposes a minimum bound on $p_i > \frac{k-\beta_j(1-L+k)}{(\frac{Lk}{1-L+k}+(1-\beta_A)(1-L+k))}$, $\frac{k}{1-L+k} > \beta_j$.

Lemma A.21 *If $\omega_i^\gamma \in (0,1)$, then the following strategy profile is an equilibrium $s^A(pr(a_1 = c) = \omega_A^\gamma | A_s, a_1 = d | A_g, pr(a_2 = c) = \omega_A^* | A_s \& a_1 = b_1 = c, a_2 = d |$ otherwise), $s^B(pr(b_1 = c) = \omega_B^\gamma | B_s, b_1 = d | B_g, pr(b_2 = c) = \omega_B^* | B_s \& a_1 = b_1 = c, b_2 = d |$ otherwise).*

We've shown in the second period, i_s cannot profitably deviate if $\sigma_2^i > \frac{1}{1-L+k}$. Note that $\sigma_2^i|j_1 = c = 1$, as desired. Also note that we can always support second period mutual defection, which we assert in all cases other than $a_1 = b_1 = 1 \& i_s$.

Working backwards, we solve for ω_i^γ. i_s is held indifferent given ω_i^γ. Focusing on A_s, $\omega_B p_B(1+\frac{Lk}{1-L+k}) + (1-\omega_B p_B)(\beta_A(1+k)-k) = \omega_B p_B L + (1-\omega_B p_B)\beta_A L$, which gives us $\omega_B = \omega_B^\gamma$, as desired.

The greedy type cannot profitably deviate from $a_1 = d \to c$ if: $p_B\omega_B H + (1-p_B\omega_B)\beta_A H > p_B\omega_B(1+\frac{Hk}{1-L+k}) + (1-p_B\omega_B)(\beta_A(k+1)-k)$. This solves for $p_B\omega_B((H-k-1)(1-\beta_A) - \frac{Hk}{1-L+k}) > -k-\beta_A(H-1-k)$, always true. This completes the proof.

Finally, recall the mixing probability,

$$\omega_i^x = \frac{k-\beta_j(1+k-L)}{p_i(1+k-L)(1-\beta_j)}$$

Lemma A.22 *If* $\omega_i^x \in (0,1)$, *and*

$$p_i > \frac{2k-(1+k-L)(k^2+\beta_j)}{(1+k-L)(1-\beta_j)} > \frac{k}{1+k-L}$$

hold, then the following strategy profile is an equilibrium $s^A(pr(a_1=c)=\omega_A^x|A_s, a_1=d|A_g, pr(a_2=c)=\omega_A^*|A_s, a_2=d|A_g)$, $s^B(pr(b_1=c)=\omega_B^x|B_s, b_1=d|B_g, pr(b_2=c)=\omega_B^x|B_s, b_2=d|B_g)$.

The analyses above are enough to show that i_g cannot profitably deviate given that i_s will mix regardless. Thus, we analyze i_s strategy. We've shown in the second period, i_s cannot profitably deviate if $\sigma_2^i > \frac{k}{1+k-L}$. A's posterior belief is $\sigma_2^A|b_1 = d = \frac{p_B(1-\omega_B^x)}{1-\omega_B^x p_B}$. Plugging in ω_B^x, A_s is only willing to cooperate even after observing defection if $p_B > \frac{2k-(1+k-L)(k^2+\beta_A)}{(1+k-L)(1-\beta_A)}$, as written. We emphasize that $\frac{2k-(1+k-L)(k^2+\beta_j)}{(1+k-L)(1-\beta_j)} > \frac{k}{1+k-L}$ to make clear that there must be so much initial trust, that i_s is still willing to mix even after being cheated in the first period.

Working backwards, we solve for ω_i^x. i_s is held indifferent given ω_i^x. Focusing on A_s, $\omega_B p_B(1+\frac{Lk}{1-L+k}) + (1-\omega_B p_B)(\beta_A(1+k)-k+\frac{Lk}{1-L+k}) = \omega_B p_B(L+\frac{Lk}{1-L+k}) + (1-\omega_B p_B)(L\beta_A + \frac{Lk}{1-L+k})$, which gives us $\omega_B = \omega_B^x$, as desired.

Remark Both of these equilibria are Pareto dominated by the mixed strategy trust building equilibria characterized in Lemma A.17, and is contained within its parameter ranges.

A.3.7 Pareto efficient equilibria

To begin, we summarize the total expected utilities of candidate equilibria. That is, all equilibria we have not yet shown are Pareto dominated in the parameter ranges where we can sustain them. When strategies are symmetric we denote all total expected utilities from A's perspective. When they are asymmetric, we denote utilities for both players. Finally, we add descriptors when there are different variants of each equilibria. When there is only a pure strategy symmetric equilibrium to consider, we do not add a descriptor.

Remark The first period total expected utilities from on path play in all candidate PBE are:

1. Symmetric, trust-building (prop A.6): $EU_1^A|A_s = 2p_B+(1-p_B)(\beta_A-(1-\beta_A)k), EU_1^A|A_g = H(p_B + (1-p_B)\beta_A)$.

2. Asymmetric, trust-building (prop A.12) $EU_1^A|A_s = (\beta_A+p_B)(1+k)-2k$, $EU_1^A|A_g = \beta_A H$, $EU_1^B|B_s = p_A[1 + L(1-\beta_B)] + \beta_B L$, $EU_1^B|B_g = H[p_A(2-\beta_B) + \beta_B]$.

3. Symmetric mixed strategy trust building (prop A.17) $EU_1^A|A_s = p(1 + L\frac{Lk}{1+k-L}) + (1-p)(\beta - (1-\beta)k)$, $EU_1^A|A_g = pH + (1-p)\beta H$,

4. Suckers (prop A.8): $EU^A|A_s = 1 - k + p(1+k)$, $EU^A|A_g = 1 + p_B H$.

5. No learning (prop A.9): $EU^A|A_s = \beta_A L - k + p(1+k)$, $EU^A|A_g = \beta_A H + p_B H$

6. Tragic (prop A.7): $EU_1^A|A_s = \beta_A L$, $EU_1^A|A_g = \beta_A H$

7. Symmetric, semi-tragic (prop A.11)) $EU_1^A|A_s = p_B+(1-p_B)(\beta_A-(1-\beta_A)k)$,$EU_1^A|A_g = p_B H + (1-p_B)\beta_A H$

8. Asymmetric, semi-tragic (prop A.14) $EU_1^A|A_s = \beta_A - (1-\beta_A)k$, $EU_1^A|A_g = \beta_A H$, $EU_1^B|B_s = \beta_B L + p_A L(1-\beta_B)$, $EU_1^B|B_g = \beta_B H + p_A H(1-\beta_B)$.

Proposition A.23 *Pure strategy equilibria are Pareto efficient with the following additional parameter constraints.*

1. *The symmetric trust-building equilibrium (prop A.6) is dominated by the suckers equilibrium if $\beta_i < \frac{1}{(1+k)}$ holds. Thus, when this condition holds, the symmetric trust building survives refinement with the additional restriction, $p_i < 1 - \frac{1}{H} \sim \frac{k}{1+k-L}$.*

2. *The asymmetric trust-building equilibrium (prop A.12) is not dominated by any other equilibrium and survives refinement under stated conditions.*

3. *The symmetric, mixed strategy trust-building equilibrium (prop A.12) is Pareto dominated by the pure strategy trust-building equilibrium and the suckers equilibrium. Thus, it survives refinement with the additional restriction, $\frac{\beta_j(H-1-k)+1}{\beta_j(H-1-k)+1+k} < p_i < 1 - \frac{1}{H} \sim \frac{k}{1+k-L}$. Note it cannot survive refinement when $\beta_i = 0$.*

4. *The suckers equilibria (prop A.8) is not dominated by any other equilibrium and survives refinement under stated conditions.*

5. *The no learning equilibrium (prop A.9) is dominated by the suckers. Thus, it survives refinement with the additional parameter restriction $p_i < 1 - \frac{1}{H}$.*

6. *The tragic equilibrium is Pareto dominated by all other candidate equilibria in the conditions we can sustain it. Thus, it only survives refinement when no other candidate equilibria survive.*

7. *The symmetric semi-tragic equilibria is Pareto dominated by all symmetric equilibria except the tragic equilibrium in the conditions we can sustain it. Thus, it only survives refinement when no other candidate equilibria survive.*

8. *The asymmetric semi-tragic equilibria, is Pareto dominated by the asymmetric trust-building equilibria. Thus, it only survives refinement if either of the additional parameter restrictions are met,* $\frac{p_B H-k}{H-1-k} > \beta_A$ *or* $\beta_B > \frac{k-p_A(1-L+k)}{(1-p_A)(1-L+k)}$.

The results follow from a simple comparison of the total expected utilities of security seekers and greedy types. Some notable points. First, the suckers equilibrium dominates even trust building when β_i is low. The reason is that both equilibria run the risk of being exploited, but in the trust building equilibria mutual cooperation only follows in the condition that players don't cheat each other. By contrast, mutual cooperation is assured in the suckers equilibrium. As β_i increases, it is increasingly attractive to run the risk of exploitation in the first period, and this can offset the loss of conditional cooperation relative to assured cooperation. Second, the asymmetric and symmetric equilibria do not dominate each other in their respective parameter ranges. The reason is that asymmetric equilibria always have a quality where one state knows they will be cheated by their rival in the first period, and the other is certain they will get to cheat their rival. Thus, we can always find at least one type that prefers the symmetric over asymmetric and vice versa.

Third, when $\beta_A, \beta_B > \frac{k}{1+k-L}$, then we can sustain the minmax in the semi-tragic equilibrium given any levels of trust. When $\beta_A, \beta_B < \frac{k}{1+k-L}$, then we can sustain the minmax in the tragic equilibrium given any levels of trust. Since all other equilibria leave at least one player with more than their minmax, (and trivially, no player can do worse), these are always dominated.

A.4 Robustness: Introducing salience

We now allow $\theta > 0$ to vary. Since our question is, does trust-building operate at different levels of salience, we focus on that equilibrium.

Proposition A.24 *If*

$$\theta > \frac{k - [\beta_j + p_i(1-\beta_j)](1+k-L)}{p_i} \tag{6}$$

and

$$\theta < \frac{k + (H-k-1)(\beta_j + p_i(1-\beta_j)}{p_i H} \tag{7}$$

holds, then there is a pure strategy trust-building equilibrium with the same strategy profile as written in Proposition A.3.

We start with A_s's strategy. Because this is a complete separating equilibrium, $\sigma_2^A|b_1 = d = 0, \sigma_2^A|b_1 = c = 1$. By Lemma A.1, we can support second-period cooperation. Turning to the first period, A_s prefers to cooperate, rather than defect if:

$$p_B(1+\theta) + (1-p_B)(\beta_A - (1-\beta_A)k) > p_B L + (1-p_B)\beta_A L$$

This assumes that if A defects, A gets the second-period value 0 from $b_2 = a_2 = d$. This re-arranges to equilibrium condition 3, as desired.

Turning to A_g's strategy. We've already shown we can support second-period defection for any set of beliefs. Turning to the first period, A_g prefers to defect, rather than cooperate if:

$$p_B H + (1-p_B)\beta_A H > p_B(1+\theta H) + (1-p_B)(\beta_A(1+k) - k)$$

This re-arranges to condition 7, as desired.

There are no off-path beliefs. This completes the proof.

A.4.1 Result 2: Implications of salience

In the manuscript, we make the claim that if independence is sufficiently large, then there is no Goldilocks problem, and grand gestures are useful tools for trust-building. We support this with two remarks.

Remark Condition 6 is certainly satisfied if the dependence threshold ($\beta_j > \frac{k}{1+k-L}$) is met.

This assures the LHS is negative. As a result, there is no upper bound on the relative importance of a domestic choice.

Remark For any β_A, there always exists a θ sufficiently large to violate Condition 7.

This suggests that domestic choices must be at least non-trivial. If they are very unimportant relative to foreign rivalries, then the equilibrium degenerates.

Remark At full dependence, we can satisfy condition 7 if $\theta < \frac{H-1}{p_B H}$, which must be satisfied if $\theta \to 0$

This in combination with the other remarks assures that if we reach a certain level of independence, then we can always find a domestic action that is sufficiently important to sustain the trust building equilibrium.

To be clear, this does not always mean that increasing independence assures trust building arises under greater conditions:

Remark If $H > k + 1$ Condition 7 is increasingly easier to satisfy as β_A increases. If $H < k + 1$ Condition 7 is increasingly harder to satisfy as β_A increases.

A.5 Robustness: Similarity

This section explores the impact of variation the similarity, or correlation, of domestic and international preferences.

We introduce similarity as a random variable $\alpha > 0.5$. We draw $pr(\alpha_i = 1) = \alpha$. If $\alpha_A = 1$ then the payoffs are as they are in Table 4 for player A. If $\alpha_A = 0$, then player A's first-period payoffs are reversed. The greedy type gets the security-seeking type's payoffs and the security-seeking type gets the greedy type's payoffs. B's payoffs are defined the same way. Here α represents how similar the two choices are in that when $\alpha = 1$ players are certain that first- and second-period preferences are aligned. When $\alpha = 0.5$ it means that there is an even chance that payoffs align or do not align across periods. In other words, 0.5 is the value of α at which domestic choices provide the least information.

The sequence of moves is as follows:

- Nature draws player types i.i.d from p_i (private)

- Nature draws α_i i.i.d. (private)
- A first trust problem arises in which A and B simultaneously select $s_{i1} = c, d$.
- A second trust problem arises in which A and B simultaneously select $s_{i2} = c, d$.
- Payoffs are realized.

A.5.1 Analysis

Our goal is to show that the trust-building equilibrium can survive under this condition. Thus, we solve for equilibria that are close to the pure strategy trust-building equilibria reported in Proposition A.24. Specifically, we are looking for equilibria where A_s plays $a_2 = c|b_1 = c$, and defects otherwise. There are two. In one, A_s follows her direct first-period incentives, in the other A_s always cooperates in the first period no matter her direct incentives.

First, we solve for the former

Proposition A.25 *When*

$$p > \frac{k(1-\alpha)}{\alpha(1-L)} \tag{8}$$

$$p > \frac{\alpha\left[k - \beta(1+k-L)\right] + (1-\alpha)\left[\theta k - 1 + L\right]}{\alpha\left[\theta + (1+k-L)(1-\beta)\right] + (1-\alpha)\left[\theta k - (1+k-L)(1-\beta)\right]} \tag{9}$$

and either

$$p < \frac{\alpha\left[k + \beta(H-k-1)\right] + (1-\alpha)\left[\theta k + H - 1\right]}{\alpha\left[\theta - (H-k-1)(1-\beta)\right] + (1-\alpha)\left[\theta k - (H-k-1)(1-\beta)\right]} \tag{10}$$

or the denominator of 10 is negative, and

$$p < \frac{\alpha\left[k - \beta(H-1-k)\right] + (1-\alpha)\left[H-1\right]}{\alpha\left[\theta - (H-k-1)(1-\beta)\right] + (1-\alpha)(H-k-1)(1-\beta)} \tag{11}$$

holds. Then A can support the following strategies in a symmetric PBE. $s^{A_g}(a_1 = c|1-\alpha, a_1 = d|1-\alpha, a_2 = d), s^{A_s}(a_1 = c|\alpha, a_1 = d|1-\alpha, a_2 = d|b_1 = d, a_2 = c|b_1 = c)$. *B's condition and strategies are defined symmetrically.*

We showed in Lemma A.1 that A_g always defects (as desired), and A_s cooperates iff $\sigma_2^A > \frac{1}{1+a}$. In equilibrium, A's posterior beliefs after observing $b_1 = 1$ are:

$$\sigma_2^A|b_1 = c = \frac{p_B\alpha}{p_B + (1-\alpha_B)(1-p_B)}$$

Setting $\sigma_2^A > \frac{k}{1+k-L}$ this solves for equilibrium condition 12.

We now turn to first-period strategies. On path, A_s cooperates in the $\alpha = 1$ condition (i.e, the good type cooperates if she has good preferences). A^s cannot profitably deviate to defect in this condition iff:

$$p\alpha(1+\theta)+p(1-\alpha)(\beta(1+k)-k)+(1-p)\alpha(\beta(1+k)-k)+(1-p)(1-\alpha)(1-\theta k) > p\alpha L+p(1-\alpha)\beta L+(1-p)\alpha\beta L+(1-p)(1-\alpha)L$$

This solves for equilibrium condition 9. We deliberately disaggregated the denominator and numerator on $\alpha, 1-\alpha$. Note that when $\alpha = 1, \theta = 1$, the inequality converges to our main trust-building result.

On path, A_s defects in the $\alpha = 0$ condition (i.e, the good type defects if she has bad preferences). A^s cannot profitably deviate to cooperate in this condition iff:

$$p\alpha(1+\theta)+p(1-\alpha)(\beta(1+k)-k)+(1-p)\alpha(\beta(1+k)-k)+(1-p)(1-\alpha)(1-\theta k) < p\alpha H+p(1-\alpha)\beta H+(1-p)\alpha\beta H+(1-p)(1-\alpha)H$$

This solves for equilibrium 10. Note this inequality has no analog in the baseline model because it assumes good types defect.

Finally, we turn to A_g first period incentive. The binding constraint is the $\alpha = 1$ case.[30] We conjecture that A_g defects (the bad type defects if she has bad preferences). A_g cannot profitably deviate if:

$$p\alpha(1+\theta H)+p(1-\alpha)(\beta(1+k)-k)+(1-p)\alpha(\beta(1+k)-k)+(1-p)(1-\alpha) < p\alpha H+p(1-\alpha)\beta H+(1-p)\alpha\beta H+(1-p)(1-\alpha)H$$

This solves for condition 11. Note when $\alpha = 1, \theta = 1$, this condition converges to the baseline.

There are no off-path beliefs. This completes the proof.

We now solve the latter.

Proposition A.26 *When*

$$p_B > \frac{k(1-\alpha_B)}{1-L+k(1-\alpha_B)} \tag{12}$$

and

$$\frac{\alpha(k+\beta(H-k-1)}{H\theta-H+1+\alpha(k+\beta(H-k-1))} > p > \frac{\theta k+H-1-\alpha[(H-k-1)(1-\beta)+\theta k]}{\theta+k-\alpha[(H-k-1)(1-\beta)+\theta k]} \tag{13}$$

holds. Then A can support the following strategies in a symmetric PBE. $s^{A_g}(a_1 = c|1-\alpha, a_1 = d|\alpha, a_2 = d), s^{A_s}(a_1 = c, a_2 = d|b_1 = d, a_2 = c|b_1 = c)$. *B's condition and strategies are defined symmetrically.*

In this variant of trust-building, A_s cooperates no matter what A's first-period motivations are. But A_g's strategy depends on α.

We start with second-period strategies. We showed in Lemma A.1 that A_g always defects (as desired), and A_s cooperates iff $\sigma_2^A > \frac{1}{1+a}$. In equilibrium, A's posterior beliefs after observing $b_1 = 1$ are:

$$\sigma_2^A|b_1 = c = \frac{p_B}{p_B+(1-\alpha_B)(1-p_B)}$$

[30] In the other case, the bad type achieves her maximum possible payoff because she strictly prefers first period cooperation to exploiting the other side, and doing so gives the opportunity to exploit in the second period.

Setting $\sigma_2^A > \frac{k}{1+k-L}$ this solves for equilibrium condition 12, as desired.

We now turn to first-period strategies. From A_s's perspective, clearly first-period cooperation is hardest to sustain in the $1-\alpha$ case (rather than α case). Focusing on the $1-\alpha$ case, A_s prefers first-period cooperation to defection iff:

$$p_B(1+\theta)+(1-p_B)\alpha_B(\beta_A(1+k)-k)+(1-\alpha_B)(1-p_B)(1-\theta k) > p_B H+(1-p_B)\alpha_B\beta_A H+(1-\alpha_B)(1-p_B)H$$

This solves for the RHS of the equilibrium condition 13.

Turning to A_g's incentives. In equilibrium $a_1 = c|1-\alpha$, and $a_1 = d|\alpha$. Clearly, it is easier to sustain $a_1 = c|1-\alpha$ because this type gets the maximum expected value in the second period and a strictly higher payoff from cooperation in the first. Focusing on the α case, A_g prefers first-period defection iff:

$$(pH+(1-p)\alpha\beta H+(1-p_B)(1-\alpha_B)H > p_B(1+H\theta)+(1-p_B)\alpha_B(\beta_B(1+k)-k))+(1-p_B)(1-\alpha_B)H$$

This solves for the LHS of the equilibrium condition 13. There are no off-path beliefs.

B Supporting Empirical Tables

U.S.-Soviet Mapping

Table B.1 presents our mapping of the model to symmetric trust-building at the end of the Cold War.

Table B.1: Parameter Values in US-Soviet Case

Item	Description
Equilibrium	Symmetric trust-building
p_{USSR}	Low: US believed Soviets were immoral, bent on global domination, cheated on agreements.
p_{US}	Low: Soviets feared U.S. nuclear first strike and broader "anti-Soviet crusade"
β_{USSR}	Moderate-to-High: Benefits of political, economic, and emigration reforms (vs. status quo) mostly did not depend on what the US did.
β_{US}	Moderate: Encouraging Gorbachev's initiatives (vs. dismissing them) would still bring important benefits even if Soviet reform proved minimal.
θ_{USSR}	High: Glasnost, perestroika, and emigration were highly salient, transformative policies.
θ_{US}	Moderate: Stance of encouragement and support to Soviet Union was a key foreign policy choice with domestic political implications for Reagan.

Table B.2 presents the game matrix for a stylized "period 1" in the Cold War case, with the cells showing the choices corresponding to cooperation and defection for either side. Table B.3 presents the game matrix for a stylized "period 2" in the Cold War case. Bold cells indicate the choices taken.

Table B.2: Period 1: US-Soviet Symmetric Trust-Building

		USA: Cooperate	USA: Defect
USSR	Cooperate	**USSR: Liberalizing reform; US: Encourage Gorbachev**	USSR: Liberalizing reform; US: Dismiss Gorbachev
	Defect	USSR: ¬ Liberalizing reform; US: Encourage Gorbachev	USSR: ¬Liberalizing reform; US: Dismiss Gorbachev

Table B.3: Period 2: US-Soviet International Cooperation

		USA: Cooperate	USA: Defect
USSR	Cooperate	**USSR: Implement INF Treaty, troop reductions in Europe, etc. ; US: Implement INF Treaty, economic aid to USSR, etc.**	USSR: Implement INF Treaty, troop reductions in Europe, etc.; US: Cheat on INF Treaty, no aid to Soviets, etc.
	Defect	USSR: Cheat on INF Treaty, keep troops in Europe, etc.; US: Implement INF Treaty, economic aid to USSR, etc.	USSR: Cheat on INF Treaty, keep troops in Europe, etc.; US: Cheat on INF Treaty, no aid to Soviets, etc.

U.S.-South Korea Mapping

Table B.4 presents our mapping of the model to asymmetric trust-building in the U.S.-South Korea case.

Item	Description
Equilibrium	Asymmetric trust-building
p_{US}	High: South Korea confident in U.S. foreign policy goal of containing Communism internationally
p_{ROK}	Low: U.S. feared Park was a Communist.
β_{ROK}	High: Benefits of economic modernization, anti-Communism (vs. not) mostly depended on Park regime's intrinsic values, vision for ROK.
β_{US}	Low: Benefits of recognizing and legitimizing Park (vs. not) depended greatly on whether Park was anti- or pro-Communist.
θ_{ROK}	High: Park's reforms and anti-Communism represented important choices for future of South Korea.
θ_{US}	Moderate: As a key ally and client, U.S. ties to South Korea's government were important for the U.S.

Table B.4: Parameter Values in US-ROK Case

Table B.5 presents the game matrix for a stylized "period 1" in the U.S.-ROK case, with the cells showing the choices corresponding to cooperation and defection for either side. Table B.6 presents the game matrix for a stylized "period 2" in the U.S.-ROK case case. Bold cells indicate the choices taken.

Table B.5: Period 1: US-ROK Asymmetric Trust-Building

		USA: Cooperate	USA: Defect
ROK	Cooperate	ROK: Modernizing reforms, domestic anti-Communism ; US: Recognize Park regime	**ROK: Modernizing reforms, domestic anti-Communism; US: Withhold recognition of Park regime**
	Defect	ROK: ¬Modernizing reforms, lax on domestic Communism; US: Recognize Park regime	ROK: ¬Modernizing reforms, lax on domestic Communism; US: Withhold recognition of Park regime

Table B.6: Period 2: US-ROK International Cooperation

		USA: Cooperate	USA: Defect
ROK	Cooperate	**ROK: Support U.S. foreign policy (e.g., send forces to Vietnam); US: Extended deterrence, provide aid**	ROK: Support U.S. foreign policy (e.g., send forces to Vietnam); US: Abandon ROK as ally, end aid
	Defect	ROK: ¬Support U.S. foreign policy goals; US: Extended deterrence, provide aid	ROK: ¬Support U.S. foreign policy goals; US: Abandon ROK as ally, end aid