Why Arms Control Is So Rare

Online Appendix

Andrew J. Coe and Jane Vaynman

This Appendix provides supplementary material for the article, "Why Arms Control Is So Rare." Section 1 of this appendix offers proofs of the propositions and corollaries stated in the article. Section 2 sets up an extension of our model in which both sides may arm, and demonstrates that the main results of the article still hold. Section 3 contains additional documentation of our empirical claims, especially about the superpowers' arms control negotiations during the Cold War.

1 Proofs

We begin by defining exactly what constitutes an equilibrium "in the absence of a deal," referred to here as a no-deal equilibrium, and a deal equilibrium. A no-deal equilibrium must satisfy two properties. First, at any decision node, A's equilibrium strategy prescribes a choice that is the same regardless of the signals received up to that point about whether B has invested. Second, as long as B is not newly-armed, A never makes an offer that gives him more than his per-period costly conflict value $W_c^B(1-\delta)$. By contrast, a deal equilibrium is one in which either A's choice at at least one node depends on whether B has been observed to invest in past periods, or A offers B more than his per-period costly conflict value at some node. Intuitively, A need only condition her choice on signals of B investing, or offer

B more than this, if A wishes to punish B for investing or reward him for not doing so in the context of a deal.

We will see below that the no-deal equilibrium always exists and is generically unique, but there may be many possible deal equilibria. These deals can differ in the equilibrium offers made by A, and also in whether B never invests on the equilibrium path or does so sometimes but not always. The first distinction is irrelevant to our analysis, which is about whether deals exist (for any offers) rather than the particular offers made. We also ignore the second distinction, because if a deal exists in which B sometimes invests and sometimes doesn't in equilibrium, then a deal in which B never invests in equilibrium will also exist. As we show below, a deal's viability requires that B's temptation to invest in a period in which he is supposed to refrain be overcome by the risk of losing the surplus that will be realized under the deal relative to the no-deal equilibrium. This surplus is maximized when B never invests on the equilibrium path, so that any deal which features B investing in some periods must rely on a surplus no larger than the one which supports a no-investment deal. Hence, a partial-investment deal will exist only under conditions at least as stringent as those for the existence of a no-investment deal.

Next we prove a simple lemma that determines equilibrium once B is newly-armed.

Lemma 1. *Suppose that B is newly-armed. In equilibrium, A's continuation value will be $V_n^A \equiv W_n^A + D_n^w$ and B's will be $V_n^B \equiv W_n^B$.*[1]

Proof. Subgame perfection implies that B will accept any $q < V_n^A(1-\delta)$, because rejecting it yields his costly conflict value, while accepting it and reject the offer in the next round guarantees a higher payoff. Because of this, A would strictly prefer offering any $q \in (W_n^A(1-\delta), V_n^A(1-\delta))$ to costly conflict. For any such q that is less than $V_n^A(1-\delta)$, there is a higher q that both A and B would strictly prefer to costly conflict. Thus, in equilibrium A makes

[1] We assume throughout that $W_n^B > 0$. If this did not hold, then B would have no bargaining power, with or without arming, and so no incentive to invest in arms, and A would have no reason to try to stop B from getting them. We discard this uninteresting case.

the offer that renders B indifferent to costly conflict ($q = V_n^A(1-\delta)$) and B accepts this or any higher offer and rejects any lower offer. □

In the main paper, we assumed that, in the absence of a deal, B would invest in equilibrium. Formally stated:

Assumption 1. *In the absence of a deal, B would invest, given the chance: $k < \delta\lambda\left[W_n^B - W_c^B\right]$.*

The intuition for the condition is as follows. B must weigh the potential benefit of arming against the cost of investing in arms and the chance that his investment will not succeed. The left side of the inequality is the cost of investing (k), while the right is the benefit in terms of larger offers from A, relative to the unarmed status quo ($W_n^B - W_c^B$), weighted by the probability that investment works (λ) and discounted (δ) since B will not realize this benefit until the next round. Assumption 1 means that any no-deal equilibrium will be inefficient: it will feature either costly investment in arming by B, or a costly action by A to preempt this.

Now we state a more precise version of Proposition 1 from the main paper.

Proposition 4. *In the absence of a deal, B remains closed to monitoring. War or an arms race occurs if and only if $W_c^A + \overline{V_c^B} > \frac{1}{1-\delta} - \frac{k}{1-\delta(1-\lambda)}$, where $\overline{V_c^B} \equiv \max\left\{W_c^B, \frac{-k+\delta\lambda W_n^B}{1-\delta(1-\lambda)}\right\}$.*[2]

To see the intuition for the condition, observe that the right side is the total value of the game ($\frac{1}{1-\delta}$), less the expected cost of B's attempts to arm himself ($\frac{k}{1-\delta(1-\lambda)}$). This is how much value the two sides have to divide between them if they remain at peace while B arms. The higher the expected costs of peaceful arming, the harder it is to satisfy the inequality and the more likely war becomes. Peace is too costly to be tolerable for both sides. The left side is the sum of the minimum values the two sides would peacefully accept. For A, this is just her war value (W_c^A). For B, this value includes the offers A will make before he is

[2]Here we ignore the knife-edge case of equality, in which both a conflictual and a peaceful equilibrium exist in the absence of a deal. We also assume that if B is indifferent, he chooses to be closed to monitoring.

newly-armed, and the possibility of larger values once B has armed. If the latter is unlikely enough, there will be an offer A can make that renders B just indifferent between acceptance and rejection, so that B simply receives his war value (W_c^B). If becoming newly-armed is likely enough, A may not be able to make offers low enough to hold B to his war value, and so will simply offer B nothing at all, so that B receives the expected value of his investments ($\frac{-k+\delta\lambda W_n^B}{1-\delta(1-\lambda)}$). The higher the expected value of these investments, the harder it is to satisfy the inequality and the more likely war becomes. Peace is too likely to lead to a substantial shift in power against A for A to tolerate B's arming.

Remarkably, B's decision to remain closed to monitoring has nothing to do with concealing his arming or intention to do so. Indeed, because in equilibrium it is common knowledge that B will invest, there is never any uncertainty about his doing so, whether he is closed to monitoring or not. Concealment here is instead entirely about preserving advantage in conflict. Because opening to monitoring undermines B's prospects in conflict ($W_c^B \geq W_o^B$), doing so would only lead to A making less generous offers or resorting to conflict, with B doing worse either way. Thus, B remains closed to avoid A exploiting his openness.

Now we turn to proving this proposition.

Proposition 1

We first prove that the inequality stated in Assumption 1 in the paper is a necessary and sufficient condition for B to invest, given the chance, in any no-deal equilibrium. We then demonstrate the proposition.

First suppose that $k > \delta\lambda\left[W_n^B - W_c^B\right]$. Because, in a no-deal equilibrium, A does not react to signals of investment, an investment that fails gives B the same continuation value starting from the next period as not investing. Let this value be X. Then the continuation value of investing is $-k + \delta\left[\lambda V_n^B + (1-\lambda)X\right]$, while that of not investing is δX. If B is closed to monitoring, then in equilibrium it must be that $X \geq W_c^B$, since B could always

guarantee himself at least this value by rejecting any offer A made in the next period. But then we have $k > \delta\lambda\left[W_n^B - W_c^B\right] \geq \delta\lambda\left[V_n^B - X\right]$, which implies that B strictly prefers not to invest. If instead B were open to monitoring, then A will always act to ensure $X \leq W_c^B$ in equilibrium. If A were to offer $W_c^B(1-\delta)$ to B in every period, B would accept this and not invest by the above argument, yielding $X = W_c^B$. A would only make a different offer or resort to costly conflict if it led to at least as large a continuation value for her, which is only possible if such an action also led to a continuation value for B of no more than W_c^B. Hence in equilibrium B (possibly weakly) prefers closure. Given that B will remain closed and not invest, A's best strategy is to offer $W_c^B(1-\delta)$ in every round, since this is the least offer B will accept and yields a continuation value for A greater than her costly conflict value.

Now suppose that $k < \delta\lambda\left[W_n^B - W_c^B\right]$. Observe first that there cannot be a no-deal equilibrium in which B never invests on the equilibrium path. Investing at any particular point would yield a continuation value of $-k + \delta\left[\lambda W_n^B + (1-\lambda)X\right]$, while not investing would yield δX, so that in a no-deal equilibrium with no investment, it must be that $k \geq \delta\lambda\left[W_n^B - X\right]$. Since A never offers any more than $W_c^B(1-\delta)$ to B in such an equilibrium, it must be that $X \leq W_c^B$. Thus we have $k \geq \delta\lambda\left[W_n^B - X\right] \geq \delta\lambda\left[W_n^B - W_c^B\right]$, contradicting our supposition. By the same logic, there cannot be a no-deal equilibrium in which B permanently stops investing on the equilibrium path.

By way of contradiction, suppose there is a no-deal equilibrium in which at round t, B invests in equilibrium, but at round $t-1$ he does not, given the chance. Letting V_t^B and V_{t-1}^B be B's continuation values and q_1 and q_2 be the offers A makes in rounds t and $t-1$ respectively, we have:

$$V_t^B = 1 - q_2 + \delta\left[\lambda W_n^B + (1-\lambda)X\right] \geq 1 - q_2 + \delta X$$
$$V_{t-1}^B = 1 - q_1 + \delta V_t^B \geq 1 = q_1 + \delta\left[\lambda W_n^B + (1-\lambda)V_t^B\right]$$

Rearranged, the first line is equivalent to $X \leq \frac{-k}{\delta\lambda} + W_n^B$, and the second is equivalent to $V_t^B \geq \frac{-k}{\delta\lambda} + W_n^B$. Substituting the equation for V_t^B into this last inequality, we have:

$$\begin{aligned} 1 - q_2 + \delta\left[\lambda W_n^B + (1-\lambda)X\right] &\geq \frac{-k}{\delta\lambda} + W_n^B \\ \Rightarrow 1 - q_2 + \delta\lambda W_n^B + \delta(1-\lambda)\left[\frac{-k}{\delta\lambda} + W_n^B\right] &\geq \frac{-k}{\delta\lambda} + W_n^B \\ \Rightarrow 1 - q_2 &\geq (1-\delta)\left[\frac{-k}{\delta\lambda} + W_n^B\right] \\ \Rightarrow W_c^B &\geq \frac{-k}{\delta\lambda} + W_n^B \end{aligned}$$

where the last line uses the fact that A never offers B more than $W_c^B(1-\delta)$ in a no-deal equilibrium, and contradicts our supposition on k. Thus, as Assumption 1 states, whenever $k < \delta\lambda\left[W_n^B - W_c^B\right]$, B must always invest, given the chance, in any no-deal equilibrium.

Given Assumption 1, next observe that in any peaceful no-deal equilibrium, B must receive a continuation value of at least $\overline{V_c^B} \equiv \max\left\{W_c^B, \frac{-k+\delta\lambda W_n^B}{1-\delta(1-\lambda)}\right\}$. If he received less than the first term, he would do better to close and reject any offer A made. He cannot receive less than the second term even if A offers $q = 1$ and he invests in every round before he becomes newly-armed. If his continuation value were higher than $\overline{V_c^B}$, then A could profitably deviate by reducing at least one offer she will make to him, without that offer or any other being rejected and without altering B's investment behavior. It follows that, if the equilibrium path is peaceful, B's equilibrium continuation value must equal $\overline{V_c^B}$. Peace also requires that A receive at least her costly conflict value. If B is closed, this is W_c^A, and if B is open, it is $W_o^A \geq W_c^A$. If equilibrium is peaceful, the total value of the game is $V \equiv \frac{1}{1-\delta} - \frac{k}{1-\delta(1-\lambda)}$, where the second term is the expected cost of B's investments. If $\overline{V_c^B} + W_c^A > V$, then peace cannot be in equilibrium because A would strictly prefer to resort to conflict rather than make any offer B would accept, regardless of whether B is open or closed, and hence B will (possibly weakly) prefer closure to openness and costly conflict must occur in equilibrium.

Suppose instead $\overline{V_c^B} + W_c^A < V$. If B is closed, then A would strictly prefer offering the minimum B would accept, or nothing if that is infeasible, to making a different offer or initiating conflict. If B is open, A would either do the same or resort to conflict, leaving B no better off than if he were closed, so that B again (possibly weakly) prefers closure to openness and peace results.

Proposition 2

First we suppose a deal equilibrium featuring B remaining closed to monitoring exists, and demonstrate that the condition stated in the proposition must hold. Then we assume that condition holds, and prove that a closed deal exists.

Suppose a closed deal exists. Then in any given period along the equilibrium path, B must prefer staying closed and not investing to remaining closed and investing or rejecting A's offer. (B would obviously prefer either of these deviations to instead opening and then investing or rejecting, since opening would make his investment more likely to be detected and (weakly) reduce his costly conflict value.) Let B's continuation value from abiding by the deal be V_d^B, and let the equilibrium offer A makes under the deal be q_d. The analogous value if he instead invests in the current period is $1-q_d-k+\delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_c V_p^B + (1-\tau_c)V_d^B\right]\right]$, where V_p^B is B's continuation value if his investment is unsuccessful and detected by A. His value if he instead rejects A's offer is W_c^B. Thus, for the deal to be viable, it must be that:

$$\delta V_d^B \geq \max\left\{-k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_c V_p^B + (1-\tau_c)V_d^B\right]\right], \delta W_c^B\right\} \tag{1}$$

For her part, A must prefer abiding by the deal to costly conflict or making a different offer. The former deviation would yield the value W_c^A. If A instead deviates by making a higher offer, she will be left worse off than under the deal even if B does not punish this deviation. If A instead deviates by making a lower offer, the worst punishment B

could credibly impose on A for reneging on the deal is to resort to playing the no-deal equilibrium. If A's offer is low enough, B can credibly reject it, giving A a value of W_c^A. If A's deviant offer is not that low, B cannot credibly reject it but can begin investing, giving A a value of at most V_c^A. Any other punishment by B would involve neither costly conflict nor investing with positive probability at some point, and so would yield a higher value for A. By Proposition 1, the no-deal equilibrium will feature whichever of these deviations has the highest continuation value for A, so that the best of A's possible deviations is to resort to the no-deal equilibrium, yielding the continuation value V_c^A. Since $V_c^A \geq W_c^A$, for the deal to be in equilibrium it must be that $V_d^A \geq V_c^A$. Since $V_d^A + V_d^B = \frac{1}{1-\delta}$, it follows that we must have $V_d^B \leq \frac{1}{1-\delta} - V_c^A$.

Similarly, the worst punishment A could credibly impose on B for reneging on the deal by investing is the no-deal equilibrium. If this equilibrium features costly conflict, then it is the worst punishment for B that could possibly be credible, since any prospective punishment worse than that would give B a profitable deviation to costly conflict. Similarly, if the no-deal equilibrium features A peacefully tolerating B's investments, then she cannot credibly threaten to do anything worse to him. Hence $V_p^B \geq V_c^B$, so that inequality 1 implies

$$\begin{aligned}
&\delta V_d^B \geq -k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_c V_c^B + (1-\tau_c)V_d^B\right]\right] \\
\Rightarrow\ &\delta\left[\frac{1}{1-\delta} - V_c^A\right] \geq -k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_c V_c^B + (1-\tau_c)\left(\frac{1}{1-\delta} - V_c^A\right)\right]\right] \\
\Rightarrow\ &\delta\left[\lambda + (1-\lambda)\tau_c\right]\left[\frac{1}{1-\delta} - V_c^A - V_c^B\right] \geq -k + \delta\lambda\left[W_n^B - V_c^B\right]
\end{aligned}$$

This establishes the condition stated in the proposition.

Now instead suppose that the condition is true, and consider the following strategy profile as a candidate for a deal equilibrium. Along the equilibrium path, A offers q^*, which B accepts and then does not invest. If A ever makes an offer $q \neq q^*$, or B is ever detected to have invested, the two players subsequently play the no-deal equilibrium for the rest of

the game. The equilibrium offer is given by:

$$1 - q^* = \frac{1-\delta}{\delta} \max\left\{\delta W_c^B, -k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_c V_c^B + (1-\tau_c)V_d^B\right]\right]\right\}$$

with $V_d^B = \frac{1-q^*}{1-\delta}$.

By construction, B has no profitable deviation to either rejecting an offer or investing. So we need only check that A does not wish to renege on the deal by resorting to costly conflict or making a stingier offer.[3]

First suppose that $1 - q^* = (1-\delta)W_c^B$. Then the value of the deal for A is $\frac{1}{1-\delta} - W_c^B$, which is guaranteed to be at least V_c^A, so that A has no incentive to initiate costly conflict or to reduce her offer to him, since each of these would yield at most V_c^A by Proposition 1.

Now suppose that $1 - q^* = \frac{1-\delta}{\delta}\left[-k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_c V_c^B + (1-\tau_c)V_d^B\right]\right]\right]$. Then the value of the deal for A is $\frac{1}{1-\delta} - \frac{1}{\delta}\left[-k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_c V_c^B + (1-\tau_c)V_d^B\right]\right]\right]$. It is easily shown that the requirement that this value be at least V_c^A is equivalent to the condition in the proposition. This completes the proof.

Corollary 1

This follows quickly from the inequality in Proposition 2. Using our assumption from the model setup that k and λ are arbitrarily small, the right side of the inequality can be made arbitrarily small (but positive, by Assumption 1), but the left side is bounded below by $l(\tau_c) \equiv \delta\tau_c\left[\frac{1}{1-\delta} - V_c^A - V_c^B\right]$. Because, by Proposition 1, the no-deal equilibrium involves either costly conflict or costly investment, it must be that $l(\tau_c) > 0$ for $\tau_c > 0$ and that $l(\cdot)$ is strictly increasing in τ_c. Suppose the fixed but small enough value of the right side is x. Then the equation $l(\underline{\tau}) = x$ will have a unique, interior solution, and for any $\tau_c \geq \underline{\tau}$, Proposition 2 guarantees a viable deal.

[3]Note that we can treat A making an offer she knows B will reject as equivalent to her resorting to costly conflict; if either deviation is profitable, both will be, so it suffices to check only the latter.

Proposition 3

The proof is similar to that for Proposition 2. First we suppose such a deal featuring B opening to monitoring exists, and demonstrate that the conditions stated in the proposition must hold. Then we assume those conditions hold, and prove that an open deal exists.

Suppose an open deal exists. In any given period along the equilibrium path, B's continuation value from abiding by the deal is $1 - q_d + \delta V_d^B$, where q_d is the equilibrium offer A makes under the deal. B's continuation value if he remains open but deviates to investing in the current period is $1 - q_d - k + \delta \left[\lambda W_n^B + (1-\lambda)\left[\tau_o V_{p,o}^B + (1-\tau_o)V_d^B\right]\right]$, where $V_{p,o}^B$ is B's continuation value if his investment is unsuccessful and detected by A. His value if he instead closes to monitoring and then invests is $1 - q_d - k + \delta \left[\lambda W_n^B + (1-\lambda)V_{p,c}^B\right]$, where $V_{p,c}^B$ is B's continuation value if his first investment after overtly reneging is not successful. Thus, for the deal to be viable, it must be that:

$$\delta V_d^B \geq \max\left\{-k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_o V_{p,o}^B + (1-\tau_o)V_d^B\right]\right], -k + \delta\left[\lambda W_n^B + (1-\lambda)V_{p,c}^B\right]\right\} \tag{2}$$

For her part, A must prefer abiding by the deal to costly conflict or making a different offer. The former would yield W_o^A, so that we must have $V_d^A \geq W_o^A$. If A instead deviates by making a higher offer, she will be left worse off than under the deal even if B does not punish this deviation. If A instead deviates by making a lower offer, the worst punishment B could credibly impose on A for reneging on the deal is to resort to playing the no-deal equilibrium. If A's offer is low enough, B can credibly close himself to monitoring and reject the offer, giving A a value of W_c^A, the worst she could possibly receive in any equilibrium. If A's deviant offer is not that low, B cannot credibly reject it but can credibly close and begin investing, giving A a value of at most V_c^A. Any other punishment by B would involve either remaining open or neither rejecting nor investing with positive probability at some point,

and so would yield at least as high a value for A. By Proposition 1, the no-deal equilibrium will feature whichever of the deviant offers has the highest continuation value for A, so that the best of A's possible deviations is to resort to the no-deal equilibrium. Thus, for the deal to be in equilibrium it must be that $V_d^A \geq \max\left\{W_o^A, V_c^A\right\}$.

The worst punishment A could credibly impose on B depends on whether B remains open or has closed to monitoring. If B remains open, then A can resort to costly conflict, yielding W_o^A for A and imposing on B the worst value he could possibly receive in equilibrium, W_o^B. A can instead resort to the no-deal equilibrium, if it differs (i.e., does not feature immediate costly conflict), yielding V_c^A and V_c^B for the two players. A can only credibly impose whichever of these has the higher continuation value for A, so that A will initiate costly conflict only if $W_o^A > V_c^A$, and resort to the no-deal equilibrium otherwise. If B is closed to monitoring, the worst punishment A can credible impose is the no-deal equilibrium, by the same argument as in Proposition 2. Let $V_p^A = \max\left\{W_o^A, V_c^A\right\}$ and $V_p^B = W_o^A$ if $W_o^A > V_c^A$ and $V_p^B = V_c^B$ otherwise. Then it must be that $V_{p,o}^B \geq V_p^B$ and $V_{p,c}^B \geq V_c^B$. Also, we established earlier that $V_d^A \geq V_p^A$, and since $V_d^A + V_d^B = \frac{1}{1-\delta}$, it must be that $V_d^B \leq \frac{1}{1-\delta} - V_p^A$.

We can derive the conditions stated in the proposition by substituting these bounds on B's deal and punishment values into each of the terms in the maximum in inequality 2 in turn. Starting with the first term, we have:

$$\begin{aligned}
&\delta V_d^B \geq -k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_o V_{p,o}^B + (1-\tau_o)V_d^B\right]\right] \\
\Rightarrow\ &\delta\left[\frac{1}{1-\delta} - V_p^A\right] \geq -k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_o V_p^B + (1-\tau_o)\left(\frac{1}{1-\delta} - V_p^A\right)\right]\right] \\
\Rightarrow\ &\delta\left[\lambda + (1-\lambda)\tau_o\right]\left[\frac{1}{1-\delta} - V_p^A - V_p^B\right] \geq -k + \delta\lambda\left[W_n^B - V_p^B\right]
\end{aligned}$$

This establishes the first condition stated in the proposition.

Moving to the second term, we have:

$$\begin{aligned} \delta V_d^B &\geq -k + \delta\left[\lambda W_n^B + (1-\lambda)V_{p,c}^B\right] \\ \Rightarrow \delta V_d^B &\geq -k + \delta\left[\lambda W_n^B + (1-\lambda)V_c^B\right] \\ \Rightarrow \frac{1}{1-\delta} - V_p^A &\geq -\frac{k}{\delta} + \lambda W_n^B + (1-\lambda)V_c^B \end{aligned}$$

This is the second condition in the proposition.

Now instead suppose that the two conditions hold, and consider the following strategy profile as a candidate for a deal equilibrium. Along the equilibrium path, A offers q^*, which B accepts and then does not close or invest. If A ever makes an offer $q \neq q^*$, or B ever closes, the two players subsequently play the no-deal equilibrium for the rest of the game. If B remains open but is detected to have invested, A initiates costly conflict if $W_o^A > V_c^A$ and both players revert to the no-deal equilibrium otherwise. The equilibrium offer is given by:

$$1 - q^* = \frac{1-\delta}{\delta}\max\left\{-k + \delta\left[\lambda W_n^B + (1-\lambda)V_c^B\right], -k + \delta\left[\lambda W_n^B + (1-\lambda)\left[\tau_o V_p^B + (1-\tau_o)V_d^B\right]\right]\right\}$$

with $V_d^B = \frac{1-q^*}{1-\delta}$.

By construction, B has no profitable deviation to remaining open but investing or to closing and then investing. Remaining open but then rejecting A's offer leaves B no better off than closing then rejecting, so we need only check the latter deviation. Observe that $-k + \delta\left[\lambda W_n^B + (1-\lambda)V_c^B\right] \geq -k + \delta\lambda\left[W_n^B - W_c^B\right] + W_c^B \geq W_c^B$, where the first inequality follows from $V_c^B \geq W_c^B$ and the second from Assumption 1. Hence the first term in the maximum above also ensures that B cannot profit from rejecting A's offer.

It remains to check that A does not wish to renege on the deal by initiating costly conflict (which yields a value of W_o^A) or making a stingier offer (the best of which yields V_c^A). First suppose the latter is more tempting. Then $V_p^A = V_c^A$ and $V_p^B = V_c^B$, which implies that the

second term in the above maximum is at least as large as the first. By the same argument as in the proof of Proposition 2, the first condition in the statement of Proposition 3 implies that A prefers the deal to making a stingier offer.

Finally suppose that costly conflict is a more tempting deviation for A. Then $V_p^A = W_o^A$ and $V_p^B = W_o^B$. Observe that $V_d^A = \frac{q^*}{1-\delta}$. If the first term in the above maximum is the larger one, simple algebra demonstrates that the second condition in the statement of this proposition implies that A prefers the deal. If the second term in the above maximum is the larger one, the first condition in the statement of the proposition implies A prefers the deal, completing the proof.

Corollary 2

By definition, because there is no transparency-security tradeoff, we have $W_c^A = W_o^A$ and $W_c^B = W_o^B$. Then the second condition stated in Proposition 3 for an open deal to exist becomes $V_c^A + V_c^B \leq \frac{1}{1-\delta}$, which is trivially true—the players' continuation values from the no-deal equilibrium cannot sum to more than the total value of the game. The first condition in Proposition 3 becomes identical to that of Proposition 2, except that τ_c is replaced with τ_o. But then the same argument as in the proof of Corollary 1 suffices to show that an open deal will exist if and only if $\underline{\tau} \leq \tau_o < 1$.

Corollary 3

If the interval $[\underline{\tau}, \overline{\tau_o}]$ is not empty, then any τ_o in it must by construction satisfy both conditions in Proposition 3, so that an open deal exists at the level of monitoring τ_o. If that interval is empty, then for any choice of $\tau_o \in [0, 1]$, it must be that either $\tau_o < \underline{\tau}$ or $\tau_o > \overline{\tau_o}$, so that either the first or the second condition of Proposition 3 are not satisfied. Either way, Proposition 3 implies that no open deal at the level of monitoring τ_o exists.

2 Both Sides Can Arm and Monitor

In this section, we describe the setup of an extension of the game described in the paper, in which both sides may engage in arming and monitoring. The changes consist of adding choices for A to open or close to monitoring and to invest or not in arming and defining the consequences of these choices. We then demonstrate that the key results in this environment are qualitatively similar to those from the one-sided model discussed in the paper.

Setup

As before, A (referred to as feminine) and B (masculine) bargain over revisions to a prior division of a composite of disputed issues, represented by $[0, 1]$. The two players have linear preferences over the interval, with A favoring settlements closer to 1 and B favoring those closer to 0, and discount future payoffs by a factor $\delta < 1$.

In the first of infinitely many discrete periods of time, A first chooses whether to open or close herself to monitoring. She then chooses either to take a costly action against B, which ends the game as described subsequently, or to make a peaceful offer of a settlement for that round. If A makes an offer, B must first choose whether to open or close himself to monitoring. He can then reject A's offer, ending the game, or accept it, in which case it is implemented for that period. If B accepts A's offer, then A and B simultaneously choose whether to invest in arming, which imposes a cost on A of $k_A > 0$ and on B of $k_B > 0$, and succeeds with probabilities $\lambda_A, \lambda_B > 0$ respectively and otherwise fails. If an investment succeeds, the investing state is newly-armed and this immediately becomes common knowledge. If it fails, then a state must invest again in the future to have a chance of its efforts succeeding. The period then ends. We assume that the length of a period is relatively short, so that k_A, k_B, λ_A, and λ_B are all relatively small: the expected arming that will occur in a single period of investment is modest, though over many periods of repeated

investment it may become quite large. This ensures that the discrete-time structure of the game does not artificially prevent either side from being able to react quickly to the other's arming. We also assume that both states begin the game in the condition of being closed to monitoring.

The structure of the game in subsequent periods depends only on whether a state became newly-armed at some point in the past. Once a state has done so, subsequent periods entail only repeated bargaining: in each round, A has only to take her costly action or make an offer which B then accepts or rejects.[4] If neither state has done so, then each subsequent period until one does is the same as the first, except that it begins with the receipt of new information by each state on whether the other invested in the previous round, in the form of a public signal. If i is closed to monitoring and has invested, then with probability τ_c^i the signal indicates that it did, and otherwise indicates that it did not. If state i is open to monitoring and has invested, then with probability $\tau_o^i \geq \tau_c^i$ the signal indicates that it did, and otherwise indicates that it did not. If i did not invest, then the signal indicates that it did not regardless of whether it was open or closed to monitoring. Thus, each state's information on the other's investment is prone to false negatives but not false positives. If B receives a signal that A has invested, then B chooses whether to open or close to monitoring and whether to initiate costly conflict with A. If B does not receive such a signal, or does but chooses not to start a conflict, then the game proceeds as before. These additional moves for B are necessary only to ensure that the possibilities for punishing the other side when it is caught cheating are the same for both sides. All choices except for each state's investment, and all parameters of the game, are common knowledge.

If A takes her costly action or B rejects an offer, the game ends in costly conflict. The

[4]We could instead allow the state that is not yet newly-armed to continue investing in arms, but this would not alter the qualitative character of equilibrium once at least one side is newly-armed. All that matters for our argument is that this situation is worse for the not-yet-newly-armed side than the situation in which neither side is newly-armed.

expected value of this outcome for each state depends on whether either is newly-armed and, if neither is, whether each is open to monitoring. If only i is newly-armed, the values are W_{ni}^A and W_{ni}^B. If neither is newly-armed, the values are W_{oi}^A and W_{oi}^B if only i is open to monitoring, W_o^A and W_o^B if both are open to monitoring, and W_c^A and W_c^B if neither is. We assume $W_{nA}^A > W_{oB}^A \geq W_o^A, W_c^A \geq W_{oA}^A > W_{nB}^A$ and $W_{nA}^B < W_{oB}^B \leq W_o^B, W_c^B \leq W_{oA}^B < W_{nB}^B$. It will sometimes be convenient to treat W_o^i, W_{oi}^i, and W_{oj}^i as functions of $\left(\tau_o^A, \tau_o^B\right)$, with $W_o^i(\tau_c^i, \tau_c^j) = W_{oi}^i(\tau_c^i) = W_{oj}^i(\tau_c^j) = W_c^i$. We assume that $W_{oj}^i(\cdot), W_o^i(\tau_o^i, \cdot)$ are nondecreasing and $W_{oi}^i(\cdot), W_o^i(\cdot, \tau_o^j)$ are nonincreasing: a state's opening up to increasing levels of monitoring does not improve its value from conflict or lessen the other's. All these values are assumed to be non-negative. Conflict is costly: the value destroyed by it is $D_x^w \equiv \frac{1}{1-\delta} - W_x^A - W_x^B > 0$, for any $x \in \{nA, nB, oA, oB, o, c\}$, with $D_o^w \leq D_{oi}^w \leq D_c^w$.

Analysis

Observe that once either state is newly-armed, the continuation game is identical to the game analyzed in the main body of the paper. We therefore simply assume a set of continuation values for each such game, denoted V_{ni}^i and V_{nj}^i, which serve the same purpose as Lemma 1 above in determining the value of the game once a state's investment has succeeded. Similarly to Assumption 1 in the main body of the paper, we assume that, in the absence of a deal, both states would invest in arms, given the chance, and simply let the continuation values of the no-deal equilibrium be V_c^i. (Equilibrium behavior in the absence of a deal can be characterized via analogous arguments to those used to prove Proposition 1 above, but this characterization is not necessary for the results we state below and so we skip it in the interest of brevity.) We then state and prove analogues of Propositions 2 and 3 in the main body of the paper, establishing the conditions for closed or open deals to be viable. We focus on deal equilibria in which neither side ever arms on the equilibrium path. There may also be one-sided deal equilibria, in which only one side refrains from arming, but these are

essentially similar to the deal equilibria we characterized in the one-sided game.

Proposition 5. *Let* $V_p^A \equiv V_c^A$, $V_p^B \equiv V_c^B$, *and* $S \equiv \frac{1}{1-\delta} - V_p^A - V_p^B$. *There is a deal in which both sides are closed to monitoring if and only if* $\delta S \geq \frac{-k_A + \delta\lambda_A\left[V_{nA}^A - V_p^A\right]}{\lambda_A + \tau_c^A(1-\lambda_A)} + \frac{-k_B + \delta\lambda_B\left[V_{nB}^B - V_p^B\right]}{\lambda_B + \tau_c^B(1-\lambda_B)}$.

Proof. First we suppose a deal equilibrium featuring both sides remaining closed to monitoring exists, and demonstrate that the condition stated in the proposition must hold. Then we assume that condition holds, and prove that a closed deal exists.

Suppose a closed deal exists. Then in any given period along the equilibrium path, B must prefer staying closed and not investing to remaining closed and investing or rejecting A's offer. (B would obviously prefer either of these deviations to instead opening and then investing or rejecting, since opening would make his investment more likely to be detected and (weakly) reduce his costly conflict value.) Let B's continuation value from abiding by the deal be V_d^B, and let the equilibrium offer A makes under the deal be q_d. The analogous value if he instead invests in the current period is $1 - q_d - k_B + \delta\left[\lambda_B V_{nB}^B + (1-\lambda_B)\left[\tau_c^B V_p^B + \left(1-\tau_c^B\right)V_d^B\right]\right]$, where V_p^B is B's continuation value if his investment is unsuccessful and detected by A. His value if he instead rejects A's offer is W_c^B. Thus, for the deal to be viable, it must be that:

$$\delta V_d^B \geq \max\left\{-k_B + \delta\left[\lambda_B V_{nB}^B + (1-\lambda_B)\left[\tau_c^B V_p^B + \left(1-\tau_c^B\right)V_d^B\right]\right], \delta W_c^B\right\} \tag{3}$$

For her part, A must prefer abiding by the deal to initiating costly conflict, making a different offer, or covertly cheating by investing while remaining closed (by the same argument as for B, one of these options would always be preferred to opening). The first deviation would yield the value W_c^A. If A instead deviates by making a higher offer, she will be left worse off than under the deal even if B does not punish this deviation. If A instead deviates by making a lower offer, the worst punishment B could credibly impose on A for reneging on the deal is to resort to playing the no-deal equilibrium. If A's offer is low enough, B can credibly reject it, giving A a value of W_c^A. If A's deviant offer is

not that low, B cannot credibly reject it but can begin investing, giving A a value of at most V_c^A. Any other punishment by B would involve neither costly conflict nor investing with positive probability at some point, and so would yield a higher value for A. As with Proposition 1, the no-deal equilibrium will feature whichever of these deviations has the highest continuation value for A, so that the best of A's possible overt deviations is to resort to the no-deal equilibrium, yielding the continuation value V_c^A. Finally, if A instead cheats covertly by investing while remaining closed, his continuation value will be $q_d - k_A + \delta\left[\lambda_A V_{nA}^A + (1-\lambda_A)\left[\tau_c^A V_p^A + \left(1-\tau_c^A\right) V_d^A\right]\right]$, where V_p^A is A's continuation value if his investment is unsuccessful and detected by B. Thus, for the deal to be viable, it must be that:

$$\delta V_d^A \geq \max\left\{-k_A + \delta\left[\lambda_A V_{nA}^A + (1-\lambda_A)\left[\tau_c^A V_p^A + \left(1-\tau_c^A\right) V_d^A\right]\right], \delta V_c^A\right\} \tag{4}$$

The worst punishment each side could credibly impose on the other for reneging on the deal by being caught investing is the no-deal equilibrium. If this equilibrium features costly conflict, then it is the worst punishment for the cheating side that could possibly be credible, since any prospective punishment worse than that would give that side a profitable deviation to costly conflict. Similarly, if the no-deal equilibrium features one side peacefully tolerating the other's investments, then the former cannot credibly threaten to do anything worse to the latter. Hence $V_p^A \geq V_c^A$ and $V_p^B \geq V_c^B$, so that inequalities 3 and 4, together with the fact that $V_d^A + V_d^B = \frac{1}{1-\delta}$, imply:

$$\delta V_d^A + \delta V_d^B \geq \frac{-k_A + \delta\lambda_A V_{nA}^A + \delta(1-\lambda_A)\tau_c^A V_c^A}{\lambda_A + \tau_c^A(1-\lambda_A)} + \frac{-k_B + \delta\lambda_B V_{nB}^B + \delta(1-\lambda_B)\tau_c^B V_c^B}{\lambda_B + \tau_c^B(1-\lambda_B)}$$

$$\Leftrightarrow \frac{\delta}{1-\delta} \geq \frac{-k_A + \delta\lambda_A\left[V_{nA}^A - V_c^A\right]}{\lambda_A + \tau_c^A(1-\lambda_A)} + \delta V_c^A + \frac{-k_B + \delta\lambda_B\left[V_{nB}^B - V_c^B\right]}{\lambda_B + \tau_c^B(1-\lambda_B)} + \delta V_c^B$$

$$\Leftrightarrow \delta\left[\frac{1}{1-\delta} - V_c^A - V_c^B\right] \geq \frac{-k_A + \delta\lambda_A\left[V_{nA}^A - V_c^A\right]}{\lambda_A + \tau_c^A(1-\lambda_A)} + \frac{-k_B + \delta\lambda_B\left[V_{nB}^B - V_c^B\right]}{\lambda_B + \tau_c^B(1-\lambda_B)}$$

This establishes the condition stated in the proposition.

Now instead suppose that the condition is true, and consider the following strategy profile as a candidate for a deal equilibrium. Choose S_A, S_B such that $S_A + S_B = S$ and $S_i \geq \frac{-k_i + \delta\lambda_i\left[V_{ni}^i - V_c^i\right]}{\lambda_i + \tau_c^i(1-\lambda_i)}$; the condition guarantees some such S_A, S_B exist. Along the equilibrium path, A offers $q^* = (1-\delta)\left[V_c^A + S_A\right]$, which B accepts and then does not invest. If A ever makes an offer $q \neq q^*$, or either side is ever detected to have invested, the two players subsequently play the no-deal equilibrium for the rest of the game. Since A's continuation value under the deal will be $V_c^A + S_A$, by construction, A has no profitable deviation to making a different offer, to initiating costly conflict, or to investing in arms. Similarly, the continuation value of the deal for B will be $V_c^B + S_B$. By construction, B has no profitable deviation to either rejecting an offer or investing. □

An analogue to Corollary 1 in the main body of the paper follows from Proposition 4:

Corollary 4. *A closed deal is possible if and only if unilateral monitoring of both sides is good enough.*

Proof. This follows by arguments very similar to those given in the proof of Corollary 1. Given that the k_i and λ_i are arbitrarily small, the numerators on the right side of the inequality in Proposition 4 can be made arbitrarily close to zero. Using the fact that $S > 0$, so long as the τ_c^i are not too small, the right side can be made to be no greater than the left and the result follows. □

Now on to mutually-open deals.

Proposition 6. *Let* $V_p^i \equiv W_{oi}^i$ *if* $W_{oi}^j > V_c^j$ *and* $V_p^i \equiv V_c^i$ *otherwise.*
Let $V_A' = \max\left\{W_{oB}^A, V_c^A, \frac{\frac{-k_A}{\delta}+\lambda_A V_{nA}^A+(1-\lambda_A)\tau_o^A V_p^A}{\lambda_A+\tau_o^A(1-\lambda_A)}\right\}$.
Let $V_B' = \max\left\{W_{oA}^B, \frac{-k_B}{\delta}+\lambda_B V_{nB}^B+(1-\lambda_B)V_c^B, \frac{\frac{-k_B}{\delta}+\lambda_B V_{nB}^B+(1-\lambda_B)\tau_o^B V_p^B}{\lambda_B+\tau_o^B(1-\lambda_B)}\right\}$. *There is a deal in which both sides are open to monitoring if and only if* $V_A' + V_B' \leq \frac{1}{1-\delta}$.

Proof. As usual, we first suppose such a deal exists and demonstrate that the condition in the proposition holds, and then prove the converse.

Consider all the possible deviations for the two players from an open deal. B could close to monitoring and reject A's offer, which would offer at least as high a continuation value as remaining open to monitoring but rejecting A's offer. B could instead close to monitoring and then invest in arming, or he could remain open but then invest in arming. A could close to monitoring and initiate conflict, which she would prefer to remaining open before initiating conflict. A could instead close to monitoring, make an offer, and then invest in arming if the offer were not rejected; the optimal offer under this deviation would be that associated with the no-deal equilibrium, since any higher offer would not be necessary to avoid B's rejection and any lower one would be rejected. Finally, A could remain open to monitoring, make the offer specified under the deal, and then invest in arming. A could also remain open and make a different offer, but this would be overt cheating on the deal and might lead to punishment by B, so that A would prefer to have closed before making a different offer.

Hence, for a deal to be in equilibrium, its continuation values for the two sides V_d^B, V_d^A

must satisfy:

$$V_d^B \geq \max\left\{W_{oA}^B, -\frac{k_B}{\delta} + \lambda_B V_{nB}^B + (1-\lambda_B)V_{p,c}^B, \right.$$
$$\left. -\frac{k_B}{\delta} + \lambda_B V_{nB}^B + (1-\lambda_B)\left[\tau_o^B V_{p,o}^B + (1-\tau_o^B)V_d^B\right]\right\}$$
$$V_d^A \geq \max\left\{W_{oB}^A, V_{p,c}^A, -\frac{k_A}{\delta} + \lambda_A V_{nA}^A + (1-\lambda_A)\left[\tau_o^A V_{p,o}^A + (1-\tau_o^A)V_d^A\right]\right\},$$

where $V_{p,c}^i$ are the two sides' continuation values if they close before investing in arming and $V_{p,o}^i$ are the same if a side is caught investing in arming having remained open to monitoring.

By the same arguments as in previous proofs, the worst punishment each side can impose on the other for closing to monitoring and investing in arms is to revert to the no-deal equilibrium. Thus we have $V_{p,c}^i \geq V_c^i$. If instead one side catches the other investing in arms while remaining open to monitoring, the former can react by closing to monitoring, initiating costly conflict, making different offers, and/or investing in arms itself. The worst credible punishment the former side can impose on the latter must be either immediate costly conflict or the no-deal equilibrium, by the same arguments as in the proof of Proposition 3. Thus it must be that $V_{p,o}^i \geq V_p^i$, where V_p^i is W_{oi}^i if $W_{oi}^j > V_c^j$ and is V_c^i otherwise. Hence the above inequalities imply:

$$V_d^B \geq \max\left\{W_{oA}^B, -\frac{k_B}{\delta} + \lambda_B V_{nB}^B + (1-\lambda_B)V_c^B, \right.$$
$$\left. -\frac{k_B}{\delta} + \lambda_B V_{nB}^B + (1-\lambda_B)\left[\tau_o^B V_p^B + (1-\tau_o^B)V_d^B\right]\right\}$$
$$V_d^A \geq \max\left\{W_{oB}^A, V_c^A, -\frac{k_A}{\delta} + \lambda_A V_{nA}^A + (1-\lambda_A)\left[\tau_o^A V_p^A + (1-\tau_o^A)V_d^A\right]\right\},$$

Since $V_d^A + V_d^B = \frac{1}{1-\delta}$, the condition stated in the proposition follows.

Now instead suppose the condition is true, and consider the following strategy profile as a candidate for a deal equilibrium. Choose V_A, V_B such that $V_A + V_B = \frac{1}{1-\delta}$ and $V_i \geq V_i'$; the

condition guarantees such a pair exists. Along the equilibrium path, A chooses to be open to monitoring and offers $q^* = (1-\delta)V_A$; B responds by choosing to be open to monitoring, accepting this offer, and then not investing. If either side ever deviates from this and closes to monitoring, both sides revert to the no-deal equilibrium. If either side is ever detected to have deviated while remaining open to monitoring, the other side initiates costly conflict or reverts to the no-deal equilibrium, whichever gives it a higher continuation value. By construction, the condition assures that neither side has a profitable deviation from this deal. □

Analogues to Corollaries 2 and 3 in the main body of the paper follow quickly from Proposition 5:

Corollary 5. *If there is no transparency-security tradeoff, then there is always a pair of levels of open monitoring high enough to support a deal, and neither level need be perfect.*

Proof. By definition, because there is no transparency-security tradeoff, we have $W_c^i = W_{oi}^i = W_{oj}^i$, and so it must be that $V_c^j \geq W_{oi}^j$ and $V_p^i = V_c^i$. It follows easily that $V_i' = \frac{\frac{-k_i}{\delta} + \lambda_i V_{ni}^i + (1-\lambda_i)\tau_o^i V_c^i}{\lambda_i + \tau_o^i(1-\lambda_i)}$, which implies that the condition stated in Proposition 5 becomes identical to that of Proposition 4, except that τ_c is replaced with τ_o. But then the same argument as in the proof of Corollary 4 suffices to show that an open deal will exist if and only if the τ_o^i are high enough, but need not be 1. □

Corollary 6. *If the transparency-security tradeoff is mild enough for both sides, then there is a viable open deal. If the transparency-security tradeoff is too severe for either side, then no deal exists.*

Proof. The condition stated in Proposition 5 requires that a sum of two maxima be no greater than the total value of the game. Observe how the components of those maxima vary in the τ_o^i. The first two components of V_A' are independent of τ_o^A and the third decreases

in it, while the first component of V'_B increases in it and the second and third are constant. The first component of V'_A increases in τ_o^B and the second and third do not depend on it, while the first two components of V'_B are constant but the third decreases in it.

Hence, as one of the τ_o^i increases from τ_c^i while the other is held constant, V'_i decreases but V'_j does not decrease and may increase. If the W_{oi}^j increase slowly enough in the τ_o^i, then the tradeoff is mild and there will be pairs of values of τ_o^A, τ_o^B that satisfy the condition in Proposition 5. However, if one or both of the W_{oi}^j increase rapidly enough in τ_o^i, it will not be possible to satisfy that condition with any values of the τ_o^i. □

3 Additional Empirical Documentation

The Present Costs of Arming

Here, we document the claim in the paper's introduction that the world's military spending in 2016 is large enough to cure certain of the world's ills were it diverted to those uses. The World Bank estimated that in 2013, 767 million people lived on less than $1.90 a day (World Bank 2016, 36). On the assumption that these people average half that income, raising all of them to that threshold would cost no more than $266 billion per year. The World Health Organization estimated that in 2015, 90% of all malaria cases occurred in Sub-Saharan Africa, and 43% of that region's population were not covered by insecticide-treated bed nets (World Health Organization 2016, 14, 16). The World Bank puts the population of that region in 2015 at 1.006 billion, so that the implied unmet need in the region is for 433 million bed nets. Conservatively assuming that coverage and incidence are the same in the rest of the world, the unmet world need is for 481 million bed nets. GiveWell estimates the total cost of providing a bed net at under $10, so that the cost of providing this total is $4.81 billion per year (GiveWell 2018). UNAIDS estimated in 2015 that the peak annual cost of a program to identify at least 90% of people with HIV, and treat at least 90% of them,

would be $31 billion (UNAIDS 2015, 5). Assuming constant returns to scale, identifying and treating everyone would require $38.4 billion. Finally, UNESCO put worldwide gross expenditures on research and development at about $1.7 trillion in 2017 (UNESCO Institute for Statistics 2018). These figures sum to less than $1.2 trillion, which is less than half of the $2.72 trillion the world spent on arms in 2016.

Interwar Arms Control

A possible alternative explanation for the pattern of agreements during this period is that in an effort to avoid future conflict, states sought to limit offensive capabilities while allowing defensive ones. However, we found little support for this explanation in the primary sources we canvassed, including transcripts of the official arms conference meetings and internal discussions among the British admiralty and cabinet and among the US Navy and negotiators).

Some secondary sources (Goldman 1994; Kitching 2003; Carr 2016) discuss this issue, but they do not document that it actually featured substantively in policymakers' calculus. In *The Twenty Years' Crisis, 1919–1939*, historian E.H. Carr points out disputes in later negotiations over which arms were defensive or offensive, with each state arguing that the capabilities on which it most relied were defensive. Other historical texts support this finding: states were interested in identifying "offensive" capabilities to limit, including within the category of naval armaments, but failed to establish mutually-agreed definitions (Hall 1987, 7,123–125). The historical accounts suggest that the issue came up in negotiations, but was not itself a reason for why agreements were made on some armaments but not others. The arms conferences discussed a wide variety of military capabilities, and there was no consensus that the arms ultimately limited by the Washington Naval Treaties were indeed purely offensive. Finally, there is also no theoretical reason to think that the offensive/defensive distinction in and of itself constituted an impediment to arms control. While there might

be a somewhat stronger motivation to limit offensive arms that are seen as destabilizing and increasing the likelihood of costly war, states also have incentives to limit defensive capabilities. Limiting arms saves money and can therefore make all sides better off, regardless of whether the arms are offensive or defensive.

Superpowers' Understanding of the Transparency-Security Trade-off

In the first two empirical settings—US-Iraq and interwar negotiations—we consider essentially the full range of deals that could have been made. During the Cold War, the range of possible bargains across weapon types and across time is much larger, so we take a slightly different empirical approach. We first show that the tradeoff played a role in considerations of comprehensive disarmament in the earliest part of the Cold War. This evidence suggests that the tradeoff was an underlying consideration across numerous weapons types throughout the later Cold War as well. Second, we look at two pairs of cases from different points in the Cold War: Freeze (1964) negotiations vs. SALT I (1979); and failed attempts to limit intermediate missiles (1983-1985) vs. the INF treaty (1987). In both pairs, we show that the theory can account for why negotiations failed in the first instance but were successfully concluded in the second.

In our examination of arms control negotiations between the superpowers during the Cold War, we found ample evidence that policymakers were well aware of the transparency-security trade off and considered it among the most important dimensions of any mooted agreement. We proceed to document these claims, and also to characterize the wider pattern of US and Soviet positions on monitoring in arms control negotiations. This appendix documents and expands on the claims presented in the main body of the article.

Passages contained in the main body are highlighted in bold below, with additional evidence and full citations presented in following paragraphs.

The US and USSR undertook the most expensive and most dangerous arms race in human history. Though there is no one definitive number, indicators of the magnitude of the US-Soviet arms race are not hard to find. Due to a wide range of programs and different ways of segregating spending, the estimates on how much the US spent on its nuclear program during the Cold War vary. The most cited source calculates that the US spent at least 5.5 trillion dollars on the nuclear weapons program, adjusted for 1996 dollars (Schwartz 1998).

There are no formal figures on Soviet nuclear or general military spending. In 1978 the CIA estimated Soviet overall military spending at 11-13 percent of GNP (Central Intelligence Agency, National Foreign Assessment Center 1978). In the late 1980's, independent scholars estimated Soviet spending at 15-17 percent of GNP (Swain 1990). In his work on the Soviet side of the arms race, David Holloway documents estimates of Soviet military spending as rising from the 1950-1970s, relatively stable but possibly still increasing through the detente of the 1970's, and facing greater economic pressure going into the 1980s (Holloway 1983).

From the very beginning of the Cold War, both sides saw cooperative monitoring as generally necessary to meet the transparency requirement for limits on nuclear and conventional forces, but also perceived serious security risks in such monitoring. The 1951 CIA Special Estimate estimate focused on the implication of broad limits on armed forces between the US and USSR. It noted that the types of verification procedures used in an agreement,

> "would be of vital importance in minimizing the risks to the US and compelling maximum disclosure by the USSR. It is apparent that these two objectives are largely contradictory. The more frequent, thorough, and unrestricted the inspec-

> tion the less would be the possibility of Soviet concealment, but the greater would be the risk of sensitive disclosures by the US" (Central Intelligence Agency 1951).

The report's recommendations identify a necessary compromise. There should be limitations on inspection of technical specifications and data in an agreement because some verification approaches, such as unrestricted inspection, exacerbate the problem of protecting US information, even though they are the only way to ensure full Soviet disclosure:

> "Although this limitation on the freedom of inspection would reduce the amount of intelligence the US would acquire, it would be necessary in order to guard against exposure of the highly sensitive information excluded from the census [agreement] and against premature disclosure in an earlier phase of information reserved for a later phase" (Central Intelligence Agency 1951)

.

The clearest sign of the trade off is the report's extensive analysis of which side would gain greater military advantage from monitoring. The report considers factors such as the amount of information each side already has on the other, the relative openness of the two societies, and upcoming technological developments. For example, on atomic research, the report assesses that detailed inspection would allow the USSR to "gain heavily" due to their limited unilateral intelligence in this area at the time. The report also notes that, "In any census and verification, each side would be bound to gain a certain amount of 'collateral' information (data on general conditions in the country, economic conditions, quality of weapons, standards of training, level of maintenance of equipment, etc.) in fields not directly covered by any census phase." In this respect, the US was assessed to gain relatively more from inspections due to the closed nature of the Soviet system. While the conclusions were mixed, the very effort to do this analysis is highly informative, as it shows

that the CIA saw monitoring as having implications for national security which went well beyond the scope of the agreement itself.[5]

Internal Soviet discussions are not readily available, but occasional public statements also reveal an appreciation of the T-S tradeoff. In 1959 the USSR submitted a proposal on disarmament to the UN General Assembly, which sought to address the problem of having inspections before disarmament was completed; though needed, such inspections would also reveal information detrimental to security. The proposal noted the "contradiction between the need to offer foreign inspectors access to military and other facilities and the apprehension that, so long as disarmament was not entirely completed, the information obtained during such inspections might be used to the detriment of national security and that inspections would become a kind of foreign patronage" (Kokeyev and Androsov 1990, 6). The Soviet proposal did not progress towards an agreement, but suggested creating stages of disarmament with a link between the degree of disarmament and degree of verification to be used, including intrusive inspections.

Throughout much of the Cold War, the two sides adopted consistent, opposing positions on monitoring. Typically, the US demanded more thorough monitoring while the Soviet Union refused to countenance any intrusive measures. The historical record provides important indications that this observed behavior was not the product of habit or ideological commitment, but rather explicit, though different, assessments of the transparency-security tradeoff. The following examples support this assessment:

- In 1957 the USSR rejected Eisenhower's early proposals for a Comprehensive Test Ban Treaty and cessation of fissile material production with inspections for monitoring. (Graham and LaVera 2003, 1375) Through early Test Ban negotiations, the US (and

[5]This report was revisited a few years later, in a 1954 memo assessing relative advantages from inspections in key areas such as nuclear and conventional military research. The memo states that the "USSR would attempt to exploit any system of census and verification to its advantage and there is a grave danger that the USSR might succeed in so exploiting it" (Central Intelligence Agency 1954).

UK) pushed for on-site inspections as a way to collect information in cases when a seismic detection system was not able to identify the difference between nuclear tests and an earthquake. The USSR acknowledged the principle of extra information being needed for more effective verification of tests, but argued that inspections should be accepted on a voluntary basis. The USSR believed that mandatory inspections, "would endanger state security by obtaining information not relevant to the purpose of verification – in short, espionage" (Heckrotte 1988, 248–249).

- During negotiations on Limited Test Ban Treaty in the early 1960's, the two sides agreed to inspections (for underground tests) in principle but the US argued for more and the USSR demanded less. The states ended up signing a treaty which banned nuclear tests in the atmosphere, under water, and in space, but did not address underground tests where there was no agreement on verification (Graham and LaVera 2003, 30).

- The USSR rejected early proposals for an Open Skies Treaty, made in 1955 by Eisenhower before the advance of photo-reconnaissance satellites. The USSR claimed that that overflights would be used for spying. (Graham and LaVera 2003, 822).

- The USSR agreed to on-site inspections in the Antarctica Treaty (1959) though notably these were intended to be carried out on Soviet territory and this did not threaten the security of other Soviet military capabilities.(Graham and LaVera 2003, 13). This case also suggests that the USSR did not reject other inspections proposals out of lack of familiarity or knowledge about the process, on principle, or due to ideology.

- In early negotiations on the Chemical Weapons Convention, the USSR accepted the principle of on-site inspections, but argued that they should be voluntary. This position shifted toward accepting inspections from the mid-1980 through CWC signature in 1992 (Leitenberg and Zilinskas 2012, 546).

The USSR assessed particularly high risks in allowing for greater openness, while the US saw greater risks in the possibility of missing Soviet cheating on an agreement. In describing the Soviet position on monitoring in 1963, US analysts noted that the primary Soviet concern was military vulnerability and the additional targeting information that monitoring would reveal. A top secret paper prepared for the Arms Control and Disarmament Agency (ACDA) in 1963 noted that the primary Soviet concern with inspections was military, as "territorial access might yield additional target information and presumably improve capabilities for a counter-force strike." A secondary concern involved the Soviet leadership's "fundamental objection" to allowing an alternative source of authority on Soviet territory, over which they would not have direct control and which would have an explicit goal of limiting the freedom of Soviet government behavior (Arms Control and Disarmament Agency 1963).

US policymakers, on the other hand, were primarily concerned with transparency—would the US detect Soviet cheating on an agreement? Most US assessments of monitoring focus on transparency, evaluating concerns over whether a particular monitoring arrangement would suffice to catch or deter Soviet violations. For example in recent historical work on verification and trust in the US-Soviet relationship, Weever et al. (2016) document a constant concern over Soviet cheating on verification schemes in the early 1980's; and Schors (2016) notes that the SALT II ratification debate in the Senate was focused on disputes of whether the administration would be able to detect Soviet cheating.

The US did not need to devote much effort to assessing the security side of the tradeoff because the Soviets were expected to refuse inspections and so eliminate any security concerns for the US. Indeed, the US sometimes made insincere proposals for inspections that the US would not have considered safe for itself, confident that the Soviets would oppose these and thereby yield the "moral high ground" to the US (Gottemoeller 1988, 35).

However, some internal discussions are indicative of ongoing concerns about security as well. For example, in discussing proposed inspections for the Chemical Weapons Convention

in 1984, the Defense Department supported a proposal for extensive inspections of military or government facilities, while intelligence community, Joint Chiefs, and State Department opposed arrangements that would give the other states access to sensitive information beyond the CW context. The Office of the Secretary of Defense supported a proposal for "unimpeded [inspections] access to all military or government-owned facilities." The intelligence community, Joint Chiefs, and State Department opposed "an arrangement by which the US would be obligated by treaty to provide the Soviet Union, or any other state, access to our sensitive, non-CW related military or intelligence facilities simply by lodging a compliance complaint under the CW treaty. These agencies do not believe the US should table proposals it cannot itself accept, and they oppose making all government facilities subject to mandatory challenge on-site inspection" (Central Intelligence Agency 1984).

Another example, a highly redacted declassified article from the intelligence community's internal journal, *Studies in Intelligence*, notes that while most arms control issues are "policy" decisions made by the National Security Council, the Director of Central Intelligence "is an influential player in these decisions when they have an intelligence dimension, as do all decisions related to verification." The article notes the reality that "final treaty provisions may not always allow maximum feasible monitoring" and so the intelligence community plays a key role in informing Congress on the tradeoffs between observing violations and other intelligence considerations (Central Intelligence Agency 2014).[6]

In an extensive study of the Soviet biological weapons program and the BWC negotiations, Leitenberg and Zilinkas state that in 1991 the US opposed strengthening monitoring provisions for the agreement for security reasons: "Despite the discovery of the enormous Soviet BW program, and the continued expression of uncertainty all through the 1990s as to whether it was all gone, the US government discarded the opportunity to obtain a regime

[6]The article is undated, but was definitely written after 1990, as it refers to multiple treaties including on-site inspections.

that would undertake routine inspection of Russian biodefense facilities as well as those of any other nation that it had concerns about, and the opportunity to mount challenge inspections, primarily to protect its own biodefense program" (Leitenberg and Zilinskas 2012, 561).

Though the USSR appeared to be more concerned with security than transparency, and did not accept an agreement with inspections on its territory prior to the late 1980's, it did sometimes demand inspections. For example, during one stage of the mid-1960's Freeze negotiations discussed subsequently, the USSR proposed inspections of declared missiles and their manufacturing facilities (Arms Control and Disarmament Agency 1963). Presumably this proposal was motivated by the perceived need for inspections to ensure that the USSR would catch US cheating on the Freeze.

Divisions in US domestic politics has traditionally been cited as the culprit arms control efforts. But, this argument can be best understood as internal differences on estimates of both the likelihood and the military significance of Soviet cheating, or in the terms of our model, different assessments of whether temptation to cheat was low or high relative to monitoring. Doves were therefore more likely to accept Soviet demands for low intrusiveness in monitoring, while hawks found these insufficient for ensuring compliance. If it were not for the TS tradeoff, the US and USSR could have simply accepted the monitoring level preferred by the most skeptical side or domestic group in order to reach agreement.

Improvements in Unilateral Monitoring from the Freeze to SALT I

Closed deals were made only under the conditions of Corollary 1: when the temptation to cheat was low relative to unilateral monitoring, so that unilateral monitoring sufficed to assure compliance and the tradeoff was therefore not operative. In the online appendix, we argue that the 1964 Freeze negotiations failed while the subsequent, similar Strategic Arms Limitation Talks (SALT I)

succeeded because the development of satellite surveillance improved unilateral monitoring enough to make a closed deal viable.

The superpowers' negotiations on arms control yielded both open deals and closed deals over the course of the Cold War. The main body of the paper discusses an open deal, the INF Treaty, and we provide additional documentation on this agreement subsequently. Here, we discuss an example of a closed deal the two superpowers made: SALT I. It is instructive to compare the negotiations over this agreement to those over a very similar mooted agreement, the Freeze, that took place fifteen years earlier but failed to produce a deal. Unilateral monitoring was considered insufficient to enforce the Freeze, but enough to make SALT I mutually acceptable. What changed in the intervening fifteen years was the launch of US spy satellites into orbit: these increased the US's ability to unilaterally monitor Soviet compliance and so made a deal possible.

The idea of the 1964 Freeze Proposals was to limit US and Soviet levels of ICBMs, MRBMs, bombers, and missile defense systems at their current levels. (For details on the Freeze Proposal see also Lall (1964)). The US proposals called for extensive on-site inspections of weapons in the territory of the adversary. Consistent with its general pattern throughout the Cold War, the USSR rejected inspections.[7]

Internally, US policymakers also questioned the provisions which US negotiators were supporting publicly in the UN. In his account of negotiations, John Newhouse writes, "A senior American official closely involved with this affair says that the idea of mutual inventory controls by inspection was really no more acceptable to the Americans than to the Russians" (Newhouse 1989). See also (Seaborg 1987, 392–393). US officials noted that inspections to verify a ban on re-entry vehicles "might disclose sensitive U.S. information on warhead or penetration aid characteristics" (Arms Control and Disarmament Agency 1964). US officials

[7]The USSR opposed the freeze on other grounds as well. In 1964, a freeze would have left it at a disadvantage in numbers of weapons (Graybeal and Krepon 1988, 96).

ultimately decided that inspection of these kinds of production facilities would not be in the US interest, but also that without additional observation it would be impossible to assess Soviet compliance, ruling out any deal (Arms Control and Disarmament Agency 1964). At this time, the tradeoff was too severe for either side to accept an agreement, even though inspections would have enabled the two sides to meet the transparency requirement for a Freeze deal.

Subsequent developments in satellite technology increased perceptions of the efficacy of unilateral monitoring, so that each side could observe key elements of the other's arming without inspections. The resulting negotiations led to the 1979 SALT I agreement. In principle, the unrealized Freeze proposals were quite similar to SALT I, differing only in two respects. First, the verification regime was downgraded from onsite inspections to satellite surveillance. During the negotiations, the US dropped its insistence on inspections, a key source of Soviet opposition to an agreement (Bunn 1992, 107). The US became willing to rely on unilateral monitoring because, with the availability of satellite surveillance, the US now perceived this to suffice to ensure Soviet compliance (Parados 1999, 22). Second, and revealingly, Freeze and SALT I differed in precisely what was to be limited. US photo-reconnaissance satellites could reliably observe Soviet launchers, but not the missiles themselves, and so the negotiations shifted to limiting launchers. The SALT I proposals placed limits on silos and submarines, not the missiles themselves, which had been the focus of the Freeze proposals (Graybeal and Krepon 1988, 96). The ability to unilaterally monitor treaty compliance from space rendered a closed deal viable, but it also shifted the precise dimension of arming which could be controlled.

Additional Evidence on the INF Treaty

The negotiations over the INF Treaty, which eliminated all land-based intermediate–range missiles, allow us to test Corollary 2. The INF treaty committed the US and

USSR to eliminate all land-based intermediate-range missiles, as well as some shorter range missiles. The existing and planned deployments of these missiles posed significant material costs for both sides, and were also considered destabilizing in that they increased the likelihood of war (Matlock 2004, 38). The Soviet SS-20 could reach numerous targets in Europe, while the US Pershing 2 could target Soviet leadership in Moscow, with very short flight times that undermine early warning and thereby generated first-strike incentives. To monitor compliance with the ban, the treaty included provisions for reporting, satellite surveillance, and, most importantly, on-site inspections.

Unilateral monitoring was judged insufficient to overcome the temptation to cheat on the INF limits. Earlier agreements covered missiles housed in fixed silos, and so could be reliably monitored unilaterally via satellite. By contrast, INF missiles were smaller and carried on mobile launchers, which could move around easily and at night, escaping satellite detection. Then Director of the U.S. Arms Control and Disarmament Agency explained that while satellite monitoring was a sufficient way to verify the number of silos under SALT, the advances towards smaller, mobile missiles make verification more difficult. "Unlike silos, mobile launchers can be moved around frequently and at night. They can be far more easily concealed"(Adelman 1987).

In the first round of negotiations (1980–1983), the US proposed "anytime anywhere inspections" which would allow nearly unfettered access to any missile-related facility to achieve transparency (Grier 1987).

However, these same inspections also risked exposing important military secrets, so that the transparency-security tradeoff was perceived to be severe. Inspections of Soviet missile production facilities might reveal the technical details and quantity of missiles not prohibited under the agreement. The Soviet SS-25 missile, which was not banned by the treaty and would continue to stay in production, was similar in outward appearance to the banned SS-20 and manufactured in the same facility.

The US wanted to verify that the Soviet facility was not manufacturing banned missiles, but manually inspecting all missiles leaving the facility would reveal far more information about Soviet capabilities than the USSR was willing to accept—and more than the US would want to accept at its own facilities being inspected by the USSR. This issue was a potential deal-breaker for the treaty (Pifer et al. 2012, 12).

On the US side, stealth technology, radar, and industrial processes would be subject to Soviet espionage. Inspections in the US posed the risk of revealing the details of stealth technology and advanced radar, or of industrial processes used in military production, all of which would be of great interest to the Soviets (Toth 1988).

The USSR refused early inspections proposals, believing that they would be used to "discover the strengths, weaknesses, and vital characteristics of Soviet weapons" (Savelev and Detinov 1995, 152).

Nongovernmental experts traditionally supportive of arms control raised concerns that highly intrusive inspections would mean "Soviets crawling around our most sensitive production facilities" and leave the US at a relative disadvantage (Grier 1987). Jack Mendelson posited that, "The administration was slow to appreciate that extensive on-site inspection would have some very unpleasant side effects. To begin with, it would apply to the US and its allies as well as to the Soviet Union. As an INF agreement became an imminent possibility, the press began to report on an intense debate within the administration over the wisdom and feasibility of allowing Soviet inspectors access to sensitive U.S. installations" (Mendelson 1987).

The CIA even recommended doing away with inspections. The US's original position of would have allowed inspectors access to any facility that might have a role in banned missile production. However, in the internal US debate, there were serious concerns about the access this would provide to the Soviets. In light of this risk, the CIA advocated doing away with on-site inspections altogether (Aaron 1987). For more information on the

shift away from anytime anywhere inspections see Garthoff (1994, 327) and Krepon (1989, xvi). At the height of the US internal debate, calls for anytime anywhere inspections were dubbed the "Fruit Loops" proposal because under such terms the USSR would be allowed to inspect facilities as unrelated to the treaty as the Kellogg's Fruit Loop cereal plant. Following the "Fruit Loops" discussions, it became clear to chief INF negotiator Maynard Glitman and others working on verification that the US push for more intrusiveness had to stop short of anytime anywhere inspections (Glitman 2006, 210). Major General William Burns, then director of the Arms Control and Disarmament Agency, testified to Congress in 1988 about the earlier proposals, and noted that: "An anywhere, anytime type of inspection unnecessarily places at risk U.S. national security interests" (Bunn 1992, 161).

During the second round of negotiations (1985–1987), a new approach to inspections was devised and both sides reassessed the severity of the tradeoff. In 1986, the US began to focus on the prospects for more limited inspections, and subsequently proposed intrusive inspections but with access to only one production facility. The Office of the Secretary of Defense requested proposal on ways to monitor production facility portals from Sandia National Laboratory in late 1983, after the first round of negotiations broke off. The research began to focus on specific design concepts in 1986. During this period the lab built a full scale demonstration site to test inspections technologies (U.S. Congress, Office of Technology Assessment 1991).

The US also devised a way to monitor this facility with a sensor that would reveal whether an exiting missile was of the banned type, but not reveal the technical details of non-banned missiles produced at the same location. The US pursued tailored technologies which intentionally limited the information collected in a key Soviet location. Technical experts devised a process of using a type of X-ray sensor to monitor the exit portal of the Soviet production facility which would reveal the presence of a banned missile leaving the facility, but provide no additional information about other

physically similar but non-banned missiles produced at the same location. The SS-20 and SS-25 missiles were identical for the first stage of each missile, but then differed in the dimensions of the second stage. The radiation measurements allowed the the US to confirm the dimensions of the second stage, thus differentiating the two missiles, without opening canisters to visually inspect the missile components (Toivanen 2017, 17). The sensor had the capacity to distinguish materials and was considered resistant to deception (John Adam, "Verification Keeps Ivan Honest," Washington Post, January 24, 1988).

Although it was understood that the USSR would be able to hide some violations under this inspections proposal (Harrison 2012, 80), **the required level of transparency could be still be achieved, and with much lower security risks than the earlier proposal of "anytime anywhere" inspections.**

The emerging availability of these technical tools made it clear to USSR negotiators that risks to security were low. The technology was developed by the US rather than the USSR, so negotiations required US experts to explain and present technical details on these methods to their Soviet counterparts (Toivanen 2017, 18). Ongoing concern about exactly how the scanner was to be implemented even well into treaty implementation indicates the level of attention the Soviets devoted to the reevaluation of the tradeoff. Continued Soviet sensitivity to revealing secrets is evident in later accusations that this "CargoScan" device collected images that were several centimeters larger than allowed (Russell 2001; Harahan 1993).

The USSR also overcame its long-standing opposition to inspections, with serious proposals for on-site inspections emerging during this period. Despite a long history of opposing inspections, in the second round of INF negotiations the Soviet Union actually suggested a highly intrusive protocol, giving inspectors full access to all facilities involved in intermediate-range missile production. Internal Soviet discussions in preparation for the 1986 Reykjavik Summit indicate that the USSR was interested in on-site

inspections for missile production facilities (Kataev 1986).

At one point, the USSR was even supporting a more intrusive option than the US would accept. Gorbachev himself proposed highly intrusive verification in an April 1987 meeting with US Secretary of State George Shultz (Memorandum of Conversation 1987). Some of the Soviet proposal were likely bluffs, but the inclusion of inspection discussion in Kataev's documents does indicate serious consideration; Vitalii Leonidovich Kataev was a high level advisor to President Mikhail Gorbachev, a participant in arms negotiations, senior adviser to the Secretary for the Defense Industry, and the organizer for the Interdepartmental Working Group ("Group of Five" or "Big Five"), a minister or deputy minister level interagency commission tasked with arms control policy in the late 1980's. For more on this group, see Savelev and Detinov (1995). Specifically on monitoring production facilities, Yuli Vorontsov, First Deputy Foreign Minister and head of the Soviet INF negotiations team, was quoted as saying: "working out measures for verifying the fulfillment of the agreement ... we are going even further than the United States ... more specifically, in the field of verification of the enterprises in which medium-range missiles are produced. According to the U.S. proposal the inspectors will be sent only to the gates of the enterprise. We propose that they should also be granted access inside the enterprise" (Mendelson 1987).

As documented in memoirs and Politburo accounts, the political leadership concluded that the military's resistance to inspections had more to do with obfuscating bureaucratic excess and incompetence than with protecting Soviet military power vis-á-vis the US. Inspections of the kind proposed by INF negotiators might embarrass the military, but would have little effect on the balance of power. By the second period of negotiations, analysis done by a number of Soviet agencies came to the new conclusion that earlier fears of inspections had been exaggerated (Savelev and Detinov 1995, 152-153). In late summer 1986, Minister of Foreign Affairs Shevardnadze began to openly argue with military leaders that American demands for inspections (being

made at the time across nuclear and conventional negotiations) were "harmless enough" (Service 2015, 203). In an August 1986 Politburo meeting Oleg Grinevski, the head of European security talks and a security adviser to Gorbachev, argued that inspections in the conventional context would be beneficial in allowing the USSR to observe important information about US forces. The head of the General Staff, General Akhromeev, voiced opposition, noting that inspections of forces in Europe would in particular reveal the weakness and disorganization of Soviet divisions there, and "see all that shame." However, his views were outvoted by the political members of the Politburo, and Akhromeev departed to discuss inspections in conventional arms negotiations with the Americans (Service 2015, 204). As noted above, that fall Gorbachev's advisers began working on inspections proposals for INF in preparation for the Reykjavik summit.

Other sources suggest that political leaders began to see the problem of inspections as not so much about revealing military secrets, but rather about revealing the the poor organizational state of the Soviet military. The Minister of Foreign Affairs, Eduard Shevardnadze, wrote in his memoir that the opposition to inspections could not be "attributed entirely to our desire to gain strategic advantages. Rather, it was basically an effort to hide our weaknesses and inadequacies" (Shevardnadze (1991), cited in Charles (2010, 167)). Shevardnadze is speaking here not about weaknesses that the US might exploit in a future conflict, but rather weaknesses that threatened the positions and power of some Soviet leaders. This posed a problem for specific actors within the state, in particular of course the military leadership. Rather than revealing secrets, allowing inspections would even allow Gorbachev greater oversight over the military-industrial complex. For the fullest discussion, see Charles (2010, 167). See also: Service (2015, 204), Grachev (2008, 99), Kataev (2000).

A mutual understanding that the tradeoff was mild was critical to making the deal. Debates over this issue continued though the very last stages of negotiation, delaying the treaty even though other terms—such as the level and scope of

limitations—were already settled. In November 1987, less than a month before the completion of the agreement, the highest-level officials from both sides met to address the remaining issues, which were centered on verification (Akhromeev and Kornienko 1992, 136). According to the memoirs of key participants, the most contentious issues sat at the crux of the tradeoff: the US wished to exclude some of its own sites from Soviet monitoring (Palazhchenko 1997, 71).

Chief US negotiator Maynard Glitman explicitly referred in his memoir to the verification debate as a tradeoff between observing Soviet behavior and tolerating threats to US military secrets. Glitman writes:

> The United States, of course, had to accept the same monitoring regime that it wished to see applied to the Soviet Union. At times this began to look like a very expensive trade off. Indeed, the closer one attempts to reach 100 percent certainty about the other side's behavior, the greater the need for highly intrusive measures that can put your own non-treaty interests at risk (Glitman 2006, 210).

Glitman continues:

> This situation created the possibility that the effort to reach a negotiated outcome could fail because of the inability to develop a monitoring regime that was sufficiently stringent to at least approach 100 percent certainty that the United States would be able to detect Soviet cheating, but which at the same time did not create an unacceptable risk that the Soviets would, in the course of carrying out their INF inspections of U.S. facilities, obtain critical information on U.S. defenses (Glitman 2006, 211).

The ultimate INF Treaty eschewed anytime-anywhere inspections because they achieved transparency but created unacceptable risks to security. Instead,

this open deal featured inspections which, while still highly intrusive, protected the secrets of key military installations and were therefore safer, in accordance with Corollary 2. The final INF treaty includes some of the most intrusive provisions used in arms control agreements between the US and USSR (and later Russia). The types of inspections are: baseline inspections, elimination inspections; closeout inspections; short-notice inspections; portal monitoring of one production facility on each side. For a detailed account of each inspection type and its implementation, see Harahan (1993).

References

Aaron, David (1987). Verification: Will it work. *The New York Times* (October 11, 1987).

Adelman, Kenneth L (1987). *"Verification In an Age of Mobile Missiles," Address before the City Club, San Diego, California, June 26, 1987.* Washington, D.C.: U.S. Dept. of State, Bureau of Public Affairs, Office of Public Communication, Editorial Division.

Akhromeev, S. F. and G. M. Kornienko (1992). *Glazami marshala i diplomata : kriticheskii vzgliad na vneshniu politiku SSSR do i posle 1985 goda.* Moscow, Russia: Mezhdunarodnye otnosheniia.

Arms Control and Disarmament Agency (1963). *Meeting on Arms Control Issues, 14-15 December, Camp David, Maryland.* Number Entry 5180, Box 2, Folder 2. National Archives at College Park: Office of Politico-Military Affairs – Records Relating to Disarmament and Arms Control, 1961–66.

Arms Control and Disarmament Agency (1964). Meeting of the committee of principals. Office of Politico-Military Affairs – Records Relating to Disarmament and Arms Control, 1961–66.

Bunn, George (1992). *Arms Control by Committee: Managing Negotiations with the Russians.* Stanford University Press.

Carr, E.H. (2016). *The Twenty Years' Crisis, 1919-1939.* London, UK: Palgrave Macmillan.

Central Intelligence Agency (1951). *Special Estimate: Intelligence Implications of a Census and Verification of Armed Forces and Armaments.* FOIA Electronic Reading Room: National Intelligence Council (NIC) Collection. Document Number (FOIA) /ESDN (CREST): 0000170378.

Central Intelligence Agency (1954). *Memorandum for the Director of Central Intelligence: Review of SE-4, Intelligence Implications of Census and Verification of Armed Forces and Armaments, 12 May 1951.* FOIA Electronic Reading Room: General CIA Records. Document Number (FOIA) /ESDN (CREST): CIA-RDP80B01083A000100080066-5.

Central Intelligence Agency (1984). *Memorandum for Director of Central Intelligence: NSC Meeting on Our Draft CW Treaty, 29 March 1984.* FOIA Electronic Reading Room: General CIA Records. Document Number (FOIA) /ESDN (CREST): CIA-RDP86B00420R000200300005-2.

Central Intelligence Agency (2014). Arms control compliance and intelligence. *Studies in Intelligence: The IC's Journal for the Intelligence Professional*. Document Number (FOIA) /ESDN (CREST): 0006002219.

Central Intelligence Agency, National Foreign Assessment Center (1978). *Estimated Soviet Defense Spending: Trends and Prospects*. FOIA Electronic Reading Room: FOIA Collection. Document Number (FOIA) /ESDN (CREST): 0000969746.

Charles, Elizabeth C. (2010). *The Game Changer: Reassessing the Impact of SDI on Gorbachev's Foreign Policy, Arms Control, and US-Soviet Relations*. Ph. D. thesis, The George Washington University.

Garthoff, Raymond (1994). *The Great Transition: American-Soviet Relations and the End of the Cold War*. Washington DC: The Brookings Institution.

GiveWell (2018). Mass distribution of long-lasting insecticide-treated nets (llins).

Glitman, Maynard W (2006). *The Last Battle of the Cold War: An Inside Account of Negotiating the Intermediate Range Nuclear Forces Treaty*. New York, NY: Palgrave Macmillan.

Goldman, Emily O. (1994). *Sunken Treaties: Naval Arms Control Between the Wars*. University Park, PA: Pennsylvania State University Press.

Gottemoeller, Rose (1988). Verifying controls on cruise missiles. In M. Krepon and M. Umberger (Eds.), *Verification and Compliance: A Problem-Solving Approach*. London, UK: The Macmillan Press.

Grachev, Andrei S. (2008). *Gorbachev's Gamble: Soviet foreign policy and the end of the Cold War*. Cambridge, UK: Polity Press.

Graham, Thomas and Damien J. LaVera (2003). *Cornerstones of Security: Arms Control Treaties in the Nuclear Era*. Seattle, WA: University of Washington Press.

Graybeal, Sidney N. and Michael Krepon (1988). On-site inspections. In M. Krepon and M. Umberger (Eds.), *Verification and Compliance: A Problem-Solving Approach*. London, UK: The Macmillan Press.

Grier, Peter (1987). Arms treaty could let 'sanctioned spies' see us weapons plants. *Christian Science Monitor* (March 11, 1987).

Hall, Christopher (1987). *Britain, America and Arms Control 1921-37*. London, UK: Palgrave Macmillan.

Harahan, Joseph P. (1993). *On-Site Inspections Under the INF Treaty*. Washington DC: On-Site Inspection Agency, US Department of Defense.

Harrison, Roger (2012). The endgame in washington. In D. T. Jones (Ed.), *The Reagan-Gorbachev Arms Control Breakthrough: The Treaty Eliminating Intermediate-Range Nuclear Force (INF) Missiles*. Washington, DC: Vellum New Academia Publishing.

Heckrotte, Warren (1988). On-site inspection to check compliance. In Goldblat and Cox (Eds.), *Nuclear Weapon Tests: Prohibition or Limitation?* Oxford, UK: Stockholm Inter-

national Peace Research Institute, Canadian Institute for International Peace and Security, Oxford University Press.

Holloway, David (1983). *The Soviet Union and the Arms Race*. New Haven, CT: Yale University Press.

Kataev, Vitalii Leonidovich (1986). Facts on the soviet position in negotiations on nuclear and space armaments including mikhail gorbachev suggestions on the reykjavik meeting. Vitalii Leonidovich Kataev Papers 1966–1999, Box 5, Folder 22, Hoover Institution, Stanford University.

Kataev, Vitalii Leonidovich (2000). Interview transcript. The Hoover Institution and the Gorbachev Foundation (Moscow) collection, Box 2, Folder 4, Hoover Institution Archives, Stanford University.

Kitching, Carolyn J. (2003). *Britain and the Problem of International Disarmament: 1919–1934*. New York, NY: Routledge.

Kokeyev, Mikhail and Andreĭ Androsov (1990). *Verification–The Soviet Stance: Its Past, Present, and Future*. New York: UNIDIR.

Krepon, Michael (1989). *Arm Control in the Reagan Administration*, Volume Volume X of *W. Alton Jones Foundation Series on Arms Control*. Lanham, MD: University Press of America.

Lall, Betty Goetz (1964). Questions and answers on the us production freeze proposal. *Bulletin of Atomic Scientists 20*(10).

Leitenberg, Milton and Raymond A. Zilinskas (2012). *The Soviet Biological Weapons Program: A History*. Cambridge, Massachusetts: Harvard University Press.

Matlock, Jack F. (2004). *Reagan and Gorbachev: How the Cold War Ended*. New York, NY: Random House.

Memorandum of Conversation (1987). M. S. Gorbachev and U.S. Secretary of State George Shultz. National Security Archive Electronic Briefing Book 238: The INF Treaty and the Washington Summit: 20 Years Later.

Mendelson, Jack (1987). INF verification: a guide for the perplexed. *Arms Control Today 17*(7), 25–29.

Newhouse, John (1989). *Cold Dawn: The Story of SALT*. New York, NY: Holt, Rinehart and Winston.

Palazhchenko, Pavel (1997). *My Years with Gorbachev and Shevardnadze: The Memoir of a Soviet Interpreter*. University Park, PA: Pennsylvania State University Press.

Parados, John (1999). Prague spring and salt: Arms limitation setbacks in 1968. In H. Brands (Ed.), *The Foreign Policies of Lyndon Johnson: Beyond Vietnam*. College Station, Texas: Texas A&M University Press.

Pifer, Steven , Avis T. Bohlen, William F. Burns, and John Woodworth (2012). *The Treaty on Intermediate-Range Nuclear Forces: History and Lessons Learned*. Washington, DC: Brookings Institution Press.

Russell, John (2001). On-site inspections under the INF treaty: A post-mortem. VERTIC Breifing Paper 01/02, London, UK.

Savelev, Alexander G. and Nikolay N. Detinov (1995). *The Big Five: Arms Control Decision-Making in the Soviet Union*. Westport, Conn.: Praeger.

Schors, Arvid (2016). Trust and mistrust and the american struggle for verification of the strategic arms limitation talks, 1969-1979. In M. Klimke, R. Kreis, and C. F. Ostermann

(Eds.), *Trust, but Verify : The Politics of Uncertainty and the Transformation of the Cold War order, 1969-1991*. Stanford, CA: Stanford University Press.

Schwartz, Stephen I. (1998). *Atomic Audit: The Costs and Consequences of U.S. Nuclear Weapons since 1940*. Washington, DC: Brookings Institution Press.

Seaborg, Glenn T. (1987). *Stemming the Tide: Arms Control in the Johnson Years*. Lexington, MA: Lexington Books.

Service, Robert (2015). *The End of the Cold War 1985–1991*. New York, NY: Public Affairs.

Shevardnadze, Eduard (1991). *The Future Belongs to Freedom*. London, UK: Sinclair-Stevenson LTD.

Swain, D. Derk (1990). The soviet military sector. In H. S. Rowen and J. Charles Wolf (Eds.), *The Impoverished Superpower: Perestroika and the Soviet Military Burden*. San Francisco: Institute for Contemporary Studies.

Toivanen, Henrietta (2017). The significance of strategic foresight in verification technologies: A case study of the INF treaty. Technical Report LLNL-TR-738786, Center for Global Security and Research, Lawrence Livermore National Laboratory, Livermore, CA.

Toth, Robert C. (January 10, 1988). Verification of arms pact may open spy door. *Los Angeles Times*.

UNAIDS (2015). *AIDS by the numbers: 2015*. UNAIDS.

UNESCO Institute for Statistics (2018). How much does your country invest in r and d?

U.S. Congress, Office of Technology Assessment (1991). Verification technologies: Managing research and development for cooperative arms control monitoring measures. Technical Report OTA-EK-488, Washington, DC.

Weever, Nicholas J. , Joshua Baker, and Laura Considine (2016). Trust or verification? accepting vulnerability in the making of the INF treaty. In M. Klimke, R. Kreis, and C. F. Ostermann (Eds.), *Trust, but Verify : The Politics of Uncertainty and the Transformation of the Cold War Order, 1969-1991*. Stanford, CA: Stanford University Press.

World Bank (2016). *Poverty and Shared Prosperity 2016: Taking on Inequality*. Washington: World Bank.

World Health Organization (2016). *2016 World Malaria Report: Summary*. Geneva: World Health Organization.