Disarmament Games

Yuan Deng and **Vincent Conitzer**
Department of Computer Science
Duke University
Durham, NC 27708, USA
{ericdy,conitzer}@cs.duke.edu

Abstract

Much recent work in the AI community concerns algorithms for computing optimal mixed strategies to commit to, as well as the deployment of such algorithms in real security applications. Another possibility is to commit *not* to play certain actions. If only one player makes such a commitment, then this is generally less powerful than completely committing to a single mixed strategy. However, if players can alternatingly commit not to play certain actions and thereby iteratively reduce their strategy spaces, then desirable outcomes can be obtained that would not have been possible with just a single player committing to a mixed strategy. We refer to such a setting as a *disarmament game*. In this paper, we study disarmament for two-player normal-form games. We show that deciding whether an outcome can be obtained with disarmament is NP-complete (even for a fixed number of rounds), if only pure strategies can be removed. On the other hand, for the case where mixed strategies can be removed, we provide a folk theorem that shows that all desirable utility profiles can be obtained, and give an efficient algorithm for (approximately) obtaining them.

Introduction

Disarmament is often a desired objective in international relations, but it is not always easy to reach the end goal. A key problem is that by removing military assets, a country may leave itself vulnerable to attack if the other country does not disarm. Therefore, disarmament typically happens in a sequence of carefully designed stages, so that neither country is ever too exposed at any stage. Besides reductions in military assets, we can also take disarmament as a metaphor for other strategic situations. For example, two companies may each hold a portfolio of patents that could be used to inflict significant damage on the other company, and the companies may wish to make a sequence of legal agreements to reduce the risk on both sides. In these situations, once one of the players deviates from the disarmament protocol, the possible actions each side has remaining can strategically interact in complex ways to determine the final payoffs realized.

In order to achieve a high level of generality, in this paper, we consider disarmament in general two-player normal-form games, as illustrated by the following example.

Example 1 (Extended Prisoner's Dilemma)**.** *Consider the following modified version of the prisoner's dilemma.*

	Cooperate	Defect	Painful
Cooperate	3,3	0,4	0.1,0
Defect	4,0	1,1	0.5,0.5
Painful	0,0.1	0.5,0.5	0,0

Table 1: Payoff matrix of Extended Prisoner's Dilemma

Strategy Defect strictly dominates strategies Cooperate and Painful for both players. Thus, the only Nash (or even correlated, or coarse correlated) equilibrium of this game is (Defect, Defect), with utilities $(1, 1)$*. If a single player can commit, this does not help, because the other player would still play Defect. On the other hand, suppose both players can alternatingly remove their strategies and act according to the following protocol. In the first round, Row removes Defect. Since Defect is still Column's dominant strategy, the only Nash equilibrium of the reduced game is (Painful, Defect) with utilities* $(0.5, 0.5)$*. Next, Column removes his strategy Defect. In the remaining game, Cooperate has become a dominant strategy, resulting in utilities* $(3, 3)$*. At each step in this protocol of removing strategies, deviating from the protocol and playing the game remaining at that point is dominated by* $(3, 3)$*. Thus, both players are best off following the disarmament protocol.*

In this paper, we first formalize the idea of a disarmament game, as played on top of a game represented in normal form (as illustrated in Example 1). We introduce the computational problem DISARM, which asks whether there is an equilibrium of the disarmament game leading to some desired specified outcome, and a variant K-DISARM in which there are only K rounds of disarmament. We show both problems to be NP-complete. We then introduce a *mixed disarmament* variant that allows the removal of mixed strategies, by upper-bounding the probabilities on individual pure strategies. Here our results are positive: we show a type of folk theorem holds (without repetition of the game!), namely that for any feasible utilities that exceed players' security levels, there is an equilibrium achieving at least those utilities. Our proof is constructive, and in fact shows that we can approximately obtain the desired result in approximate equilibrium using only few rounds of disarmament.

Related Work

Game-theoretic commitment, especially to mixed strategies, has received significant recent attention in the multiagent systems literature, in large part due to their application in various security domains (Tambe 2011). In a two-player normal-form game, an optimal mixed strategy to commit to can be found in polynomial time (Conitzer and Sandholm 2006; von Stengel and Zamir 2010), though there are hardness results for Bayesian and extensive-form games (Conitzer and Sandholm 2006; Letchford, Conitzer, and Munagala 2009; Letchford and Conitzer 2010; Bošanský et al. 2015). There is also significant interest in other types of action one can take before the game in order to make the outcome more favorable. This includes promising payments if a certain outcome is reached (Monderer and Tennenholtz 2004; Anderson, Shoham, and Altman 2010; Deng, Tang, and Zheng 2016) or getting to choose from multiple possible utilities in the entries (Brill, Freeman, and Conitzer 2016). Also, *mechanism design* involves a special player, the designer, committing to the mechanism beforehand. But in all these cases, there is only one commitment step before the game is played, unlike in this paper.

Definitions

We define disarmament games based on the normal-form representation, as in the example provided in the introduction. We restrict attention to two-player games throughout. A normal-form game is defined by $G = \langle S_0, S_1, u_0, u_1 \rangle$, such that for each player b, his set of pure strategies is S_b and his utility function is $u_b : S_0 \times S_1 \to \mathbb{R}$, where $u_b(s_0, s_1)$ denotes player b's utility when player 0 plays s_0 and player 1 plays s_1. Moreover, for $T_0 \subseteq S_0$ and $T_1 \subseteq S_1$, the game induced by T_0 and T_1 is the two-player normal-form game $G_{T_0,T_1} = \langle T_0, T_1, u_0, u_1 \rangle$, where u_0 and u_1 are restricted to $T_0 \times T_1$. As usual, we use $-b$ to denote the player other than b. For convenience, the utility for each player is normalized into the interval $[0, 1]$.

We now define the disarmament game $G_D(G)$ on top of this normal-form game. This disarmament game consists of a disarmament stage during which players alternatingly remove nonempty sets of strategies from S_0 and S_1, and a game play stage—triggered when a player removes nothing—during which they play whatever normal-form game remains. Note that many disarmament sequences can result in the same state $[T_0, T_1, b]$ (where b is the player to move); rather than duplicate this state many times in the game tree, we represent the tree as a directed acyclic graph (DAG) in which each state occurs only once. Note the game is one of perfect information, except that the players move simultaneously in the game play stage. We now present the extensive form of the game precisely.[1]

Definition 1 (Disarmament Game)**.** *The disarmament game $G_D(G)$ is defined as an extensive-form game as follows.*

- *The set of disarmament actions A: $\{X_0 \mid \emptyset \neq X_0 \subsetneq S_0\} \cup \{X_1 \mid \emptyset \neq X_1 \subsetneq S_1\} \cup \{Play\}$, where X_b denotes the set of strategies to keep and Play denotes ending the disarmament stage;*
- *The set of non-terminal nodes H: $\{[T_0, T_1, b] \mid \emptyset \neq T_0 \subseteq S_0, \emptyset \neq T_1 \subseteq S_1, b \in \{0, 1\}\}$;*
- *The set of terminal nodes Z: $\{[T_0, T_1] \mid \emptyset \neq T_0 \subseteq S_0, \emptyset \neq T_1 \subseteq S_1\}$;*
- *The player selection function $\rho : H \to \{0, 1\}$: $\rho([T_0, T_1, b]) = b$;*
- *The available-actions function $\chi : H \to 2^A$, where : $\chi([T_0, T_1, b]) = \{X_b \mid \emptyset \neq X_b \subsetneq T_b\} \cup \{Play\}$;*
- *The successor function $\gamma : H \times A \to H \cup Z$:*
  - *$\gamma([T_0, T_1, 0], X_0) = [X_0, T_1, 1]$;*
  - *$\gamma([T_0, T_1, 1], X_1) = [T_0, X_1, 0]$;*
  - *$\gamma([T_0, T_1, b], Play) = [T_0, T_1]$;*
- *The root of the game is root $= [S_0, S_1, 0]$.*

In the terminal nodes $z = [T_0, T_1]$, players 0 and 1 play the normal-form game G_{T_0,T_1}, resulting in their final utilities.

Definition 2 (Strategy in G_D)**.** *A strategy $\sigma_b = (\alpha, \beta)$ for G_D consists of disarmament strategy $\alpha \in \prod_{h \in H \mid \rho(h)=b} \chi(h)$ for non-terminal nodes and play strategy $\beta \in \prod_{z=[T_0,T_1] \in Z} \Delta(T_b)$ for terminal nodes (where $\Delta(T_b)$ is the set of distributions over T_b).*

Note we restrict our attention to deterministic behavior during the (perfect-information) disarmament stage.

Definition 3 (On-path history & outcome)**.** *For a strategy profile (σ_0, σ_1) with $\sigma_0 = (\alpha_0, \beta_0)$ and $\sigma_1 = (\alpha_1, \beta_1)$, denote the* on-path history *by $P = (h_0 = root, h_1, \cdots, h_K, h_{K+1} = z = [T_0, T_1])$ where for all i, $\gamma(h_i, \alpha_{\rho(h_i)}(h_i)) = h_{i+1}$. The* outcome *of the strategy profile (σ_0, σ_1) is $(\beta_0(z), \beta_1(z))$.*

We say an on-path history has length K if it contains K non-terminal nodes, excluding the root. In a slight abuse of notation, let $u_b(o)$ be player b's (expected) utility for outcome o. A strategy profile (σ_0, σ_1) forms a Nash equilibrium if and only if no player can increase his utility by deviating to another strategy.

We consider the following computational problem:

Definition 4 (DISARM problem)**.** *In* DISARM*, given a disarmament game $G_D(G)$ and an outcome $o^* = (\beta_0^*, \beta_1^*)$, the objective is to determine whether there exists a Nash equilibrium (σ_0^*, σ_1^*) such that the outcome is o^*.*

We also consider a variation of DISARM problem, called K-DISARM.

Definition 5 (K-DISARM problem)**.** *In K-*DISARM*, given a disarmament game $G_D(G)$ and an outcome $o^* = (\beta_0^*, \beta_1^*)$, the objective is to determine whether there exists a Nash equilibrium (σ_0^*, σ_1^*) such that the outcome is o^* and the length of its induced on-path history is at most K.*

[1] Formally, what we present is not exactly the extensive form because (1) we use a DAG rather than a tree and (2) the terminal nodes are associated with the game to be played in the game-play stage rather than directly with utilities, but it is straightforward to extract the formal extensive form from this.

Computational Complexity

Since the number of game states is exponential, an efficient algorithm cannot output the entire strategy profile under the standard representation that lists each player's action at every game state. We next show that a restricted class of strategies that can be represented efficiently suffices. These strategies directly specify the on-path behavior and require that minimax strategies are used off-path.

On-path Histories are Sufficient

Recall that the *on-path history* of a profile of strategies in an extensive-form game consists of the actions that the players take when nobody deviates. For every non-terminal node in the on-path history, one can define the security level for each player.

Definition 6 (Security level)**.** *The security level sec_b for player b in a two-player normal-form game $G = \langle T_0, T_1, u_0, u_1 \rangle$ is the utility that player b can guarantee himself no matter how the other player plays. Formally,*

$$sec_b(G_{T_0,T_1}) = \max_{\beta_b \in \Delta(T_b)} \min_{\beta_{-b} \in \Delta(T_{-b})} u_b(\beta_b, \beta_{-b})$$

We are going to show that the security level is the essential quantity to determine whether an on-path history can be induced by a Nash equilibrium strategy profile or not. In order for a specific outcome to be reached in Nash equilibrium, every player must have strong enough incentive to follow the on-path history, rather than deviating, at any point in the on-path history. To minimize incentive to deviate, we may assume that if someone deviates, the other player will act to minimize the deviator's utility. This does not prevent the latter player's strategy from being a best response, because such punishment will not occur on the path of play. The most effective punishment will be to bring the deviator down to his security level, which is possible by the minimax theorem (von Neumann 1928).

Lemma 1. *P is an on-path history induced by a Nash equilibrium strategy profile with outcome o if and only if for each non-terminal node $[T_0, T_1, b] \in P$, $sec_b(G_{T_0,T_1}) \leq u_b(o)$.*

Proof. "⇒": If there exists a non-terminal node $h = [T_0, T_1, b] \in P$ such that $sec_b(G_{T_0,T_1}) > u_b(o)$, then at the node h, player b can deviate to choose Play and play a strategy in $\arg\max_{\beta_b \in \Delta(T_b)} \min_{\beta_{-b} \in \Delta(T_{-b})} u_b(\beta_b, \beta_{-b})$ in the induced game to guarantee himself a utility at least $sec_b(G_{T_0,T_1})$, which is larger than $u_b(o)$.

"⇐": If for each non-terminal node $h = [T_0, T_1, b] \in P$, $sec_b(G_{T_0,T_1}) \leq u_b(o)$, consider a strategy profile that specifies to:

- For nodes in the on-path history, choose the action to follow the on-path history; and for every non-terminal node not in the on-path history, choose Play;
- For an off-path terminal node $z = [X_b, T_{-b}]$ with $\emptyset \neq X_b \subseteq T_b$ (where this node was reached because b deviated from the path), player $-b$ plays $\arg\min_{\beta_{-b} \in \Delta(T_{-b})} \max_{\beta_b \in \Delta(X_b)} u_b(\beta_b, \beta_{-b})$.

We claim that such a strategy profile forms a Nash equilibrium with outcome o. Suppose, for the sake of contradiction, that player b benefits from deviating to another strategy resulting in induced on-path history P'. Let the longest common prefix of P and P' be P_{pre}, ending with $h^* = [T_0, T_1, b]$. Thus, player b deviates at h^*. Then, the disarmament stage ends in at most one round after h^*, since either player b deviates to choose Play, or player b chooses a different subset X_b to keep but player $-b$ chooses Play immediately after that. Therefore, the induced game is $G_{X_b,T_{-b}}$ with $\emptyset \neq X_b \subseteq T_b$, and $-b$ will act to minimize b's utility. By the minimax theorem,

$$\begin{aligned} & \min_{\beta_{-b} \in \Delta(T_{-b})} \max_{\beta_b \in \Delta(X_b)} u_b(\beta_b, \beta_{-b}) \\ = & \max_{\beta_b \in \Delta(X_b)} \min_{\beta_{-b} \in \Delta(T_{-b})} u_b(\beta_b, \beta_{-b}) \\ \leq & \max_{\beta_b \in \Delta(T_b)} \min_{\beta_{-b} \in \Delta(T_{-b})} u_b(\beta_b, \beta_{-b}) \leq u_b(o) \end{aligned}$$

□

This proves that we can restrict our attention to strategy profiles that explicitly specify on-path play and implicitly assume immediate Play and minimax punishment of deviators off-path. Moreover, note that the length of an on-path history is $O(|S_0| + |S_1|)$ and the security level can be computed via a linear programming. Hence, an on-path history serves as a polynomial-length, polynomially verifiable certificate for the DISARM (or K-DISARM) problem, which is hence in NP.

Complexity of K-DISARM and DISARM

In an on-path history with length K, the total number of turns in which a player chooses an action other than Play is at most K. For $K \leq 2$, K-DISARM is in P since there is a unique on-path history leading to the node o^*, namely the one where each player removes all the strategies he needs to remove in his one (non-Play) disarmament turn. We now show that 3-DISARM is NP-complete by a reduction from the BALANCED-VERTEX-COVER problem.

Definition 7. *In* VERTEX-COVER*, the objective is to check in graph (V, E) whether there exists a subset of the vertices $V' \subseteq V$, with $|V'| = L$, such that every edge $e \in E$ has at least one of its endpoints in V'.* BALANCED-VERTEX-COVER *is the special case of* VERTEX-COVER *in which $L = |V|/2$.*

BALANCED-VERTEX-COVER is NP-complete via a reduction from VERTEX-COVER (Conitzer and Sandholm 2006).

Theorem 1. 3-DISARM *is* NP*-complete.*

Proof. Given an instance of BALANCED-VERTEX-COVER with $|V| = n$ and $|E| = m$, we construct a two-player normal-form game $G(V, E) = \langle S_0, S_1, u_0, u_1 \rangle$, in which $S_0 = \{\ell\} \cup V$ and $S_1 = \{\ell\} \cup V \cup E$. The utilities are defined as follows:

- $U_0(\ell, \ell) = U_1(\ell, \ell) = 1 - \frac{2}{n}$;
- $U_0(v, \ell) = 2$ for all $v \in V$;

- $U_0(v, v') = 1$ for all $v, v' \in V$ and $v \neq v'$;
- $U_1(\ell, e) = 2$ for all $e \in E$;
- $U_1(v, e) = 1$ for all $e \in E$, $v \in V$ and $v \notin e$;

and unspecified utilities are simply 0. The desired outcome o^* is (ℓ, ℓ).

If there exists a BALANCED-VERTEX-COVER with vertices V^* then the on-path history with $h_1 = [\{\ell\} \cup V^*, S_1, 1]$, $h_2 = [\{\ell\} \cup V^*, \{\ell\} \cup V, 0]$ and $h_3 = [\{\ell\}, \{\ell\} \cup V, 1]$ is induced by a Nash equilibrium strategy profile with outcome o^* by Lemma 1. This is because if player 0 deviates in the first round, player 1 can punish player 0 by playing any strategy $e \in E$; if player 1 deviates in the second round, player 0 can punish player 1 by playing uniformly from V^*, which will result in a utility of 0 for player 1 at least $2/n$ of the time because V^* is a vertex cover of size $n/2$; and if player 0 deviates in the third round, player 1 can punish player 0 by playing uniformly from V^*, which will result in a utility of 0 for player 0 at least $2/n$ of the time.

For the other direction, suppose there exists an on-path history leading to o^* induced by a Nash equilibrium strategy profile. Note that (ℓ, ℓ) cannot be a Nash equilibrium in the induced game if for the terminal node $z = [T_0, T_1]$, $V \cap T_0 \neq \emptyset$ or $E \cap T_1 \neq \emptyset$. Therefore, we have $T_0 = \{\ell\}$ and $T_1 \subseteq \{\ell\} \cup V$. In 3-DISARM, there are at most 4 different induced games in the on-path history. These must be G_{S_0,S_1}, G_{S_0',S_1}, G_{S_0',T_1} and $G_{\{\ell\},T_1}$, where $S_0' = \{\ell\} \cup V'$ for some V'. According to Lemma 1, we must have $sec_1(G_{S_0',S_1}) \leq 1 - \frac{2}{n}$ and $sec_0(G_{S_0',T_1}) \leq 1 - \frac{2}{n}$.

In G_{S_0',S_1}, since $U_1(v, e) = 1$ for all $e \in E$, $v \in V$ and $v \notin e$, if there exists an $e' \in E$ uncovered by V', then player 1 can deviate to Play and then play strategy e' to guarantee himself utility 1. Thus, V' must be a vertex cover. As for G_{S_0',T_1}, if $|V'| > n/2$, then player 0 can play uniformly among the strategies in V' to guarantee himself utility $1 - \frac{1}{|V'|} > 1 - \frac{2}{n}$. Thus, $|V'|$ cannot be larger than $n/2$. Therefore, V' is a solution to the BALANCED-VERTEX-COVER problem. □

The fact only 3 rounds are available is essential for the reduction above to work: if there are more rounds, disarmament may be possible even without a balanced vertex cover, by alternatingly removing vertices for player 0 and edges for player 1 while ensuring that the remaining vertices for player 0 form a vertex cover *for the remaining edges only*. However, with a number of modifications to the reduction, we can prove that any successful disarmament of the modified game requires a balanced vertex cover for all the edges at some point in the process. We defer all remaining proofs to the full version due to space limitations,

Theorem 2. DISARM *is* NP*-complete.*

Mixed Disarmament

So far, we have considered only removing pure strategies. But pure disarmament has its limitations. Consider the standard prisoner's dilemma:

If either player removes his Defect strategy, the other player has no motivation to do the same: he would prefer to play Defect. The former player, anticipating this, would not remove Defect, either.

	C	D
C	3,3	0,4
D	4,0	1,1

Table 2: Payoff matrix for the prisoner's dilemma

However, now suppose that the players are able to remove *mixed* strategies. It turns out under this setting, there is a way to get to cooperation.

Example 2. *Consider the prisoner's dilemma as presented above, and suppose players can reduce their strategy spaces by limiting the maximum probability they can put on D. No matter how players reduce their strategy space, it is always a dominant strategy to put as much probability on Defect as possible. Therefore, in any sequence of disarmament steps, the player who should take the* last *disarmament step that reduces the probability he can put on D will have no incentive to do so. As a result it is impossible to reduce the probability that either player puts on D at all.*

This is reminiscent of how in a finitely *repeated prisoner's dilemma, cooperation cannot be attained. But the same is not true for* infinitely *repeated prisoner's dilemma, per the folk theorem. It turns out we can make a similar move here. Suppose that after each disarmament step, we flip a coin that comes up Heads with probability δ. If it comes up Heads, the players stop disarming and play the game. Otherwise, disarmament continues. In this way, neither player ever knows that she is about to take the last disarmament step.*

Specifically, consider the following procedure. In each disarmament step, each player reduces the maximum probability he can put on D by a factor $k \in (0, 1)$, so that once each player has taken t disarmament steps, he can put at most probability k^t on D. Once the coin comes up Heads, of course both players will put maximum probability on D. For player 0, the expected utility of continuing to follow the protocol, given that both players have already reduced to k^t, is

$$\sum_{t'=0}^{\infty} (1-\delta)^{2t'} \delta u_0(k^{t+1+t'}, k^{t+t'}) + \sum_{t'=0}^{\infty} (1-\delta)^{2t'+1} \delta u_0(k^{t+1+t'}, k^{t+1+t'})$$

where $u_0(p, q) = 1pq + 0(1-p)q + 4p(1-q) + 3(1-p)(1-q)$ is the utility to player 0 when the players play the game with limits p and q on D, respectively. On the other hand, if player 0 deviates at this point and is immediately punished by player 1 choosing Play, then 0 obtains utility at most $u_0(k^t, k^t)$. In order for player 0 to be best off following the protocol, the former needs to be no less than the latter; after simplification, we obtain $(3\delta - 2)(k - 1) \geq 0$. Since $k < 1$, we must have $\delta \leq \frac{2}{3}$. Similarly, for player 1, after simplification, we have $(k-1)(k - \frac{1}{3(1-\delta)}) \leq 0$. Since $\delta \leq \frac{2}{3}$, we have $1 - \delta \geq \frac{1}{3}$. Therefore, $\frac{1}{3(1-\delta)} \leq k < 1$.

To avoid dealing with nasty expressions involving δ in what follows, we observe that in the limit as $\delta \to 0$, the utility for following the protocol converges to 3, which is the utility that would be obtained after "infinitely many rounds of disarmament." We can use this fiction—that following the protocol results in playing the game after infinitely many rounds of disarmament—to facilitate the analysis. Specifically, in this infinite-length model, all that is needed to ensure that nobody deviates is that the utility for deviating never exceeds 3. (In Lemma 2, we prove formally that as $\delta \to 0$, we reach the infinite-length model in the limit.)

Example 3. *Consider now the following slightly different version of the prisoner's dilemma in which a player receives utility* 12 *instead of* 4 *when he plays* D *while the other player plays* C*. As it turns out, in this game it is* not *possible to reach the* (C, C) *outcome with the infinite-disarmament model. The problem is that if we were reach a point where both players have achieved certain values for the upper bound on* D*—for example,* $1/2$*—then deviation at that point would give at least* $0.25 \cdot 3 + 0.25 \cdot 12 + 0.25 \cdot 1 = 4$*, which is more than the* 3 *from reaching* (C, C)*. Moreover we cannot "jump over" these values. On the other hand, this is not really a problem, in the sense that both players would actually be better off with this profile than with* (C, C)*. Thus, this example is consistent with the following claim: any sufficiently high utilities that occur in the game can be obtained via disarmament.*

Specifically, we can attain utilities $(4, 4)$ *for the players with the following disarmament sequence (in the infinite-length model). Let* p_t *be the maximum probability on* D *after* $2t$ *disarmament steps, initialized to* $p_0 = 1$*. Both players take turns to reduce the maximum probability to* $p_{t+1} = \frac{9p_t - 1}{8p_t + 3}$*, which converges to* $\frac{1}{2}$ *as* $t \to \infty$*. In this case, player* 1*'s utility from deviation after* t *steps is exactly* 4 *(so player* 1 *is indifferent*[2]*) while player* 0*'s incentive to deviate is smaller.*

Motivation. In practice, it is easy to imagine how one could commit to not play a given pure strategy, but perhaps this is more difficult for the case of mixed strategies. Of course, the issue is similar for the case of committing *to* a mixed strategy, and yet this has found plenty of applications. Commitment to a mixed strategy is usually achieved by reputation, for example by the follower observing the leader's actions many times before acting. The same might work in our model. Alternatively, one could look for some unrelated random variable, such as the weather. If one can make it so that one cannot play a given pure strategy if it is (say) cloudy, then this upper-bounds the probability of the strategy at 70% (assuming it is cloudy 30% of the time). Finally, rather than interpreting mixed strategies literally as probability distributions, one can think of these probabilities as representing a quantitative choice, such as how much of one's continuum of resources to devote to a particular course of action, and then placing upper bounds on this.

[2]Player 1 can be made to strictly prefer not deviating by reducing the upper bounds at a slightly slower rate.

Formal Definition and Basic Properties

In mixed disarmament, each player can reduce his strategies by reducing the maximum probabilities on each of his pure strategies (with the requirement that at least one of these decreases strictly), or choose *Play*. For convenience, assume $S_0 = S_1 = [n]$ and let $\mathbf{0}$ (and $\mathbf{1}$) denote a vector of all zeroes (ones). Therefore, the representation of the game state becomes $[p_0, p_1, b]$ where $p_b \in \Delta^+(S_b)$ is the vector of probability upper bounds for player b. Here, $\Delta^+(S_b)$ denotes vectors of probabilities summing to *at least* 1. In the induced game G_{p_0,p_1}, player b's strategy β_b must satisfy $\beta_b \leq p_b$ point-wisely.

Strategies, on-path histories, and outcomes in mixed disarmament games are defined similarly as in pure disarmament games. For convenience, let $C(p) = \{\mu \mid |\mu|_1 = 1 \text{ and } \mathbf{0} \leq \mu \leq p\}$ denote the set of probability distributions consistent with upper bounds p. The security level in a mixed disarmament game is defined as follows.

Definition 8 (Security level in mixed disarmament game)**.** *The security level* sec_b *for player* b *in a two-player normal-form game* $G = \langle p_0, p_1, u_0, u_1 \rangle$ *is the utility that player* b *can guarantee himself no matter how the other player plays. Formally,*

$$sec_b(G_{p_0,p_1}) = \max_{\beta_b \in C(p_b)} \min_{\beta_{-b} \in C(p_{-b})} u_b(\beta_b, \beta_{-b})$$

Note that in game G_{p_0,p_1}, $C(p_0) \times C(p_1)$ is a compact convex set. Therefore, the minimax theorem still holds for G_{p_0,p_1} (von Neumann 1928), and so does Lemma 1. Now, we are ready to demonstrate that as $\delta \to 0$, we reach the infinite-length model in the limit. Throughout, ε-equilibrium denotes approximate Nash equilibrium in the (standard) additive sense.

Lemma 2. *Consider a Nash equilibrium* $(\sigma_0, \sigma_1) = ((\alpha_0, \beta_0), (\alpha_1, \beta_1))$ *of the infinite-length disarmament game, leading to*[3] *terminal node* $z = [p_0^*, p_1^*]$*. Then for any* $\epsilon > 0$*, there exists some* $D > 0$ *such that for any* $0 < \delta < D$*, we have that* (σ_0, σ_1) *is an* ϵ*-equilibrium of the* δ*-coin-toss disarmament game (where* $(\beta_0(z), \beta_1(z))$ *is played when the coin toss lands Heads).*

A Folk Theorem for Mixed Disarmament

The folk theorem for repeated games shows that any utilities that exceed the players' security levels can be obtained as an equilibrium of the infinitely repeated game (as the discount factor approaches 1). We now show a similar result for (non-repeated) mixed disarmament games. (For pure disarmament games, the prisoner's dilemma provides a counterexample.) Let $G_D^M(G)$ denote the infinite-length mixed disarmament game resulting from normal-form game G, and let utilities (v_0, v_1) be *feasible* for G if there exists a mixed-strategy profile of G that results in these utilities.

Theorem 3. *In mixed disarmament games* $G_D^M(G)$*, for all feasible utilities* (v_0, v_1) *with* $v_0 > sec_0(G)$ *and* $v_1 >$

[3]Here, *leading to* means either that we terminate at this node after finitely many rounds, or that the disarmament stage continues forever but the upper bounds converge to $[p_0^*, p_1^*]$.

$sec_1(G)$, there exists a Nash equilibrium of $G_D^M(G)$ such that the terminal node $z = [p_0, p_1]$ of its induced on-path history satisfies $sec_0(G_{p_0,p_1}) = v_0$ and $sec_1(G_{p_0,p_1}) = v_1$.

For any (β_0, β_1) constituting an equilibrium of G_{p_0,p_1}, we have $u_b(\beta_0, \beta_1) \geq sec_b(G_{p_0,p_1})$; otherwise, player b can deviate to the strategy that guarantees his security level.

Corollary 1. *In mixed disarmament games $G_D^M(G)$, for all feasible utilities (v_0, v_1) with $v_0 > sec_0(G)$ and $v_1 > sec_1(G)$, there exists a Nash equilibrium of $G_D^M(G)$ such that its outcome o satisfies $u_0(o) \geq v_0$ and $u_1(o) \geq v_1$.*

In Theorem 3 and Corollary 1, the induced on-path history of the Nash equilibrium may have infinite length. Of course, by Lemma 2, we can come arbitrarily close to this with a game that will terminate in finite time with probability 1.

A Constructive Proof

In this subsection, we provide a constructive proof for Theorem 3 via an algorithm to generate the corresponding on-path history. Following the intuition provided in Example 3, in each round, the player reduces his upper bounds as much as possible, i.e., until his opponent's security level equals his target utility.

Formally, given target utilities (v_0, v_1) with corresponding strategy profile (β_0, β_1), for each player b, let $\beta'_b = \mathbf{1} - \beta_b$. We introduce two parameters θ_0 and θ_1, initialized to 1, such that the upper bound vectors are $p_0 = \beta_0 + \theta_0\beta'_0$ and $p_1 = \beta_1 + \theta_1\beta'_1$. For convenience, let $G_{\beta_0,\beta_1,\theta_0,\theta_1} = G_{\beta_0+\theta_0\beta'_0,\beta_1+\theta_1\beta'_1}$. Define an update function f:

$$f(b, \theta_b, \theta_{-b}) = \inf\{\theta'_b \mid sec_{-b}(G_{\beta_b,\beta_{-b},\theta'_b,\theta_{-b}}) \leq v_{-b}\}$$

That is, when $p_0 = \beta_0 + \theta_0\beta'_0$, $p_1 = \beta_1 + \theta_1\beta'_1$ and it is player b's turn, function f returns the minimum θ_b such that player $-b$'s security level in the induced game is still lower than or equal to the target utility.

The function $f(b, \theta_b, \theta_{-b})$ can be computed efficiently by linear programming:

$$\begin{array}{ll} \textbf{minimize} & \theta'_b \\ \textbf{s.t.} & u_{-b}(\mu_b, \mu) \leq v_{-b} \quad \forall \mu \in C(\beta_{-b}, \theta_{-b}) \\ & \mu_b \in C(\beta_b, \theta'_b) \end{array}$$

where $C(\beta, \theta) = C(\beta + \theta(\mathbf{1} - \beta))$. This program has infinitely many constraints of the first type, so we need to show that an efficient separation oracle exists. The first type of constraints is equivalent to

$$\max_{\mu \in C(\beta_{-b},\theta_{-b})} u_{-b}(\mu_b, \mu) \leq v_{-b}$$

whose left-hand side can be computed by a simple water-filling method that puts as much probability as possible on the remaining strategy that provides player $-b$ the most utility, given that player b plays mixed strategy μ_b. Therefore, an efficient separation oracle exists.

We now present the algorithm for generating the on-path history (Algorithm 1). Of course, if we require an infinite-length history, this algorithm will not terminate, but every point in the history will be eventually generated by it.

Lemma 3. *Algorithm 1 produces an on-path history that is part of a Nash equilibrium and leads to a terminal node $z = [p_0, p_1]$ where $sec_0(G_{p_0,p_1}) = v_0$ and $sec_1(G_{p_0,p_1}) = v_1$.*

Algorithm 1: Generate on-path strategy for feasible utilities that exceed security levels

Input: G and a target strategy profile (β_0, β_1)
Output: An on-path history P for $G_D^M(G)$

```
Let β'_0 = 1 − β_0;
Let β'_1 = 1 − β_1;
Let h_0 = [1, 1, 0], b = 0, t = 0;
Let θ_0 = θ_1 = 1;
while θ_b − f(b, θ_b, θ_{−b}) > 0 do
    θ_b = f(b, θ_b, θ_{−b});
    h_{t+1} = [β_0 + θ_0 · β'_0, β_1 + θ_1 · β'_1, 1 − b];
    b = 1 − b;
    t = t + 1;
h_{t+1} = [β_0 + θ_0 · β'_0, β_1 + θ_1 · β'_1];
return h;
```

Convergence Rate

Theorem 3 guarantees the existence of a Nash equilibrium with the desired utilities, but its induced on-path history may have infinite length. From Lemma 2, we know we can approximate this with a path that (with probability 1) has finite length, but this still does not tell us whether this is a reasonable number of rounds. We next show that we can in fact approximate it in a reasonable number of rounds. Instead of using coin tosses, here we simply stop after $O(T)$ rounds, resulting in an $O(n/T)$ approximate equilibrium where the security levels are within $O(n/T)$ of the desired values.

Theorem 4. *In mixed disarmament games $G_D^M(G)$, for all feasible utilities (v_0, v_1) with $v_0 > sec_0(G)$ and $v_1 > sec_1(G)$, and $\varepsilon > 0$, there exists an $n\varepsilon$-Nash equilibrium of $G_D^M(G)$ such that the length of its on-path history is $O(1/\varepsilon)$ and the terminal node $z = [p_0, p_1]$ satisfies $sec_0(G_{p_0,p_1}) > v_0 - n\varepsilon$ and $sec_1(G_{p_0,p_1}) > v_1 - n\varepsilon$.*

Conclusion

We have shown that while disarmament with pure strategies is NP-hard, with mixed strategies a type of folk theorem holds and an efficient algorithm exists. Since this is, to our knowledge, the first paper on this topic, there are many directions for future research. These including studying the topic for representations other than the normal form and solution concepts other than Nash equilibrium. One could also consider different types of mixed disarmament, where the mixed strategy space is reduced in a way that is different from putting upper bounds on the probabilities. Of course, our result shows that upper bounds already allow us to attain everything that can reasonably be expected. In any case, as long as the disarmament procedure ensures that the space of remaining mixed strategies stays compact and convex and there exists an efficient separation oracle for the LP to compute the update function f, our results should continue to hold. Another direction is to design algorithms for the pure disarmament case that work well in practice; in the full version of our paper, we give a mixed-integer linear program formulation for this. Finally, we can look for new applications of this framework.

Acknowledgement

We are thankful for support from ARO under grants W911NF-12-1-0550 and W911NF-11-1-0332, NSF under awards IIS-1527434 and CCF-1337215, the Future of Life Institute, and a Guggenheim Fellowship. We also thank Josh Letchford for helpful early discussions.

References

Anderson, A.; Shoham, Y.; and Altman, A. 2010. Internal implementation. In *Proceedings of the Ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 191–198.

Bošanský, B.; Brânzei, S.; Hansen, K. A.; Miltersen, P. B.; and Sørensen, T. B. 2015. Computation of Stackelberg equilibria of finite sequential games. In *Proceedings of the Eleventh Conference on Web and Internet Economics (WINE-15)*, 201–215.

Brill, M.; Freeman, R.; and Conitzer, V. 2016. Computing possible and necessary equilibrium actions (and bipartisan set winners). In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 369–375.

Conitzer, V., and Sandholm, T. 2006. Computing the optimal strategy to commit to. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 82–90.

Deng, Y.; Tang, P.; and Zheng, S. 2016. Complexity and algorithms of K-implementation. In *Proceedings of the Fifteenth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 5–13.

Letchford, J., and Conitzer, V. 2010. Computing optimal strategies to commit to in extensive-form games. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 83–92.

Letchford, J.; Conitzer, V.; and Munagala, K. 2009. Learning and approximating the optimal strategy to commit to. In *Proceedings of the Second Symposium on Algorithmic Game Theory (SAGT)*, 250–262.

Monderer, D., and Tennenholtz, M. 2004. K-Implementation. *Journal of Artificial Intelligence Research* 21:37–62.

Tambe, M. 2011. *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge University Press.

von Neumann, J. 1928. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen* 100:295–320.

von Stengel, B., and Zamir, S. 2010. Leadership games with convex strategy sets. *Games and Economic Behavior* 69:446–457.