Trembling-Hand Perfection in Extensive-Form Games with Commitment

Gabriele Farina[1], **Alberto Marchesi**[2], **Christian Kroer**[1], **Nicola Gatti**[2], **Tuomas Sandholm**[1]

[1] Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

[2] Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133, Milan, Italy

gfarina@cs.cmu.edu, alberto.marchesi@polimi.it, ckroer@cs.cmu.edu,
nicola.gatti@polimi.it, sandholm@cs.cmu.edu

Abstract

We initiate the study of equilibrium refinements based on trembling-hand perfection in extensive-form games with commitment strategies, that is, where one player commits to a strategy first. We show that the standard *strong* (and *weak*) Stackelberg equilibria are not suitable for trembling-hand perfection, because the limit of a sequence of such strong (weak) Stackelberg commitment strategies of a perturbed game may not be a strong (weak) Stackelberg equilibrium itself. However, we show that the universal set of all Stackelberg equilibria (*i.e.*, those that are optimal for at least some follower response function) is natural for trembling-hand perfection: it does not suffer from the problem above. We also prove that determining the existence of a Stackelberg equilibrium—refined or not—that gives the leader expected value at least ν is NP-hard. This significantly extends prior complexity results that were specific to strong Stackelberg equilibrium.

1 Introduction

In interactions, it can make a significant difference if one player can commit to a strategy first. This was originally studied in the context of using pure strategies to commit to quantity [Cournot, 1838] or price [Bertrand, 1883] in deterring market entry [von Stackelberg, 1934]. Further power can be gained by committing to a mixed (*i.e.*, randomized) strategy [Conitzer and Sandholm, 2006; von Stengel and Zamir, 2010]. In a *mixed Stackelberg equilibrium (SE)*, a *leader* commits to a (potentially mixed) strategy, and a *follower* then best responds. When multiple follower actions are best-responses, tie-breaking needs to be considered. Two notable cases are the *strong (i.e., optimistic) SE (SSE)* and *weak (i.e., pessimistic) SE (WSE)*, in which the follower breaks ties to maximize or minimize, respectively, the leader's utility [Breton *et al.*, 1988]. The SE is now a well-known solution concept in game theory, and has been deployed to practical settings such as airport patrol scheduling, air marshal allocation to flights, and train security [Tambe, 2011; Yin *et al.*, 2012]. Usually, these scenarios are modeled as specialized classes of normal-form games, and an SSE is

sought. An SSE and a WSE in 2-player normal-form games can be found in polynomial time [Conitzer and Sandholm, 2006; von Stengel and Zamir, 2010]. This is Poly-APX-complete in Bayesian games when the number of types is not fixed [De Nittis *et al.*, 2018].

The Stackelberg paradigm can also be applied to *extensive-form games (EFGs)*. EFGs can provide a richer description of a strategic situation. They can be used to model sequential multi-step settings, such as patrol scheduling with interaction between patrols and adversaries. Letchford and Conitzer [2010] show that finding an SSE in an EFG is NP-hard. Furthermore, because Bayesian games are a special case of EFGs with Nature moves the results of De Nittis *et al.* [2018] imply Poly-APX hardness for EFGs with Nature moves as well. Worst-case exponential-time algorithms have been developed for computing an SSE [Bošanskỳ and Cermak, 2015; Kroer and Sandholm, 2015; Cermak *et al.*, 2016]. Kroer *et al.* [2018] propose *robust* SEs, where a solution is found against the worst-case follower's utility model.

In this paper we extend the extensive-form Stackelberg paradigm to settings where players might take off-equilibrium actions with low-but-non-zero probability (a.k.a. *trembles*). Trembles are common in other types of games for refining solution concepts to play well even in parts of the game tree that are reached only through one or more mistakes [van Damme, 1991]. Trembles that cause every information set of the game to be reached with positive probability guarantee off-equilibrium-path optimality. In extensive-form games, the two best-known trembling-hand-perfection-based refinements of *Nash equilibrium (NE)* are the *quasi-perfect equilibrium (QPE)* [van Damme, 1984], where players play their best response at every information set taking into account only the future trembles of the opponent(s), and the *extensive-form perfect equilibrium (EFPE)* [Selten, 1975], in which they account for future trembles of themselves also.[1]

SEs, including SSEs and WSEs, may be suboptimal in presence of trembles. In particular, an SE may be suboptimal due to a leader's mistake (Fig. 1(a)), a follower's mistake (Fig. 1(b)), or both (Fig. 1(c)). The robust SE of Kroer *et*

---

[1] A further NE refinement, defined on the normal form and known as the *normal-form perfect equilibrium (NFPE)*, does not guarantee off-equilibrium-path optimality. The same holds for SE: in Fig. 1(c), when uniform trembles are applied over the normal-form actions, an SE requires players to pick a_ℓ^1, while the only optimal choice is a_ℓ^2.
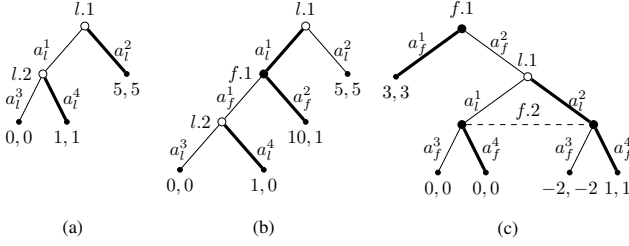
Figure 1: (a) Any strategy at node $l.2$ is optimal in an SE, while only a_ℓ^4 is optimal off the equilibrium path. (b) Any strategy at node $l.2$ is optimal in an SE, while only a_ℓ^4 is optimal off the equilibrium path. (c) Any strategy profile at nodes $l.1$ and $f.2$ is optimal in an SE, while only (a_ℓ^2, a_f^4) is optimal off the equilibrium path.

al. [2018] does not guard against trembles either. Consider the game in Fig. 1(b): the leader should commit to (a_ℓ^1, a_ℓ^4) in order to get utility 10. However, (a_ℓ^1, a_ℓ^3) also achieves utility 10, but it is a worse strategy when trembles may happen. Adding payoff uncertainty on the $(1, 0)$-payoff node such that the follower has a utility function where she picks a_f^1 does not solve this problem. The robust solution would pick a_ℓ^2 initially, since the worst-case follower picks a_f^1. In contrast, our perturbed SEs will uniquely identify (a_ℓ^1, a_ℓ^4) as the solution.

Contributions of this paper. We extend the extensive-form Stackelberg paradigm to settings where players might tremble. We show that for every perturbation scheme, the set of limit points of SEs for perturbed games with vanishing perturbations is always a nonempty subset of the SEs of the non-perturbed game. This does not hold when focusing only on SSEs (or WSEs): for a given game, the set of SSEs (or WSEs) in the non-perturbed game may be disjoint from the set of limit points of SSEs (or WSEs) in the perturbed game. We resort to the perturbation schemes used for QPEs and EFPEs to define *quasi-perfect SEs (QP-SEs)* and *extensive-form perfect SEs (EFP-SEs)*—and their strong and weak versions—as refinements of SE in which, at every information set, the follower plays a best response and the leader's commitment is optimal. We prove that the sets of SSEs and QP-SSEs may be disjoint and that the same holds for the sets of SSEs and EFP-SSEs and for the sets of WSEs and EFP-WSEs. Furthermore, we show that the problem of deciding the existence of an SE—refined or not—that gives the leader expected value at least ν is NP-hard.

2 Preliminaries

Following the notation in the textbook by Shoham and Leyton-Brown [2008], we define a *Stackelberg extensive-form game (SEFG)* with imperfect information as a tuple $(N, H, Z, A, \rho, \chi, u_\ell, u_f, \mathcal{I}_\ell, \mathcal{I}_f)$, where: $N = \{\ell, f\}$ is the set of players (ℓ being the leader and f the follower), $H = H_\ell \cup H_f$, with H_ℓ and H_f being finite sets of leader's and follower's nonterminal decision nodes, respectively, Z is a finite set of terminal nodes, A is a finite set of actions, $\rho : H \to 2^A$ is the action function that assigns to each nonterminal node a set of available actions, $\chi : H \times A \to H \cup Z$ is the successor function that defines the node that is reached when an action is performed in a nonterminal node (if possible), u_ℓ and u_f, with $u_\ell, u_f : Z \to \mathbb{R}$, specify the leader's and follower's pay-

offs, respectively, in each terminal node, and $\mathcal{I}_\ell$ and $\mathcal{I}_f$ define a partition of H_ℓ and H_f, respectively, into information sets, *i.e.*, groups of nonterminal nodes that are indistinguishable by the player. For every player $i \in N$, information set $I_i \in \mathcal{I}_i$, and nonterminal nodes $h, h' \in I_i$, it must be the case that $\rho(h) = \rho(h')$, otherwise the nodes in I_i would not be indistinguishable for player i. Therefore, we can overload notation and let $\rho(I_i)$ be the set of actions available to player i in all nodes belonging to $I_i \in \mathcal{I}_i$. Moreover, we focus on *perfect-recall* games in which no player forgets what she did or knew in the past, *i.e.*, for every $i \in N$ and $I_i \in \mathcal{I}_i$, all decision nodes belonging to I_i share the same sequence of player i's moves on their paths from the root.

In an imperfect-information extensive-form game, a player's *pure strategy* (akin to a *contingent plan* in AI terminology) specifies an action at each information set in which it is the player's turn to move. A *mixed strategy* is a probability distribution over a player's pure strategies. In games with perfect recall, we can restrict the attention to *behavioral strategies* [Kuhn, 1953], which define, for every player $i \in N$ and information set $I_i \in \mathcal{I}_i$, a probability distribution over the actions $\rho(I_i)$ available at I_i. For every $i \in N$, let β_i be a player i's behavioral strategy, where $\beta_i : I_i \in \mathcal{I}_i \mapsto \Delta(\rho(I_i))$. For simplicity, we use $\beta_i(a)$ to denote the probability that β_i assigns to action $a \in A$. In the following, we overload the notation and use $u_i(\beta_\ell, \beta_f)$ to denote player i's expected utility when β_ℓ and β_f are played.

Perfect-recall extensive-form games admit an equivalent representation, the *sequence form* [Romanovskii, 1962; Koller *et al.*, 1996; von Stengel, 1996]. A sequence for player $i \in N$, defined by a node $h \in H \cup Z$, is the ordered set of player i's actions on the path from the root to h. Let Q_ℓ and Q_f be the sets of leader's and follower's sequences, respectively. As usual, we assume that each Q_i, for $i \in N$, contains a fictitious sequence $q_\emptyset$ that corresponds to the root node. Moreover, given a sequence $q \in Q_i$ and an action $a \in A$, we denote as qa the *extended* sequence obtained by appending a to q. With an overloading of notation, u_ℓ and u_f, with $u_\ell, u_f : Q_\ell \times Q_f \to \mathbb{R}$, denote players' utilities in the sequence form, *i.e.*, $u_\ell(q_\ell, q_f)$ and $u_f(q_\ell, q_f)$ are the leader's and the follower's payoff, respectively, in the terminal node reached when playing the actions in q_ℓ and q_f. If q_ℓ and q_f do not identify a terminal node, then $u_\ell(q_\ell, q_f)$ and $u_f(q_\ell, q_f)$ are assumed to be zero. In the sequence form, a strategy, called a *realization plan*, assigns each sequence with its probability of being played. For $i \in N$, let $r_i : Q_i \to \mathbb{R}$ be a player i's realization plan. In order to be well-defined, a realization plan r_i must be such that $r_i(q_\emptyset) = 1$ and, for every $I_i \in \mathcal{I}_i$ and sequence q leading to I_i, $r_i(q) = \sum_{a \in \rho(I_i)} r_i(qa)$. In the following, we denote with L and F the strategy polytopes of all leader's and follower's realization plans, respectively. Overloading the notation, we use u_i as if it were defined over $L \times F$ instead of $Q_\ell \times Q_f$, denoting player i's expected utility.

Given a realization plan r_i, we can easily recover an equivalent behavioral strategy β_i by setting $\beta_i(a) = \frac{r_i(qa)}{r_i(q)}$, where q is the sequence leading to the information set in which a is played. The equivalence is in terms of probabilities that the

two strategies induce on terminal nodes. Similarly, every β_i has an equivalent r_i.

Stackelberg solution concepts. A Stackelberg equilibrium prescribes the leader to commit to a utility-maximizing strategy, given that the follower best-responds to it. Given any strategy for the leader, $r_\ell \in L$, a *best response* for the follower is any strategy $r_f \in F$ that maximizes $u_f(r_\ell, r_f) = \sum_{q_\ell \in Q_\ell, q_f \in Q_f} r_\ell(q_\ell) r_f(q_f) u_f(q_\ell, q_f)$. Let $\mathsf{BR}(r_\ell) = \arg\max_{r_f \in F} u_f(r_\ell, r_f)$ be the set of follower's best-responses to $r_\ell \in L$. If there are multiple follower best responses to the leader's strategy, in order for the leader to determine an optimal commitment, the leader needs to make an assumption about which one of the follower's best responses the follower will choose. A *follower response function* specifies how the follower responds to any possible (mixed) strategy of the leader.

Definition 1. *A* follower response function *is a function τ : $L \to F$ such that $\tau(r_\ell) \in \mathsf{BR}(r_\ell)$ for all $r_\ell \in L$.*

Two classes of follower response functions are popular in the literature [von Stengel and Zamir, 2010]:

- *strong* (*i.e.*, optimistic) follower response functions τ^{st} break ties in favor of the leader: $u_\ell(r_\ell, \tau^{\text{st}}(r_\ell)) \geq u_\ell(r_\ell, r_f) \,\forall r_f \in \mathsf{BR}(r_\ell), r_\ell \in L$;
- *weak* (*i.e.*, pessimistic) follower response functions τ^{we} break ties to minimize the leader's utility: $u_\ell(r_\ell, \tau^{\text{we}}(r_\ell)) \leq u_\ell(r_\ell, r_f) \,\forall r_f \in \mathsf{BR}(r_\ell), r_\ell \in L$.

We formally define the concept of Stackelberg equilibrium.

Definition 2. *Given an SEFG and a follower response function τ, a τ-Stackelberg equilibrium (τ-SE), if it exists, is any leader-utility-maximizing strategy profile, that is, any $(r_\ell, \tau(r_\ell)) \in L \times F$ such that $u_\ell(r'_\ell, \tau(r'_\ell)) \leq u_\ell(r_\ell, \tau(r_\ell))$ for all $r'_\ell \in L$.*

Definition 3. *Given an SEFG, (r_ℓ, r_f) is a Stackelberg equilibrium (SE) if there exists a follower response function τ such that (r_ℓ, r_f) is a τ-SE.*

The following are two well-known examples of SEs.

Definition 4. *A* strong Stackelberg equilibrium *(SSE) is a τ^{st}-SE.*[2] *A* weak Stackelberg equilibrium *(WSE) is a τ^{we}-SE.*

It is known that an SSE always exists, while a WSE might not exist [von Stengel and Zamir, 2010].

3 Game Perturbation Schemes

As discussed in the introduction, many of the most important Nash equilibrium refinement concepts are based on the idea that the player and/or opponent makes mistakes at every decision point (i.e., information set) with some small, vanishing probability. In this section, we introduce a more general family of such perturbation, of which the typical prior schemes, EFPEs and QPEs, are subfamilies. Then in the following sections we develop theory for the whole family and the subfamilies in the context of Stackelberg extensive-form games.[3]

---

[2]Some authors only mean SSE when they talk about Stackelberg equilibrium [Bošanský *et al.*, 2017].

[3]Our perturbation family applies to any strategy polytope, not just Stackelberg extensive-form games, and not even just the se-

In the following definition, X is any strategy polytope, where no distinction is made based on to which player the polytope belongs.

Definition 5. *An ϵ-perturbation scheme for a strategy polytope X is a function $\epsilon \mapsto X(\epsilon)$ defined over $\epsilon \in (0, 1]$ such that*

- $X(\epsilon) \subseteq X(\epsilon')$ *for all $\epsilon \geq \epsilon'$, and*
- $\text{cl}(\bigcup_{\epsilon \in (0,1]} X(\epsilon)) = X$.

The closure operation $\text{cl}(\cdot)$ assumes that a topology is defined for the space containing X. We will always assume that the strategy polytopes X live in a Euclidean space where the usual metric induces open balls $B_\delta(\bar{x}) = \{x : \|x - \bar{x}\| < \delta\}$. The classic EFPE- and QPE-type perturbations (which we define formally in the next section) are two notable subfamilies of ϵ-perturbation schemes.

As a direct consequence of the conditions in Definition 5, every point in X is eventually "reached" by $X(\epsilon)$ when ϵ is small enough:

Lemma 1. *Given $\bar{x} \in X$ and $\delta > 0$, there exists $\hat{x} \in X$ and $\bar{\epsilon} \in (0, 1]$ such that $\hat{x} \in X(\epsilon) \cap B_\delta(\bar{x})$ for all $\epsilon \leq \bar{\epsilon}$.*

A perturbed Stackelberg game is now simply an SEFG augmented with a perturbation scheme for each player:

Definition 6. *A* perturbed SEFG *is an SEFG, together with two ϵ-perturbation schemes $\epsilon \mapsto L(\epsilon)$ and $\epsilon \mapsto F(\epsilon)$ for the leader's and the follower's strategy polytope, respectively.*

Given a perturbed SEFG $(\Gamma, \epsilon \mapsto L(\epsilon), \epsilon \mapsto F(\epsilon))$, we denote by $\Gamma(\epsilon)$ the SEFG obtained from Γ by letting the leader and follower strategy polytopes be $L(\epsilon)$ and $F(\epsilon)$, respectively.

4 EFP and QP Perturbations

In this section we analyze perturbed SEFGs with EFPE- and QPE-type perturbations. They are SEFG ϵ-perturbation schemes with specific structure.

In an EFP-ϵ-perturbation scheme, each player takes into account the possibility that all players, including herself, may make mistakes in the future. Players are constrained to placing at least a minimum probability α on every action at each information set, and those lower bounds α are functions of ϵ that go to zero as $\epsilon \to 0$. A formal definition, using sequence-form strategy polytopes, follows.

Definition 7. *An EFP-ϵ-perturbation scheme for a sequence-form strategy polytope X is an ϵ-perturbation scheme $\epsilon \mapsto X^{EFP}(\epsilon)$ where a strategy r belongs to $X^{EFP}(\epsilon)$ if*

- $r(q) \geq \alpha(q, \epsilon) r(q') \,\forall q, q' \in Q : q = q'a$;
- $\alpha(q, \epsilon) \geq 0 \,\forall q \in Q$ *and* $\lim_{\epsilon \to 0} \alpha(q, \epsilon) = 0 \,\forall q \in Q$;
- $\sum_{q \in Q | q = q'a} \alpha(q, \epsilon) \leq 1 \,\forall q' \in Q$.

In a QP-ϵ-perturbation scheme, each player takes into consideration only the possibility of opponent's errors, assuming she will not make mistakes in future. This is modeled by requiring that sequences q be played with probabilities at least $\omega(q, \epsilon)$, which are polynomials of ϵ whose degrees increase as sequences are extended with additional actions. Formally:

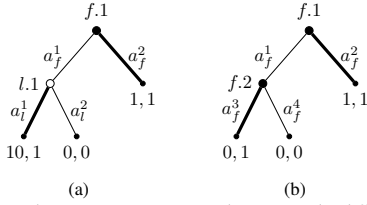quence form. That said, in this paper, we assume that the game is in sequence form.

Figure 2: Game that we use to prove that perturbed SSEs and WSEs are not refinements of SSEs and WSEs.

Definition 8. *A QP-ϵ-perturbation scheme for a sequence-form strategy polytope X is an ϵ-perturbation scheme $\epsilon \mapsto X^{QP}(\epsilon)$ where a strategy r belongs to $X^{QP}(\epsilon)$ if*

- $r(q) \geq \omega(q, \epsilon) \ \forall q \in Q$;
- $\omega(q, \epsilon) \geq 0 \ \forall q \in Q$ and $\lim_{\epsilon \to 0} \omega(q, \epsilon) = 0 \ \forall q \in Q$;
- $\text{maxdeg}\omega(q', \epsilon) < \text{mindeg}\omega(q, \epsilon) \ \forall q, q' \in Q : q = q'a$.

Now, we introduce new solution concepts defined as limit points of sequences of SSEs and WSEs for perturbed game instances $\Gamma(\epsilon)$ as $\epsilon \to 0$, given particular perturbation schemes.

Definition 9. *Given a perturbed SEFG $(\Gamma, \epsilon \mapsto L^{EFP}(\epsilon), \epsilon \mapsto F^{EFP}(\epsilon))$, (r_ℓ, r_f) is an EFP-SSE (resp., EFP-WSE) if it is a limit point of SSEs (resp., WSEs) of $\Gamma(\epsilon)$ as $\epsilon \to 0$.*

Definition 10. *Given a perturbed SEFG $(\Gamma, \epsilon \mapsto L^{QP}(\epsilon), \epsilon \mapsto F^{QP}(\epsilon))$, (r_ℓ, r_f) is a QP-SSE (resp., QP-WSE) if it is a limit point of SSEs (resp., WSEs) of $\Gamma(\epsilon)$ as $\epsilon \to 0$.*

Since SSEs always exist in an SEFG, and since the strategy spaces are compact sets, EFP-SSEs and QP-SSEs always exist. The same is not true for EFP- and QP-WSEs, as a WSE need not exist in an SEFG.

The following observation shows that the sets of EFP- or QP-SSEs can be disjoint from the set of SSEs, thereby showing that the EFP- and QP-SSE solution concepts are *not* refinements of the SSE solution concept!

Observation 1. *There are perturbed SEFGs in which an EFP-SSE is not an SSE, a QP-SSE is not an SSE, and an EFP-WSE is not a WSE.*

Proof. Consider the game of Figure 2(a). The SSE prescribes the leader and the follower to play a_ℓ^1 and a_f^1, respectively. On the other hand, in any perturbed instance the leader has to place positive probability on a_ℓ^2, and the follower's best response is a_f^2.

Consider the game of Figure 2(b). The follower plays $a_f^1 a_f^3$, while in any perturbed instance resulting from an EFP-ϵ-perturbation scheme, the follower has to put positive probability on a_f^4 and her best response at the root becomes a_f^2. $\square$

We leave as an open problem the determination of whether a QP-WSE is also a WSE (assuming it exists) or not.

5 Stackelberg Trembling-Hand Refinements

As we showed in the previous section, SSEs and WSEs are not refinable by trembling. In this section we remedy this problem by showing that the universal set of all Stackelberg equilibria is natural for trembling-hand perfection: it does not suffer from the problem above. In other words, the set of SEs is closed under trembling-hand refinement. Formally, we prove that any limit point of SEs for the perturbed game $\Gamma(\epsilon)$ as $\epsilon \to 0$ is an SE of the original, unperturbed SEFG Γ.

Theorem 1. *Let $\{\epsilon_i\} \to 0$ and let $\{(r_{\ell_i}, r_{f_i})\}$ be a sequence of SEs for the perturbed game instances $\{\Gamma(\epsilon_i)\}$. Then:*

- $\{(r_{\ell_i}, r_{f_i})\}$ *has at least one limit point, and*
- *all limit points of $\{(r_{\ell_i}, r_{f_i})\}$ are SEs.*

We now present three lemmas, and at the end of this section we present the proof of the theorem using these lemmas.

One can think of SEs as "minimally-rational" for the leader: any strategy for the leader is acceptable as long as there is no other strategy for the leader that is better *no matter how the follower breaks ties*. We now formalize this by giving the following alternative characterization of SEs.

Lemma 2. *A strategy profile (r_ℓ, r_f) is an SE if and only if $r_f \in \mathsf{BR}(r_\ell)$ and for all $r'_\ell \in L$ there exists $r'_f(r'_\ell) \in \mathsf{BR}(r'_\ell)$ such that $u_\ell(r'_\ell, r'_f(r'_\ell)) \leq u_\ell(r_\ell, r_f)$.*

Proof. ($\Leftarrow$) Construct the follower response function τ defined as $\tau(r_\ell) = r_f$ and $\tau(r'_\ell) = r'_f(r'_\ell)$. Then, for all $r'_\ell \in L$, $u_\ell(r'_\ell, \tau(r'_\ell)) \leq u_\ell(r_\ell, \tau(r_\ell))$, and thus (r_ℓ, r_f) is a τ-SE.

($\Rightarrow$) Assume that (r_ℓ, r_f) is a τ-SE. Then $r_f = \tau(r_\ell) \in \mathsf{BR}(r_\ell)$. Furthermore, by definition of τ-SE, for all $r'_\ell \in L$, $r'_f = \tau(r'_\ell)$ is such that $u_\ell(r'_\ell, r'_f) \leq u_\ell(r_\ell, \tau(r_\ell)) = u_\ell(r_\ell, r_f)$. $\square$

Lemma 3. *Let $\{\epsilon_i\} \to 0$ and let $\{(r_{\ell_i}, r_{f_i})\}$ be a sequence of strategy profiles for the SEFG instances $\{\Gamma(\epsilon_i)\}$. Then $\{(r_{\ell_i}, r_{f_i})\}$ has at least one limit point.*

Proof. The conclusion follows directly from the Bolzano-Weierstrass theorem since $L(\epsilon) \times F(\epsilon) \subseteq L \times F$ for all $\epsilon \in (0, 1]$ and $L \times F$ is a compact set. $\square$

Lemma 4. *Let $\{\epsilon_i\} \to 0$ and let $\{(r_{\ell_i}, r_{f_i})\}$ be a sequence of strategy profiles for the SEFG instances $\{\Gamma(\epsilon_i)\}$ where r_{f_i} is a best-response to r_{ℓ_i}. Then any limit point $(\bar{r}_\ell, \bar{r}_f)$ of $\{(r_{\ell_i}, r_{f_i})\}$ is such that $\bar{r}_f$ is a best-response to $\bar{r}_\ell$.*

Proof. Existence of at least one limit point for $\{(r_{\ell_i}, r_{f_i})\}$ is guaranteed by Lemma 3. Without loss of generality, assume that $(r_{\ell_i}, r_{f_i}) \to (\bar{r}_\ell, \bar{r}_f) \in L \times F$. Suppose, for contradiction, that $\bar{r}_f$ is *not* a best response to $\bar{r}_\ell$ which means that there exists $\hat{r}_f \in F$ such that $u_f(\bar{r}_\ell, \hat{r}_f) > u_f(\bar{r}_\ell, \bar{r}_f)$. By continuity of u_f, there exists $\delta > 0$ such that $u_f(r_\ell^>, r_f^>) > u_f(r_\ell^<, r_f^<)$ for all $(r_\ell^>, r_f^>) \in B_\delta(\bar{r}_\ell) \times B_\delta(\hat{r}_f)$ and $(r_\ell^<, r_f^<) \in B_\delta(\bar{r}_\ell) \times B_\delta(\bar{r}_f)$. From Lemma 1 we know that there exist $\bar{\epsilon} \in (0, 1]$ and $\tilde{r}_f$ such that $\tilde{r}_f \in F(\epsilon) \cap B_\delta(\hat{r}_f)$ for all $\epsilon \leq \bar{\epsilon}$. Considering the three converging sequences $\epsilon_i \to 0$, $r_{\ell_i} \to \bar{r}_\ell$ and $r_{f_i} \to \bar{r}_f$, we know that there exists an index $j \in \mathbb{N}$ such that $\epsilon_j \leq \bar{\epsilon}, r_{\ell_j} \in B_\delta(\bar{r}_\ell)$, and $r_{f_j} \in B_\delta(\bar{r}_f)$. Furthermore, from $\epsilon_j \leq \bar{\epsilon}$ we deduce that $(r_{\ell_j}, \tilde{r}_f) \in L(\epsilon_j) \times F(\bar{\epsilon}) \subseteq L(\epsilon_j) \times F(\epsilon_j)$. Thus $(r_{\ell_j}, \tilde{r}_f)$ is a valid strategy profile for $\Gamma(\epsilon_j)$. Yet, $(r_{\ell_j}, \tilde{r}_f) \in B_\delta(\bar{r}_\ell) \times B_\delta(\hat{r}_f)$ and $(r_{\ell_j}, r_{f_j}) \in B_\delta(\bar{r}_\ell) \times B_\delta(\bar{r}_f)$, implying $u_f(r_{\ell_j}, \tilde{r}_f) > u_f(r_{\ell_j}, r_{f_j})$ and contradicting the fact that r_{f_j} is a best response to r_{ℓ_j}. $\square$

Proof of Theorem 1. The first bullet is by Lemma 3. We now prove the second one. Let $\mathsf{BR}_\Gamma(r_\ell)$ and $\mathsf{BR}_{\Gamma(\epsilon_i)}(r_\ell)$ be the sets of follower's best-responses to r_ℓ in Γ and $\Gamma(\epsilon_i)$, respectively. Without loss of generality, assume that $\{r_{\ell_i}, r_{f_i}\} \to (\bar{r}_\ell, \bar{r}_f) \in L \times F$. By Lemma 4, $\bar{r}_f$ is a best response to $\bar{r}_\ell$. Therefore, by Lemma 2, we only need to prove that $\forall r'_\ell$ there exists $r'_f \in \mathsf{BR}_\Gamma(r'_\ell)$ with $u_\ell(r'_\ell, r'_f) \le u_\ell(\bar{r}_\ell, \bar{r}_f)$.

Suppose for contradiction that there exists r'_ℓ such that $u_\ell(r'_\ell, r'_f) > u_\ell(\bar{r}_\ell, \bar{r}_f)$ for all $r'_f \in \mathsf{BR}_\Gamma(r'_\ell)$. Let g_i be the family of functions with the property that, for all i, $g_i(r_\ell)$ is equal to one of the $r_f \in \mathsf{BR}_{\Gamma(\epsilon_i)}(r_\ell)$ such that $u_\ell(r_\ell, r_f) \le u_\ell(r_{\ell_i}, r_{f_i})$; existence is guaranteed by Lemma 2 and the fact that (r_{ℓ_i}, r_{f_i}) is an SE by hypothesis. Construct any sequence $\{(r'_{\ell_i}, r'_{f_i})\} \to (r'_\ell, \hat{r}'_f)$ such that (r'_{ℓ_i}, r'_{f_i}) are valid profiles for $\Gamma(\epsilon_i)$ with $r'_{f_i} = g_i(r'_{\ell_i})$. From Lemma 4 we know that $\hat{r}'_f \in \mathsf{BR}_\Gamma(r'_\ell)$. However, $u_\ell(r'_{\ell_i}, r'_{f_i}) \le u_\ell(r_{\ell_i}, r_{f_i})$, and by continuity we have $u_\ell(r'_\ell, \hat{r}'_f) \le u_\ell(\bar{r}_\ell, \bar{r}_f)$. But then $\hat{r}'_f \in \mathsf{BR}_\Gamma(r'_\ell)$ while having value no larger than $u_\ell(\bar{r}_\ell, \bar{r}_f)$, contradicting our assumption. $\square$

Remark. Because SSEs and WSEs are SEs, we have from Theorem 1 that the limit of perturbed SSEs (and perturbed WSEs when they exist) is *guaranteed* to be an SE. This means that, even though the limits are not SSEs or WSEs in general, they preserve minimal rationality of the commitment—as per Lemma 2.

6 Computational Complexity

In this section we study the computational complexity of deciding the existence of an SE (refined or not) that gives the leader expected value at least ν. This problem (in the unrefined case) is known to be polynomial in constant-sum settings, where all SEs give the same expected utility to the leader, equal to the value of the game. We show that this problem is NP-hard in general-sum settings, using a reduction from 3SAT. In particular, given a 3SAT formula, we construct a polynomially-large SEFG instance such that:

- If the 3SAT formula is satisfiable, all SEs of the SEFG give an expected utility of 1 to the leader.
- If the 3SAT formula is *not* satisfiable, all SEs of the SEFG give an expected utility strictly less than 1 to the leader. [4]

Since the 3SAT decision problem is NP-hard [Garey and Johnson, 1979], this implies the following theorems.

Theorem 2. *Deciding the existence of an SE (refined or not) that gives the leader expected value at least ν in an SEFG is* NP-*hard.*

Theorem 3. *Given a follower response function τ, deciding the existence of a τ-SE (refined or not) that gives the leader expected value at least ν in an SEFG is* NP-*hard.*

---

[4] Our reduction is based on the construction of Letchford and Conitzer [2010]. However, their construction only proves the NP-hardness for the special case of SSEs, since, whenever the 3SAT formula is satisfiable, there are SEs of the SEFG that provide the leader with an expected utility strictly less than 1. We suitably modify players' payoffs so that the result holds for all SEs.

6.1 SEFG Instance Construction

Definition 11. *We are given a 3SAT formula (C, V), where C is a set of 3-literal clauses defined over a set V of variables, with $|V| = n$. We let the variable corresponding to literal l_k be denoted by $v(l_k)$. We construct a perfect-recall SEFG $\Gamma(C, V)$ as follows:*

- *The root is $h_f^0 \in H_f$ such that $\rho(h_f^0) = \{a_f^t\} \cup \{a_f^v : v \in V\} \cup \{a_f^\phi : \phi \in C\}$, $\chi(h_f^0, a_f^x) = h_\ell^{1,x} \in H_\ell$.*

- *All nodes $h_\ell^{1,x}$ belong to $I \in \mathcal{I}_\ell$. The available actions at the information set are $\rho(I) = \{a_\ell^v : v \in V\}$.*

- *For all $h_\ell^{1,v}$ and a_ℓ^w $(v, w \in V)$, we let $\chi(h_\ell^{1,v}, a_\ell^w) = z_{vw} \in Z$. Furthermore, $u_\ell(z_{vw}) = 0$, and $u_f(z_{vw}) = \left(\frac{n+2}{(n+1)^2} + 1 \right) - \mathbf{1}\{v = w\}$.*

- *For all a_ℓ^v $(v \in V)$, $\chi(h_\ell^{1,t}, a_\ell^v) = h_f^{2,v} \in H_f$, $\rho(h_f^{2,v}) = \{a_f^{v,\mathsf{T}}, a_f^{v,\mathsf{F}}\}$, $\chi(h_f^{2,v}, a_f^{v,x}) = h_\ell^{3,vx} \in I_v \in \mathcal{I}_\ell$, $\rho(I_v) = \{a_\ell^{v,\mathsf{T}}, a_\ell^{v,\mathsf{F}}\}$, $\chi(h_\ell^{3,vx}, a_\ell^{v,y}) = z_{vxy} \in Z$, $u_\ell(z_{vxy}) = u_f(z_{vxy}) = \mathbf{1}\{x = y\}$.*

- *For all $h_\ell^{1,\phi}$ and a_ℓ^v $(\phi \in C, v \in V)$,*
  $u_\ell(z_{\phi v}) = 0, u_f(z_{\phi v}) = 0, \chi(h_\ell^{2,\phi v}, a_\ell^{v,x}) = z_{\phi vx} \in Z,$

  $$\chi(h_\ell^{1,\phi}, a_\ell^v) = \begin{cases} h_\ell^{2,\phi v} \in I_v & \text{if } v \text{ is in } \phi \\ z_{\phi v} \in Z & \text{otherwise,} \end{cases}$$

  and $u_\ell(z_{\phi vx}) = 0, u_f(z_{\phi v\mathsf{T}}) = \frac{n+1}{3}$ (resp., $u_f(z_{\phi v\mathsf{F}}) = \frac{n+1}{3}$) if v appears negated (resp., not negated) in ϕ, $u_f(z_{\phi vx}) = 0$ otherwise.

Figure 3 shows an example of a game $\Gamma(C, V)$. Intuitively, the leader looks for a strategy such that the follower's best-response is to play action a_f^t, thus achieving an expected utility of 1. The leader's strategy at information sets I_v $(v \in V)$ defines a truth assignment to the variables such that, whenever a clause $\phi \in C$ is not satisfied, then the follower best-responds playing action a_f^ϕ. Thus, the leader's goal is to find a strategy that defines a truth assignment satisfying all clauses. First we show that when the 3SAT formula is satisfiable there exists a leader's strategy that guarantees a payoff of 1.

Lemma 5. *If (C, V) is satisfiable, then there exists a leader's strategy β_ℓ such that for all follower's best-responses $\beta_f \in \mathsf{BR}(\beta_\ell)$ it holds $u_\ell(\beta_\ell, \beta_f) = 1$.*

Proof. Let T be a truth assignment satisfying all clauses. Take β_ℓ such that $\beta_\ell(a_\ell^v) = \frac{1}{n} \, \forall \, v \in V$ and $\beta_\ell(a_\ell^{v,\mathsf{T}}) = 1$ if $T(v) = \mathsf{T}$, while $\beta_\ell(a_\ell^{v,\mathsf{F}}) = 1$ whenever $T(v) = \mathsf{F}$. Clearly, at each node $h_f^{2,v}$, the unique follower's best-response is to play the action corresponding to that played by the leader in I_v. As a result, the follower gets a utility of 1 by playing a_f^t at the root, i.e, $u_f(\beta_\ell, a_f^t) = 1$. Now, let us prove that playing action a_f^t at the root is the unique follower's best-response to β_ℓ. Let us consider actions a_f^v, for all $v \in V$, we have:

$$u_f(\beta_\ell, a_f^v) = \left(\frac{n+2}{(n+1)^2} + 1 \right) \frac{n-1}{n} + \left(\frac{n+2}{(n+1)^2} \right) \frac{1}{n}$$
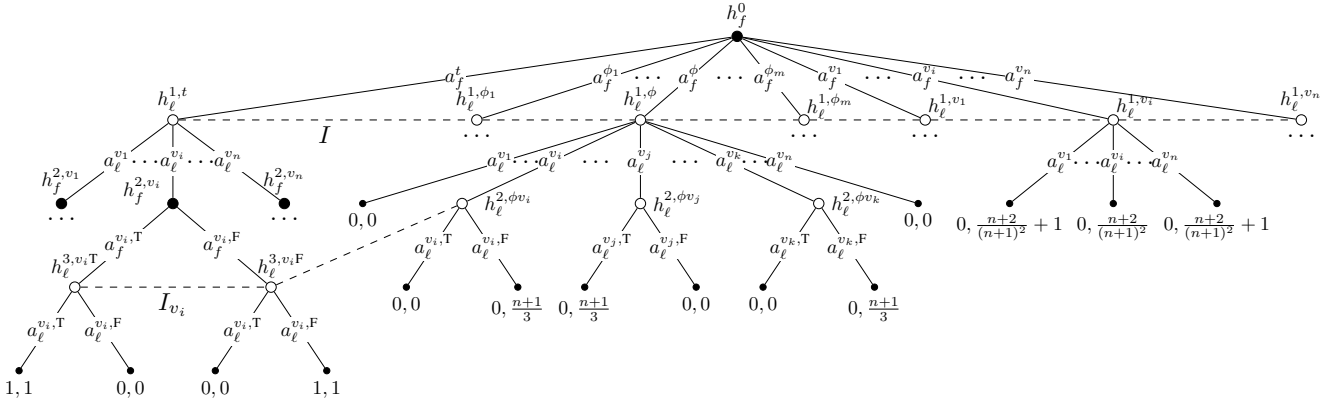$$= \left(\frac{n+2}{(n+1)^2} + 1 \right) - \frac{1}{n} < 1.$$

Figure 3: Game $\Gamma(C,V)$, where $V = \{v_1, \ldots, v_n\}$, $C = \{\phi_1, \ldots, \phi_m\}$, and clause $\phi \in C$ is such that $\phi = v_i \bar{v}_j v_k$. Hollow nodes belong to the leader, while solid ones belong to the follower.

Thus, playing a_f^v is not a best-response, for all $v \in V$. Now, we analyze actions a_f^ϕ, for all $\phi \in C$. Since T satisfies all clauses, each clause $\phi \in C$ has a literal l_k that is true under T and, thus, $\beta_\ell(a_\ell^{v(l_k),\mathrm{T}}) = 1$ if l_k requires the corresponding variable to be true, or $\beta_\ell(a_\ell^{v(l_k),\mathrm{F}}) = 1$ if it requires false. Assume, without loss of generality, that l_k requires the variable to be true for all $l_k \in \phi$. By playing a_f^ϕ, the follower gets utility $u_f(\beta_\ell, a_f^\phi) = \beta_\ell(a_\ell^{v(l_1)})\beta_\ell(a_\ell^{v(l_1),\mathrm{F}})\frac{n+1}{3} + \beta_\ell(a_\ell^{v(l_2)})\beta_\ell(a_\ell^{v(l_2),\mathrm{F}})\frac{n+1}{3} + \beta_\ell(a_\ell^{v(l_3)})\beta_\ell(a_\ell^{v(l_3),\mathrm{F}})\frac{n+1}{3}$. Three cases are possible.

1. There exists unique $l_k \in \phi$ such that $\beta_\ell(a_\ell^{v(l_k),\mathrm{T}}) = 1$, for instance literal l_1. Thus, since $\beta_\ell(a_\ell^{v(l_2),\mathrm{F}}), \beta_\ell(a_\ell^{v(l_3),F}) \leq 1$, it holds $u_f(\beta_\ell, a_f^\phi) \leq \frac{n+1}{3}(\beta_\ell(a_\ell^{v(l_2)}) + \beta_\ell(a_\ell^{v(l_3)}))$. Also, $\beta_\ell(a_\ell^v) = \frac{1}{n}\ \forall v \in V$ implies $u_f(\beta_\ell, a_f^\phi) < \frac{n+1}{3}\left(1 - \frac{1}{n}(n-2)\right) = \frac{2}{3}\frac{n+1}{n} < 1$, for n sufficiently large ($n > 2$).

2. Exactly two literals l_k in ϕ are such that $\beta_\ell(a_\ell^{v(l_k),\mathrm{T}}) = 1$. With a similar reasoning, we conclude that $u_f(\beta_\ell, a_f^\phi) < \frac{n+1}{3}\left(1 - \frac{1}{n}(n-1)\right) < 1$, for every n.

3. $\beta_\ell(a_\ell^{v(l_k),\mathrm{T}}) = 1$ for all literals $l_k \in \phi$, and $u_f(\beta_\ell, a_f^\phi) < 1$. Therefore, it must be $u_f(\beta_\ell, a_f^\phi) < 1$ and a_f^ϕ is not a follower's best-response to β_ℓ.

In conclusion, the unique follower's best-response is to play a_f^t at node h_f^0, and $u_\ell(\beta_\ell, a_f^t) = 1$. $\qquad \square$

Given that 1 is the maximum leader's payoff in $\Gamma(C,V)$, we can conclude the following:

Corollary 1. *If (C,V) is satisfiable, then all SEs of $\Gamma(C,V)$ give the leader an expected utility of 1.*

We now show that a utility of 1 for the leader implies the existence of a truth assignment satisfying the 3SAT formula.

Lemma 6. *If there exists a leader's strategy β_ℓ and a follower's best-response $\beta_f \in \mathsf{BR}(\beta_\ell)$ such that $u_\ell(\beta_\ell, \beta_f) = 1$, then (C,V) is satisfiable.*

Proof. Since $u_\ell(\beta_\ell, \beta_f) = 1$, it must be the case that in β_f the follower plays a_f^t at the root node h_f^0, or else the leader

would not get a utility of 1. Moreover, the leader's strategy β_ℓ must be such that either $\beta_\ell(a_\ell^{v,\mathrm{T}}) = 1$ or $\beta_\ell(a_\ell^{v,\mathrm{F}}) = 1$, for every $v \in V$, and, at each node $h_f^{2,v}$, the follower must play the action corresponding to that played by the leader in I_v. Because a_f^t is a best response, it must be that $u_f(\beta_\ell, a_f^v) \leq 1$ for every $v \in V$, otherwise a_f^t would not be a follower best response. From $u_f(\beta_\ell, a_f^v) \leq 1$, it follows that $u_f(\beta_\ell, a_f^v) = \frac{n+2}{(n+1)^2} + 1 - \beta_\ell(a_\ell^v) \leq 1$, so that $\beta_\ell(a_\ell^v) \geq \frac{n+2}{(n+1)^2} > \frac{1}{n+1}$ for every n. For every $\phi \in C$ we have $u_f(\beta_\ell, a_f^\phi) \leq 1$, otherwise playing a_f^t is not a best response for the follower. As a consequence, for every $\phi \in C$, there must exist at least one literal $l_k \in \phi$ such that $\beta_\ell(a_\ell^{v(l_k),\mathrm{T}}) = 1$ if l_k requires true, or $\beta_\ell(a_\ell^{v(l_k),\mathrm{F}}) = 1$ if l_k requires false. By contradiction, suppose such a literal l_k does not exist, and assume, without loss of generality, that l_k requires true for all $l_k \in \phi$. Thus, $u_f(\beta_\ell, a_f^\phi) = \beta_\ell(a_\ell^{v(l_1)})\beta_\ell(a_\ell^{v(l_1),\mathrm{F}})\frac{n+1}{3} + \beta_\ell(a_\ell^{v(l_2)})\beta_\ell(a_\ell^{v(l_2),\mathrm{F}})\frac{n+1}{3} + \beta_\ell(a_\ell^{v(l_3)})\beta_\ell(a_\ell^{v(l_3),\mathrm{F}})\frac{n+1}{3} = \frac{n+1}{3}(\beta_\ell(a_\ell^{v(l_1)}) + \beta_\ell(a_\ell^{v(l_2)}) + \beta_\ell(a_\ell^{v(l_3)})) > 1$, as $\beta_\ell(a_\ell^v) > \frac{1}{n+1}$ for all $v \in V$. This contradicts the fact that $u_f(\beta_\ell, a_f^\phi) \leq 1$. It follows that ϕ must be satisfied. Since ϕ was arbitrary, this shows that all clauses are satisfied. In conclusion, a variable assignment T such that $T(v) = \mathsf{T}$ if $\beta_\ell(a_\ell^{v,\mathrm{T}}) = 1$, while $T(v) = \mathsf{F}$ whenever $\beta_\ell(a_\ell^{v,\mathrm{F}}) = 1$, satisfies all clauses. $\qquad \square$

It directly follows from Lemma 6 that:

Corollary 2. *If (C,V) is not satisfiable, then all SEs of $\Gamma(C,V)$ give an expected utility smaller than 1 to the leader.*

7 Conclusions

We initiated the study of equilibrium refinement based on trembling-hand perfection in Stackelberg extensive-form games, that is, games where one player commits to a strategy first. To our knowledge, this is the first solution concept that guarantees off-equilibrium-path optimality in extensive-form Stackelberg games. We studied the equilibrium space of all Stackelberg equilibria (containing both strong and weak Stackelberg equilibria), and showed that it is complete with

respect to the limit points induced by perturbation schemes. We showed that this is not the case for strong and weak Stackelberg equilibria. Finally, we showed that deciding the existence of any Stackelberg equilibrium—refined or not—giving the leader expected value of at least ν is NP-hard.

Acknowledgements

References

[Bertrand, 1883] Joseph Louis François Bertrand. Theorie mathematique de la richesse sociale. *Journal des Savants*, pages 499–508, 1883.

[Bošanskỳ and Cermak, 2015] Branislav Bošanskỳ and Jiri Cermak. Sequence-form algorithm for computing stackelberg equilibria in extensive-form games. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2015.

[Bošanskỳ *et al.*, 2017] Branislav Bošanskỳ, Simina Brânzei, Kristoffer Arnsfelt Hansen, Troels Bjerre Lund, and Peter Bro Miltersen. Computation of Stackelberg equilibria of finite sequential games. *ACM Transaction on Economics and Computation (TEAC)*, 5(4):23:1–23:24, December 2017.

[Breton *et al.*, 1988] Michèle Breton, Abderrahmane Alj, and Alain Haurie. Sequential Stackelberg equilibria in two-person games. *Journal of Optimization Theory and Applications*, 59(1):71–97, October 1988.

[Cermak *et al.*, 2016] Jiri Cermak, Branislav Bošanskỳ, Karel Durkota, Viliam Lisy, and Christopher Kiekintveld. Using correlated strategies for computing Stackelberg equilibria in extensive-form games. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2016.

[Conitzer and Sandholm, 2006] Vincent Conitzer and Tuomas Sandholm. Computing the optimal strategy to commit to. In *Proceedings of the ACM Conference on Economics and Computation (EC)*, 2006.

[Cournot, 1838] Antoine Augustin Cournot. *Recherches sur les principes mathématiques de la théorie des richesses (Researches into the Mathematical Principles of the Theory of Wealth)*. Hachette, Paris, 1838.

[De Nittis *et al.*, 2018] Giuseppe De Nittis, Alberto Marchesi, and Nicola Gatti. Computing the strategy to commit to in polymatrix games. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[Garey and Johnson, 1979] Michael Garey and David Johnson. *Computers and Intractability*. W. H. Freeman and Company, 1979.

[Koller *et al.*, 1996] Daphne Koller, Nimrod Megiddo, and Bernhard von Stengel. Efficient computation of equilibria for extensive two-person games. *Games and Economic Behavior*, 14(2):247–259, 1996.

[Kroer and Sandholm, 2015] Christian Kroer and Tuomas Sandholm. Limited lookahead in imperfect-information games. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[Kroer *et al.*, 2018] Christian Kroer, Gabriele Farina, and Tuomas Sandholm. Robust Stackelberg equilibria in extensive-form games and extension to limited lookahead. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[Kuhn, 1953] Harold W. Kuhn. Extensive games and the problem of information. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games*, volume 2 of *Annals of Mathematics Studies, 28*, pages 193–216. Princeton University Press, Princeton, NJ, 1953.

[Letchford and Conitzer, 2010] Joshua Letchford and Vincent Conitzer. Computing optimal strategies to commit to in extensive-form games. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*, 2010.

[Romanovskii, 1962] I. Romanovskii. Reduction of a game with complete memory to a matrix game. *Soviet Mathematics*, 3:678–681, 1962.

[Selten, 1975] Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4(1):25–55, March 1975.

[Shoham and Leyton-Brown, 2008] Yoav Shoham and Kevin Leyton-Brown. *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008.

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

[van Damme, 1984] Eric van Damme. A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. *International Journal of Game Theory*, 1984.

[van Damme, 1991] Eric van Damme. *Stability and perfection of Nash equilibria*, volume 339. Springer, 1991.

[von Stackelberg, 1934] Heinrich von Stackelberg. *Marktform und Gleichgewicht*. Springer, Vienna, 1934.

[von Stengel and Zamir, 2010] Bernhard von Stengel and Shmuel Zamir. Leadership games with convex strategy sets. *Games and Economic Behavior*, 69(2):446–457, 2010.

[von Stengel, 1996] Bernhard von Stengel. Efficient computation of behavior strategies. *Games and Economic Behavior*, 14(2):220–246, 1996.

[Yin *et al.*, 2012] Zhengyu Yin, Albert Xin Jiang, Milind Tambe, Christopher Kiekintveld, Kevin Leyton-Brown, Tuomas Sandholm, and John P. Sullivan. TRUSTS: Scheduling randomized patrols for fare inspection in transit systems using game theory. *AI Magazine*, 33(4):59–72, 2012.