Robust Commitments and Partial Reputation

Vidya Muthukumar Anant Sahai

Department of Electrical Engineering and Computer Sciences, UC Berkeley
{vidya.muthukumarsahai}@eecs.berkeley.edu

May 10, 2019

Abstract

Agents rarely act in isolation – their behavioral history, in particular, is public to others. We seek a non-asymptotic understanding of how a leader agent should shape this history to its maximal advantage, knowing that follower agent(s) will be learning and responding to it. We study Stackelberg leader-follower games with finite observations of the leader commitment, which commonly models security games and network routing in engineering, and persuasion mechanisms in economics. First, we formally show that when the game is not zero-sum and the vanilla Stackelberg commitment is mixed, it is not robust to observational uncertainty. We propose observation-robust, polynomial-time-computable commitment constructions for leader strategies that approximate the Stackelberg payoff, and also show that these commitment rules approximate the maximum obtainable payoff (which could in general be greater than the Stackelberg payoff).

1 Introduction

Consider a selfish, rational agent (designated as the "leader") who is interacting non-cooperatively with another selfish, rational agent (designated as the "follower"). If the agents are interacting *simultaneously*, and know the game that they are playing, they will naturally play a Nash equilibrium(a) of the two-player game. This is the traditionally studied solution concept in game theory. Now, let's say that the leader has the ability to reveal its strategy in advance – in the form of a *mixed strategy commitment*, and the follower has the ability to observe this commitment and respond to it. The optimal strategy for the leader now corresponds to the Stackelberg equilibrium of the two-player leader-follower game. The Stackelberg solution concept enjoys application in several engineering settings where commitment is the natural framework for the leader, such as security [PPM+08], network routing [Rou04] and law enforcement [MS17]. The solution concept is interpreted in a broader sense as the ensuing strategy played by a patient leader that wishes to build a *reputation* by playing against an infinite number of myopic followers [KW82, MR82, FL89, FL92]. Crucially, one can show rigorously that the leader will benefit significantly from commitment power, i.e. its ensuing Stackelberg equilibrium payoff is at least as much as the simultaneous equilibrium payoff [VSZ10]. Further, several *mechanism design* problems that involve private information revelation - this includes signalling games [CS82] and persuasion games [KG11] - can be thought of as Stackelberg games, and the optimal mechanism can be interpreted as the Stackelberg commitment[1].

But whichever interpretation one chooses, the Stackelberg solution concept assumes a very idealized setting (even over and above the assumptions of selfishness and infinite rationality) in which the mixed strategy commitment is *exactly revealed* to the follower. Further, the follower 100% believes that the leader will actually stick to her commitment. What happens when these assumptions are relaxed? What if the leader could only demonstrate her commitment in a finite number of interactions – how would she modify her commitment to maximize payoff, and how much commitment power would she continue to enjoy? Is she even incentivized to help the follower estimate her commitment effectively?

What changes in a finite-interaction regime is that the follower only observes a part of the leader behavioral history and needs to *learn about the leader's strategic behavior* – to the extent that he is able to respond as optimally as possible. By restricting our attention to finitely repeated play, we

[1] albeit sometimes with multiple followers.

arrive at a problem setting that is fairly general: these follower agents in general will not know about the preferences of the leader agent. When provided with historical context, we assume that they will use it rather than ignore it. A broad umbrella of problems that has received significant attention in the machine learning literature is learning of strategic behavior from samples of play; for example, learning the agent's utility function through *inverse reinforcement learning* [ZMBD08], learning the agent's level of rationality [WZB13], and inverse game theory [KS15]. While significant progress has been made in this goal of learning strategic behavior, attention has been restricted to the *passive learning setting* in which the leading agent is unaware of the presence of the learner, or agnostic to the learner's utility. In many situations, the agent himself will be invested in the outcome of the learning process. In this paper, we put ourselves in the shoes of an agent who is shaping her historical context and is aware of the learner's presence as well as preferences, and study her choice of optimal strategy[2]. As we will see, the answer will depend on her utility function itself, as well as what kind of response she is able to elicit from the learner.

1.1 Related work

The Stackelberg solution concept is used in the engineering and economics literature to model a number of scenarios. For one, the *Stackelberg security game* is played between a defender, who places different utility levels on different targets to be protected and accordingly uses her resources to defend some subset of targets; and an attacker, who observes the defender strategy and wishes to attack some of the targets depending on whether he thinks they are left open as well as how much he values those targets. Stackelberg games can also be modeled with a single leader and multiple followers, such as in computer network routing applications [Rou04]. Many mechanism design problems involve computing an optimal mechanism to commit to, or an optimal way of *revealing private information* - this includes auctions and, more recently, Bayesian persuasion mechanisms [KG11].

Economists have established an important link between the Stackelberg solution concept and the asymptotic limit of *reputation building*. Reputation effects were first observed in the *chain-store paradox*, a firm-competition game where an *incumbent* firm would often deviate from Nash equilibrium behavior and play its aggressive Stackelberg (pure, in this case) strategy [Sel78]. Theoretical justification was provided for this particular application [KW82, MR82] by modeling a $(N+1)$-player interaction between a leader and multiple followers, and studying the Nash equilibrium of an ensuing game as $N \to \infty$. It was shown that the leader would play its *pure Stackelberg strategy*[3] in the Bayes-Nash equilibrium of this game endowed with a common prior on the leader's payoff structure[4]. This model was generalized to such leader-follower ensembles for a general two-player game, and considering the possibility of mixed-strategy reputation, still retaining the asymptotic nature of results [FL89, FL92]. The "first-player" advantage, and the entire Stackelberg solution concept, rely on an important assumption: *that the commitment is perfectly revealed to the follower.* This is usually not the case: in security games, the attacker will usually observe a finite number of deployments of the defender's resource, as opposed to the allocation strategy itself (which is often mixed). In theoretical models for Bayesian persuasion, the persuader conveys a conditional distribution on her signal given the privately observed state of the world - but what will be practically observed is her history, and thus *realizations* of the signal, not the distribution itself. In all of these models, the leader establishes her reputation *only partially*, and the manifestation of the revelation is itself random. It is natural to ask how she should plan to optimally reveal her information under this constraint.

The idea of a robust solution concept in game theory is certainly not new. The concept of *trembling-hand-perfect-equilibrium* [Sel75] explicitly studies how robust mixed-strategy Nash equilibria are to slight perturbations in the mixtures themselves, and a similar concept was proposed for Stackelberg [VDH97]. Another solution concept, *quantal-response-equilibrium* [MP95], studies agents that

[2]It is worth mentioning the recent paradigm of *cooperative inverse reinforcement learning* [HMRAD16] which studies the problem of agent investment in principal learning where the incentives are not completely aligned, but the setting is cooperative. In contrast, we focus on non-cooperative settings.

[3]This is clearly more restrictive than the mixed strategy Stackelberg solution concept, and not necessarily advantageous over Nash, but it turns out to be so in the firm competition case.

[4]A more nuanced model considered a Bayesian prior over a leader being constrained to play its pure Stackelberg strategy as opposed to unconstrained play.

are *boundedly rational*, an orthogonal but important source of uncertainty in response. In the Stackelberg setting, it was noted that robust commitments *exist* that preserve the Stackelberg guarantee for small enough amounts of noise in the commitment; however, this is still an asymptotic perspective and does not directly help us answer our key computational questions: can we construct a robust commitment efficiently when the game is multi-dimensional, and does the leader want to use the noise to reveal or obfuscate her commitment?

The problem of computing the optimal commitment under *finitely* limited observability corresponds to a robust optimization problem that is, in fact, NP-hard [AKK+12, SAY+12]; so directly reasoning about the optimal commitment is not easy. Whether there exists a polynomial-time approximation scheme for this problem was also unclear. A duo of papers [AKK+12, SAY+12] considered a model of full-fledged observational uncertainty with a Bayesian prior and posterior update based on samples of behavior, and proposed heuristic algorithmic techniques to compute the optimum. In fact, they also factored for quantal responses using a bounded rationality model [MP95]. This work showed through simulations that there could be a positive return over and above Stackelberg payoff. In one important piece of analytical work, the problem was also considered for the special case of zero-sum games [BHP14], and it was shown that the Stackelberg commitment itself approximated the optimal payoff. In this result, the extent of approximation actually depends on the amount of observational uncertainty itself - the results we prove for all non-zero-sum games have a similar flavor.

The problem of *communication constraints* in the commitment has also received a lot of interest in the recent algorithmic persuasion literature, but with quantitatively different models for the uncertainty. Communication constraints on signaling in bilateral trading games [DKQ16] and auction design [DIR14] have been studied from a *compression perspective*, where the leading agent can design the observation channel - while in our model the observation channel is even more constrained to be a finite number of *random* realizations of the mixed commitment. Further, in many of these settings, the principal is naturally incentivized to reveal the private information and the problem primarily becomes about the communication complexity, and whether the social welfare of the optimal mechanism can be approximated[5]. In security games, the possibility of the mixed commitment being either fully observed or not observed at all has been considered [KCP11]; as well as different ways of handling the uncertainty, eg: showing that for some security games the Stackelberg and Nash equilibria coincide and observation of commitment does not matter [KYK+11]. Pita et al [PJT+10] first proposed a model for the defender (leader) to account for attacker (follower) observational uncertainty by allowing the follower to anchor [RT97] to a certain extent on its uniform prior. While they showed significant returns from using their model through extensive experiments, they largely circumvented the algorithmic and analytical challenges by not explicitly considering random samples of defender behavior, thus keeping the attacker response deterministic but shifted. Our work limits observation in the most natural way for the applications that we consider (i.e. number of samples of leader behavior), and *because the manifestation of the uncertainty is itself random*, our results have distinct and new implications.

1.2 Our contributions

Our main contribution is to understand the extent of reputational advantage when interaction is finite, and prescribe approximately optimal commitment rules for this regime.

We study Stackelberg leader-follower games in which a follower obtains a limited number of observations of the leader commitment. We first prove that in most non-zero-sum games the payoff of the Stackelberg commitment is not robust to even an *infinitesmal amount of* observational uncertainty. Therefore, the Stackelberg commitment is suboptimal in its payoff[6]. Next, we propose robust commitment rules for leaders and show that we can approach the Stackelberg payoff as the number of observations increases. The robust commitment construction involves optimizing a tradeoff between preserving the follower best response and staying close to the ideal Stackelberg commitment, by moving the commitment a little bit into the interior of an appropriate convex polytope [CS06]. The analysis of

[5]These are clearly interesting algorithmic questions in themselves, especially in the case of multiple receivers and private vs public signaling [DX16], but do not directly address the questions we have raised.

[6]This property has actually been proved for special examples of Stackelberg games [VDH97], but it was unclear whether it holds for all or most games.

payoff of the commitment construction is inspired by interior point convex geometry [KN12]. Finally, we show that a possible advantage for the leader from limited observability is only related to follower response mismatch, and show that this advantage is limited. Computationally speaking, the corollary is that we are able to approximate the optimal payoff through a simple construction which can be obtained in constant time from computation of the Stackelberg commitment (itself a polynomial-time operation [CS06]). Philosophically, this result implies that a leader can gain to a very limited extent by misrepresenting her commitment and eliciting a suboptimal response from the follower. We corroborate our theoretical results with simulations on illustrative examples and random ensembles of security games.

2 Problem statement

2.1 Preliminaries

We represent a two-player leader-follower game in normal form by the pair of $d \times n$ matrices (T_1, T_2), where $T_1 \in \mathbb{R}^{d\times n}$ denotes the leader payoff matrix and $T_2 \in \mathbb{R}^{d\times n}$ denotes the follower payoff matrix. We denote the leader mixed strategy space by Δ_d (where Δ_k for any k represents the k-dimensional probability simplex) and the follower mixed strategy space by Δ_n. From now on, we define an *effective dimension* of a game as a number $m < d$ for which the effective payoff matrices of leader and follower respectively are $A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n\end{bmatrix} \in \mathbb{R}^{m\times n}$, $B = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n\end{bmatrix} \in \mathbb{R}^{m\times n}$, and the effective set of leader strategies is given by a *convex polytope* $K \subseteq \Delta_m$[7].

We consider a setting of *asymmetric private information* in which the leader knows about the follower preferences (i.e. she knows the matrix B) while the follower does not know about the leader preferences (i.e. he possesses no knowledge of the matrix A)[8].

With infinite experience, the well-established effect from the follower's point of view is that the leader has established *commitment*, or developed a *reputation*, for playing according to some mixed strategy $\mathbf{x} \in K$. We denote the follower's set of theoretically best *pure-strategy* responses to a mixed strategy commitment $\mathbf{x}$ by $\mathcal{K}^*(\mathbf{x}) \subseteq [m]$. Explicitly, we have

$$\mathcal{K}^*(\mathbf{x}) := \arg\max_{j\in[n]} \langle \mathbf{x}, \mathbf{b}_j \rangle.$$

An important assumption that we make (and that has been made in the classical literature [CS06]) is the follower actually responds with the pure strategy in the set $\mathcal{K}^*(\mathbf{x})$ that is most beneficial to the leader[9]. That is, the follower responds with pure strategy

$$j^*(\mathbf{x}) := \arg\max_{j\in\mathcal{K}^*(\mathbf{x})} \langle \mathbf{x}, \mathbf{a}_j \rangle.$$

Then, we also define *best-response regions* as the set of leader commitments that would elicit the pure strategy response j from the follower, i.e. $\mathcal{R}_j := \{\mathbf{x} \in K : j^*(\mathbf{x}) = j\}$.

With these definitions, we can define the leader's ideal payoff to be expected with an infinite reputation:

Definition 1. *A leader with an infinite reputation of playing according to the strategy* $\mathbf{x}$ *should expect payoff*

$$f_\infty(\mathbf{x}) := \langle \mathbf{x}, \mathbf{a}_{k^*} \rangle.$$

Therefore, the leader's ***Stackelberg payoff*** *is the solution to the program*

$$f^*_\infty := \max_{\mathbf{x}\in\Delta_m} f_\infty(\mathbf{x}).$$

[7]This definition is important in the context of Stackelberg security games, for which the leader strategy space looks exponential in the number of targets m - but the actual manifestation of all leader strategies is in fact m-dimensional. In particular, a defender strategy manifests as a distribution over different targets being covered.

[8]This is an important assumption for the paper, and is in fact used in traditional reputation building frameworks. In future, we will want to better understand situations of repeated interaction where the ∞-level leader and 1-level follower are both learning about one another.

[9]The technical reason for this tie-breaking rule is to be able to explicitly define the Stackelberg commitment as an explicit *maximum* - this in itself gives a subtle clue of its fragility.

The argmax of this program is denoted as the ***Stackelberg commitment*** $\mathbf{x}^*_\infty$. *Further, we denote the best response faced in Stackelberg equilibrium by* $j^* := j^*(\mathbf{x}^*_\infty)$.

It is clear that the Stackelberg commitment is optimal for the leader under two conditions: *the leader is 100% known to be committed to a fixed strategy*, and *the follower knows exactly the leader's committed-to strategy.* For a finite number of interactions, neither is true.

2.2 Observational uncertainty with established commitment

Even assuming that there is a shared belief in commitment, there is uncertainty. In particular, with a finite number of plays, the follower does not know the exact strategy that the leader has committed to, and only has an estimate.

Consider the situation where a leader can only reveal its commitment $\mathbf{x}$ through N *pure strategy plays* $I_1, I_2, \ldots, I_N$ i.i.d $\sim \mathbf{x}$. The commitment is known (to both leader and followers) to come from a set of mixed strategies $\mathcal{X} \subseteq K$. We denote the maximum likelihood estimate of the leader's mixed strategy, as seen by the follower, by $\widehat{\mathbf{X}}_N$. It is reasonable to expect, under certainty of commitment, that a "rational"[10] follower would best-respond to $\widehat{\mathbf{X}}_N$, i.e. play the pure strategy

$$j^*(\widehat{\mathbf{X}}_N). \tag{1}$$

We can express the expected leader payoff under this learning rule.

Definition 2. *A leader which will have N plays according to the hidden strategy $\mathbf{x}$ can expect payoff in the Nth play of*

$$f_N(\mathbf{x}) := \mathbb{E}\left[\langle \mathbf{x}, \mathbf{a}_{j^*(\widehat{\mathbf{X}}_N)} \rangle\right]$$

against a follower that plays according to (1). The maximal payoff a leader can expect is

$$f^*_N := \max_{\mathbf{x} \in \Delta_m} f_N(\mathbf{x})$$

and it acquires this payoff by playing the argmax strategy $\mathbf{x}^*_N$.

Ideally, we want to understand how close f^*_N is to f^*_∞, and also how close $\mathbf{x}^*_N$ is to $\mathbf{x}^*_\infty$. An answer to the former question would tell us how observational uncertainty impacts the first-player advantage. An answer to the latter question would shed light on whether the best course of action deviates significantly from Stackelberg commitment. We are also interested in algorithmic techniques for *approximately computing* the quantity f^*_N, as doing so exactly would involve solving a non-convex optimization problem.

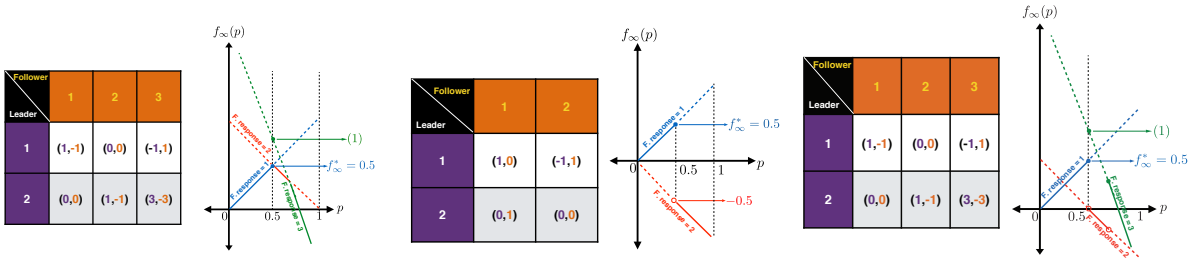
3 Main Results

3.1 Robustness of Stackelberg commitment to observational uncertainty

A natural first question is whether the Stackelberg commitment, which is clearly optimal if the game were being played infinitely (or equivalently, if the leader had infinite commitment power and exact public commitment), is also suitable for finite play. In particular, we might be interested in evaluating whether we can do better than the baseline Stackelberg performance f^*. We show through a few paradigmatic examples that the answer can vary.

Example 1. *We consider a 2×3 zero-sum game, represented in normal form in Figure 1(a), in which we can express the leader strategy according to the probability p with which she will pick strategy* 1*, and*

[10] Rational is in quotes because the follower is not necessarily using expected-utility theory (although there is an expected-utility-maximization interpetation to this estimate if the mixed strategy were uniform drawn from $\mathcal{X}$).



(a) 2×3 zero-sum game. (b) 2×2 non-zero-sum game. (c) 2×3 non-zero-sum game.

Figure 1. Illustration of examples of zero-sum game and non-zero-sum games in the form of normal form tables and ideal leader payoff function $f_\infty(.)$. p denotes the probability that the leader will play strategy 1, and fully describes leader mixed commitment for these $2 \times n$ games.

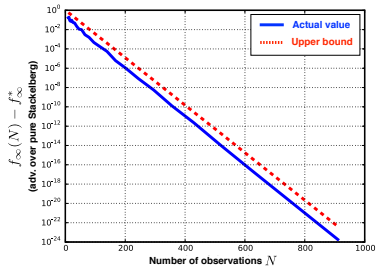


Figure 2. Semilog plot of extent of advantage over Stackelberg payoff as a function of N in the 2×3 zero-sum game depicted in Figure 1(a).

leader payoff is as follows:

$$\begin{cases} f(p;1) := p \text{ if follower best responds with strategy 1} \\ f(p;2) := 1-p \text{ if follower best responds with strategy 2} \\ f(p;3) := 3-4p \text{ if follower best responds with strategy 3} \end{cases}$$

Since the game is zero-sum, the follower responds in a way that is worst-case for the leader. This means that we can express the leader payoff as

$$f_\infty(p) = \min\{f(p;1), f(p;2), f(p;3)\}.$$

This leader payoff structure is depicted in Figure 1(a). Therefore, we can express the Stackelberg payoff as

$$f^*_\infty = \max_{p\in[0,1]} f_\infty(p) = 1/2,$$

attained at $p^*_\infty = 1/2$. *We wish to evaluate* $f_N(p^*_\infty)$. *It was noted [BHP14] that* $f_N(p^*_\infty) \geq f^*_\infty(p^*_\infty)$ *by the minimax theorem, but not always clear whether strict inequality would hold (that is, if observational uncertainty gives a strict advantage). For this example, we can actually get a sizeable improvement! To see this, look at the simple example of* $N = 1$. *Denoting* $\widehat{P} = \frac{1}{N}\sum_{j=1}^{N} \mathbb{I}[I_j = 1]$, *we have*

$$\begin{aligned} f_1(1/2) &= \Pr[\widehat{P}_1 = 0].f(1/2;1) + \Pr[\widehat{P}_1 = 1].f(1/2;3) \\ &= 1/2 \times 1/2 + 1/2 \times 1 = 3/4. \end{aligned}$$

The semilog plot in Figure 2 shows that this improvement persists for larger values of N, *although the*

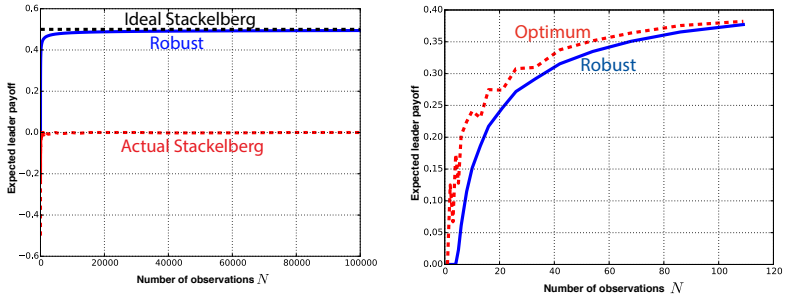
extent of improvement decreases exponentially with N. We can show that

$$\begin{aligned} f_N(1/2) - f^*_\infty &= 1/2 \Pr[\widehat{P}_N > 2/3] \\ &= 1/2 \Pr[\widehat{P}_N - 1/2 > 1/6] \\ &\leq \exp\{-N D_{\mathsf{KL}}(2/3 \parallel 1/2)\} \end{aligned}$$

where $D_{\mathsf{KL}}(. \parallel .)$ denotes the Kullback-Leibler divergence, and the last inequality is due to Sanov's theorem [CK11].

This shows analytically that the advantage does indeed decrease exponentially with N. Naturally, this is because we see a decrease in the stochasticity that elicits the more favorable follower response with action 3 with the number of observations N.

Example 1 showed us how Stackelberg commitment power could be increased by stochastically eliciting more favorable responses. We now see an example illustrating that the commitment power can disappear completely.



(a) Plot depicting the performance of the sequence of robust commitments $\{\mathbf{x}_N\}_{N\geq 1}$ and the Stackelberg commitment $\mathbf{x}^*_\infty$ as a function of N. The benchmark for comparison is idealized Stackelberg payoff f^*_∞.

(b) Plot showing the performance of sequence of robust commitments $\{\mathbf{x}_N\}_{N\geq 1}$ as compared to the optimum performance f^*_N (brute-forced).

Figure 3. Example of the 2×2 non-zero-sum game depicted in Figure 1(b), for which observational uncertainty is always undesirable.

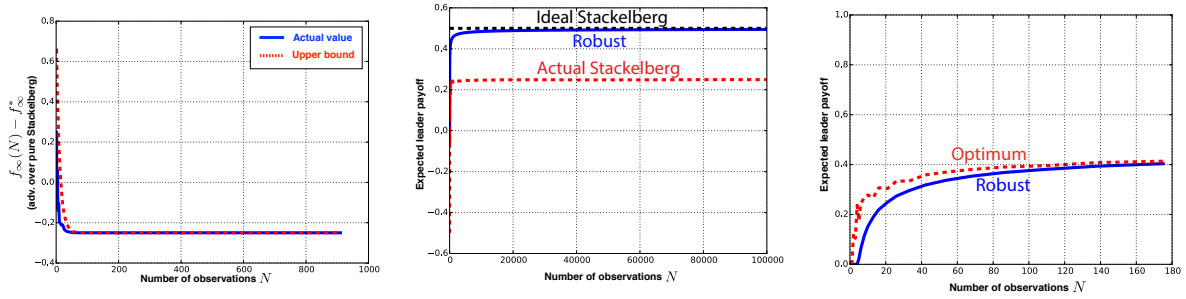
Example 2. *We consider a 2×2 non-zero-sum game, represented in normal form and leader payoff structure in Figure 1(b). Explicitly, the ideal leader payoff function is*

$$f_\infty(p) = \begin{cases} p \text{ if } p \leq 1/2 \\ -p \text{ if } p > 1/2. \end{cases}$$

This is essentially the example reproduced in [BHP14], which we repeat for storytelling value. Notice that $f^*_\infty = 1/2, p^*_\infty = 1/2$, but the advantage evaporates with observational uncertainty. For any finite N, we have

$$\begin{aligned} f_N(1/2) &= \Pr[\widehat{P}_N \leq 1/2](1/2) + \Pr[\widehat{P}_N > 1/2](-1/2) \\ &= 1/2 \times 1/2 - 1/2 \times 1/2 = 0. \end{aligned}$$

Remarkably, this implies that $f^*_\infty - f_N(p^*_\infty) = 1/2$ and so $\lim_{N\to\infty} f^*_\infty - f_N(p^*_\infty) \neq 0$! This is clearly a very negative result for the robustness of Stackelberg commitment, and as a very pragmatic matter tells us that the idealized Stackelberg commitment p^*_∞ is far from ideal in finite-observation settings. This example shows us a case where stochasticity in follower response is not desired, principally because of the discontinuity in the leader payoff function at p^*_∞.



(a) Plot of the extent of (dis)advantage over Stackelberg payoff as a function of N.

(b) Plot depicting the performance of the sequence of robust commitments $\{\mathbf{x}_N\}_{N\geq 1}$ and the Stackelberg commitment $\mathbf{x}^*_\infty$ as a function of N. The benchmark for comparison is idealized Stackelberg payoff f^*_∞.

(c) Plot showing the performance of sequence of robust commitments $\{\mathbf{x}_N\}_{N\geq 1}$ as compared to the optimum performance f^*_N (brute-forced).

Figure 4. Example of the 2×3 non-zero-sum game depicted in Figure 1(c), in which observational uncertainty could either help or hurt the leader.

Example 2 displayed to the fullest the significant disadvantage of observational uncertainty. The game considered was special in that there was no potential for limited-observation gain, while in the game presented in Example 1 there was only potential for limited-observational gain. What could happen in general? Our next and final example provides an illustration.

Example 3. *Our final example considers a* 2×3 *non-zero-sum game, whose normal form and leader payoff structure are depicted in Figure 1(c). The ideal leader payoff function is*

$$f_\infty(p) = \begin{cases} p \text{ if } p \leq 1/2 \\ 1/2 - p \text{ if } 1/2 < p \leq 5/7 \\ 3 - 4p \text{ if } p > 5/7. \end{cases}$$

As in the other examples, $f^*_\infty = 1/2, p^*_\infty = 1/2$. *Notice that this example captures both positive and negative effects of stochasticity in response. On one hand, follower response* 2 *is highly undesirable (a la Example 2) but follower response* 3 *is highly desirable (a la Example 1). What is the net effect? We have*

$$\begin{aligned} f_N(1/2) &= \Pr[\widehat{P}_N \leq 1/2](1/2) + \Pr[1/2 < \widehat{P}_N < 5/7](0) + \Pr[\widehat{P}_N \geq 5/7](1) \\ &= (1/2)(1/2) + \Pr[\widehat{P}_N \geq 5/7](1) \\ &\leq 1/4 + 1/2 \exp\{-N D_{\mathsf{KL}}(5/7 \parallel 1/2)\}. \end{aligned}$$

A quick calculation thus tells us that $f_N(p^*_\infty) <= f^*_\infty$ *if* $N \geq 8$, *showing that Stackelberg in fact has poor robustness for this example. Intuitively, the probability of the "bad" stochastic event remains constant while the probability of the "good" stochastic event decreases exponentially with* N. *Even more damningly, we see that* $\lim_{N\to\infty} f^*_\infty - f_N(p^*_\infty) \geq \lim_{N\to\infty} 1/4 - 1/2 \exp\{-N D_{\mathsf{KL}}(5/7 \parallel 1/2) = 1/4$, *again showing that the Stackelberg commitment is far from ideal. We can see the dramatic decay of leader advantage over and above Stackelberg, and ensuing disadvantage even for a very small number of observations, in Figure 3(a).*

While the three examples detailed above provide differing conclusions, there are some common threads. For one, in all the examples it is the case that committing to the Stackelberg mixture $\mathbf{x}^*_\infty$ can result in the follower being agnostic between more than one response. Only one of these responses, the pure strategy j^*, is desirable for the leader. A very slight misperception in the estimation of the true value $\mathbf{x}^*_\infty$ can therefore lead to a different, worse-than-expected response and this misperception happens with a sizeable, non-vanishing probability. On the flipside, a different response could also lead to better-than-expected payoff, raising the potential for a gain over and above f^*. However, these

better-than-expected responses cannot share a boundary with the Stackelberg commitment, and we will see that the probability of eliciting them decreases exponentially with N. The net effect is that the Stackelberg commitment is, most often, not robust – and critically, this is even the case for small amounts of uncertainty.

Our first result is a formal statement of instability of Stackelberg commitments for a general $2 \times n$ game. We denote the leader probability of playing strategy 1 by $p \in [0,1]$, and the Stackelberg commitment's probability of playing strategy 1 by p^*_∞.

Furthermore, let $\phi(t)$ denote the CDF of the standard normal distribution $\mathcal{N}(0,1)$. We are now ready to state the result.

Theorem 1. *For any $2 \times n$ leader-follower game in which $p^*_\infty \in (0,1)$ and $f_\infty(p)$ discontinuous at $p = p^*_\infty$, we have*

$$f_N(p^*_\infty) \leq f^*_\infty - C\left(\phi(\sqrt{N}C') - \frac{1}{2} - \frac{C'}{\sqrt{N}}\right) + \exp\{-NC''^2\} \tag{2}$$

where C, C', C'' are strictly positive constants depending on the parameters of the game. This directly implies the following:

1. *For some $N_0 > 0$, we have $f_N(p^*_\infty) < f^*_\infty$ for all $N > N_0$.*
2. *We have $\lim_{N\to\infty} f_N(p^*_\infty) < f^*_\infty$.*

The proof of Theorem 1 is contained in Section A.1. The technical ingredients in the proof are the Berry-Esseen theorem [Ber41, Ess42], used to show that the detrimental alternate responses on the Stackelberg boundary are non-vanishingly likely – and the Hoeffding bound, used to tail bound the probability of potentially beneficial alternate responses not on the boundary[11]

For non-robustness of Stackelberg commitment to hold, the two critical conditions for the game are that there is a discontinuity at the Stackelberg boundary, and that the Stackelberg commitment is mixed. For a zero-sum game, the first condition does not hold and the Stackelberg commitment stays robust as we saw in Example 1.

The theorem directly implies that the ideal Stackelberg payoff is only obtained for the exact case of $N = \infty$ (when the commitment is perfectly observed), and that for any value of $N < \infty$ there is a *non-vanishing reduction* in payoff. In the simulations in Section 4, we will see that this gap is empirically significant.

3.2 Robust commitments achieving close-to-Stackelberg performance

The surprising message of Theorem 1 is that, in general, the Stackelberg commitment $\mathbf{x}^*$ is undesirable. The commitment $\mathbf{x}^*$ is pushed to the *extreme point* of the best-response-region $\mathcal{R}_{j^*}$ to ensure optimality under idealized conditions; and this is precisely what makes it sub-optimal under uncertainty. What if we could move our commitment a little bit into the interior of the region $\mathcal{R}_{j^*}$ instead, such that we can get a *high-probability-guarantee* on eliciting the expected response, while staying sufficiently close to the idealized optimum? Our next result quantifies the ensuing tradeoff and shows that we can cleverly construct the commitment to approximate Stackelberg performance.

Theorem 2. *Let the best-response polytope $\mathcal{R}_{j^*}$ be non-empty in $\mathbb{R}^{m-1}$. Then, provided that the number of samples $N = \widetilde{\mathcal{O}}(m)$, we can construct commitment $\mathbf{x}_{N,p}$ for every $0 < p < 1/2$ such that*

$$f^*_\infty - f_N(\mathbf{x}_N) = \widetilde{\mathcal{O}}\Big(\Big(\frac{m}{N}\Big)^p + e^{-\omega(1)\cdot N^{1-2p}}\Big). \tag{3}$$

Furthermore, these constructions are computable in $\mathcal{O}(1)$ time with knowledge of the Stackelberg commitment $\mathbf{x}^*_\infty$. (The $\widetilde{\mathcal{O}}(\cdot)$ contains constant factors that depend on both the local and global geometry of the best-response-region $\mathcal{R}_{j^*}$. For a fully formal statement that includes these factors, see Lemma 6.)

[11] It is worth noting that a similar argument as presented here could be extended to a general $m \times n$ game, using iid random vectors instead of random variables and considering a demarcation into best-response regions as illustrated in Figure 7. We restrict attention to the $2 \times n$ case for ease of exposition.

The full proof of Theorem 2, deferred to Appendix A.2, involves some technical steps to achieve as good as possible a scaling in N. The caveat of Theorem 2 is that commitment power can be robustly exploited in this way only if there are enough observations of the commitment. One obvious requirement is that the best-response-region $\mathcal{R}_{j^*}$ needs to be non-empty in $\mathbb{R}^{m-1}$. Second, the number of observations N needs to be greater than the *effective dimension* of the game for the leader, m. This is a natural requirement to ensure that the follower has learned at least a meaningful estimate of the commitment. Third, the "constant" factors in Theorem 2 actually reflect properties about both the local and global geometry of the polytope; see Appendix A.2 for more details. Intuitively, geometric properties that lead to undesirable scaling in the constant factors in the robustness guarantee are listed below:

1. The Stackelberg commitment being a "pointy" vertex: this can lead to a commitment being far away from the boundary in certain directions, but closer in others, making it more likely for a different response to be elicited.

2. Local constraints being very different from global constraints, which implies that commitments too far in the interior of the local feasibility set will no longer satisfy all the constraints of the best-response-region.

Even with these caveats, Theorem 2 provides an attractive general framework for constructing robust commitments by making a natural connection to interior-point methods in optimization[12]. We observe significant empirical benefit from the constructions in the simulations in Section 4.

We also mention a couple of special cases of leader-follower games for which the robust commitment constructions of Theorem 2 are not required; in fact, it is simply optimal to play Stackelberg.

Remark 1. *For games in which the mixed-strategy Stackelberg equilibrium coincides with a pure strategy, the follower's best response is always as expected regardless of the number of observations. There is no tradeoff and it is simply optimal to play Stackelberg even under observational uncertainty.*

Remark 2. *For the zero-sum case, it was observed [BHP14] that a Stackelberg commitment is made assuming that the follower will respond in the worst case. If there is observational uncertainty, the follower can only respond in a way that yields payoff for the leader that is better than expected. This results in an expected payoff greater than or equal to the Stackelberg payoff f^*_∞, and it simply makes sense to stick with the Stackelberg commitment $\mathbf{x}^*_\infty$. As we have seen, this logic does not hold up for non-zero-sum games because different responses can lead to worse-than-expected payoff. One way of thinking of this is that the function $f_\infty(.)$ can generally be discontinuous in $\mathbf{x}$ for a non-zero-sum game, but is always continuous for the special case of zero-sum.*

3.3 Approximating the maximum possible payoff

So far, we have considered the limited-observability problem and shown that the Stackelberg commitment $\mathbf{x}^*_\infty$ is not a suitable choice. We have constructed robust commitments that come close to idealized Stackelberg payoff f^*_∞ and shown that the guarantee fundamentally depends on the number of observations scaling with the effective dimension of the game. Now, we turn to the question of whether we can approximate f^*_N, the actual optimum of the program. Note that since the problem is in general non-convex in $\mathbf{x}$, it is NP-hard to exactly compute.

Rather than the traditional approach of constructing a polynomial-time-approximation-algorithm, our approach is approximation-theoretic[13]. We first show that in the large-sample case, we cannot do much better than the actual Stackelberg payoff f^*_∞; informally speaking, our ability to fool the follower into responding *strictly-better-than-expected* is limited. Combining this with the robust commitment construction of Theorem 2, we obtain an approximation to the optimum payoff.

The main result of this section is stated below.

[12]Noting that interior point methods are provably polynomial-time algorithms to solve LPs, it is plausible to think that in fact, stopping the interior point method appropriately early would also give us a robustness guarantee - which would imply that finding optimal *robust* commitments is even easier than finding optimal commitments!

[13]In other words, the extent of approximation is measured by the *number of samples* as opposed to the runtime of an algorithm. This is very much the flavor of previously-obtained results on Stackelberg zero-sum security games [BHP14].

Theorem 3. *We have*

$$f_N^* \leq f_\infty^* + Cn\sqrt{\frac{m}{N}}.$$

for some constant $C > 0$ depending on the parameters of the game (A, B).

As a corollary the commitment construction defined in Theorem 2 provides a $\widetilde{\mathcal{O}}(\sqrt{\frac{1}{N}})$-additive-approximation algorithm to f_N^*. The proof of Theorem 3 is provided in the appendix.

Intellectually, Theorem 3 tells us that the robust commitments are essentially optimal. The practical benefit that Theorem 3 affords us is that we now have an approximation to the optimum payoff the leader could possibly obtain, which can be computed in constant time after computing the Stackelberg equilibrium, which itself is polynomial time [CS06]. This is because the robust commitment is obtained by first computing Stackelberg equilibrium $\mathbf{x}_\infty^*$, and then deviating away from $\mathbf{x}_\infty^*$ in the magnitude and direction specified. We will now study the empirical benefits of our robust commitment constructions.

4 Simulations

4.1 Example 2×2 and 2×3 games

First, we return to the non-zero-sum games described in Examples 2 and 3. These were 2×2 and 2×3 games respectively, and the Stackelberg commitment was non-robust for both games. Now, armed with the results in Theorem 2, we can employ our robust commitment constructions and study their performance. To construct our robust commitments, we first computed the Stackelberg commitment using the LP solver in scipy (scipy.optimize.linprog), and then used the construction in Theorem 2.

Figures 3(a) and 4(b) compares the expected payoff obtained by our robust commitment construction scheme $\{\mathbf{x}_N\}_{N \geq 1}$ for different numbers of samples N, and for the games described in Examples 2 and 3 respectively. The benchmark with respect to which we measure this expected payoff is the Stackelberg payoff f_∞^* (obtained by Stackelberg commitment under *infinite observability* and *tie-breakability in favor of the leader*). We also observe a significant gap between the payoffs obtained by these robust commitment constructions and the payoff obtained if we used the Stackelberg commitment $\mathbf{x}_\infty^*$. We showed in theory that there is significant benefit for choosing the commitment to factor in such observational uncertainty, and we can now see it in practice.

Furthermore, for the case of 2 leader actions we were able to brute-force the maximum possible obtainable payoff[14] f_N^*, and compare the value to the robust commitment payoff. This comparison is particularly valuable for smaller values of N, as shown in Figures 3(b) and 4(c). We notice that the values are much closer even than our theory would have predicted, and even for small values of N. Thus, our constructions have significant practical benefit as well: we are able to get close to the optimum while drastically reducing the required computation (to just solving n LPs!).

Since these examples involved $2 \times n$ games, the commitment construction became trivial (i.e. only one direction to move along) – next, we test our commitment constructions for $m \times m$ security games. Instead of looking at specific examples, we now look at a random ensemble to see what behavior ensues.

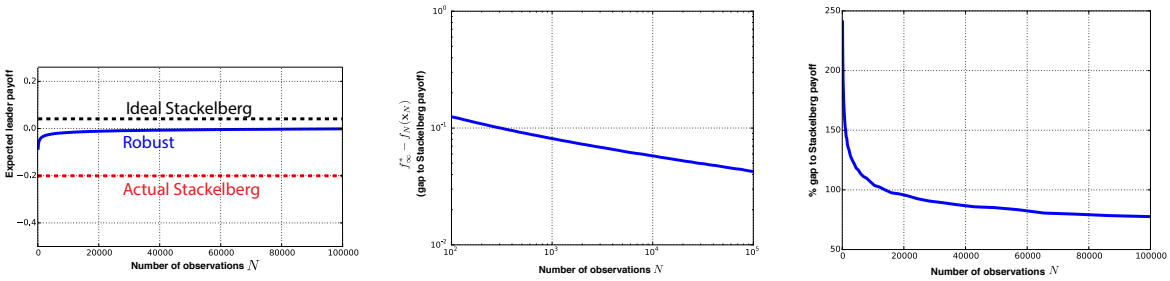
4.2 Random security games

Our next set of simulations is inspired by the security games framework. We create a random ensemble of 5×5 security games in which the defender can defend one of 5 targets, and the attacker can attack one of these 5 targets. The defender and attacker rewards are chosen to be uniformly at random between $[0, 1]$, and their penalties are uniform at random between $[-1, 0]$. This is essentially the random ensemble that was created in previous empirical work on security games [AKK+12]. Figure 5 shows the construction of this ensemble.

[14] First we used scipy.optimize.brute with an appropriate grid size to initialize, and then ran a gradient descent at that initialization point. This was feasible for the case of 2 pure strategies.

Target / Reward	1	2	3	4	5
Defender (if protected)	Unif[0,1]	Unif[0,1]	Unif[0,1]	Unif[0,1]	Unif[0,1]
Defender (if unprotected)	Unif[-1,0]	Unif[-1,0]	Unif[-1,0]	Unif[-1,0]	Unif[-1,0]
Attacker (if protected)	Unif[-1,0]	Unif[-1,0]	Unif[-1,0]	Unif[-1,0]	Unif[-1,0]
Attacker (if unprotected)	Unif[0,1]	Unif[0,1]	Unif[0,1]	Unif[0,1]	Unif[0,1]

Figure 5: Illustration of random ensemble of 5×5 security game.



(a) Plot of expected defender payoff when defender uses robust commitments – compared to Stackelberg commitment as well as idealized Stackelberg payoff.

(b) Log-log plot of the gap between robust commitment payoff and idealized Stackelberg payoff.

(c) Percentage plot of the gap between robust commitment payoff and idealized Stackelberg payoff.

Figure 6. Illustration of performance of robust commitments and Stackelberg commitment in random 5×5 Stackelberg security games for a finite number of observations of defender commitment.

The purpose of random security games is to show that the properties we observed above – unstable Stackelberg commitment, robust commitment payoff approximating the optimum – are the norm rather than the exception. Figure 6 illustrates the results for random security games. The performance of the sequence of robust commitments $\{\mathbf{x}_N\}_{N \geq 1}$, as well as the Stackelberg commitment $\mathbf{x}^*_\infty$ is plotted in Figure 6(a) against the benchmark of idealized Stackelberg performance f^*_∞. Figure 6(b) depicts the rate of convergence of the gap in robust commitment performance to the idealized Stackelberg payoff – we can clearly see the $\mathcal{O}(\frac{1}{\sqrt{N}})$ rate of convergence in this plot. Finally, Figure 6(c) plots the *percentage gap* between robust commitment payoff and idealized Stackelberg payoff as a function of N.

We can make the following conclusions from these plots:

1. The Stackelberg commitment is extremely non-robust *on average*. In fact we noticed that this was the case with high probability. This happens because the Stackelberg commitment, although it can vary widely for different games in the random ensemble, is very likely on a boundary shared with other responses and therefore unstable.

2. The robust commitments are doing much better *on average* than the original Stackelberg commitment even for very large values of N. The stark difference in payoff between the two motivates the construction of the robust commitment, which was as easy to compute as the Stackelberg commitment.

5 Proof sketches

In this section we describe briefly the philosophy for the proofs of our main theorems. To understand the strong lack of robustness in Stackelberg equilibrium, it is essential to visualize the best-response-regions of the leader, i.e. subsets of the mixed strategy space for which the follower best response is a particular pure strategy. (Note that there are n such best-response regions.) Figure 7 depicts an

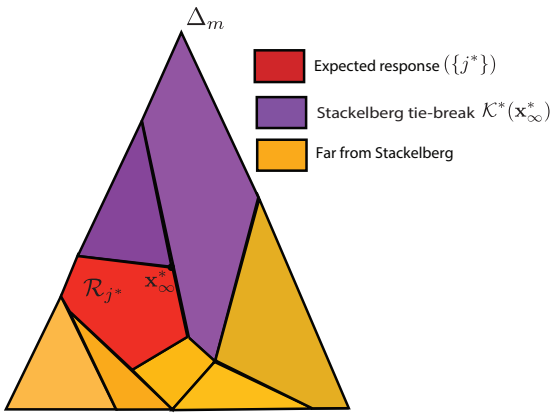


Figure 7. Illustration of partition of the set of follower responses, $[n]$, into sets $\{j^*\}$ (red region), *alternate best responses* (purple regions) and everything else (orange regions).

illustration of these best-response-regions, with the region corresponding to the follower's best response to the Stackelberg commitment highlighted in red. The figure shows the Stackelberg commitment at a vertex (extreme-point) of the best-response polytope $\mathcal{R}_{j^*}$; this is generally the case [CS06].

First, the reason for the strong instability of Stackelberg commitment to even an infinitesmal amount of uncertainty can be seen from Figure 7: an infinitesmal amount of fluctuation in how the leader commitment is observed will make the follower respond with a different pure strategy *with constant probability*, corresponding to the regions depicted in purple. Because of the tie-breaking assumption, it turns out that the expected payoff from any of these alternate responses is strictly worse than the Stackelberg payoff. These facts are proved formally using the Berry-Esseen theorem. Note that an uncertainty in commitment could also lead to a response from one of the yellow regions in the figure (which could either hurt or benefit the leader), but the probability of this happening turns out to decay exponentially.

This observation implied that the optimality of the Stackelberg commitment *under ideal assumptions* was exactly what made it suboptimal under a small amount of uncertainty; we exploit this to construct the robust commitment constructions of Theorem 2. The qualitative idea is to push the commitment to a small extent into the interior of the best-response-region $\mathcal{R}_{j^*}$ so that it simultaneously satisfies a property of being "close" to the Stackelberg commitment, while also ensuring that the fluctuations in its empirical estimate are highly likely to stay in $\mathcal{R}_{j^*}$ (which guarantees that the identity of the best response of the follower is preserved). For the special case of $m = 2$, this is a simple tradeoff to navigate as there is only one direction in which one can move into the interior. For higher dimensions, we take inspiration from the rich literature on interior-point methods and, in fact, use Dikin ellipsoids [KN12] for both the commitment construction and analysis. Ensuring that the fluctuations of the commitment preserve the follower best response with high probability, in particular, requires sophisticated tail bounds on discrete distribution learning and a careful consideration of the best-response-polytope geometry.

The proof of Theorem 3 ties several facts that we have seen formally, as well as alluded to, together. First, a generalization of Theorem 1 tells us that we cannot improve sizeably over Stackelberg by committing to any mixed strategy *on the boundary* between two or more best-response regions. Second, we show that the improvement gained by a *fixed commitment* in the interior of any best-response-region decreases exponentially with N, simply because the probability of eliciting a better-than-expected response decreases exponentially with N. Putting these two facts together, the natural thing to try would be commitments that approach a boundary as N increases (much like our robust commitment constructions, but now with a different motive). This should happen fast enough that we maintain a sizeable probability of eliciting a different response for every N, while simultaneously ensuring that that different response is actually better-than-expected. We then show that the ensuing gain over and above Stackelberg would have to decrease with N according to the rate specified.

6 Conclusions and Discussion

We constructed robust commitment constructions with several advantages. First, we are able to effectively preserve the Stackelberg payoff by ensuring a high-probability guarantee on the follower responding as expected. An oblique, but significant philosophical advantage to our robust commitments is that their guarantees hold even if we removed the pivotal assumption of follower breaking ties in favor of the leader. We essentially showed that as the number of observations N grows, our construction naturally converges to the Stackelberg commitment at a specific rate. We also motivated that the constructions, which were inspired by interior point geometry, are computable in polynomial time given the Stackelberg commitment.

Second, we established fundamental limits on the ability of the leader to gain over and above Stackelberg payoff. We formally showed that this ability disappears in the large-sample regime, and in a certain sense that our robust commitments are approximately optimal. Our results established a formal connection between leader payoff and follower discrete distribution learning, and in the context of these limits, both players are mutually incentivized to increase learnability under limited samples, even though the setting is non-cooperative – which was a rather surprising conclusion.

Our work provides implications for both leader and follower payoffs when the leader is known to be committed to a fixed strategy, but the commitment can only be revealed partially. However, our model took commitment establishment for granted, i.e. the follower assumed that the leader would indeed be drawing its pure strategies iid from the same mixture in every round. The partial reputation setting should most generally be modeled as a repeated game (either with a finite-horizon or discounted model), in which the belief in commitment needs to be built up over time. Studying the problem of finite *observability* of commitment in isolation is, in our view, an important first step towards eventually solving this problem, which poses many modeling challenges in itself. In earlier rounds, directly responding to the empirical estimate of leader commitment will be suboptimal for the follower. Instead, he may want to maintain a possibility that the leader will play minimax/Nash and respond accordingly. From the point of view of the leader, if the iid assumption is removed, an interesting question is whether the leader could choose to play more deterministically, in such a way to increase strategy learnability while maintaining a follower impression of iid commitment. By doing this, the leader could establish commitment faster but also run the risk of looking too deterministic/predictable in time, in which case the follower would take undue advantage. Conversely, the leader may not even be incentivized to increase follower learnability in the finitely repeated, or discounted setting.

Finally, it is interesting to think about the applicability of the robust commitment perspective to algorithmically more difficult problems like Bayesian persuasion and public/private signalling games with multiple followers, which can have observational limitations on information transfer in much the same way as has been described for the applications in this paper.

Acknowledgments

We thank the anonymous reviewers for valuable feedback. We gratefully acknowledge the support of the NSF through grant AST-1444078, and the Berkeley ML4Wireless research center.

References

[AKK+12] Bo An, David Kempe, Christopher Kiekintveld, Eric Shieh, Satinder Singh, Milind Tambe, and Yevgeniy Vorobeychik. Security games with limited surveillance. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 1241–1248. AAAI Press, 2012.

[Ber41] Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, 49(1):122–136, 1941.

[BHP14] Avrim Blum, Nika Haghtalab, and Ariel D Procaccia. Lazy Defenders Are Almost Optimal against Diligent Attackers. In *AAAI*, pages 573–579, 2014.

[CK11] Imre Csiszar and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[CS82] Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982.

[CS06] Vincent Conitzer and Tuomas Sandholm. Computing the optimal strategy to commit to. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, pages 82–90. ACM, 2006.

[Dev83] Luc Devroye. The equivalence of weak, strong and complete convergence in l1 for kernel density estimates. *The Annals of Statistics*, pages 896–904, 1983.

[DIR14] Shaddin Dughmi, Nicole Immorlica, and Aaron Roth. Constrained signaling in auction design. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1341–1357. Society for Industrial and Applied Mathematics, 2014.

[DKQ16] Shaddin Dughmi, David Kempe, and Ruixin Qiang. Persuasion with limited communication. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 663–680. ACM, 2016.

[DX16] Shaddin Dughmi and Haifeng Xu. Algorithmic bayesian persuasion. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 412–425. ACM, 2016.

[Ess42] Carl-Gustaf Esseen. *On the Liapounoff limit of error in the theory of probability*. Almqvist & Wiksell, 1942.

[FL89] Drew Fudenberg and David K Levine. Reputation and equilibrium selection in games with a patient player. *Econometrica: Journal of the Econometric Society*, pages 759–778, 1989.

[FL92] Drew Fudenberg and David K Levine. Maintaining a reputation when strategies are imperfectly observed. *The Review of Economic Studies*, 59(3):561–579, 1992.

[HMRAD16] Dylan Hadfield-Menell, Stuart J Russell, Pieter Abbeel, and Anca Dragan. Cooperative inverse reinforcement learning. In *Advances in neural information processing systems*, pages 3909–3917, 2016.

[KCP11] Dmytro Korzhyk, Vincent Conitzer, and Ronald Parr. Solving Stackelberg games with uncertain observability. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1013–1020. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[KG11] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *The American Economic Review*, 101(6):2590–2615, 2011.

[KN12] Ravindran Kannan and Hariharan Narayanan. Random walks on polytopes and an affine interior point method for linear programming. *Mathematics of Operations Research*, 37(1):1–20, 2012.

[KS15] Volodymyr Kuleshov and Okke Schrijvers. Inverse game theory. *Web and Internet Economics*, 2015.

[KW82] David M Kreps and Robert Wilson. Reputation and imperfect information. *Journal of economic theory*, 27(2):253–279, 1982.

[KYK+11] Dmytro Korzhyk, Zhengyu Yin, Christopher Kiekintveld, Vincent Conitzer, and Milind Tambe. Stackelberg vs. Nash in security games: An extended investigation of interchangeability, equivalence, and uniqueness. *Journal of Artificial Intelligence Research*, 2011.

[MP95] Richard D McKelvey and Thomas R Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

[MR82] Paul Milgrom and John Roberts. Predation, reputation, and entry deterrence. *Journal of economic theory*, 27(2):280–312, 1982.

[MS17] Vidya Muthukumar and Anant Sahai. Fundamental limits on ex-post enforcement and implications for spectrum rights. In *Dynamic Spectrum Access Networks (DySPAN), 2017 IEEE International Symposium on*, pages 1–10. IEEE, 2017.

[PJT+10] James Pita, Manish Jain, Milind Tambe, Fernando Ordóñez, and Sarit Kraus. Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition. *Artificial Intelligence*, 174(15):1142–1171, 2010.

[PPM+08] Praveen Paruchuri, Jonathan P Pearce, Janusz Marecki, Milind Tambe, Fernando Ordonez, and Sarit Kraus. Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pages 895–902. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[Rou04] Tim Roughgarden. Stackelberg scheduling strategies. *SIAM Journal on Computing*, 33(2):332–350, 2004.

[RT97] Yuval Rottenstreich and Amos Tversky. Unpacking, repacking, and anchoring: advances in support theory. *Psychological review*, 104(2):406, 1997.

[SAY+12] Eric Shieh, Bo An, Rong Yang, Milind Tambe, Craig Baldwin, Joseph DiRenzo, Ben Maule, and Garrett Meyer. Protect: A deployed game theoretic system to protect the ports of the United States. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 13–20. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[Sel75] Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory*, 4(1):25–55, 1975.

[Sel78] Reinhard Selten. The chain store paradox. *Theory and decision*, 9(2):127–159, 1978.

[VDH97] Eric Van Damme and Sjaak Hurkens. Games with imperfectly observable commitment. *Games and Economic Behavior*, 21(1-2):282–308, 1997.

[VSZ10] Bernhard Von Stengel and Shmuel Zamir. Leadership games with convex strategy sets. *Games and Economic Behavior*, 69(2):446–457, 2010.

[WZB13] Kevin Waugh, Brian D Ziebart, and J Andrew Bagnell. Computational rationalization: The inverse equilibrium problem. *arXiv preprint arXiv:1308.3506*, 2013.

[ZMBD08] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. Maximum Entropy Inverse Reinforcement Learning. In *AAAI*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.

A Proofs

Before moving into the proofs themselves, we define some additional notation.

Definition 3. *The set of* ***alternate follower best response*** *to the mixed commitment* $\mathbf{x}$ *is denoted by*

$$\mathcal{K}^*_{\mathsf{alt}}(\mathbf{x}) := \mathcal{K}^*(\mathbf{x}) - \{j^*\}.$$

We will be particularly interested in this set for the Stackelberg commitment, that is, $\mathcal{K}^*_{\mathsf{alt}}(\mathbf{x}^*_\infty)$. In general, the set will be non-empty as the follower could be agnostic between more than one pure strategy in response – it is only responding with the pure strategy j^* to break ties in the leader's favor. Figure 7 shows this demarcation of follower responses into the expected response j^*, and alternate responses to the Stackelberg commitment $\mathbf{x}^*_\infty$.

Further, we denote maximum and minimum obtainable leader payoffs respectively by

$$f_{max} := \max_{i\in[m],j\in[n]} A_{ij}$$
$$f_{min} := \min_{i\in[m],j\in[n]} A_{ij}.$$

A.1 Proof of Theorem 1

We consider a general $2 \times n$ game and denote the Stackelberg probability of leader playing pure strategy 1 by p^*_∞. Recall that $p^*_\infty \in (0,1)$ (since we have assumed for the proof that the Stackelberg commitment is mixed). Let j_{alt} be *the* alternate response to the Stackelberg commitment, i.e. we have $\mathcal{K}^*(p^*_\infty) = \{j^*_\infty, j_{\mathsf{alt}}\}$. Without loss of generality, the best-response regions can be described as

$$\mathcal{R}_{j^*_\infty} = [p^-, p^*_\infty]$$
$$\mathcal{R}_{j_{\mathsf{alt}}} = (p^*_\infty, p^+].$$

Finally, we define $f^{(2)} := \lim_{\epsilon\to 0} f_\infty(p^*_\infty + \epsilon)$. Since we are considering leader-follower games for which the function $f_\infty(.)$ is discontinuous at p^*_∞, by the tie-breaking assumption on Stackelberg commitment we will have $f^{(2)} < f^*_\infty$.

Now, we consider the quantity $f_N(p^*_\infty)$. Denoting $\widehat{P}_N$ as the empirical estimate of the quantity p^*, we have

$$\begin{aligned}
f_N(p^*_\infty) &\le \Pr\left[\widehat{P}_N \in \mathcal{R}_{j^*_\infty}\right] f^*_\infty + \Pr\left[\widehat{P}_N \in \mathcal{R}_{j_{\mathsf{alt}}}\right] f^{(2)} + \left(1 - \Pr\left[\widehat{P}_N \in \mathcal{R}_{j^*_\infty}\right] - \Pr\left[\widehat{P}_N \in \mathcal{R}_{j_{\mathsf{alt}}}\right]\right) f_{max} \\
&= \Pr\left[\widehat{P}_N \in (p^-, p^*_\infty]\right] f^*_\infty + \Pr\left[\widehat{P}_N \in (p^*_\infty, p^+]\right] f^{(2)} + \Pr\left[\widehat{P}_N \in [0,p^-] \cup (p^+, 1]\right] f_{max} \\
&= f^*_\infty - \underbrace{\Pr\left[\widehat{P}_N \in (p^*_\infty, p^+]\right]}_{T_1(N)} (f^*_\infty - f^{(2)}) + \underbrace{\Pr\left[\widehat{P}_N \in [0,p^-] \cup (p^+, 1]\right]}_{T_2(N)} (f_{max} - f^*_\infty).
\end{aligned}$$

We will now proceed to bound the probabilities $T_1(N)$ and $T_2(N)$.

First, we deal with the quantity $T_2(N)$, which reflects the probability of a mismatched response that is neither Stackelberg nor the alternate response on the boundary. By the Hoeffding bound, we have

$$\begin{aligned}
T_2(N) &:= \Pr\left[\widehat{P}_N \in [0,p^-] \cup (p^+, 1]\right] \\
&= \Pr\left[\widehat{P}_N \in [0,p^-]\right] + \Pr\left[\widehat{P}_N \in (p^+, 1]\right] \\
&\le \exp\{-2N(p^*_\infty - p^-)^2\} + \exp\{-2N(p^+ - p^*_\infty)^2\}.
\end{aligned}$$

Denoting $C'' := 2\left(\min\{p^+ - p^*_\infty, p^*_\infty - p^-\}\right)^2$, we then have

$$T_2(N) \le 2\exp\{-NC''\} \tag{4}$$

and as expected, this probability decays exponentially with N.

Next, we deal with the quantity $T_1(N)$, which reflects the probability of eliciting the alternate response on the Stackelberg boundary. We show that this event is non-vanishingly probable.

We define the following quantities

$$S_N := N\widehat{P}_N \tag{5}$$

$$Z_N := \frac{S_N - Np^*_\infty}{\sqrt{Np^*_\infty(1-p^*_\infty)}}. \tag{6}$$

Recall that Z_N is a real-valued random variable. We denote its cumulative distribution function by $F_N(.)$.

By a simple change of variables, we then have

$$\begin{aligned} T_1(N) &= \Pr\left[\widehat{P}_N \in (p^*_\infty, p^+]\right] \\ &= \Pr\left[Z_N \in \left(0, \frac{\sqrt{N}(p^+ - p^*_\infty)}{\sqrt{p^*_\infty(1-p^*_\infty)}}\right]\right] \\ &= F_N(\frac{\sqrt{N}(p^+ - p^*_\infty)}{\sqrt{p^*_\infty(1-p^*_\infty)}}) - F_N(0). \end{aligned}$$

Now, recall that $S_N = \sum_{j=1}^N I_j$ for iid random variables $I_j \sim \text{Ber}(p^*_\infty)$. Also note that since we have considered games with mixed Stackelberg commitment, we have $0 < p^*_\infty < 1$. We now invoke the first half of the classical Berry-Esseen theorem [Ber41, Ess42] stated here as a lemma.

Lemma 1. *There exists a positive constant C such that if $I_1, I_2, \ldots$ are iid random variables with $\mathbb{E}[I_1] = \mu < \infty$, $\text{var}(I_1) = \sigma^2 > 0$ and $\mathbb{E}[|I_1 - \mu|^3] = \rho < \infty$, we have*

$$|F_N(x) - \phi(x)| \leq \frac{C\rho}{\sigma^3\sqrt{N}}$$

for all $x \in \mathbb{R}$, where $\phi(.)$ denotes the CDF of the standard normal distribution $\mathcal{N}(0,1)$.

It is easy to verify that the distribution $I_1 \sim \text{Ber}(p^*_\infty)$ satisfies the above conditions. Therefore, we can directly apply Lemma 1 and get

$$F_N\left(\frac{\sqrt{N}(p^+ - p^*_\infty)}{\sqrt{p^*_\infty(1-p^*_\infty)}}\right) \geq \phi(C\sqrt{N}) - \frac{C'}{\sqrt{N}} \text{ and}$$

$$F_N(0) \leq \frac{1}{2} + \frac{C'}{\sqrt{N}}$$

for positive constant $C > 0$, thus giving

$$T_1(N) \geq \left(\phi(C'\sqrt{N}) - \frac{1}{2}\right) - \frac{C'}{\sqrt{N}}. \tag{7}$$

Substituting for the expressions for $T_1(N)$ and $T_2(N)$, we now have

$$f^*_\infty - f_N(p^*_\infty) \geq \left(\left(\phi(C'\sqrt{N}) - \frac{1}{2}\right) - \frac{C'}{\sqrt{N}}\right) C - 2C\exp\{-NC''\},$$

which corresponds exactly to Equation (2). Clearly, the right hand side of this equation is decreasing in N and so the first corollary – that $f_N(p^*_\infty) \leq f^*_\infty$ for $N \geq N_0$ – holds. Precisely, we have

$$\lim_{N\to\infty} \phi(\sqrt{N}C') = 1$$

$$\lim_{N\to\infty} \frac{C'}{\sqrt{N}} = 0$$

$$\lim_{N\to\infty} 2C\exp\{-NC''\} = 0,$$

and so we have

$$f^*_\infty - \lim_{N\to\infty} f_N(p^*_\infty) \geq \frac{f^*_\infty - f^{(2)}}{2}.$$

This is the second corollary from Theorem 1 and completes the proof. □

A.2 Proof of Theorem 2

A.2.1 Notation

For this proof, it will be convenient to consider the $(m-1)$-dimensional representation of the probability simplex, i.e.

$$\Delta_{m-1} := \{\mathbf{y} \succeq \mathbf{0} \text{ and } \langle \mathbf{y}, \mathbf{1}\rangle \leq 1\}.$$

Then, we can represent a commitment $\mathbf{x} \in \Delta_m$ by its $(m-1)$-dimensional representation $\mathbf{y} = \begin{bmatrix} x_1 & x_2 & \dots & x_{m-1}\end{bmatrix}$, and the *leader payoff* if the follower were to respond with pure strategy $j \in [n]$ by

$$\langle \mathbf{y}, \mathbf{c}_j\rangle + d_j$$

where we have

$$\mathbf{c}_j := \begin{bmatrix} a_{j,1} - a_{j,m} \\ a_{j,2} - a_{j,m} \\ \vdots \\ a_{j,m-1} - a_{j,m}\end{bmatrix}$$

$$d_j = a_{j,m}.$$

Similarly, we can represent the corresponding *follower payoff* by

$$\langle \mathbf{y}, \mathbf{b}'_j\rangle + d'_j$$

where we have

$$\mathbf{b}'_j := \begin{bmatrix} b_{j,1} - b_{j,m} \\ b_{j,2} - b_{j,m} \\ \vdots \\ b_{j,m-1} - b_{j,m}\end{bmatrix}$$

$$d'_j = b_{j,m}.$$

We can also represent this representation of the empirical estimate of $\mathbf{y}$ from N samples by $\widehat{\mathbf{Y}}_N$, and this representation Stackelberg commitment by $\mathbf{y}^*_\infty$.

Now, we can consider all the functions introduced in Section 2.2 in terms of the commitment $\mathbf{x}$ and equivalently define them in terms of the $(m-1)$-dimensional representation of the commitment, $\mathbf{y}$.

We also denote the pth operator norm of a matrix by $\|\!|.|\!\|_p$.

A.2.2 The commitment construction

We consider the $(m-1)$-dimensional representation of the best-response-region corresponding to the Stackelberg commitment, $\mathcal{R}_{j^*}$. There are many things to consider while constructing a robust commitment. The first, and obvious, one would be that the follower should respond the same way as it would to Stackelberg when it observes the full mixture. That is, we would have $j^*(\mathbf{y}_N) = j^*$ or alternatively stated, $\mathbf{y}_N \in \mathcal{R}_{j^*}$.

Intuitively, the expected payoff of a leader commitment under observational uncertainty, particularly in terms of gap to the optimal Stackelberg payoff, will depend on two factors: one, how likely the follower is to respond the same as it would if it observed the full commitment; and two, how "far" the leader commitment mixture is from the optimal Stackelberg commitment mixture. We qualitatively show this dependence in the following lemma.

Lemma 2. *Consider a commitment $\mathbf{y}_N$ for which we can provide the following guarantee:*

$$\Pr[\widehat{\mathbf{Y}}_N \notin \mathcal{R}_{j^*}] \leq \epsilon_N.$$

We then have

$$f^*_\infty - f_N(\mathbf{y}_N) \leq 2(1-\epsilon_N) f_{max} \|\mathbf{y}_N - \mathbf{y}^*_\infty\|_1 + \epsilon_N (f^*_\infty - f_{min})$$

Proof. We have

$$\begin{aligned} f_N(\mathbf{y}_N) &= \sum_{j=1}^{n} \Pr[\widehat{\mathbf{Y}}_N \in \mathcal{R}_j](\langle \mathbf{y}_N, \mathbf{c}_j \rangle + d_j) \\ &\geq \Pr[\widehat{\mathbf{Y}}_N \in \mathcal{R}_{j^*}](\langle \mathbf{y}_N, \mathbf{c}_{j^*} \rangle + d_{j^*}) + (1 - \Pr[\widehat{\mathbf{Y}}_N \in \mathcal{R}_{j^*}]) f_{min} \\ &\geq (1-\epsilon_N)\left(\langle \mathbf{y}_N, \mathbf{c}_{j^*} \rangle + d_{j^*} - f_{min}\right) + f_{min} \\ &= (1-\epsilon_N)\left(\langle \mathbf{y}_N, \mathbf{c}_{j^*} \rangle + d_{j^*}\right) + \epsilon_N f_{min}. \end{aligned}$$

Recall that we have $f^*_\infty = \langle \mathbf{y}^*_\infty, \mathbf{c}_{j^*} \rangle + d_{j^*}$. Therefore, the gap from Stackelberg is bounded as

$$\begin{aligned} f^*_\infty - f_N(\mathbf{y}_N) &\leq (1-\epsilon_N)\langle \mathbf{y}^*_\infty - \mathbf{y}_N, \mathbf{c}_{j^*} \rangle + \epsilon_N (f^*_\infty - f_{min}) \\ &\leq (1-\epsilon_N)\|\mathbf{c}_{j^*}\|_\infty \|\mathbf{y}_N - \mathbf{y}^*_\infty\|_1 + \epsilon_N (f^*_\infty - f_{min}) \\ &\leq 2(1-\epsilon_N) f_{max} \|\|\mathbf{y}_N - \mathbf{y}^*_\infty\|_1 + \epsilon_N (f^*_\infty - f_{min}), \end{aligned}$$

where the second inequality follows from Holder's inequality. This proves the lemma. □

This lemma implies that we want a commitment construction $\mathbf{y}_N$ with the following two-fold guarantee[15].

1. $\|\mathbf{y}_N - \mathbf{y}^*_\infty\|_1$ is bounded (and ideally vanishes with N).
2. $\widehat{\mathbf{Y}}_N \in \mathcal{R}_{j^*}$ with high probability.

A.2.3 Commitment construction using localized geometry

We will leverage the special structure of the Dikin ellipsoid [KN12] used in interior-point methods to make our commitment constructions. Observe that $\mathbf{y}^*_\infty$ is always going to be on an extreme point (vertex) of the best-response-polytope[16] $\mathcal{R}_{j^*}$. We now collect the $k = |\mathcal{K}^*(\mathbf{x}^*_\infty)|$ constraints that are satisfied *with equality* at $\mathbf{x}^*_\infty$:

$$\langle \mathbf{y}, \mathbf{b}'_j \rangle + d'_j \leq \langle \mathbf{y}, \mathbf{b}'_{j^*} \rangle + d'_{j^*} \text{ for all } j \in \mathcal{K}^*(\mathbf{x}^*_\infty).$$

This is simply the constraint set for commitments such that the follower prefers to respond with pure strategy j^* over any pure strategy $j \in \mathcal{K}^*(\mathbf{y}^*_\infty)$ (i.e. any pure strategy whose corresponding best-response-polytope shares a boundary with the Stackelberg best-response-polytope at point $\mathbf{y}^*$), and can be thought of as the set of *local constraints to the Stackelberg vertex* in the best-response polytope $\mathcal{R}_{j^*}$. We also collect the other constraints that describe $\mathcal{R}_{j^*}$:

$$\begin{aligned} \langle \mathbf{y}, \mathbf{b}'_j \rangle + d'_j \leq \langle \mathbf{y}, \mathbf{b}'_{j^*} \rangle + d'_{j^*} \text{ for all } j \notin \mathcal{K}^*(\mathbf{x}^*_\infty) \cup \{j^*\} \\ \mathbf{y} \succeq 0 \\ \langle \mathbf{1}, \mathbf{y} \rangle \leq 1, \end{aligned}$$

[15] Interestingly, the fact that $\mathbf{y}^*_\infty$ is on an extreme point of $\mathcal{R}_{j^*}$ will imply that the two conditions are at odds with one another, and we will need to trade them off. For instance, choosing $\mathbf{y}_N = \mathbf{y}^*_\infty$ would satisfy the second condition perfectly by being as close as possible to the Stackelberg commitment, but there would be no guarantee on the best-response as it lies on the boundary of the best-response region.

[16] Recall that the Stackelberg equilibrium is the solution to the LP defined on the best-response-polytope [CS06].

and together with the local constraints at the Stackelberg vertex, these describe the global constraints for the polytope.

We represent the system of inequalities in matrix form as: $B\mathbf{y} \preceq \mathbf{c}$ for some $B \in \mathbb{R}^{k\times(m-1)}$ and some $\mathbf{c} \in \mathbb{R}^k$. We leverage the following useful fact about a general set of linear constraints.

Fact 1. *For any parameterization of linear constraints $(B, \mathbf{c})$, there exists an affine transformation $\mathbf{y}' = T_1\mathbf{y} + T_2$ (where $T_1 \in \mathbb{R}^{(m-1)\times(m-1)}$ is invertible and $T_2 \in \mathbb{R}^{m-1}$) and a matrix $B' \in \mathbb{R}^{k\times(m-1)}$ such that*

$$B\mathbf{y} \preceq \mathbf{c} \iff B'\mathbf{y}' \preceq \mathbf{1}.$$

We denote the transformation function by $T(\cdot)$ and its inverse by $T^{-1}(\cdot)$. In particular, we note the relationship $B = B'T_1$.

The above fact is useful[17] because it is most convenient to define our class of commitments in the transformed space $\mathbf{y}' = T(\mathbf{y})$.

Definition 4. *For a particular value of $\delta \in (0, 1)$, Stackelberg commitment $\mathbf{y}^*_\infty$, and local constraints modeled by $(B, \mathbf{c})$, we define a δ-deviation commitment by*

$$\begin{aligned}\mathbf{y}(\delta; \mathbf{y}^*_\infty) &:= T^{-1}(\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty)) \text{ where}\\ \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty) &:= (1-\delta)(\mathbf{y}^*)'_\infty.\end{aligned}$$

Our robust commitments $\{\mathbf{y}_N\}_{N\geq 1}$ are going to be taken out of the set of δ-deviation commitments, with appropriately chosen values of $\{\delta_N\}_{N\geq 1}$. *Clearly, the computational complexity of constructing any δ-deviation commitment is equivalent to the complexity of computing the Stackelberg equilibrium itself.*

To understand how to set these values, we will turn to the question of how to satisfy the three conditions above.

First, we observe that $\mathbf{y}(\delta; \mathbf{y}^*_\infty)$ satisfies the *local constraints* $B\mathbf{y} \preceq \mathbf{c}$ for any $\delta \in (0, 1)$. Because of Fact 1, it suffices to show that its affine transformation $\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty)$ satisfies the local constraints $B'\mathbf{y}' \preceq \mathbf{1}$. Recall that $(\mathbf{y}^*)'_\infty$ satisfies *all the local constraints* with equality, i.e. we have $B'(\mathbf{y}^*)'_\infty = \mathbf{1}$. From the definition of the commitment, we thus have

$$\begin{aligned}B'\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty) &= (1-\delta)B'(\mathbf{y}^*)'_\infty\\ &= (1-\delta)\mathbf{1} \preceq \mathbf{1}.\end{aligned}$$

Next, we turn to the question of how close such a defined commitment would be from the Stackelberg commitment $\mathbf{y}^*_\infty$, in terms of the ℓ_1 norm. For this, we have

$$\begin{aligned}\|\mathbf{y}(\delta; \mathbf{y}^*_\infty) - \mathbf{y}^*_\infty\|_1 &= \|T_1^{-1}(\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty) - (\mathbf{y}^*)'_\infty)\|_1\\ &= \delta\|T^{-1}(\mathbf{y}^*)'_\infty\|_1\\ &= \delta\|\mathbf{y}^*_\infty\|_1 \leq \delta.\end{aligned}$$

Therefore, we have

$$\|\mathbf{y}(\delta; \mathbf{y}^*_\infty) - \mathbf{y}^*_\infty\|_1 \leq \delta. \tag{8}$$

In lieu of Lemma 2, we wish to choose values $\{\delta_N\}_{N\geq 1}$ (to create commitments $\{\mathbf{y}_N\}_{N\geq 1}$) such that δ_N decreases with N sufficiently fast, while maintaining a high probability of staying in the best-response polytope $\mathcal{R}_{j^*}$. To understand the rate at which we can decrease δ_N, we need to prove a high-probability best-response guarantee.

[17] A subtle point is that there do exist special cases of polytope constraints for which Fact 1 is true only with an augmentation of the variable space from m to $2m$ dimensions. Then, defining the invertible map becomes trickier. Nevertheless, for ease of exposition and clarity in the proof, we assume that we can indeed carry out the affine transformation without augmenting the dimension.

A.2.4 Using the local Dikin ellipsoid as a confidence ball

For a (affine-transformed) commitment $\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty)$, we make use of the *local Dikin ellipsoid* centered at $\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty)$, defined below for an arbitrary point $\mathbf{y}'$.

Definition 5 ([KN12]). *For constraint set $B'\mathbf{y}' \preceq \mathbf{1}$, the **Dikin ellipsoid** of radius r centered at $\mathbf{y}'$ is given by*

$$\mathbb{B}_{B',\mathbf{1},\mathbf{y}'}(r) := \{\mathbf{z}' : (\mathbf{z}' - \mathbf{y}')^\top H(\mathbf{y}')(\mathbf{z}' - \mathbf{y}') \leq r\}, \tag{9}$$

where we define

$$H(\mathbf{y}') := \sum_{i=1}^{k} \frac{(\mathbf{b}')_i (\mathbf{b}')_i^\top}{(1 - \langle (\mathbf{b}')_i, \mathbf{y}' \rangle)^2}. \tag{10}$$

The Dikin ellipsoid has two special properties [KN12]:

1. ***Affine invariance**: (using the notation from Fact 1) For transformation $\mathbf{y}' = T(\mathbf{y})$, the Dikin ellipsoid of radius r centered at the point $\mathbf{y}$ for the polytope $B\mathbf{y} \preceq \mathbf{c}$ is $\mathbb{B}_{B,\mathbf{c},\mathbf{y}'}(r) = T^{-1}(\mathbb{B}_{B',\mathbf{1},\mathbf{y}'}(r))$.*
2. ***Interior guarantee**: For any interior point $\mathbf{y}'$ (according to the constraint set $B'\mathbf{y}' \preceq \mathbf{1}$), the Dikin ellipsoid of radius 1 centered at $\mathbf{y}'$ is contained in the feasibility set, that is,*

$$\mathbf{z}' \in \mathbb{B}_{B',\mathbf{1},\mathbf{y}'}(1) \implies B'\mathbf{z}' \preceq \mathbf{1}.$$

We center our Dikin ellipsoid at $\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty)$, and observe that the constraint takes on a particularly nice form, as stated by the following simple lemma.

Lemma 3. *For any $\delta \in (0, 1)$, the Dikin ellipsoid can be expressed as*

$$\mathbb{B}_{B',\mathbf{1},\mathbf{y}'(\delta;(\mathbf{y}^*)'_\infty)}(1) = \{\mathbf{z}' : \|B'(\mathbf{z}' - \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty))\|_2 \leq \delta\}. \tag{11}$$

Furthermore, in the original space we can write

$$\mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1) = \{\mathbf{z} : \|B(\mathbf{z} - \mathbf{y}(\delta; \mathbf{y}^*_\infty))\|_2 \leq \delta\}. \tag{12}$$

Proof. From Definition 4, we observe that $B'\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty) = (1 - \delta)B'(\mathbf{y}^*)'_\infty = (1 - \delta)\mathbf{1}$. This implies that

$$1 - \langle (\mathbf{b}')_i, \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty) \rangle = 1 - (1 - \delta) = \delta,$$

and thus we have

$$\begin{aligned} H(\mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty)) &= \frac{\sum_{i=1}^{k} (\mathbf{b}')_i (\mathbf{b}')_i^\top}{\delta^2} \\ &= \frac{(B')^\top B'}{\delta^2} \end{aligned}$$

where in the last equality step, we have used $(B')^\top B' = \sum_{i=1}^{k} (\mathbf{b}')_i (\mathbf{b}')_i^\top$, noting that $(\mathbf{b}')_i$ denotes the i^{th} row of B'.

Thus, the ellipsoid constraint in Equation (9) can be rewritten as

$$\begin{aligned} \frac{1}{\delta^2} (\mathbf{z}' - \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty))^\top (B')^\top B' (\mathbf{z}' - \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty)) \leq 1 \\ \implies \|B'(\mathbf{z}' - \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty))\|_2^2 \leq \delta^2 \\ \implies \|B'(\mathbf{z}' - \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty))\|_2 \leq \delta, \end{aligned}$$

thus completing the first part of the proof (Equation (11)).

For the second part of the proof, we use the affine invariance property of the Dikin ellipsoid, which tells us that

$$\begin{aligned}\mathbf{z} \in \mathbb{B}_{B,\mathbf{c},\mathbf{y}}(1) &\implies \mathbf{z}' = T_1\mathbf{z} + T_2 \in \mathbb{B}_{B',\mathbf{1},\mathbf{y}'}(1) \\ &\implies \|B'(\mathbf{z}' - \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty))\|_2 \leq \delta.\end{aligned}$$

Now, observe that

$$\begin{aligned}B'(\mathbf{z}' - \mathbf{y}'(\delta; (\mathbf{y}^*)'_\infty)) &= B'(T_1\mathbf{z} + T_2 - T_1\mathbf{y}(\delta; \mathbf{y}^*_\infty) - T_2) \\ &= (B'T_1)(\mathbf{z} - \mathbf{y}(\delta; \mathbf{y}^*_\infty)) \\ &= B(\mathbf{z} - \mathbf{y}(\delta; \mathbf{y}^*_\infty))\end{aligned}$$

where in the last step we have used the relationship $B = B'T_1$ from Fact 1. Putting these observations together, we have

$$\mathbf{z} \in \mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1) \implies \|B(\mathbf{z} - \mathbf{y}(\delta; \mathbf{y}^*_\infty))\|_2 \leq \delta,$$

completing the second part of the proof. □

At this stage, it is worth remembering that the commitment is *mixed*, and the payoff from using a δ-deviation commitment $\mathbf{y}(\delta; \mathbf{y}^*_\infty) \in \Delta_{m-1}$ under a finite number of observations N depends on the guarantee that its observed empirical distribution $\widehat{\mathbf{Y}}_N$ (typically) stays inside the best-response region. As a starting point we need to guarantee that at least the *local vertex constraints* are not violated.

Note that $\mathbf{y}(\delta; \mathbf{y}^*_\infty) \in \Delta_{m-1}$ is an interior point for any $\delta > 0$, and thus the interior guarantee property of the Dikin ellipsoid can be applied. We thus know that if the empirical distribution of the commitment stays inside the Dikin ellipsoid centered at the actual commitment, it will stay inside the local constraint feasibility set. Thus, it makes sense to use the Dikin ellipsoid as a confidence ball and tail bound the probability that the empirical estimate lies outside this ball. Because of the weighted ℓ_2-ball structure on the particular ellipsoid corresponding to a δ-deviation commitment that we proved in Lemma 3, this is not difficult to do. We state this formally in the following lemma.

Lemma 4. *For a given $\delta > 0$, let $\widehat{\mathbf{Y}}_N$ be the empirical distribution of N samples drawn from the δ-deviation commitment $\mathbf{y}(\delta; \mathbf{y}^*_\infty)$. Then, we have*

$$\Pr\left[\widehat{\mathbf{Y}}_N \notin \mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1)\right] \leq 3\exp\{-\frac{N\delta^2}{25\|\!|B|\!\|^2_{\mathsf{op}}}\}$$

provided that $N \geq \frac{20m\|\!|B|\!\|^2_{\mathsf{op}}}{\delta^2}$.

Proof. The proof is a simple consequence of Devroye's lemma [Dev83], which tail bounds the total variation between the empirical estimate of a discrete distribution and the true distribution.

Lemma 5 ([Dev83])**.** *Let $\widehat{\mathbf{Y}}_N$ be the empirical distribution of N samples drawn from any distribution $\mathbf{y} \in \Delta_{m-1}$. Then, as long as $\delta \geq \sqrt{\frac{20m}{N}}$ we have*

$$\Pr\left[\|\widehat{\mathbf{Y}}_N - \mathbf{y}\|_1 \geq \delta\right] \leq 3\exp\{\frac{N\delta^2}{25}\}.$$

We note from Lemma 3 that

$$\widehat{\mathbf{Y}}_N \notin \mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1) \implies \|B(\widehat{\mathbf{Y}}_N - \mathbf{y}(\delta; \mathbf{y}^*_\infty))\|_2 > \delta,$$

and thus, we have

$$\begin{aligned}\Pr\left[\widehat{\mathbf{Y}}_N \notin \mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1)\right] &= \Pr\left[\|B(\widehat{\mathbf{Y}}_N - \mathbf{y}(\delta;\mathbf{y}^*_\infty))\|_2 > \delta\right] \\ &\overset{\text{(i)}}{\leq} \Pr\left[\|B\|_{op}\|\widehat{\mathbf{Y}}_N - \mathbf{y}(\delta;\mathbf{y}^*_\infty))\|_2 > \delta\right] \\ &\overset{\text{(ii)}}{\leq} \Pr\left[\|B\|_{op}\|\widehat{\mathbf{Y}}_N - \mathbf{y}(\delta;\mathbf{y}^*_\infty))\|_1 > \delta\right] \\ &= \Pr\left[\|\widehat{\mathbf{Y}}_N - \mathbf{y}(\delta;\mathbf{y}^*_\infty))\|_1 > \delta/\|B\|_{op}\right]\end{aligned}$$

where inequality (i) uses the definition of the operator norm and inequality (ii) uses the fact that $\|\mathbf{v}\|_2 \leq \|\mathbf{v}\|_1$ for any finite-dimensional vector $\mathbf{v}$. Applying Lemma 5 directly then gives us

$$\Pr\left[\widehat{\mathbf{Y}}_N \notin \mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1)\right] \leq 3\exp\{-\frac{N\delta^2}{25\|\!|B|\!\|^2_{\mathsf{op}}}\}$$

as long as

$$\begin{aligned}&\frac{\delta}{\|\!|B|\!\|_{\mathsf{op}}} \geq \sqrt{\frac{20m}{N}} \\ \implies\ & N \geq \frac{20m\|\!|B|\!\|^2_{\mathsf{op}}}{\delta^2}.\end{aligned}$$

This completes the proof.

□

A.2.5 Completing proof of Theorem 2: Ensuring global constraint satisfiability

Let us recap what we have proved so far about a δ-deviation commitment $\mathbf{y}(\delta;\mathbf{y}^*_\infty)$ for any $\delta \in (0,1)$.

1. For N samples from $\mathbf{y}(\delta;\mathbf{y}^*_\infty)$, we have $\Pr\left[\widehat{\mathbf{Y}}_N \notin \mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1)\right] \leq 3\exp\{-\frac{N\delta^2}{25\|\!|B|\!\|^2_{\mathsf{op}}}\}$ (from Lemma 4).
2. $\|\mathbf{y}(\delta;\mathbf{y}^*_\infty) - \mathbf{y}^*_\infty\|_1 \leq \delta$.

Thus, from Lemma 2 we have for any δ-deviation commitment,

$$f^*_\infty - f_N(\mathbf{y}(\delta;\mathbf{y}^*_\infty)) \leq 2\delta f_{max} + \Pr\left[\widehat{\mathbf{Y}}_N \notin \mathcal{R}_{j^*}\right](f^*_\infty - f_{min})$$

Thus, if we had $\mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1) \subset \mathcal{R}_{j^*}$, we would have

$$\Pr\left[\widehat{\mathbf{Y}}_N \notin \mathcal{R}_{j^*}\right] \leq \Pr\left[\widehat{\mathbf{Y}}_N \notin \mathbb{B}_{B,\mathbf{c},\mathbf{y}(\delta;\mathbf{y}^*_\infty)}(1)\right] \leq 3\exp\{-\frac{N\delta^2}{25\|\!|B|\!\|^2_{\mathsf{op}}}\}.$$

However, the set $\mathcal{R}_{j^*}$ includes *global constraints* in addition to the local constraints $B\mathbf{y} \preceq \mathbf{c}$, and all points in the *local* Dikin ellipsoid need not satisfy these constraints. This is the final technicality in the proof that we now deal with. We will see that for a small enough value of δ (that depends on how the local geometry of the polytope relates to the global geometry), we can guarantee global satisfiability. Let the constraints corresponding to the convex polytope $\mathcal{R}_{j^*}$ be represented by $C\mathbf{y} \preceq \mathbf{d}$, and the corresponding constraints after the affine transformation (T_1, T_2) be represented as $C'\mathbf{y}' \preceq \mathbf{d}'$ (where the values of $\mathbf{d}'$ corresponding the local constraints are 1). Thus, for the vertex $(\mathbf{y}^*)'_\infty$, we can define the quantity

$$\mathcal{Z}(\mathcal{R}_{j^*};(\mathbf{y}^*)'_\infty) := \sup\{\delta > 0 : \mathbf{z}' \in \mathbb{B}_{B',\mathbf{1},\mathbf{y}'(\delta;(\mathbf{y}^*)'_\infty)}(1) \implies C'\mathbf{z}' \preceq \mathbf{d}'\}.$$

Because $\mathcal{R}_{j^*}$ is *non-empty and convex*, we have $\mathcal{Z}(\mathcal{R}_{j^*};(\mathbf{y}^*)'_\infty) > 0$.

From this definition, under the condition $\delta < \mathcal{Z}(\mathcal{R}_{j^*}; (\mathbf{y}^*)'_\infty)$ we have

$$\begin{aligned} &\mathbb{B}_{B',\mathbf{1},\mathbf{y}'(\delta;(\mathbf{y}^*)'_\infty)}(1) \subset T(\mathcal{R}_{j^*}) \\ &\implies \mathbb{B}_{B,\mathbf{1},\mathbf{y}(\delta;\mathbf{y}^*_\infty}(1) \subset \mathcal{R}_{j^*}, \end{aligned}$$

where the last implication is because of the affine-invariance property of the Dikin ellipsoid.

On the other hand, we used the condition $N \geq \frac{20m\|B\|_{\mathsf{op}}^2}{\delta^2}$ to prove Lemma 4. Combining these inequalities tells us that we require $N > \frac{20m\|B\|_{\mathsf{op}}^2}{\mathcal{Z}(\mathcal{R}_{j^*};(\mathbf{y}^*)'_\infty)^2} = \mathcal{O}(m)$ to prove our result.

Then, we formally define our robust commitment for a particular value of N below, and prove this final lemma which is essentially a formal statement of Theorem 2.

Lemma 6. *For every $N > \frac{20m\|B\|_{\mathsf{op}}^2}{\mathcal{Z}(\mathcal{R}_{j^*};(\mathbf{y}^*)'_\infty)^2}$, and every $p < 1/2$, we define the p-robust commitment as a $\delta_{N,p}$-deviation commitment $\mathbf{y}_{N,p} := \mathbf{y}(\delta_{N,p}; \mathbf{y}^*_\infty)$, where*

$$\delta_{N,p} := \mathcal{Z}(\mathcal{R}_{j^*}; (\mathbf{y}^*)'_\infty) \left(\frac{m}{N}\right)^p. \tag{13}$$

We then have

$$\begin{aligned} f_N(\mathbf{y}_{N,p}) &\leq \frac{2f_{max}}{\cdot} \mathcal{Z}(\mathcal{R}_{j^*}; (\mathbf{y}^*)'_\infty) \cdot \left(\frac{m}{N}\right)^p + 3\exp\{-\frac{m^{2p} \cdot \mathcal{Z}(\mathcal{R}_{j^*}; (\mathbf{y}^*)'_\infty)^2 \cdot N^{1-2p}}{25\|B\|_{\mathsf{op}}^2}\}(f^*_\infty - f_{min}) \\ &= \mathcal{O}\left(\left(\frac{m}{N}\right)^p + \exp\{-\omega(1) \cdot N^{1-2p}\}\right). \end{aligned}$$

Proof. This is a simple consequence of everything put together. Since $N > m$, we have $\delta_{N,p} < \mathcal{Z}(\mathcal{R}_{j^*}; (\mathbf{y}^*)'_\infty)$ and thus we have $\mathbb{B}_{B,\mathbf{1},\mathbf{y}(\delta_{N,p};\mathbf{y}^*_\infty}(1) \subset \mathcal{R}_{j^*}$. This tells us that

$$\Pr\left[\widehat{\mathbf{Y}}_N \notin \mathcal{R}_{j^*}\right] \leq 3\exp\{-\frac{N\delta_{N,p}^2}{25\|B\|_{\mathsf{op}}^2}\}.$$

and thus from Lemma 2 we get the following expression:

$$f^*_\infty - f_N(\mathbf{y}_N) \leq 2\delta_{N,p} f_{max} + 3\exp\{-\frac{N\delta_{N,p}^2}{25\|B\|_{\mathsf{op}}^2}\}(f^*_\infty - f_{min}).$$

Directly substituting the expression for $\delta_{N,p}$ in Equation (13) into the above expression completes the proof. □

A.3 Proof of Theorem 3

Recall that $f^*_N := \max_{\mathbf{x}\in\Delta_m} f_N(\mathbf{x})$. To prove an upper bound on f^*_N, we will upper bound $f_N(\mathbf{x})$ for every $\mathbf{x} \in \Delta_m$.

Without loss of generality the same proof method will extend to all $\mathbf{x} \in \Delta_m$. Denoting as shorthand $p_j(\mathbf{x}) := \Pr\left[\widehat{\mathbf{X}}_N \in \mathcal{R}_j\right]$, we have

$$\begin{aligned} f_N(\mathbf{x}) &= \sum_{j=1}^n p_j(\mathbf{x}) \langle \mathbf{a}_j, \mathbf{x}\rangle &(14)\\ &= \sum_{j=1}^n T_j(\mathbf{x}) &(15) \end{aligned}$$

where we denote $T_j(\mathbf{x}) := p_j(\mathbf{x})\langle \mathbf{a}_j, \mathbf{x}\rangle$. We will proceed to upper bound the quantity $T_j(\mathbf{x})$ for every $\mathbf{x} \in \Delta_m$ and every $j \in [n]$.

To do this, we will see that it is natural to divide all the pure strategy responses that are possible to a commitment $\mathbf{x}$ into three categories. The first is the expected response $j^*(\mathbf{x})$. The second is the set of responses whose regions are *adjacent* to the expected response as defined below.

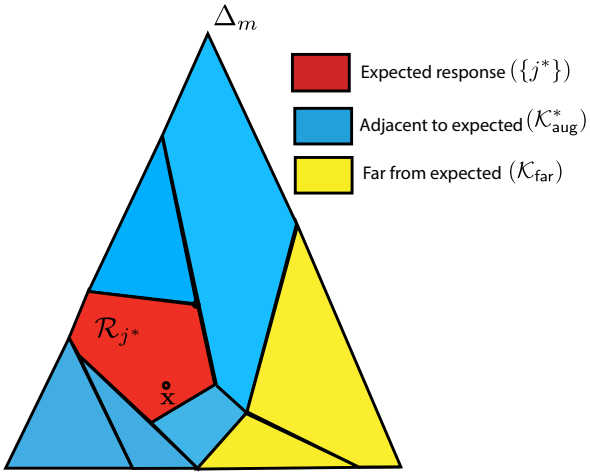


Figure 8. Illustration of partition of the set of follower responses, $[n]$, into sets $\{j^*\}$ (red region), $\mathcal{K}^*_{\text{aug}}$ (blue regions) and $\mathcal{K}^*_{\text{far}}$ (yellow regions).

Definition 6. *For a particular commitment $\mathbf{x} \in \Delta_m$, the set of **adjacent-to-expected** responses $\mathcal{K}^*_{aug}(\mathbf{x})$ is the set of all best-responses whose corresponding best-response-regions share a boundary with the best-response-region corresponding to the best response to $\mathbf{x}$. Formally, we have*

$$\mathcal{K}^*_{\text{aug}}(\mathbf{x}) := \{j \in [n] : j \neq j^*(\mathbf{x}) \text{ and } cl(\mathcal{R}_{j^*(\mathbf{x})}) \cap cl(\mathcal{R}_j) \neq \emptyset\}.$$

We also define $\mathcal{K}_{\text{far}} := [n] - (\{j^*(\mathbf{x})\} \cup \mathcal{K}^*_{\text{aug}}(\mathbf{x}))$ as the set of all follower responses that are "far" from the expected response in this sense.

The illustration in Figure 8 shows this division.

For the rest of the proof, we will drop the term $\mathbf{x}$ from the notation and denote $\mathcal{K}^*_{\text{aug}} := \mathcal{K}^*_{\text{aug}}(\mathbf{x})$ as well as $j^* := j^*(\mathbf{x})$. This is done for notational simplicity.

It is first easy to show a bound on $T_{j^*}(\mathbf{x})$. In particular, we can directly use the definition of the function $f_\infty(.)$ to obtain

$$\begin{aligned} T_{j^*}(\mathbf{x}) &= p_{j^*}(\mathbf{x})\langle \mathbf{a}_{j^*}, \mathbf{x}\rangle && (16)\\ &= p_{j^*}(\mathbf{x}) f_\infty(\mathbf{x}) && (17)\\ &\leq p_{j^*}(\mathbf{x}) f^*_\infty. && (18) \end{aligned}$$

This inequality is also intuitive because the leader would only hope to gain from eliciting a different-than-expected response. Next, we deal with this cases.

A.3.1 "Far"-from-expected responses

We collect the set of commitments that (if observed fully) would elicit a response far away from the actual expected response. Formally, we denote $\mathcal{R}_{\text{far}} := \cup_{j \in \mathcal{K}_{\text{far}}} \mathcal{R}_j$. Now, we wish to bound the term

$$T_{far} := \sup_{\mathbf{x} \in \Delta_m} \sum_{j \in \mathcal{K}_{\text{far}}} T_j(\mathbf{x}).$$

By definition, we have $\text{cl}(\mathcal{R}_{j^*}) \cap \text{cl}(\mathcal{R}_{far}) = \emptyset$. Because we are considering *finite* games, i.e. $n < \infty$, there exists a constant $C > 0$ that depends solely on the parameters of the game such that

$$\inf_{\mathbf{x} \in \mathcal{R}_{j^*}, \mathbf{x}' \in \mathcal{R}_{far}} D_{\text{KL}}(\mathbf{x}' \parallel \mathbf{x}) \geq C. \tag{19}$$

Geometrically, Figure **??** shows this separation between the expected-response-region and any far-from-expected-response-region. To understand the probability of eliciting such responses, we invoke a classical result from large-deviations theory, Sanov's theorem [CK11]. The upper bound part of the theorem is restated here as a lemma and with appropriate notation.

Lemma 7. *Let $I_1, I_2, \ldots, I_N$ be i.i.d $\sim \mathbf{x}$ for any $\mathbf{x} \in \Delta_m$ and $\widehat{\mathbf{X}}_N$ denote the empirical estimate. Then, for any region $\mathcal{R} \subseteq \Delta_m$, we have*

$$\Pr\left[\widehat{\mathbf{X}}_N \in \mathcal{R}\right] \leq (N+1)^m 2^{-N \inf_{\mathbf{x}' \in \mathcal{R}} D_{\mathsf{KL}}(\mathbf{x}' \parallel \mathbf{x})}. \tag{20}$$

Combining equations (20) and (19), we therefore get

$$\begin{aligned} T_{far} &\leq \left[\sup_{\mathbf{x} \in \Delta_m} \sum_{j \in \mathcal{K}_{\mathsf{far}}} p_j(\mathbf{x})\right] f_{max} && (21)\\ &\leq \left[(N+1)^m 2^{-N \inf_{\mathbf{x} \in \mathcal{R}_{j^*}, \mathbf{x}' \in \mathcal{R}_{far}} D_{\mathsf{KL}}(\mathbf{x}' \parallel \mathbf{x})}\right] f_{max} && (22)\\ &\leq (N+1)^m 2^{-NC} f_{max} && (23)\\ &= C(N+1)^m \exp\{-NC\} f_{max}. && (24) \end{aligned}$$

The rationale for calling these responses *far-from-expected* is now clear: there is a minimum constant separation in terms of the KL-divergence from the expected best response, and so the probability of realizing these responses decreases exponentially with N.

Dealing with the adjacent-to-expected responses is more delicate. We turn to this case next.

A.3.2 Adjacent-to-expected responses

Consider the set of adjacent-to-expected response $\mathcal{K}^*_{\mathsf{aug}}$. We wish to bound the term $\sum_{j \in \mathcal{K}^*_{\mathsf{aug}}} T_j(\mathbf{x})$. It turns out that we can no longer control the probability that one of these responses is elicited for all choices of $\mathbf{x} \in \mathcal{R}_{j^*}$ – this is because the commitment $\mathbf{x}$ could be chosen arbitrarily close to a boundary of its expected-response-region. However, we can bound the ensuing payoff as a function of how close the commitment is to a boundary. This notion of closeness is defined in terms of the ℓ_1-norm below.

Definition 7. *For a commitment $\mathbf{x} \in \mathcal{R}_{j^*}$ and a particular adjacent response $j \in \mathcal{K}^*_{\mathsf{aug}}$, we define its minimum distance to the boundary by*

$$\delta_1(\mathbf{x}; j) := \inf_{\mathbf{x}' \in cl(\mathcal{R}_j)} \|\mathbf{x} - \mathbf{x}'\|_1.$$

First, we use this notion to bound the maximum possible payoff that could be elicited.

Lemma 8. *For any commitment $\mathbf{x} \in \mathcal{R}_{j^*}$, we have*

$$T_j(\mathbf{x}) \leq p_j(\mathbf{x}) \left[f^*_\infty + f_{max} \delta_1(\mathbf{x}; j)\right].$$

Proof. Let $\widetilde{\mathbf{x}} \in \arg\min_{\mathbf{x}' \in \mathrm{cl}(\mathcal{R}_j)} \|\mathbf{x} - \mathbf{x}'\|_1$. (Note that the minimum exists because we've taken the closure of the region.) Using Holder's inequality, we have

$$\begin{aligned} \langle \mathbf{a}_j, \mathbf{x} - \widetilde{\mathbf{x}} \rangle &\leq \|\mathbf{a}_j\|_\infty \|\mathbf{x} - \widetilde{\mathbf{x}}\|_1 \\ &\leq f_{max} \delta_1(\mathbf{x}; j). \end{aligned}$$

For every $j \in \mathcal{K}^*_{aug}$ we have

$$\begin{aligned} \langle \mathbf{a}_j, \mathbf{x} \rangle &\leq \langle \mathbf{a}_j, \widetilde{\mathbf{x}} \rangle + f_{max} \delta_1(\mathbf{x}; j) \\ &\leq f_\infty(\widetilde{\mathbf{x}}) + f_{max} \delta_1(\mathbf{x}; j) \\ &\leq f^*_\infty + f_{max} \delta_1(\mathbf{x}; j). \end{aligned}$$

where we are crucially using the fact that $\widetilde{\mathbf{x}}$ lies on the boundary and the tie-breaking assumption, to tie its payoff to the function $f_\infty(.)$. Substituting the above bound into the definition of $T_j(\mathbf{x})$ completes the proof. □

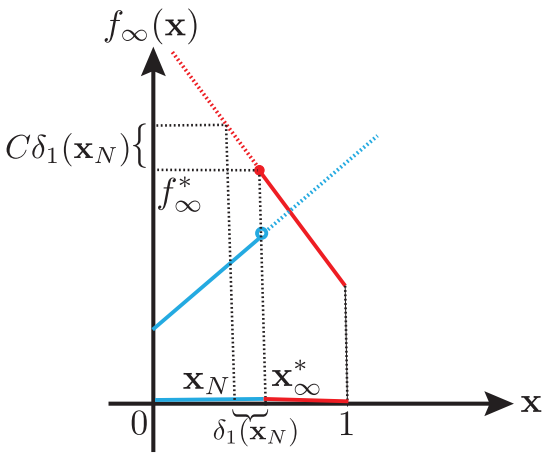


Figure 9. Illustration showing the potential gain in payoff obtainable by eliciting a different-than-expected response for a 2×2 game.

Lemma 8 is important because it limits the potential of leader gain from eliciting an adjacent follower response, even if she is able to do this with high probability, i.e. by committing very close to a boundary. Figure 9 clearly illustrates this for a 2×2 game: here, the leader might wish to elicit different-than-expected response 2 with high probability. However, to do this she would have to commit close to the boundary between regions expecting responses 1 and 2, resulting in her payoff being close to an objective function value of $f_\infty(.)$ (in the figure, depicted as the optimum payoff f^*_∞). For a general $m \times n$ game, the picture stays the same.

Since the quantity $\delta_1(\mathbf{x})$ can take values anywhere in the interval $[0, 2]$ (by the triangle inequality), we will still want to control the quantity $p_j(\mathbf{x})$ for large enough values of δ. We will again use Devroye's lemma (Lemma 5) for tail bounding the total variation between the empirical estimate of a distribution and a true distribution. Recall that the condition required for it to be applied was $\delta \geq \sqrt{\frac{20m}{N}}$.

It is natural to further divide the set $\mathcal{K}^*_{\mathsf{aug}}$ into two subsets, defined by the commitment $\mathbf{x}$.

$$\mathcal{K}^*_{\mathsf{aug},1}(\mathbf{x}) := \{j \in \mathcal{K}^*_{\mathsf{aug}} : \delta_1(\mathbf{x}) \leq \sqrt{\frac{20m}{N}}\}$$
$$\mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x}) := \{j \in \mathcal{K}^*_{\mathsf{aug}} : \delta_1(\mathbf{x}) > \sqrt{\frac{20m}{N}}\}.$$

Let's consider these subsets one-by-one. First, we use Lemma 8 and the definition of the subset $\mathcal{K}^*_{\mathsf{aug},1}(\mathbf{x})$ to get

$$\begin{aligned}
\sum_{j \in \mathcal{K}^*_{\mathsf{aug},1}(\mathbf{x})} T_j(\mathbf{x}) &= \sum_{j \in \mathcal{K}^*_{\mathsf{aug},1}(\mathbf{x})} p_j(x) \langle \mathbf{a}_j, \mathbf{x} \rangle \\
&\leq \sum_{j \in \mathcal{K}^*_{\mathsf{aug},1}(\mathbf{x})} p_j(x) \left[f^*_\infty + f_{max} \delta_1(\mathbf{x}; j) \right] \\
&\leq \sum_{j \in \mathcal{K}^*_{\mathsf{aug},1}(\mathbf{x})} p_j(x) \left[f^*_\infty + f_{max} \sqrt{\frac{20m}{N}} \right] \\
&\leq \left[\sum_{j \in \mathcal{K}^*_{\mathsf{aug},1}(\mathbf{x})} p_j(x) \right] f^*_\infty + f_{max} \sqrt{\frac{20m}{N}}. \qquad (25)
\end{aligned}$$

Next, we consider the term $\sum_{j \in \mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x})} T_j(\mathbf{x})$. We state and prove the following lemma.

Lemma 9. *For any commitment $\mathbf{x} \in \Delta_m$, we have*

$$\sum_{j \in \mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x})} T_j(\mathbf{x}) \leq \left[\sum_{j \in \mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x})} p_j(\mathbf{x}) \right] f^*_\infty + 3|\mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x})| f_{max} \sqrt{\frac{20m}{N}}. \tag{26}$$

Proof. Consider any $j \in \mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x})$. Now note that by the definition of $\delta_1(\mathbf{x}; j)$, we can denote the open ℓ_1 ball with center $\mathbf{x}$ and radius $\delta_1(\mathbf{x})$ by $B_1(\mathbf{x}; \delta_1(\mathbf{x}))$. By the definition of $\delta_1(\mathbf{x}; j)$, it follows that $B_1(\mathbf{x}; \delta_1(\mathbf{x}; j)) \cap \mathcal{R}_j = \emptyset$. Therefore, we have

$$\begin{aligned} p_j(\mathbf{x}) &= \Pr\left[\widehat{\mathbf{X}}_N \in \mathcal{R}_j\right] \\ &\leq \Pr\left[\widehat{\mathbf{X}}_N \notin B_1(\mathbf{x}; \delta_1(\mathbf{x}; j))\right] \\ &= \Pr\left[\|\widehat{\mathbf{X}}_N - \mathbf{x}\|_1 \geq \delta_1(\mathbf{x}; j)\right] \\ &\leq 3\exp\{-\frac{N\delta_1(\mathbf{x})^2}{25}\} \end{aligned}$$

where we used Lemma 5 in the last inequality since we have $\mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x})$, we have $\delta_1(\mathbf{x}) \geq \sqrt{\frac{20m}{N}}$. Combining this with Lemma 8, we then have

$$T_j(\mathbf{x}) \leq p_j(\mathbf{x}) f^*_\infty + f_{max} 3\delta_1(\mathbf{x}; j) \exp\{-\frac{N\delta_1(\mathbf{x}; j)^2}{25}\}.$$

Next, it is easy to verify that the function $g_2(\delta) = \delta \exp\{-\frac{N\delta^2}{25}\}$ is decreasing in δ over the domain $\delta \geq \sqrt{\frac{20m}{N}}$ for all $m \geq 1$. This tells us that

$$\begin{aligned} 3\delta_1(\mathbf{x}; j) \exp\{-\frac{N\delta_1(\mathbf{x}; j)^2}{25}\} &\leq 3\sqrt{\frac{20m}{N}} \exp\{-\frac{4m}{5}\} \\ &\leq 3\sqrt{\frac{20m}{N}} \end{aligned}$$

and so we have

$$T_j(\mathbf{x}) \leq p_j(\mathbf{x}) f^*_\infty + 3 f_{max} \sqrt{\frac{20m}{N}}. \tag{27}$$

Summing over all $j \in \mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x})$ and substituting Equation (27) then proves the lemma. □

A.3.3 Putting it all together

Combining Equations (16), (21), (25) and (26) into Equation (14), we have

$$\begin{aligned} f_N(\mathbf{x}) &= \sum_{j=1}^{n} T_j(\mathbf{x}) \\ &\leq p_{j^*}(\mathbf{x}) f^*_\infty + C(N+1)^m \exp\{-NC\} f_{max} + \left[\sum_{j \in \mathcal{K}^*_{aug}} p_j(\mathbf{x}) \right] f^*_\infty + (3|\mathcal{K}^*_{\mathsf{aug},2}(\mathbf{x})| + 1) f_{max} \sqrt{\frac{20m}{N}} \\ &\leq f^*_\infty + C(N+1)^m \exp\{-NC\} f_{max} + 4n f_{max} \sqrt{\frac{20m}{N}} \\ &\leq f^*_\infty + Cn f_{max} \sqrt{\frac{20m}{N}} \end{aligned}$$

for some constant $C > 0$. This inequality holds for any $\mathbf{x} \in \Delta_m$. This implies that $f^*_N \leq f^*_\infty + Cn\sqrt{\frac{m}{N}}$, thus completing the proof of Theorem 3. □