How and Why to Manipulate Your Own Agent: On the Incentives of Users of Learning Agents
— Appendix —

Yoav Kolumbus
The Hebrew University of Jerusalem
`yoav.kolumbus@mail.huji.ac.il`

Noam Nisan
The Hebrew University of Jerusalem
`noam@cs.huji.ac.il`

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

A Convergence of Regret-Minimizing Agents

We consider repeated-game settings, in which the same finite group of automated agents repeatedly play a fixed game with bounded utilities on behalf of their users, with one agent per user. We focus on learning agents that are implemented as regret-minimization algorithms. The (external) regret of player i at time T, given a history of play $(\mathbf{a}^1, ..., \mathbf{a}^T)$, is defined as the difference between the optimal utility from using a fixed action in hindsight and the actual utility: $R_i^T = \max_a \sum_{t=1}^{T} u_i(a, \mathbf{a}_{-i}^t) - u_i(a_i^t, \mathbf{a}_{-i}^t)$, where a_i^t is the action of player i at time t and $\mathbf{a}_{-i}^t$ denotes the action profile of the other players at time t. As usual, regret-minimization algorithms are stochastic, and whenever we talk about the limit behavior, we consider a sequence of algorithms with $T \to \infty$ and with probability approaching 1. An agent i is said to be "regret minimizing" if $R_i^T/T \to 0$ almost surely as $T \to \infty$. A joint distribution over the players' actions is said to be a coarse correlated equilibrium if under this distribution all players have on expectation at most zero regret.

We use the following notation to describe the empirical distributions of the agents' dynamics. We denote by Δ the space of probability distributions over action profiles, by $\mathbf{p}_t^T \in \Delta$ the empirical distribution of actions after t rounds in a sequence of $T \geq t$ repetitions of the game, and by $p_t^T(\mathbf{a})$ the empirical frequency of an action profile $\mathbf{a}$ at the end of round t of the repeated game.

It is well known that the regret-minimization property ensures that the inequalities that define the CCE condition are all satisfied for the empirical time-average utilities over T steps to within a diminishing error term. This implies that, in Δ (the space of probability distributions over the agents' joint actions), the empirical joint-action distribution, $\mathbf{p}_t^T$, must get arbitrarily close to the *polytope of CCE distributions*.[1] One may hope that the empirical distribution converges to some specific CCE, p^*, and even be encouraged by the many known examples (starting with the matching pennies game) in which, even though the mixed strategies of the players do not converge, the time average of the empirical-play dynamics does converge to a specific CCE [2, 3, 4].

However, not only is this hope not always justified [18], but our following observation shows that whenever there is more than a single CCE no convergence is guaranteed. Not only may regret-minimization dynamics not converge at all, but even the time averages of the action distribution and utilities may keep changing over time.

Proposition 1. *For every finite game in which the set of CCEs is not a singleton and for every pair of distinct CCE distributions in that set, there exist regret-minimizing algorithms for the players whose empirical time-average joint dynamics do not converge at all and oscillate between getting arbitrarily close to each of these two CCEs.*

Thus, when allowing for *general* regret-minimizing agents, the only types of games in which the question of convergence is resolved are games in which the CCE is unique; in these games, regret-minimization dynamics must converge to the single CCE. For games with multiple CCEs, it is required to have additional information on the types of regret-minimization algorithms that are applied, that will allow to analyze the convergence properties of their dynamics, as we do in the companion paper [11] for a large class of natural regret-minimization algorithms in first-price and second-price auctions.

Before proceeding to the proof of Proposition 1, we need to formally define what notions of convergence we are looking at. We propose the following definitions of convergence as general and concrete notions that are compatible with the standard models of regret-minimization algorithms that focus on algorithms ALG^T, each of which is targeted to a fixed time horizon T, and looks at a sequence of such algorithms as[2] $T \to \infty$. All the following notions of convergence concern the average-iterate (i.e., the empirical distributions of the agents' dynamics). Note that average-iterate convergence does not imply last-iterate convergence (whereas the converse is true), and so under all the following definitions, a CCE may include dynamical patterns such as cycles or recurrent sets [13, 17].

The first notion of convergence that we consider is the one closest to the definition of the regret-minimization property. As mentioned above, the dynamics of regret-minimizing agents approach the

---

[1] This follows directly from the compactness of the space of distributions, assuming that utilities are bounded.

[2] An alternative formalism would consider a single algorithm with an infinite horizon, as in [2, 3], and look at its intermediate results at different times T. We prefer following the fixed-horizon formalism as it is the most commonly used one and since in the infinite-horizon definitions one must have an appropriately decreasing "update parameter."

set of coarse correlated equilibria in the space of utilities. The following definition deals with the space of distributions over joint action profiles.

Definition 1. *The dynamics* approach a set $S \subseteq \Delta$ *of distributions if for every* $\epsilon > 0$ *there exists* $T_0(\epsilon)$ *such that for every* $T > T_0$ *with probability at least* $1-\epsilon$ *it holds that* $\inf_{\boldsymbol{p} \in S} |\boldsymbol{p}_T^T - \boldsymbol{p}| < \epsilon$.

The next definition describes a different property of the dynamics. Basically, this property means that after a sufficiently long time, the distribution of the empirical play stabilizes and does not change much. Notice that this property allows for having different outcomes for different values of T, or even for different instances of the dynamics with the same algorithms and the same T.

Definition 2. *The dynamics are* self-convergent *if for every* $\epsilon > 0$ *there exists* $T_0(\epsilon)$ *such that for every* $T > T_0$ *with probability at least* $1-\epsilon$ *it holds that for every* $\epsilon T < t \leq T$, $|\boldsymbol{p}_t^T - \boldsymbol{p}_T^T| < \epsilon$.

Finally, the following definition describes convergence to a single distribution. This definition will be useful in our analysis of meta-games of games with a single CCE, in which the dynamics must converge to that CCE for any set of regret-minimization algorithms.[3]

Definition 3. *The dynamics* converge to a distribution $\boldsymbol{p} \in \Delta$ *if for every* $\epsilon > 0$ *there exists* $T_0(\epsilon)$ *such that for every* $T > T_0$ *with probability at least* $1-\epsilon$ *it holds that for every* $\epsilon T < t \leq T$, $|\boldsymbol{p}_t^T - \boldsymbol{p}| < \epsilon$.

It is not difficult to see that the above definition of convergence to a distribution is equivalent to the combination of the first two definitions shown above, as follows.

Observation 1. *The dynamics converge to the distribution* $\boldsymbol{p}$ *if and only if the dynamics are self-convergent and approach the unit set* $\{\boldsymbol{p}\} \subset \Delta$.

Proposition 1, which considers Definition 3 of convergence, shows that in any game with more than a single CCE there exist regret-minimizing algorithms for the players whose empirical time-average joint dynamics do not converge at all. The proof takes as a starting point the fact that for any CCE in any game there exist regret-minimization dynamics that converge to it. To establish convergence to a specific CCE, as in [14], we look at dynamics in which all agents play according to a schedule that yields that CCE as its time average, and in any case of deviation by any subset of the other players, the algorithms divert to playing a standard regret-minimization algorithm in the remaining time. The idea of the proof of Proposition 1 is that instead of using an action schedule that converges to a single CCE, we construct dynamics that alternate between two such schedules, and show that if this alternation slows down at a sufficient rate, the time average oscillates between arbitrarily approaching each of the two pre-determined CCEs, while the regret-minimization property of each agent is preserved.

Proof. (Proposition 1): We start with the following claim: *for every finite game and every CCE distribution of that game there exist regret-minimizing algorithms for the players whose joint dynamics converge to the given CCE.* This result was previously shown in [14] in an infinite-horizon setting. We technically re-prove this claim here to make the proof compatible with the finite-horizon setting and, specifically, with Definition 3 of convergence given above.

Let $\mathbf{p}$ be a CCE distribution of a finite n-player game, and let $M = |A_1 \times ... \times A_n|$ denote the size of the joint action space of the players, where A_i is the action space of agent i. We assume for simplicity that every probability in $\mathbf{p}$ is a rational number. Let $\mathbf{a}^k$, $k \in [M]$, be the action profile that has the k'-th highest probability in $\mathbf{p}$ (ties are broken in favor of the action tuple with the higher index in $\mathbf{p}$) and denote by a_i^k the action of player i in the action tuple $\mathbf{a}^k$. Let $T_k = 1/\Pr(\mathbf{a}^k)$ if $\Pr(\mathbf{a}^k) > 0$ or $T_k = 0$ otherwise and let $T_0 = 0$. Define the mapping $k(t) = k$ s.t. $\sum_{s=0}^{k-1} T_s < mod(t, \sum_{s=1}^{M} T_s) \leq \sum_{s=0}^{k} T_s$.

Consider the following "$\mathbf{p}$-schedule" algorithm for agent i: at every time $t = 1, ..., T$, the algorithm plays action $a_i^{k(t)}$. After every action, the algorithm observes the actions of the other agents. If at least one of the other agents deviated from its schedule, i.e., if there exists a player j that played an action different from $a_j^{k(t)}$, then the algorithm stops playing according to $a_i^{k(t)}$ and switches to playing the "unconditional regret-matching" algorithm [8, 9] in the remaining time.

---

[3] This definition would correspond to convergence to $\mathbf{p}$ almost surely in the infinite-horizon model.

This algorithm is clearly regret-minimizing: on the one hand, if any agent deviates, then all agents play the unconditional regret-matching algorithm, which is regret-minimizing.[4] On the other hand, if all agents are playing according to this algorithm then in every period of length $\tau = \sum_{s=1}^{M} T_s$ the empirical action distribution exactly equals the CCE distribution $\mathbf{p}$. Since all the utilities are assumed to be bounded, this implies that the regret is vanishing in the limit $T \to \infty$ (since the regret accumulated in any "partial cycle" of the schedule, of length less than τ, is bounded by a constant and thus vanishes in the time average).

It also follows from a similar argument that in the dynamics in which all agents play according to the algorithm described above, the empirical distribution of play converges to $\mathbf{p}$ (we will call such dynamics in which all agents play the $\mathbf{p}$-schedule algorithm "$\mathbf{p}$-schedule dynamics"). Formally, at the end of every period τ the empirical action distribution equals $\mathbf{p}$, and thus the empirical distribution at time $t > \tau$ can be written as $\mathbf{p}_t^T = \frac{1}{t}\left(\tau \cdot \lfloor t/\tau \rfloor \cdot \mathbf{p} + mod(t, \tau) \cdot \mathbf{x}\right)$, where $\mathbf{x}$ can be any arbitrary distribution obtained in a partial cycle of the schedule, or it can be equal to $\mathbf{p}$ if t completes a full cycle. To show that Definition 3 holds, let $\epsilon > 0$ such that $T_0(\epsilon) > 2\tau/\epsilon$. To bound the distance of the empirical action distribution from $\mathbf{p}$, consider the vector $\mathbf{x} = -\mathbf{p}$ with weight τ. Then, for every $T > T_0$ with probability 1 (as there in no stochasticity in these dynamics) for every t s.t. $\epsilon T < t \leq T$ it holds that $|\mathbf{p}_t^T - \mathbf{p}| \leq |\frac{1}{t}\big((t-\tau)\mathbf{p} - \tau\mathbf{p}\big) - \mathbf{p}| \leq \frac{2\tau}{t} < \epsilon$. Thus, we have that every CCE has regret-minimization dynamics that converge to it.

Next, consider any finite game that has more than a single CCE and let $\mathbf{p}_1$ and $\mathbf{p}_2$ be two distinct distributions in the set of its CCE distributions. Notice that since the set of CCEs in every game is a convex set, also every weighted average of $\mathbf{p}_1$ and $\mathbf{p}_2$ is a CCE itself. Consider the dynamics of agents that all switch between $\mathbf{p}_1$-schedule dynamics with cycle time τ_1 and $\mathbf{p}_2$-schedule dynamics with cycle time τ_2. Let $\epsilon > 0$ such that $\epsilon < 1/\max(\tau_1, \tau_2)$ and let $\alpha = \lceil 1/\epsilon^2 \rceil$. We will define the series $\Gamma_c = (2\alpha)^c$, $c = 1, 2, ...$ to be the number of full cycles in which each dynamic is played before switching to the other dynamic. That is, the lengths of the periods in which each dynamic is played are $\tau_1\Gamma_c$ if c is odd or $\tau_2\Gamma_c$ if c is even, i.e., $(2\alpha\tau_1, 4\alpha^2\tau_2, 8\alpha^3\tau_1, 16\alpha^4\tau_2, ...)$. By the above proof for the convergence of $\mathbf{p}$-schedule dynamics we have that by the end of the first period of length $2\alpha\tau_1$ the empirical distribution of play reaches a distance of less than ϵ from the CCE distribution $\mathbf{p}_1$. Next, by the end the second period, i.e., at time $t = 2\alpha\tau_1 + 4\alpha^2\tau_2$, the distance of the empirical distribution from the CCE distribution $\mathbf{p}_2$ is $|\mathbf{p}_t^T - \mathbf{p}_2| \leq \left|\frac{1}{4\alpha^2\tau_2 + 2\alpha\tau_1}\Big((4\alpha^2\tau_2)\mathbf{p}_2 + (2\alpha\tau_1)\mathbf{p}_1\Big) - \mathbf{p}_2\right| \leq \left|\frac{1}{4\alpha^2\tau_2 + 2\alpha\tau_1}\Big((4\alpha^2\tau_2)\mathbf{p}_2 - (2\alpha\tau_1)\mathbf{p}_2\Big) - \mathbf{p}_2\right| = \left|\frac{4\alpha^2\tau_2 - 2\alpha\tau_1}{4\alpha^2\tau_2 + 2\alpha\tau_1} - 1\right| = \frac{4\alpha\tau_1}{4\alpha^2\tau_2 + 2\alpha\tau_1} < \epsilon$.

The same argument holds at the end of the subsequent periods as well. Thus, given a sufficiently long time T, the average empirical distribution $\mathbf{p}_t^T$ oscillates between getting arbitrarily close to each of the two CCE distributions $\mathbf{p}_1$ and $\mathbf{p}_2$, with an oscillation period that slows down exponentially, and hence there is no convergence of the average empirical play. ☐

B Dominance-Solvable Games

The following lemma is a formalization of a well-known result that dominance-solvable games have a single CCE. The direct implication is that these games are stable in the dynamical sense, as the empirical time average of regret-minimization dynamics must converge to the unique Nash equilibrium outcome. For completeness, we provide here the formal statement and a simple proof. Some relevant references in this context include [5, 15, 16].

Lemma 1. *In any dominance-solvable game the only CCE is the unique pure Nash equilibrium.*

Proof. Consider any dominance-solvable game. Denote the Nash equilibrium joint action distribution by σ_{NE}. In this distribution there is probability one for the pure Nash equilibrium joint action profile and probability zero for all other action profiles. First, note that dominance-solvable games have a unique Nash equilibrium, and that every Nash equilibrium is also a CCE, and so σ_{NE} is a CCE. Next, assume by way of contradiction that there is another distribution $\sigma \neq \sigma_{NE}$ that is also a CCE. Fix any order $\mathbf{s}$ of iterated elimination of strictly dominated strategies in the game, such that s_i is the i'th eliminated action. Since σ is different from the pure Nash equilibrium distribution, there

---

[4] This algorithm is especially simple here since it is does not require any parameter to be fitted to the remaining time after the step at which a deviation occurred. Other algorithms can also be used with proper adjustments.

exists an action that has the smallest index i such that s_i is played with positive probability according to σ, but played with zero probability according to σ_{NE}. If $i = 1$ then s_i is a strictly dominated strategy. Therefore, the player that plays action s_1 with a finite frequency (with high probability) must accumulate linear regret (since action $i = 1$ is specifically dominated also by the best fixed strategy in hindsight which yields, by definition, zero regret). Since the CCE condition is equivalent to the requirement that all players have, with high probability, a sub-linear regret over time, σ cannot be a CCE, a contradiction. Hence $i > 1$. However, if $i = 2$, since action 1 is played with zero probability, a player that plays s_2 with a finite frequency must also accumulate linear regret, which again violated the CCE condition, and thus $i > 2$. The same argument holds until reaching the actions in $\mathbf{s}$ which comprise together the pure Nash equilibrium profile (a single action for each player). Thus, we reach the contradiction $\sigma = \sigma_{NE}$ and so the unique pure Nash equilibrium of a dominance-solvable game is also its unique CCE. $\square$

Proof. (Theorem 1): Consider any dominance-solvable $n \times m$ game in which there is a player who's Stackelberg value (i.e., his utility in a pure-strategy Stackelberg equilibrium of the game where he plays the first action) is higher than his utility in the unique Nash equilibrium of the game. Assume for convenience and without loss of generality that this is player 1. For a sufficiently unrestricted parameter space, this player can declare his Stackelberg strategy as a dominant strategy to his agent. In this case, regret-minimization dynamics will quickly reach distributions of play where only this dominating strategy is played, regardless of the actions of the other agent. Specifically, if the opponent provides a truthful declaration to his agent, the declared game has a unique Nash equilibrium which is exactly the Stackelberg outcome. Thus a unilateral manipulation of player 1 can give him his Stackelberg value and so improve his utility. $\square$

C Cournot Competition Games

As described in the main text, we consider linear Cournot competition games [6, 7, 12], where player 1 produces quantity $q_1 \in \mathcal{R}^+$ with a per-unit production cost of c_1 (such that his total production cost is $c_1 \cdot q_1$) and player 2 produces quantity $q_2 \in \mathcal{R}^+$ with a per-unit production cost of c_2. The utilities of the players are $u_1 = q_1(a - b(q_1 + q_2) - c_1)$, and $u_2 = q_2(a - b(q_1 + q_2) - c_2)$, where a and b are commonly known positive constants. The Nash equilibrium of the game depends on the parameters as follows. If $a + c_2 - 2c_1 > 0$ and $a + c_1 - 2c_2 > 0$, then the Nash equilibrium is $q_1^{true} = \frac{1}{3b}(a + c_2 - 2c_1)$ and $q_2^{true} = \frac{1}{3b}(a + c_1 - 2c_2)$. If $a + c_1 - 2c_2 > 0$ and $c_1 < a$, the equilibrium is $q_1^{true} = \frac{a-c_1}{2b}$ and $q_2^{true} = 0$, and symmetrically, if $a + c_2 - 2c_1 > 0$ and $c_2 < a$, the equilibrium is $q_1 = 0$ and $q_2^{true} = \frac{a-c_2}{2b}$. Otherwise, in the Nash equilibrium both players produce zero.

Thus, there are four parameter regions of interest according to the four possible types of unique Nash equilibria of the game, as illustrated in Figure 2. The parameter region $A = \{c_1, c_2 | a + c_2 - 2c_1 > 0,\ a + c_1 - 2c_2 > 0,\ c_1 > 0,\ c_2 > 0\}$ shown in the figure is the region where both agents produce positive quantities (the shaded areas in region A in the figure relate to equilibria of the meta-game; see below). The parameter regions $B = \{c_1, c_2 | a + c_1 - 2c_2 > 0,\ 0 < c_1 < a,\ c_2 > 0\}$ and $C = \{c_1, c_2 | a + c_2 - 2c_1 > 0,\ 0 < c_2 < a,\ c_1 > 0\}$ are regions where only one player produces a positive quantity. In the remaining region, region $D = \{c_1, c_2 | c_1, c_2 \geq a\}$, both agents produce zero.

Consider the meta-game defined for this game where the true types of the players are their per-unit production costs c_1, c_2. Player 1 declares to his agent a value[5] $x_1 \in \mathcal{R}^+$ and player 2 declares to his agent a value $x_2 \in \mathcal{R}^+$. The agents then interact repeatedly. Since this game is known to be "socially concave" [7], the time-average dynamics of regret-minimizing agents must converge to the Nash equilibrium of the game that is defined by the parameters x_1, x_2 provided by the users.

Given the declarations x_1, x_2, we can identify the relevant parameter region in which the declared game lies, substitute the appropriate equilibrium production levels of the agents in the utility functions for the players, and obtain the utilities of the players. The utilities in the four regions as functions of the declarations x_1, x_2, are $u_1 = \frac{1}{9b}(a + x_2 - 2x_1)(a + x_1 + x_2 - 3c_1)$, $u_2 = \frac{1}{9b}(a + x_1 - 2x_2)(a + x_1 + x_2 - 3c_2)$ in region A, $u_1 = \frac{1}{4b}(a - x_1)(a + x_1 - 2c_1)$, $u_2 = 0$ in region B, and $u_1 = 0$, $u_2 = \frac{1}{4b}(a - x_2)(a + x_2 - 2c_2)$ in region C. In region D there is no production and thus zero utility.

---

[5] Notice that any declaration $x > a$ leads to zero production, and so it is equivalent to declaring $x = a$. Thus, $x_i \leq a$ suffices for a full description of the meta-game, where region D is described by the point $x_1 = x_2 = a$.

The following lemma specifies the equilibria of the meta-game for those cases where in these equilibria both agents produce positive quantities (i.e., where the equilibrium declarations are in region A).

Lemma 2. *Let* $x_1^* = \frac{1}{5}(8c_1 - 2c_2 - a)$ *and* $x_2^* = \frac{1}{5}(8c_2 - 2c_1 - a)$. *If in the equilibrium of the meta-game both players produce positive quantities, then the equilibrium declarations are* x_1^*, x_2^* *if* $x_1^*, x_2^* \geq 0$*;* $(x_1 = \frac{1}{4}(6c_1 - a)^+,\ x_2 = 0)$ *if* $x_2^* < 0, x_1^* \geq 0$*;* $(x_1 = 0,\ x_2 = \frac{1}{4}(6c_2 - a)^+)$ *if* $x_1^* < 0, x_2^* \geq 0$*; and* $(x_1 = 0, x_2 = 0)$ *otherwise.*

Proof. If in the equilibrium of the meta-game both players produce positive quantities, then the equilibrium can be found using the derivatives of the utilities as follows.

$$\frac{\partial u_1}{\partial x_1} = \frac{1}{9b}(6c_1 - a - 4x_1 - x_2) = 0, \qquad \frac{\partial u_2}{\partial x_2} = \frac{1}{9b}(6c_2 - a - x_1 - 4x_2) = 0.$$

$x_1^* = \frac{1}{5}(8c_1 - 2c_2 - a)$ and $x_2^* = \frac{1}{5}(8c_2 - 2c_1 - a)$ are the unique solution to these equations. If x_1, x_2 are positive then they are the Nash equilibrium of the meta-game. If $x_1 = \frac{1}{5}(8c_1 - 2c_2 - a) < 0$ then the utility of player 1 is decreasing in his declaration, and thus declaring $x_1 = 0$ is his best-reply to any declaration of the other player. The best-reply of player 2 is then obtained from the above derivatives by substituting $x_1 = 0$ in the second equation, yielding $x_2 = \frac{1}{4}(6c_2 - a)$. If this expression is non-negative then it is the best-reply of player 2. If this expression is negative then the utility of player 2 is decreasing in x_2 and his best-reply is $x_2 = 0$. The same argument holds for player 2; if $x_2 = \frac{1}{5}(8c_2 - 2c_1 - a) < 0$ then declaring $x_2 = 0$ is the best-reply of player 2 to any declaration of player 1 and the best-reply of player 1 is then $x_1 = \frac{1}{4}(6c_1 - a)^+$. □

Theorem 2 characterizes the types of equilibria of the meta-game when in the equilibrium of the game with the true parameters both players produce positive quantities. To prove the theorem, we consider its following technical restatement.

Theorem. *(Restatement of Theorem 2): Consider a two-player linear Cournot competition with positive linear costs where both players produce positive quantities in the Nash equilibrium.*

1. *If the production costs* c_1, c_2 *of the two players are sufficiently low such that* $c_1, c_2 < a/2$ *or sufficiently close such that* $\frac{1}{2}(3c_2 - a) < c_1 < \frac{1}{3}(2c_2 + a)$*, then in the equilibrium of the meta-game both players declare to their agents values that are lower than their true production costs, produce larger quantities than those they produce in the Nash equilibrium of the game with the truthful reports (and thus the price is lower), and have lower utilities.*

2. *If the production cost of one of the players is at least* $a/2$ *and the cost of the other player is sufficiently low, namely,* $c_2 \geq a/2$ *and* $c_2 \geq \frac{1}{3}(2c_1 + a)$*, or* $c_1 \geq a/2$ *and* $c_1 \geq \frac{1}{3}(2c_2 + a)$*, then in the equilibrium of the meta-game the player with the low production cost declares a value that is lower than his true cost and produces alone. The quantity produced by this player alone is larger than the total quantity produced by both players in the Nash equilibrium of the game with the truthful reports (and thus the price is lower), and the utility for this player is higher than his utility in the Nash equilibrium of the game with the truthful reports.*

Proof. We start with the following lemma concerning the best-replies of the players.

Lemma 3. *If* $0 < c_1 < a/2$*, the best-reply of player* 1 *to any declaration* $x_2 < a$ *of player* 2 *is strictly less than player* 1*'s true cost* c_1*.*

Proof. Assume that player 2 declares a cost $0 \leq x_2 < a$. The utility of player 1 is $u_1 = \frac{1}{9b}(a + x_2 - 2x_1)(a + x_1 + x_2 - 3c_1)$. The best-reply, $\operatorname{argmax}_{x_1}(u_1)$, is obtained from $\frac{\partial u_1}{\partial x_1} = \frac{1}{9b}(6c_1 - a - 4x_1 - x2) = 0$, resulting in $x_1 = \frac{1}{4}(6c_1 - a - x_2)$. If this expression is non-negative, then this is the best-reply; if it is negative, then the utility of player 1 is decreasing in x_1 and the best-reply is $x_1 = 0$. If the best-reply is zero then it is less than c_1 as required. If the best-reply is positive then, since $c_1 < a/2$, it holds that $x_1 = \frac{1}{4}(6c_1 - a - x_2) \leq \frac{1}{4}(6c_1 - a) < \frac{1}{4}(6c_1 - 2c_1) = c_1$. □

It thus follows that if $c_1, c_2 < a/2$, specifically, also in the Nash equilibrium declaration profile, which is a mutual best-reply profile, both declarations are less than the true costs.

Next, consider the case where the production costs of the two players, c_1, c_2, are at least $a/2$ and it holds that $\frac{1}{2}(3c_2 - a) < c_1 < \frac{1}{3}(2c_2 + a)$. The equilibrium of the meta-game is then $x_1 = \frac{1}{5}(8c_1 - 2c_2 - a)$ and $x_2 = \frac{1}{5}(8c_2 - 2c_1 - a)$ (by Lemma 2). Since $c_2 > \frac{1}{2}(3c_1 - a)$, it holds that $x_1 = \frac{1}{5}(8c_1 - 2c_2 - a) < \frac{1}{5}(8c_1 - 3c_1 + a - a) = c_1$, and similarly, since $c_1 > \frac{1}{2}(3c_2 - a)$, it holds that $x_2 = \frac{1}{5}(8c_2 - 2c_1 - a) < \frac{1}{5}(8c_2 - 3c_2 + a - a) = c_2$. Thus, the decelerations of both players are lower than their true costs, and each player produces a larger quantity than in the truthful equilibrium. The utilities obtained are strictly lower than those obtained in the truthful equilibrium, as can be verified by substituting the equilibrium declarations x_1, x_2 into the utility functions.

Next assume w.l.o.g. that player 2 has a cost $c_2 \geq a/2$ and $c_2 \geq \frac{1}{3}(2c_1 + a)$. Assume by way of contradiction that in the equilibrium of the meta-game both players produce positive quantities. On the one hand, the sufficient and necessary condition for positive production by both agents under a declaration profile (x_1, x_2) is $a + x_1 - 2x_2 > 0$ and $a + x_2 - 2x_1 > 0$. On the other hand, consider the equilibrium given by Lemma 2. If the equilibrium is $x_1 = \frac{1}{5}(8c_1 - 2c_2 - a)$, $x_2 = \frac{1}{5}(8c_2 - 2c_1 - a)$, by substituting these declarations in the positive-production condition we obtain $a + 2c_1 - 3c_2 > 0$ and $a + 2c_2 - 3c_1 > 0$, which is in contradiction to $c_2 \geq a/2$ and $c_2 \geq \frac{1}{3}(2c_1 + a)$. The other alternative is that the equilibrium is $x_1 = 0, x_2 = \frac{1}{4}(6c_2 - a)$. Substituting these declarations into the positive-production condition we obtain $c_2 < a/2$, which is a contradiction.

Therefore, it cannot be that both agents produce. Notice that player 1 who has a low cost will prefer to produce even if player 2 declares zero. Therefore, in the equilibrium of the meta-game, player 1 produces alone, and declares the value that maximizes his utility as a monopolist, subject to the constraint that player 2 still prefers not to produce, which is $x_1 = 2c_2 - a < c_1$ (since in the truthful Nash equilibrium both players produce, which implies $a + c_1 - 2c_2 > 0$). The best-reply declaration of player 2 is then any declaration $x_2 \geq c_2$.

The quantity that is produced by player 1 in the equilibrium of the meta-game where he produces alone is $\frac{a-c_2}{b}$, which is more than the total quantity of $\frac{1}{3b}(a - c_1 - c_2)$ that is produced in the Nash equilibrium of the game with the true parameters (since $\frac{1}{3b}(a - c_1 - c_2) < \frac{1}{3b}(a - (2c_2 - a) - c_2) = \frac{1}{3b}(2a - 3c_2) < \frac{a-c_2}{b}$), thus yielding a lower price. The utility of player 1 in this meta-game equilibrium is $u_1 = \frac{1}{b}(a - c - 2)(c_2 - c_1)$. The utility of player 1 in the Nash equilibrium of the game with the true parameters is $u_1^{true} = \frac{1}{9b}(a + c_2 - 2c_1)^2$. The utility difference $u_1^{true} - u_1 = \frac{1}{9b}\left(a^2 + 10c_2^2 + 4c_1^2 + 5ac_1 - 7ac_2 - 13c_1c_2\right)$ is negative for $c_2 \geq a/2$ and $c_2 \geq \frac{1}{3}(2c_1 + a)$. That is, the utility of player 1 who has a low production cost is higher in the equilibrium of the meta-game in which he "drives player 2 out of the market" than in the equilibrium of the game with the true parameters. □

Theorem 3 describes the limited set of cases where the game is manipulation-free. Basically, to establish conditions on the game parameters such that the game is manipulation-free, we require that in the equilibrium of the meta-game $x_i = c_i$. The proof is then based on the characterization of the equilibria as specified in Theorem 2 and Lemma 2.

Proof. (Theorem 3): The game is manipulation free if in the equilibrium of the meta-game $x_1 = c_1$ and $x_2 = c_2$. If in the equilibrium of the meta-game both players produce positive quantities, then by Lemma 2 it holds that the Nash equilibrium is $(x_1 = \frac{1}{5}(8c_1 - 2c_2 - a), x_2 = \frac{1}{5}(8c_2 - 2c_1 - a))$ if $x_1^*, x_2^* \geq 0$, $(x_1 = \frac{1}{4}(6c_1 - a)^+, x_2 = 0)$ if $x_2^* < 0, x_1^* \geq 0$, $(x_1 = 0, x_2 = \frac{1}{4}(6c_2 - a)^+)$ if $x_1^* < 0, x_2^* \geq 0$, and $(x_1 = 0, x_2 = 0)$ otherwise. In the first case, where $x_1 = \frac{1}{5}(8c_1 - 2c_2 - a), x_2 = \frac{1}{5}(8c_2 - 2c_1 - a)$, requiring $x_1 = c_1$ and $x_2 = c_2$ yields $x_1 = x_2 = a$, in which case the players do not produce in the truthful equilibrium. That is, if both players prefer not to produce at all under the true parameters, then they also do not have any profitable manipulation. In the second case, $x_1 = \frac{1}{4}(6c_1 - a)^+, x_2 = 0$, requiring $x_1 = c_1$ and $x_2 = c_2$ yields either $c_1 = a/2$ and $c_2 = 0$, in which case player 1 does not produce, or $x_1 = x_2 = 0$. Similarly, the symmetric case $x_1 = 0, x_2 = \frac{1}{4}(6c_2 - a)^+$ yields either $c_1 = 0$ and $c_2 = a/2$ or $x_1 = x_2 = 0$. Finally, in the case where $x_1^*, x_2^* < 0$, the equilibrium is $x_1 = x_2 = 0$ which is truthful if and only if $c_1 = c_2 = 0$.

The remaining cases are the case where under the true parameters both players produce, but in the equilibrium of the meta-game only one player produces, and the case where under the true parameters only one player produces. In the first case, as shown in Theorem 2, it must be that the player with the

	q	$1-q$
p	A_1, A_2	B_1, B_2
$1-p$	C_1, C_2	D_1, D_2

Figure 5: Notation for generic 2×2 games.

lower cost shades his declaration, and so this equilibrium is not truthful. In the other case, we assume w.l.o.g. that player 1 is a monopolist. Player 1 then maximizes his utility when declaring his true cost $x_1 = c_1$, producing a quantity q^*. Player 2 then prefers not to produce. It is not difficult to see that given that player 1 declares the truth, any declaration of player 2 that will lead him to produce a positive quantity will result in a total production grater than $q^* = \frac{a-c_1}{2b}$, and thus will give him a negative utility. Thus, player 2 prefers to declare the truth, $x_2 = c_2$, and the truthful declarations form a Nash equilibrium of the meta-game. $\square$

D Opposing-Interests Games

To set notations, Figure 5 denotes the utilities of the players in each game outcome where capital letters denote the (pure) game outcomes and the subscripts denote player indices. Thus, in opposing-interests 2×2 games $A_1, D_1 > B_1, C_1$ and $A_2, D_2 < B_2, C_2$, or vice versa, i.e., all inequalities are reversed. Additionally, the parameter p in the figure denotes the probability that the row player plays the top row in a mixed Nash equilibrium and the parameter q denotes the probability that the column player plays the left column.

We start with two technical observations that will be useful for our proofs. These observations are formalizations of standard calculations according to the notation presented in Figure 5 above. The first observation specifies the utilities of the players in any mixed strategy profile (p, q), and the second observation shows the expressions for (p, q) in a completely mixed Nash equilibrium, when such exists in the game. The proofs of these observations are straightforward.

Observation 2. *Consider a* 2×2 *game as presented in Figure 5. If the players play a mixed strategy profile* (p, q)*, i.e., both players mix their pure strategies such that the row player plays the top row w.p.* $p \in [0, 1]$ *and the column player plays the left column w.p.* $q \in [0, 1]$*, then the expected utilities* u_1 *of the row player and* u_2 *of the column player are given by* $u_1 = pq(A_1 - C_1 + D_1 - B_1) + p(B_1 - D_1) + q(C_1 - D_1) + D_1$ *and* $u_2 = pq(A_2 - B_2 + D_2 - C_2) + p(B_2 - D_2) + q(C_2 - D_2) + D_2$.

Observation 3. *The mixed Nash equilibrium profile* (p, q) *of a* 2×2 *game (as presented in Figure 5), if such exists, is given by* $p = \frac{D_2 - C_2}{A_2 - B_2 + D_2 - C_2}$, $q = \frac{D_1 - B_1}{A_1 - C_1 + D_1 - B_1}$.

Theorem 4 characterizes the equilibrium of the meta-game and the utilities obtained in it, when the players use any types of regret-minimizing agents and parameter spaces where any one of the four utilities of each player in the game is a parameter that the player can manipulate. The proof of the theorem first shows that both players will strictly prefer to use declarations that lead to a ("manipulated") game between the agents that is a fully mixed game, and then the proof characterizes the declaration profiles that form equilibria of the meta-game and shows that all such equilibria lead to the same utilities as those obtained in the truthful Nash equilibrium. The theorem technically holds not only for manipulations of any one of the parameters of each player but for a broader range of parameter spaces which we term "natural parameter spaces" that have the following properties.

Definition 4. *A parameter space* $P \subseteq \mathcal{R}^4$ *of the row player in a* 2×2 *game with (true) row-player parameters* (A_1, B_1, C_1, D_1) *is called* natural *if it has the following properties.*

1. *(Sufficient generality) The parameter space induces all possible values on the column's mixed strategy; i.e., for every* $0 < q < 1$*, there exists a declaration profile* $(A'_1, B'_1, C'_1, D'_1) \in P$ *such that* $q \cdot A'_1 + (1-q) \cdot B'_1 = q \cdot C'_1 + (1-q) \cdot D'_1$.

2. *(Analyzability) For every* $(A'_1, B'_1, C'_1, D'_1) \in P$ *either the best-replies to pure strategies do not change, i.e.,* $sign(A_1 - C_1) = sign(A'_1 - C'_1)$*, and* $sign(B_1 - D_1) = sign(B'_1 - D'_1)$*, or the player has a dominant strategy, i.e.,* $sign(A'_1 - C'_1) = sign(B'_1 - D'_1) \neq 0$.

where $sign(x) = 1$ *if* $x > 0$*,* $sign(x) = -1$ *if* $x < 0$*, and* $sign(x) = 0$ *if* $x = 0$.

The parameter space for the column player is called natural *in a similar manner, and the whole parameter space of the game is called* natural *if it is natural for both players.*

The following lemmas show examples of natural parameter spaces for opposing-interests games. The first example describes user manipulations that include all or any subset of the agent's utilities, as long as the best-reply structure of the game is preserved, and the second example includes arbitrary manipulations of any single parameter by each user.

Lemma 4. *For every opposing-interests game and every choice of a non-empty subset of the four parameters for each player, the parameter space where the non-chosen parameters are fixed to their true values and the chosen parameters are arbitrary as long as they conserve best-replies to pure strategies (i.e., for the row player* $sign(A_1 - C_1) = sign(A'_1 - C'_1)$ *and* $sign(B_1 - D_1) = sign(B'_1 - D'_1)$*, and for the column player* $sign(A_2 - B_2) = sign(A'_2 - B'_2)$ *and* $sign(C_2 - D_2) = sign(C'_2 - D'_2)$*) is a natural parameter space.*

Proof. Property (2) of a natural parameter space holds immediately in this case. To show property (1), let $q \in (0, 1)$, and we require that for some $(A'_1, B'_1, C'_1, D'_1) \in P_1$ it holds that $q \cdot A'_1 + (1-q) \cdot B'_1 = q \cdot C'_1 + (1 - q) \cdot D'_1$, and similarly for the column player let $p \in (0, 1)$, and we require that for some $(A'_2, B'_2, C'_2, D'_2) \in P_2$ it holds that $p \cdot A'_2 + (1 - p) \cdot C'_1 = p \cdot B'_1 + (1 - p) \cdot D'_2$, yielding

$$p = \frac{1}{1 + \frac{A'_2 - B'_2}{D'_2 - C'_2}}, \qquad q = \frac{1}{1 + \frac{A'_1 - C'_1}{D'_1 - B'_1}}.$$

Notice, that under the conditions of the lemma the values of these two expressions lie in the range $(0, 1)$ since both terms in the denominators are positive. We next solve for $\frac{A'_2 - B'_2}{D'_2 - C'_2}$ and for $\frac{A'_1 - C'_1}{D'_1 - B'_1}$:

$$\frac{A'_1 - C'_1}{D'_1 - B'_1} = \frac{1-q}{q}, \qquad \frac{A'_2 - B'_2}{D'_2 - C'_2} = \frac{1-p}{p}.$$

These equations can be easily satisfied by selecting declarations within the parameter space of each player, since only a single degree of freedom is required to do so. To demonstrate this for the row player (equation for q), assume w.l.o.g. that $A_1 > C_1$ and $D_1 > B_1$, denote $y = (1 - q)/q > 0$, and denote a single free parameter of this player by x, and assume that the other parameters are the true parameters of the game. We will look at the four cases where x replaces each one of the parameters of the true game. If x replaces A_1, then declaration $x = C_1 + (D_1 - B_1)y$, which is in P_1 (since $x > C_1$, and thus it preserves best-replies), satisfies the equation for q. If x replaces C_1, then declaration $x = A_1 - (D_1 - B_1)y$, which is in P_1 (since $x < A_1$), satisfies the equation for q. If x replaces D_1, then declaration $x = B_1 + (A_1 - C_1)/y$, which is in P_1 (since $x > B_1$), satisfies the equation for q. If x replaces B_1, then declaration $x = D_1 - (A_1 - C_1)/y$, which is in P_1 (since $x < D_1$), satisfies the equation for q. Similar considerations apply to the declarations of the column player and the equation of p. Additional free parameters also allow the row player to induce q as a mixed strategy (e.g., in a degenerate way of controlling only one parameter), and allow to the column player to induce p. ☐

Lemma 5. *For any opposing-interests game and choice of one of the parameters of each player, the parameter space where three parameters are fixed to the true values and the fourth parameter is arbitrary (except for exactly being equal to another parameter) is a natural parameter space.*

Proof. Consider any opposing-interests 2×2 game with utilities (A_1, B_1, C_1, D_1) to the row player and (A_2, B_2, C_2, D_2) to the column player (see Figure 5), and assume that each player can set the declaration of one of his parameters to any arbitrary value with generic parameters, i.e., without equalities in utilities between actions.

Property (1) of a natural parameter space can be obtained by the same argument given in the proof of Lemma 4 above. Regarding property (2), for the case that the player does not change his best-replies to pure strategies in his declaration, property (2) holds directly. If a player does declare a single parameter that changes his best-reply to a pure strategy, then since the game is an opposing-interests game and only a single parameter has changed, this implies that this player now has a dominant strategy, and so property (2) holds in this case as well. ☐

To prove Theorem 4 we consider the following restatement of the theorem.

Theorem. *(Restatement of Theorem 4): In any 2×2 game with opposing interests and a natural parameter space, the Nash equilibrium of the meta-game is essentially unique. That is, there is a unique strategy profile $(p,q)\in(0,1)^2$ such that every Nash equilibrium of the meta-game induces (p,q) as a unique Nash equilibrium of the agents' game. The utility of each player in that equilibrium is equal to the utility of the player when all players enter their true parameters into their agents.*

Proof. Consider any opposing-interests 2×2 game with a natural parameter space and with true utilities (A_1,B_1,C_1,D_1) of the row player and (A_2,B_2,C_2,D_2) of the column player (see Figure 5) and assume w.l.o.g. that $A_1 > B_1, C_1$ and $D_1 > B_1, C_1$ and $B_2 > A_2, D_2$ and $C_2 > A_2, D_2$, i.e., that the row player has higher utilities on the main diagonal of the game utility matrix and the column player has higher utilities off the diagonal.

With a natural parameter space the parameter declarations lead to one of the following cases: (a) no player changes the signs of his best-replies, or (b) one or both players declare parameters such that their agents have dominant strategies in the game with the declared parameters. We argue that (b) cannot be the case in a Nash equilibrium of the meta-game. If, for example, the row player declares a dominant strategy to his agent, say to play the top row, then the best-reply of the column player would be to declare the right column as a dominant strategy to his agent. However, this is not a Nash equilibrium of the meta-game since in this declaration profile the row player has utility B_1 and so he would prefer to change strategy to any mixed strategy $0\le p<1$ to obtain utility $pB_1+(1-p)D_1 > B_1$. A similar argument holds for any other dominant strategy declaration.

Thus, in a Nash equilibrium of the meta-game the players do not declare dominant strategies for their agents; i.e., the strategy of the row player in a Nash equilibrium of the meta-game is to declare parameters that induce $q\in(0,1)$, and the strategy of the column player in a Nash equilibrium of the meta-game is to declare parameters that induce $p\in(0,1)$. This also implies that in a Nash equilibrium of the meta-game no player reverses any of the directions of his best-replies in the game, and therefore the game with the declared parameters still has a single mixed Nash equilibrium to which the regret-minimizing agents converge, and so the utilities of the players in the meta-game can be analytically derived. The utilities of the two players when the agents converge to such a mixed strategy profile (p,q) are, by Observation 2,

$$u_1 = pq(A_1-C_1+D_1-B_1)+p(B_1-D_1)+q(C_1-D_1)+D_1$$
$$u_2 = pq(A_2-B_2+D_2-C_2)+p(B_2-D_2)+q(C_2-D_2)+D_2.$$

Since p is a function of only the column player's (player 2's) declared parameters and the true game parameters, and q is a function of only the row player's (player 1's) parameters and the true game parameters, and in a natural parameter space the row player can choose parameters to induce any $q\in(0,1)$, and the column player can similarly choose parameters to induce any $p\in(0,1)$, we can think of q of the Nash equilibrium of the agents' game as (essentially) the strategy of the row player, and similarly, of p as (essentially) the strategy of the column player. Thus, the condition for a Nash equilibrium in the meta-game is $\frac{\partial u_1}{\partial q}=\frac{\partial u_2}{\partial p}=0$.

Hence, the strategy profile (p,q) in any Nash equilibrium of the meta-game is

$$p=\frac{D_1-C_1}{A_1-B_1+D_1-C_1},\qquad q=\frac{D_2-B_2}{A_2-B_2+D_2-C_2}.$$

Next, we calculate the utilities in an equilibrium of the meta-game and compare them with the utilities of the Nash equilibrium of the game with the true parameters. By substituting the unique (p,q) Nash equilibrium profile of the meta-game into the equations of the utilities shown above, we get that the utilities in any Nash equilibrium of the meta-game are

$$u_1=\frac{(D_1-C_1)(B_1-D_1)}{A_1-C_1+D_1-B_1}+D_1,\qquad u_2=\frac{(D_2-C_2)(B_2-D_2)}{A_2-B_2+D_2-C_2}+D_2.$$

The Nash equilibrium strategy profile of the game with the true parameters is (by Observation 3)

$$p^{NE}=\frac{D_2-C_2}{A_2-B_2+D_2-C_2},\qquad q^{NE}=\frac{D_1-B_1}{A_1-C_1+D_1-B_1},$$

and the Nash equilibrium utilities of the game with the true parameters are

$$u_1^{NE}=\frac{(D_1-B_1)(C_1-D_1)}{A_1-C_1+D_1-B_1}+D_1,\qquad u_2^{NE}=\frac{(D_2-C_2)(B_2-D_2)}{A_2-B_2+D_2-C_2}+D_2.$$

That is, in a Nash equilibrium of the meta-game the players have the same utilities as in the Nash equilibrium of the game with the true declarations. □

Using Theorem 4 we can determine necessary and sufficient conditions for the meta-game to be manipulation-free. Before continuing to the proof of Theorem 5, we consider the following technical restatement of the theorem, where the game utilities are denoted as indicated in Figure 5.

Theorem. *(Restatement of Theorem 5): An opposing-interests* 2×2 *game with a natural parameter space is manipulation-free iff* $\frac{D_1-C_1}{A_1-B_1+D_1-C_1}=\frac{D_2-C_2}{A_2-B_2+D_2-C_2}$ *and* $\frac{D_2-B_2}{A_2-B_2+D_2-C_2}=\frac{D_1-B_1}{A_1-B_1+D_1-C_1}$.

Notice that the above condition exactly specifies that the Nash equilibrium (p, q) of the game (see Observation 3) is symmetric to permutations of player indices, as stated in the theorem.

Proof. A game is manipulation-free if the mixed strategy profile of the agents in a Nash equilibrium of the meta-game is identical to the Nash equilibrium profile of the game with the true parameters, where we used the fact shown in the proof of Theorem 4 above, that any opposing-interests 2×2 game with a natural parameter space has a unique mixed strategy profile $(p, q) \in (0, 1)^2$ that is induced by every Nash equilibrium of the meta-game, and the fact that opposing-interests games have a unique mixed Nash equilibrium.

Using Observation 3, the (true-parameters) Nash equilibrium profile is $p^* = \frac{D_2-C_2}{A_2-B_2+D_2-C_2}$ and $q^* = \frac{D_1-B_1}{A_1-C_1+D_1-B_1}$. Next, using Observation 2, we can write the utilities of the players as functions of p, q, as follows.

$$u_1 = pq(A_1 - C_1 + D_1 - B_1) + p(B_1 - D_1) + q(C_1 - D_1) + D_1,$$

$$u_2 = pq(A_2 - B_2 + D_2 - C_2) + p(B_2 - D_2) + q(C_2 - D_2) + D_2.$$

As in the proof of Theorem 4 above, since p is a function of only the column player's (player 2's) declared parameters and the true game parameters, and q is a function of only the row player's (player 1's) parameters and the true game parameters, and in a natural parameter space the row player can choose parameters to induce any $q \in (0, 1)$, and the column player can similarly choose parameters to induce any $p \in (0, 1)$, we can think of q of the Nash equilibrium of the agents' game as the strategy of the row player, and similarly, of p as the strategy of the column player, and thus the condition for a Nash equilibrium in the meta-game is $\frac{\partial u_1}{\partial q} = \frac{\partial u_2}{\partial p} = 0$.

Hence, the strategy profile (p, q) in a Nash equilibrium of the meta-game is $\tilde{p} = \frac{D_1-C_1}{A_1-B_1+D_1-C_1}$, $\tilde{q} = \frac{D_2-B_2}{A_2-B_2+D_2-C_2}$. Requiring $p^* = \tilde{p}$ and $q^* = \tilde{q}$, we obtain the condition stated in the theorem. $\square$

D.1 Opposing-interests game example

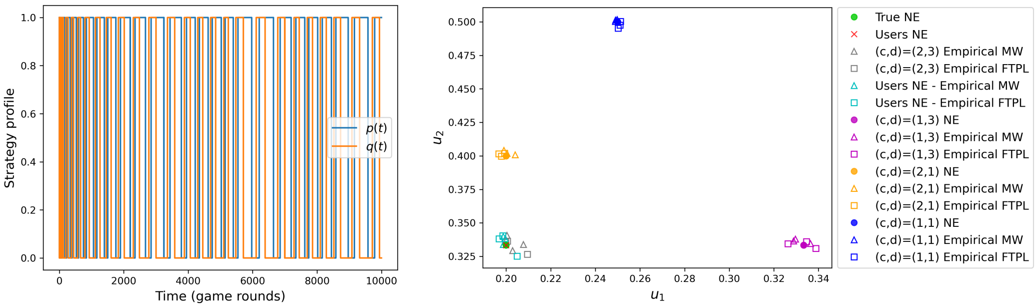
Here we provide further details on the example presented in Section 5.

Nash equilibrium of the (true) game: Using Observation 3 and the parameters of the game shown in Figure 3 (left) in the main text, we obtain that the Nash equilibrium of the game is the mixed strategy profile: $p = 2/3$ and $q = 2/5$. Using Observation 2, we obtain that the utilities in this mixed strategy profile are $u_1 = 1/5$ for the row player and $u_2 = 1/3$ for the column player.

Nash equilibrium of the manipulated game: For any declarations $c, d > -1$ of the players, the "manipulated game", i.e., the game with the declared parameters that is played by the agents, has a unique mixed Nash equilibrium, with mixed strategies (using Observation 3) $p = \frac{d+1}{d+3}$ and $q = \frac{2}{c+3}$. The true expected utilities of the users can be calculated with these values of p and q and the true payoffs of the game by using Observation 2, yielding

$$u_1 = 5pq - 2p - 2q + 1 = \frac{c(1-d) + 3d + 1}{(c+3)(d+3)}, \quad u_2 = 4q + 2p - 6pq - 1 = \frac{c(d-1) - d + 9}{(c+3)(d+3)}.$$

In this example, when the row player declares $c = 1$, and the column player declares the truth, $d = 3$, the utilities to the two players according to the above expressions are $u_1 = 1/3$ and $u_2 = 1/3$. Notice that in this example, this unilateral manipulation increased the utility to the player that manipulated his agent, while the utility to the other player remained the same as in the truthful declarations case. Next, the example describes a unilateral manipulation by the column player, in which the declarations are $c = 2$ (i.e., the truthful declaration for the row player) and $d = 1$. In this case, the utilities are $u_1 = 1/5$ and $u_2 = 2/5$. Here again, a unilateral manipulation by one player increased this player's



(a) Bid dynamics of FTPL agents

(b) Average user utilities for MW and FTPL agents

Figure 6: Dynamics and user utilities in the opposing-interests game example from Section 5, for multiplicative-weights (MW) agents and follow the perturbed leader (FTPL) agents. (a) the dynamics of the mixed strategies p and q of two FTPL agents in 10,000 game rounds. (b) average user utilities in $100{,}000$ game rounds for different manipulation profiles (c, d), where the true values of the utilities of the users are $c = 2, d = 3$ (as presented in Figure 3 in the main text), compared with the theoretical Nash Equilibrium (NE) utilities. The triangular markers show results of MW agents in three simulation runs and the square markers show results of FTPL agents in three simulation runs.

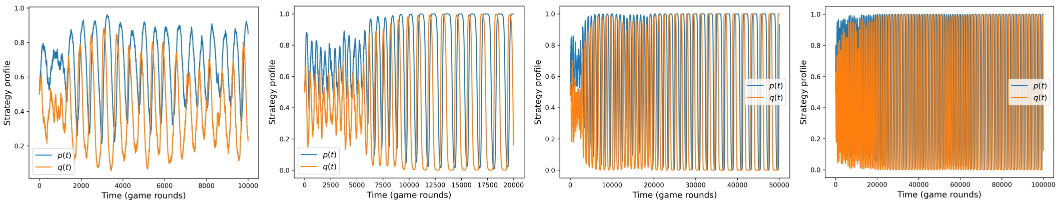
utility without changing the other player's utility compared with the truthful declarations utilities. Yet, these two declaration profiles described above are not equilibria in the meta-game.

Nash equilibrium of the meta-game: The equilibrium condition for the meta-game is $\frac{\partial u_1}{\partial c} = \frac{2-6d}{(c+3)^2(d+3)} = 0$, and $\frac{\partial u_2}{\partial d} = \frac{4(c-3)}{(c+3)(d+3)^2} = 0$. Hence, the Nash equilibrium of the meta-game is the declaration profile $c = 3$, $d = 1/3$. As can be seen using Observation 2, the utilities of the players in this distribution of play are the same as their utilities in the Nash equilibrium of the (true) game. As discussed in the main text, this result is generalized in Theorem 4.

Additional simulations: Figure 6 shows a comparison of simulations of "follow the perturbed leader" (FTPL) [10] with the multiplicative-weights (MW) algorithm [1]. Figure 6a shows an example of the dynamics of two FTPL agents playing the game example presented in Section 5 in the main text, and Figure 6b shows the utilities of the users for several manipulation profiles, including all those presented in the example. The declaration profiles are shown in the legend. It can be seen that the average utilities for the users obtained from the dynamics of the two algorithm types are similar and close to the theoretical Nash equilibrium utilities of each manipulation profile. The manipulation profile $c = d = 1$ (marked in blue in the figure) demonstrates an example of a manipulation profile which leads to increased payoffs to both users compared with the truthful declarations (however, it is not an equilibrium). Additional simulations are shown in the following figures.

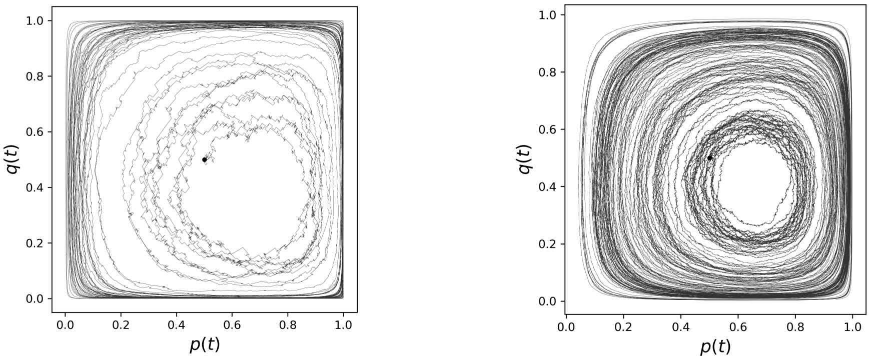
Figure 7 shows additional examples of the dynamics of multiplicative-weights agents in the same opposing-interests game. Figure 8 shows the learning dynamics in a parametric plot, showing the evolution of joint strategy profiles. Every point is the empirical mixed strategy profile of the agents, presented on the p, q plane, where consecutive points in time are connected by a line. The left panel depicts the dynamics with the parameters used in Figure 4 in the main text: $\eta = 0.01$ and $T = 50{,}000$ game repetitions, and the right panel shows the dynamics with $\eta = 0.001$ and $T = 1{,}000{,}000$.

Figure 9 shows estimates of the deviations of the time average of multiplicative-weights agents dynamics from the Nash equilibrium distribution in the opposing-interests game example. In the left panel, it can be seen that, for a fixed update step size ($\eta = 0.01$), indeed these inaccuracies in the convergence of the agents to the Nash distribution (in the time average sense) decrease as $O(1/\sqrt{T})$, as theoretically expected. The right panel shows the distribution of the mixed strategies p, q of the two agents for the case of $T = 50{,}000$ across $N = 1{,}000$ simulation repetitions, which are narrowly centered near the Nash equilibrium profile.



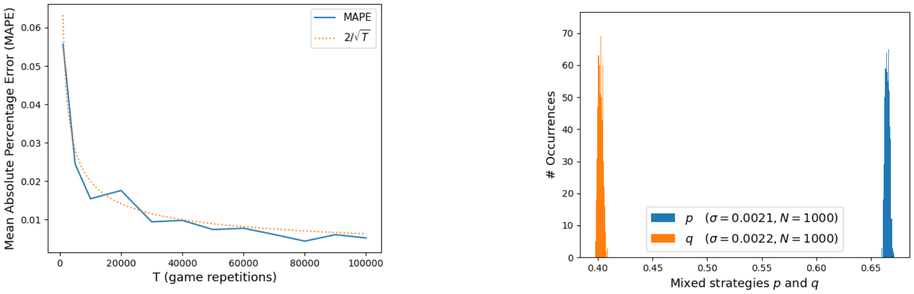
(a) $T = 10{,}000$ (b) $T = 20{,}000$ (c) $T = 50{,}000$ (d) $T = 100{,}000$

Figure 7: Simulations of multiplicative-weights agents in the opposing-interests game example from Section 5. The figures show the dynamics of the mixed strategies across game repetitions for different simulation lengths T.



(a) $T = 50{,}000$, $\eta = 0.01$. (b) $T = 1{,}000{,}000$, $\eta = 0.001$.

Figure 8: Dynamics of multiplicative-wights agents in the opposing-interests game example from Section 5. The figures show parametric plots of the mixed strategies of the two agents in time. (a) the dynamics with update step size parameter $\eta = 0.01$ and $T = 50{,}000$ game repetitions. (b) the dynamics with update step size $\eta = 0.001$ and $T = 1{,}000{,}000$ game repetitions.



(a) Mean Absolute Percentage Error (MAPE) between the empirical distribution and the NE distribution, obtained in simulations of multiplicative-weights agents playing T game rounds, as a function of T. The full line shows the average MAPE and the dotted line shows a comparison to the function $2/\sqrt{T}$, showing that the empirical error indeed decreases as $O(1/\sqrt{T})$.

(b) Histograms of the empirical average values of the action probabilities of multiplicative-weights agents in $T = 50{,}000$ game rounds. The histogram of p (the probability that the row agent plays the top row) is shown in blue, and of q (the probability that the column agent plays the left column) is shown in orange. The legend shows the sample sizes and standard deviations.

Figure 9: Estimates of the deviation of multiplicative-weights agents' time average frequency of play from convergence to the Nash equilibrium distribution in the opposing-interests game example from Section 5.

References

[1] Arora, S., Hazan, E., Kale, S.: The multiplicative weights update method: a meta-algorithm and applications. Theory of Computing 8(1), 121–164 (2012)

[2] Bailey, J.P., Nagarajan, S.G., Piliouras, G.: Stochastic multiplicative weights updates in zero-sum games. arXiv preprint arXiv:2110.02134 (2021)

[3] Bailey, J.P., Piliouras, G.: Multiplicative weights update in zero-sum games. In: Proceedings of the 2018 ACM Conference on Economics and Computation. pp. 321–338 (2018)

[4] Benaïm, M., Hofbauer, J., Hopkins, E.: Learning in games with unstable equilibria. Journal of Economic Theory 144(4), 1694–1709 (2009)

[5] Calvó-Armengol, A.: The set of correlated equilibria of 2x2 games. Working paper (2006)

[6] Cournot, A.A.: Recherches sur les principes mathématiques de la théorie des richesses. L. Hachette (1838)

[7] Even-Dar, E., Mansour, Y., Nadav, U.: On the convergence of regret minimization dynamics in concave games. In: Proceedings of the forty-first annual ACM symposium on Theory of computing. pp. 523–532 (2009)

[8] Hart, S., Mas-Colell, A.: A simple adaptive procedure leading to correlated equilibrium. Econometrica 68(5), 1127–1150 (2000)

[9] Hart, S., Mas-Colell, A.: Simple adaptive strategies: from regret-matching to uncoupled dynamics, vol. 4. World Scientific (2013)

[10] Kalai, A., Vempala, S.: Efficient algorithms for online decision problems. Journal of Computer and System Sciences 71(3), 291–307 (2005)

[11] Kolumbus, Y., Nisan, N.: Auctions between regret-minimizing agents. In: Proceedings of the ACM Web Conference 2022 (WWW '22). pp. 100–111 (2022), `https://arxiv.org/pdf/2110.11855.pdf`

[12] Mas-Colell, A., Whinston, M.D., Green, J.R., et al.: Microeconomic theory, vol. 1. Oxford university press New York (1995)

[13] Mertikopoulos, P., Papadimitriou, C., Piliouras, G.: Cycles in adversarial regularized learning. In: Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 2703–2717. SIAM (2018)

[14] Monnot, B., Piliouras, G.: Limits and limitations of no-regret learning in games. The Knowledge Engineering Review 32 (2017)

[15] Moulin, H.: Dominance solvability and cournot stability. Mathematical social sciences 7(1), 83–102 (1984)

[16] Nachbar, J.H.: "evolutionary" selection dynamics in games: Convergence and limit properties. International journal of game theory 19(1), 59–89 (1990)

[17] Papadimitriou, C., Piliouras, G.: From nash equilibria to chain recurrent sets: An algorithmic solution concept for game theory. Entropy 20(10), 782 (2018)

[18] Shapley, L.: Some topics in two-person games. Advances in game theory 52, 1–29 (1964)