Imperfect Best-Response Mechanisms

Diodato Ferraioli · Paolo Penna

Abstract Best-response Mechanisms, introduced by Nisan et al. (2011), provide a unifying framework for studying various distributed protocols in which the participants are instructed to repeatedly best respond to each others' strategies. Two fundamental features of these mechanisms are convergence and incentive compatibility. This work investigates convergence and incentive compatibility conditions of such mechanisms when players are not guaranteed to always best respond but they rather play an *imperfect* best-response strategy. That is, at every time step every player deviates from the prescribed best-response strategy according to some probability parameter. The results explain to what extent convergence and incentive compatibility depend on the assumption that players never make mistakes, and how robust such protocols are to "noise" or "mistakes".

Keywords Game dynamics · Noise · Best-response · Convergence · Incentive compatibility

1 Introduction

One of the key issues in designing a distributed protocol (algorithm) is its convergence to a stable state, also known as self-stabilization. Intuitively, starting from any initial (arbitrarily corrupted) state, the protocol should eventually converge to the "correct state" as intended by the designer. Incentive compatibility considerations

D. Ferraioli (✉)
DIAG, Sapienza Università di Roma, Rome, Italy
e-mail: ferraioli@dis.uniroma1.it

P. Penna
Institute of Theoretical Computer Science, ETH Zurich, Zurich, Switzerland
e-mail: paolo.penna@gmail.com

Springer

have been also become important in the study of distributed protocols since the participants cannot be assumed to altruistically implement the protocol if that is not beneficial for themselves.

A unifying game-theoretic approach for proving both convergence and incentive compatibility has been recently proposed by Nisan et al. [19]. They consider so-called *best-response mechanisms* or *dynamics* in which the protocol prescribes that each participant (or player) should simply best-respond to the strategy currently played by the other players. Essentially the same base *game* is played over and over (or until some equilibrium is reached), with players updating their strategies in *some* (unspecified) order. Nisan et al. [19] proved that for a suitable class of games the following happens:

– *Convergence.* The dynamics eventually reaches a unique equilibrium point (a unique pure Nash equilibrium) of the base game regardless of the order in which players respond (including concurrent responses).
– *Incentive compatibility.* A player who deviates from the prescribed best-response strategy can only worsen his/her final utility, that is, the dynamics will reach a different state that yields weakly smaller payoff.

These two conditions say that the protocol will eventually "stabilize" if implemented correctly, and that the participants are actually willing to do so. Convergence itself is a rather strong condition because no assumption is made on how players are scheduled for updating their strategies, a typical situation in *asynchronous* settings. Incentive compatibility is also non-trivial because a best-response is a *myopic* strategy which does not take into account the future updates of the other players. In fact, neither of these conditions can be guaranteed on general games.

Nisan et al. [19] showed that several protocols and mechanisms arising in computerized and economics settings are in fact best-response mechanisms over the *restricted* class of games for which convergence and incentive compatibility are always guaranteed. Their applications include: (1) the Border Gateway Protocol (BGP) currently used in the Internet, (2) a game-theoretic version of the TCP protocol, and (3) mechanisms for the classical cost-sharing and stable roommates problems studied in micro-economics.

In this work we address the following question:

> *What happens to these protocols/mechanisms if players do not always best respond?*

Is it possible that when players sometimes deviate from the prescribed protocol (e.g., by making mistakes in computing their best-response or by scarce knowledge about other players' actions) then the protocol does not converge anymore? Can such mistakes induce some other player to adopt a non-best-response strategy that results in a better payoff? Such questions arise naturally from fault tolerant considerations in protocol design, and have several connections to equilibria computation and bounded rationality issues in game theory.

Our contribution. We investigate convergence and incentive compatibility conditions of mechanisms (dynamics) described in [19] when players are not guaranteed

to always best respond but they rather play an *imperfect* best-response strategy. That is, at every time step every player deviates from the prescribed best-response strategy according to some probability parameter $p \geq 0$. The parameter p can be regarded as the probability of making a mistake every time the player updates his/her strategy.

Our results indicate to what extent convergence and incentive compatibility depend on the assumption that players never make mistakes, and provide necessary and sufficient conditions for the robustness of these mechanisms/dynamics:

– *Convergence.* Because of mistakes convergence can be achieved only in a probabilistic sense. We give bounds on the parameter p in order to guarantee convergence with sufficiently good probability.

One might think that for small values of p our dynamics behaves (approximately) as the dynamics without mistakes, i.e. it converges to an equilibrium point regardless of the order in which players respond. However, it turns out this is not the case. Indeed, our first negative result (Theorem 2) shows that even when p is exponentially small in the number n of players the dynamics does not converge, i.e., the probability of being in the equilibrium is always small (interestingly, such negative result applies also to certain instances of BGP in the realistic model of Gao and Rexford [10]).

The proof of this result shows the existence of a particularly "bad" schedule that amplifies the probability that the imperfect dynamics deviates from the perfect one. This highlights that imperfect dynamics differ from their perfect counterpart, since in the former convergence results *must consider* how players are scheduled. Indeed, we complement the negative result above with a general positive result (Theorem 3) saying that convergence can be guaranteed whenever p is polynomially small in some parameters depending both on the game and on the schedule of the players. For such values of p, the upper bound on the convergence time of dynamics without mistakes is (nearly) an upper bound for the *imperfect* best-response dynamics.

– *Incentive compatibility.* We first observe that games that are incentive compatible for dynamics without mistakes, may no longer be incentive compatible for imperfect best-response dynamics (Theorem 4). In other words, a player who deviates incidentally from the mechanism induces another player to *deliberately* deviate. A sufficient condition for incentive compatibility of imperfect best-response mechanisms (Theorem 5) turns out to be a quantitative version of the one given in [19]. Roughly speaking, if the payoffs of the Nash equilibrium are *sufficiently* larger than the other possible payoffs, then incentive compatibility holds. As the probability p of making mistakes vanishes, the class of games for which convergence and incentive compatibility holds tends to the class of games in [19].

Our focus is on the same class of (base) games of [19] since this is the only known general class for which best-response dynamics converge (regardless of the schedule of the players) and are incentive compatible. In our view this class is important as it describes accurately certain protocols that are implemented in practice and it unifies several results in game theory. In particular, the mathematical model of how the commercial relationships between Autonomous Systems (the Gao-Rexford model [10])

leads to games in this class and, ultimately, to the fact that BGP converges and is incentive compatible [17, 19]. Considering more general games for the analysis of BGP would in fact produce "wrong" results (constructing unrealistic examples for which the protocol does not converge or is not incentive compatible).

We nevertheless take one step further and apply the tools from [19] (and this work) to a natural generalization of their games. Intuitively speaking, these games guarantee only that best-response converge to a *subgame*. In this case, the dynamics of the original game can be approximated by the dynamics of the subgame (Theorem 6). Unfortunately, this "reduction" cannot be pushed further simply because the subgame can be an arbitrary game and different p-imperfect best-response dynamics lead to different equilibria (even for the same p; see Section 5.1). However, when the dynamics on the subgame are well-understood, then we can infer their behavior also on the original game.

Related work. Convergence of best-response dynamics is a main topic in game theory. It relates to the so-called problem of equilibrium selection, that is, how the players converge to an equilibrium (see the book by Harsanyi and Selten [13]). Noisy versions of best-response dynamics have been studied in order to consider the effects of bounded rationality and limited knowledge of the players (which limits their ability to compute their best responses).

Our imperfect best response dynamics are similar to the *mutation model* by Kandori et al. [15], and to the *mistakes model* by Young [20], and Kandori and Rob [16]. A related model is the *logit dynamics* of Blume [7] in which the probability of a mistake depends on the payoffs of the game. All of these works assume a particular schedule of the players (the order in which they play in the dynamics). Whether such an assumption affects the selected equilibrium is the main focus of a recent work by Alos-Ferrer and Netzer [1]. They studied convergence of these dynamics on general games when the parameter p vanishes, and provide a characterization of the resulting equilibria in terms of a kind of potential function of the game. Convergence results that take into account non-vanishing p are only known for fixed dynamics on specific class of games (see, e.g., [2, 3, 7]).

Incentive compatibility of best-response dynamics provide a theoretical justification for several protocols and auctions widely adopted in practice. Levin et al. [17] proved convergence and incentive compatibility of the intricate BGP protocol in the current Internet (based on the mathematical model by Gao and Rexford [10] that captures the commercial structure that underlies the Internet and explains convergence of BGP). The theoretical analysis of TCP-inspired games by Godfrey et al. [11] shows that certain variants of the current TCP protocol converge (the flow rate stabilizes) and are incentive compatible on arbitrary networks (this property assumes routers adopt specific queuing policy). The so-called Generalized Second-Price auctions used in many ad-auctions is another example of incentive compatible best-response mechanism as proved by Nisan et al. [18]. All of these problems (and others) and results have been unified by Nisan et al. [19] in their framework.

2 Definitions

We consider an n-player base game G in which each player i has a finite set of strategies S_i, and a utility function u_i. Each player can select a strategy $s_i \in S_i$ and the vector $\mathbf{s} = (s_1, \ldots, s_n)$ is the corresponding strategy profile, with $u_i(\mathbf{s})$ being the payoff of player i. To stress the dependency of the utility u_i on the strategy z_i of player i we adopt the standard notation $(z_i, \mathbf{s}_{-i})$ to denote the vector $(s_1, \ldots, s_{i-1}, z_i, s_{i+1}, \ldots, s_n)$.

Definition 1 (**best response**) A strategy $s_i^* \in S_i$ is a *best response* to the strategy profile $\mathbf{s}_{-i}$ if this strategy maximizes i's payoff, that is, for every other strategy $s_i' \in S_i$ it holds

$$u_i(s_i', \mathbf{s}_{-i}) \le u_i(s_i^*, \mathbf{s}_{-i}).$$

(Imperfect) Best-response dynamics. A *game dynamics* consists of a (possibly infinite) sequence of strategy profiles $\mathbf{s}^0, \mathbf{s}^1, \ldots$, where $\mathbf{s}^0$ is an arbitrarily chosen profile and the profile $\mathbf{s}^t$ is obtained from $\mathbf{s}^{t-1}$ by letting some of the players updating their strategies. Therefore a game dynamics is determined by a *schedule* of the players specifying, for each time step, the subset of players that are selected for updating their strategies, and a *response rule*, which specifies how a player updates her strategy (possibly depending on the past history and on the current strategy profile). A natural and well-studied response rule prescribes that players (myopically) choose the best response to the strategies currently chosen by the others (according to Definition 1).

In this work we focus on dynamics based on the following kind of schedules and response rules. As for the response rule, we consider a scenario in which a selected player can deviates from the (prescribed) best-response.

Definition 2 (*p*-**imperfect response rule**) A response rule is p-imperfect if a player does not update her strategy to the best-response with probability at most p.

Examples of these rules are given in the mutation [15] or mistakes models [16, 20] (see Appendix A.1 for a brief overview). The best-response rule is obviously 0-imperfect, which we also denote as *perfect*. The response rule in logit dynamics [7] (see Appendix A.2 for a brief overview) is p-imperfect with

$$p \le \frac{m-1}{m-1+e^\beta}$$

for all games in which the payoff are integers[1] and each player has at most m strategies.

---

[1]When the minimum difference in the payoff of an agent between a best-response and a non-best response is γ, this extends easily by taking $\beta_\gamma = \beta \cdot \gamma$ in place of β.

In order to avoid trivial impossibility results on convergence we need to consider a non-adaptive adversarial schedule that satisfies some reasonable fairness condition. We allow both deterministic and randomized schedules satisfying the following definition.

Definition 3 (((R, ε)-**fair schedule**) A schedule is (R, ε)-fair if there exists a non-negative integer R such that, for any interval of R time steps, all players are selected at least once in this interval with probability at least $1 - \varepsilon$, i.e., for every player i and any time step a we have
$$\Pr(SEL_{i,a,R}) \geq 1 - \varepsilon,$$
where $SEL_{i,a,R}$ is the event that player i is selected at least once in the interval $[a + 1, a + R]$.

The convergence rate of a dynamics is measured in number of rounds, where a *round* is a sequence of consecutive time steps in which each player is selected for update at least once. The definition above affects the (expected) length of a round and the probability that a round has a certain length (number of steps).

Observation 1 *If a schedule is (R, ε)-fair then, for every $0 < \delta < \varepsilon$, it is also (R', δ)-fair with*
$$R' = R \cdot \left\lceil \frac{\ln(1/\delta)}{\ln(1/\varepsilon)} \right\rceil.$$

Proof We show that, if a schedule is (R, ε)-fair, then every player is selected at least once with probability at least $1 - \delta$ in an interval of R' time steps. For any interval of R time steps, the probability that a player i is not selected is at most ε regardless of what happened before this interval. Thus, the probability that i is never selected in k consecutive intervals of R time steps each is at most ε^k. For $k = \left\lceil \frac{\ln(1/\delta)}{\ln(1/\varepsilon)} \right\rceil$, this probability is at most δ because
$$\varepsilon^k \leq \varepsilon^{\frac{\ln(1/\delta)}{\ln(1/\varepsilon)}} = \varepsilon^{\frac{\ln \delta}{\ln \varepsilon}} = e^{\ln \delta} = \delta.$$
Then, the probability that player i is never selected in $R' = kR$ time steps is at most δ. $\qquad\square$

Another important parameter of the schedule is the *maximum* number of players selected for update in one step by the schedule, denoted as η (with $\eta \leq n$).

Example 1 Scheduling players in round-robin fashion or concurrently corresponds to $(n, 0)$-fair and $(1,0)$-fair schedules, respectively. From the well-known Coupon Collector's problem, we have that selecting a player uniformly at random at each time step is (R, ε)-fair with $R = \lceil n \ln n + n \ln(1/\epsilon) \rceil = O(n \ln n)$.[2]

---

[2]The analysis of the Coupon Collector's says that, if we pick one player uniformly at random and repeat this for $R \geq n \ln n + cn$ time steps, then the probability that *some* player has never been selected is at most e^{-c}. This probability is smaller than ϵ for $c \geq \ln(1/\epsilon)$.

Henceforth, we always refer as *imperfect best-response dynamics* to any dynamics whose schedule is (R, ε)-fair and whose response rule is p-imperfect, and as *imperfect best-response mechanisms* to the class of all imperfect best-response dynamics. Note that we do not put any other constraint on the way the dynamics run. For instance, the response rule of the players *can* depend on the history of the dynamics, as long as the probability of not choosing a best response is at most p.

Convergence and incentive compatibility. We define convergence and incentive compatibility as in Nisan et al. [19]. We say that a game dynamics for a game G *converges* if it eventually converges to a (pure) *Nash equilibrium* of the game, i.e. there exists $t > 0$ such that the strategy profile of players at time step t coincides with a Nash equilibrium of G. This definition might seem limited in a setting in which players can "make mistakes" since, once a Nash equilibrium is reached, there is a positive probability of leaving again from this state. However, we choose to adopt this definition of convergence (the same as in [19]) because when the parameters of the dynamics guarantee convergence to a Nash equilibrium in a "small" number of rounds with "good probability", then the dynamics is likely to remain in the Nash equilibrium for "many" steps (see Corollary 1). This is useful in applications in which there is a "termination condition" (for instance, the dynamics ends if no player updates her strategy for a certain number of consecutive steps).

The above considerations on "termination condition" leads to the following definition of *total utility* and the resulting *incentive compatibility* condition, both defined as in [19]. Let us denote with X^t the random variable that represents the strategy profile induced by a dynamics on a game G after t time steps. If the dynamics terminates after some finite number of steps T, then the *total utility* of a player i is defined as $\Gamma_i = E\left[u_i\left(X^T\right)\right]$; otherwise, that is if the dynamics does not terminate after finite time, the total utility is defined as $\Gamma_i = \limsup_{t \to \infty} E\left[u_i\left(X^t\right)\right]$. Then, a dynamics for a game G is *incentive compatible* if playing this dynamics is a pure Nash equilibrium in a new game $G^\star$ in which players' strategies are all possible response rule that may be used in G and players' utilities are given by their total utilities. That is, a dynamics for a game G is incentive compatible if every player does not improves her total utilities by playing according to a response rule different from the one prescribed, given that each other player does not deviate from the prescribed response rule.

Never best-response and the main result in [19]. Nisan et al. [19] analyzed the convergence and incentive compatibility of the (perfect) best-response dynamics. Before stating their result, let us now recall some definitions.

Definition 4 (**never best-response**) A strategy s_i is a never best-response (NBR) for player i if, for every $\mathbf{s}_{-i}$, there exists s_i' such that $u_i(s_i, \mathbf{s}_{-i}) < u_i(s_i', \mathbf{s}_{-i})$.[3]

---

[3]Nisan et al. [19] assume that each player has also a *tie breaking rule* $\prec_i$, i.e., a total order on S_i, that depends solely on the player's private information. In the case that a tie breaking rule $\prec_i$ has been defined for player i, then s_i is a NBR for i also if $u_i(s_i, \mathbf{s}_{-i}) = u_i(s_i', \mathbf{s}_{-i})$ and $s_i \prec_i s_i'$. However, such tie-breaking

Note that according to a *p*-imperfect response rule, a player updates her strategy to a NBR with probability at most *p*.

Definition 5 (**elimination sequence**) An elimination sequence for a game G consists of a sequence of subgames

$$G = G_0 \supset G_1 \supset \cdots \supset G_r = \hat{G},$$

where any game G_{k+1} is obtained from the previous one by letting a player $i^{(k)}$ eliminate strategies which are NBR in G_k.

The length of the shortest elimination sequence for a game G is denoted with ℓ_G (we omit the subscript when it is clear from the context). It is easy to see that for each game $\ell_G \leq n(m-1)$, where m is the maximum number of strategies of a player.
Our results will focus on the following classes of games.

Definition 6 (**NBR-reducible and NBR-solvable games**) The game G is NBR-reducible to $\hat{G}$ if there exists an elimination sequence for G that ends in $\hat{G}$. The game G is NBR-solvable if it is NBR-reducible to $\hat{G}$ and $\hat{G}$ consists of a unique profile.

Example 2 Consider a 2-player game with strategies $\{0, 1, 2\}$ and the following utilities:

	0	1	2
0	0,0	0,0	0,-2
1	0,0	-1,-1	-1,-2
2	-2,0	-2,-1	-2,-2

Notice that strategy 2 is a NBR for both players. Hence, there exists an elimination sequence of length 2 that reduces the above game to its upper-left 2×2 subgame with strategy set $\{0, 1\}$ for each player. Therefore, this game is NBR-reducible. If we modify the utilities in this upper-left 2×2 subgame as follows

	0	1
0	0,0	$0,-\delta$
1	$-\delta,0$	-1,-1

then the game reduces further to the profile (0,0) and hence it is NBR-solvable. Observe that the unique profile at which the game G is reduced in an NBR-solvable game is also the unique Nash equilibrium of the original game.

While the convergence result of [19] holds for the class of NBR-solvable games, in order to guarantee incentive compatibility they introduce the following condition on the payoffs:

Definition 7 (**NBR-solvable with clear outcome**) A NBR-solvable game is said to have a clear outcome if, for every player *i*, there is a player-specific elimination

---

rule can be implemented in a game by means of suitable perturbations of the utility function: with such an implementation our definition become equivalent to the one given in [19].

sequence such that the following holds. If i appears the first time in this sequence at position k, then in the subgame G_k the profile that maximizes the utility of player $i = i^{(k)}$ is the Nash equilibrium.

Theorem 1 (**main result of** [19]) *Best-response dynamics of every NBR-solvable game G converge to a pure Nash equilibrium of the game and, if G has clear outcome, are incentive compatible. Moreover, convergence is guaranteed in ℓ_G rounds for any schedule, where a round is a sequence of consecutive time steps in which each player is selected for update at least once.*

Note that convergence and incentive compatibility holds regardless of the schedule of players. Moreover, the theorem implies that for a specific (R, ε)-fair schedule the dynamics converges in $O(R \cdot \ell_G)$ time steps. Note also that convergence does *not* require a clear outcome and this condition is only needed for incentive compatibility.

3 Convergence Properties

3.1 A Negative Result

In this section we will show that the result about convergence of the best-response dynamics in NBR-solvable games given in [19] is not resistant to the introduction of "noise", i.e., there is a NBR-solvable games and an imperfect best-response dynamics that never converges, except with small probability, to the Nash equilibrium even for values of p very small. Specifically we will prove the following theorem.

Theorem 2 *For every $0 < \delta < 1$, there exist an n-player NBR-solvable game G and an imperfect best-response dynamics with parameter p exponentially small in n such that for every integer $t > 0$ the dynamics converges after t steps with probability at most δ.*

We highlight that this theorem does not state only that the probability of being in the Nash equilibrium at any time t is low, but that the probability that the system has ever been in the Nash equilibrium at any time *before t* is low.

The game. Consider the following game: there are n players with two strategies 0 and 1. For each player i, we define the utility function of i as follows: if players $1, \ldots, i - 1$ are playing 1, then i prefers to play 1 regardless of the strategies played by players $i + 1, \ldots, n$, i.e., $u_i(\mathbf{1}_{1 \cdots i-1}, 0, \mathbf{s}_{i+1 \cdots n}) < u_i(\mathbf{1}_{1 \cdots i-1}, 1, \mathbf{s}_{i+1 \cdots n})$ for each strategy profile $\mathbf{s}$, otherwise i prefers to play 0, i.e. $u_i(\mathbf{s}_{-i}, 1) < u_i(\mathbf{s}_{-i}, 0)$ for any $\mathbf{s}$ such that $\mathbf{s}_{1 \cdots i-1} \neq \mathbf{1}_{1 \cdots i-1}$.

It is easy to check that the above game is NBR-solvable. Indeed, the elimination sequence consists of players $1, 2, \ldots, n$ eliminating strategy 0 one-by-one in this order (note that 1 is a dominant strategy for player 1 and, more in general, strategy 1 is dominant for i in the subgame in which all players $1, \ldots, i - 1$ have eliminated

0). The subgame $\hat{G}$ consists of the unique pure Nash equilibrium that is the profile $\mathbf{1} = (1, \ldots, 1)$.

The p-imperfect response rule. All players play a generic p-imperfect response rule, according to which best-response is played with probability $1 - p$. Specifically, we have:

– Player i chooses strategy 0 with probability p if all players $j < i$ are playing strategy 1;
– Player i chooses strategy 0 with probability $1 - p$ if at least one player $j < i$ is playing strategy 0.

The $(2^{n-1}, 0)$-fair schedule. Let us start by defining sequences σ_i, with $i = 1, \ldots, n$, recursively as follows

$$\sigma_1 = 1, \quad \sigma_2 = 12, \quad \sigma_3 = 1213, \quad \ldots \quad \sigma_i = \sigma_{i-1}\sigma_{i-2}\cdots\sigma_1 i.$$

Observe that each sequence has length 2^{i-1}. Then players are scheduled one at a time according to σ_n and then repeat. The resulting schedule will be denoted as σ. A key observation about the schedule σ follows.

Observation 2 *Between any two occurrences of player $i < n$ in the schedule σ there is an occurrence of a player $j \geq i + 1$.*

Proof Note that the player i can occur in the sequence only when the subsequence σ_i occurs. Anyway, any occurrence of the sequence σ_i is within a subsequence σ_j for $j = i + 1, \ldots, n$. In σ_j the occurrence of σ_i is followed by the occurrences of $\sigma_{i-1}, \ldots, \sigma_1$, in which the player i cannot occur, and then by the occurrence of j, completing in this way the proof. $\qquad\square$

Intuitively speaking, this property causes any *bad move* of some player in the sequence σ_n to propagate to the last player n, where by "bad move" we mean that at time t the corresponding player $\sigma(t)$ plays strategy 0 given that each player $j < \sigma(t)$ plays 1 (thus, a bad move occurs with probability p).

Proof (of Theorem 2) Throughout the proof we will denote 2^{n-1} as τ for sake of readability.

Let X_t be the random variable that represents the profile of the game at step t. We will denote with X_t^n the n-th coordinate of X_t, i.e. the strategy played by player n at time t. Suppose that player n plays 0 at the beginning. Then, for every $t < \tau$, the probability that at time step t the game is in a Nash equilibrium is obviously 0. Consider now $t \geq \tau$. The probability that at time step t the game is in a Nash equilibrium is obviously less than the probability that $X_t^n = 1$. Hence it will be sufficient to show that $\Pr(X_t^n = 1) \leq \delta$. Note that $X_t^n = X_{c\tau}^n$, c being the largest

integer such that $t \geq c \cdot \tau$. Since both the response rule and the schedule described above are memoryless, for every profile $\mathbf{s}$

$$\Pr\left(X^n_{c\tau} = 1 \mid X_{(c-1)\tau} = \mathbf{s}\right) = \Pr\left(X^n_\tau = 1 \mid X_0 = \mathbf{s}\right).$$

Let us use $\Pr_{\mathbf{s}}\left(X^n_\tau = 1\right)$ as a shorthand for $\Pr\left(X^n_\tau = 1 \mid X_0 = \mathbf{s}\right)$. Moreover, let $\overline{B}$ denote the event that no bad move occurs in the interval $[1, \tau]$ and let B_t denote the event that the first bad move occurs at time $t \in \{1, \ldots, \tau\}$. Then

$$\Pr_{\mathbf{s}}\left(X^n_\tau = 1\right) = \Pr_{\mathbf{s}}\left(X^n_\tau = 1 \mid \overline{B}\right) \cdot \Pr(\overline{B}) + \sum_{t=1}^{\tau} \Pr_{\mathbf{s}}\left(X^n_\tau = 1 \mid B_t\right) \Pr(B_t).$$

Note that B_t has probability at most p and $\overline{B}$ has probability $(1 - p)^\tau$. Obviously, $\Pr_{\mathbf{s}}\left(X^n_\tau = 1 \mid \overline{B}\right) = 1$. Moreover, by Observation 2, given a bad move of player $i_0 \neq n$ at time t_{i_0}, there is a sequence of time steps $t_{i_0} < t_{i_1} < t_{i_2} < \cdots < t_n$ such that player $i_j > i_{j-1}$ is selected at time t_{i_j} and it is not selected further before $t_{i_{j+1}}$. Therefore, player i_1 plays 0 at time t_{i_1} with probability $1 - p$ because at that time i_0 is still playing 0. Similarly, if player i_j at time $t_{i_{j+1}}$ is still playing 0, then player i_{j+1} will play 0 with probability $1 - p$. Hence,

$$\Pr_{\mathbf{s}}\left(X^n_\tau \neq 1 \mid B_t\right) \geq (1 - p)^n.$$

Then

$$\Pr_{\mathbf{s}}\left(X^n_\tau = 1\right) \leq (1 - p)^\tau + \tau p(1 - (1 - p)^n).$$

We next upper bound these quantities by using $1 - x \leq e^{-x} \leq (1 + x)^{-1}$ as follows. As for the first quantity,

$$(1 - p)^\tau \leq e^{-p\tau} \leq \frac{1}{1 + p\tau}.$$

As for the second quantity, observe that

$$e^{-\frac{p}{1-p}} \leq \left(1 + \frac{p}{1 - p}\right)^{-1} = 1 - p.$$

This and $1 - e^{-x} \leq x$ imply

$$1 - (1 - p)^n \leq 1 - e^{-n\frac{p}{1-p}} \leq n\frac{p}{1 - p}.$$

We have thus shown that

$$\Pr_{\mathbf{s}}\left(X^n_\tau = 1\right) \leq \frac{1}{1 + p\tau} + \tau pn\frac{p}{1 - p}.$$

By setting $p := \frac{2-\delta}{\delta\tau}$, we have

$$\frac{1}{1 + p\tau} = \frac{1}{1 + \frac{2-\delta}{\delta}} = \frac{\delta}{2} \quad \text{and} \quad \tau pn\frac{p}{1 - p} = \frac{2 - \delta}{\delta} n\frac{2 - \delta}{\delta\tau - 2 + \delta}.$$

Since $\tau = 2^{n-1}$, for every $\delta > 0$, the latter quantity tends to 0 as n goes to infinity. The theorem thus follows by observing that there exists a sufficiently large n such that $\mathrm{Pr}_{\mathbf{s}}\left(X_\tau^n = 1\right) < \delta$. □

Remark 1 (BGP games) The proof of the above theorem can be adapted to the class of BGP games [17, 19]. In these games, we are given an undirected graph with n nodes (the players) and an additional destination node d. The strategy of each player is to choose the *next hop* in the routing towards d, that is, to point to one of its neighbors in the graph. Each node has a preference order over all possible paths (and "non-valid paths" have naturally the lowest rank).

We consider the instance with $n + 2$ players and the destination node d given in Fig. 1.

Moreover, we specify the relationships among players according to the Gao-Rexford model [10], as follows: each node i, with $i = 1, \ldots, n - 1$, is a provider for node $i + 1$ and a customer of node $i - 1$; moreover, node a is a provider for each node $i = 1, \ldots, n$ (Fig. 1).

The so-called filtering policy of the Gao-Rexford model (that dictates that each node forwards traffic from one of its customers or to one of its customers only) implies that in this instance the only available path for nodes 0 and a is the one that routes directly to d. As for the player $i = 1, 2, \ldots, n$, they are left with two available choices: we denote by 0 the strategy to route to node a and by 1 the strategy to route to node $i - 1$. We can then set the preferences of player i over the paths as follows:

- Node i's top ranked path is $i \rightarrow i - 1 \rightarrow \cdots \rightarrow 1 \rightarrow 0 \rightarrow d$.
- Node i's second top ranked path is $i \rightarrow a \rightarrow d$.
- Every other path has lower rank.

The reader can check that in this instance we have that:

1. For any strategy profile $\mathbf{s}$,

$$u_i(\mathbf{1}_{1\cdots i-1}, 0, \mathbf{s}_{i+1\cdots n}) < u_i(\mathbf{1}_{1\cdots i-1}, 1, \mathbf{s}_{i+1\cdots n}),$$

   by definition of i's top ranked path above.
2. For any $\mathbf{s}$ such that $\mathbf{s}_{1\cdots i-1} \neq \mathbf{1}_{1\cdots i-1}$

$$u_i(\mathbf{s}_{-i}, 1) < u_i(\mathbf{s}_{-i}, 0),$$

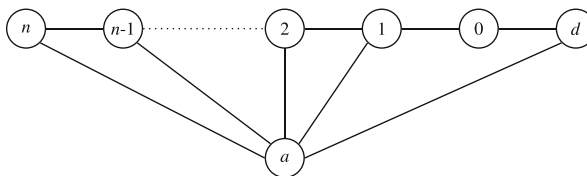   by definition of i's second top ranked path above.

Fig. 1 An instance of the BGP game in which an imperfect best-response dynamics fails to converge

Remark 2 (*logit dynamics*) It is interesting that we can instantiate the abstract game described in the proof with utilities:

$$u_i(0, \mathbf{s}_{-i}) = 0 \quad \text{and} \quad u_i(1, \mathbf{s}_{-i}) = \begin{cases} 1, & \text{if} s_1 = \cdots = s_{i-1} = 1; \\ -1, & \text{otherwise}; \end{cases}$$

where L is a large number. Similarly, the response rule described above may be instantiated as a logit response rule with noise β, that corresponds to setting

$$p = \frac{1}{1 + e^\beta}.$$

Therefore our lower bound also applies to logit dynamics (for a suitable game and a suitable noise parameter β).

The proof of Theorem 2 highlights that imperfect dynamics differ from the perfect ones, since convergence result should necessarily depend on the schedule of players. Specifically, a closer look at the proof shows that there is a game in which an imperfect best-response dynamics does not converge for $p = 1/R$ or greater. As a consequence, it may be possible to prove that imperfect best response dynamics converge for any game only by taking p smaller than $1/R$.

3.2 A Positive Result (Convergence Time)

Given the negative result above, we wonder whether there are values of p for which the convergence of perfect best-response mechanisms is restored. The following theorem states that this occurs when p is small with respect to parameters R, η and ℓ.

Theorem 3 *For any NBR-solvable game G and any small $\delta > 0$ an imperfect best-response dynamics converges to the Nash equilibrium of G in $O(R \cdot \ell \ln \ell)$ steps with probability at least $1 - \delta$, whenever $p \leq \frac{c}{\eta R \cdot \ell \ln \ell}$, for a suitably chosen constant $c = c(\delta)$.*

The following two lemmas represent the main tools in the proof of the theorem. Both lemmas hold for NBR-solvable games as for the more general class of NBR-reducible games. Moreover, in both lemmas we denote with X_t the random variable that represents the profile of the game after t steps of an imperfect best-response dynamics. Note also that, for an event E we denote with $\Pr_{\mathbf{s}}(E)$ the probability of the event E conditioned on the initial profile being $X_0 = \mathbf{s}$, i.e., $\Pr_{\mathbf{s}}(E) = \Pr(E \mid X_0 = \mathbf{s})$.

Lemma 1 *For any NBR-reducible game G, any imperfect best-response dynamics, any starting profile $\mathbf{s}$, and any t, θ with $t, \theta \geq 0$, we have*

$$\Pr_{\mathbf{s}}(X_{t+\theta} \in G_k \mid X_\theta \in G_k) \geq 1 - \eta p t, \qquad \forall k = 0, \ldots \ell; \qquad (1)$$

$$\Pr_{\mathbf{s}}(X_{R+\theta} \in G_{k+1} \mid X_\theta \in G_k) \geq 1 - \eta p R - \varepsilon, \qquad \forall k = 0, \ldots, \ell - 1. \qquad (2)$$

Proof Let the dynamics be in G_k at time θ and observe that if the dynamics is not in G_k at time $t + \theta$, then in one of the steps in the interval $[\theta + 1, \theta + t]$ some selected player played a NBR. Since at every step at most η players are selected, (1) follows from the union bound.

Similarly, if the dynamics is not in G_{k+1} at time $t + \theta$ given that player $i^{(k)}$ has been selected for update at least once during the interval $[\theta + 1, \theta + t]$, then in one of these time steps some selected player played a NBR. Hence,

$$\Pr_{\mathbf{s}} \left(X_{t+\theta} \notin G_{k+1} \mid X_\theta \in G_k \cap SEL_{i^{(k)},\theta,t} \right) \le \eta t p. \tag{3}$$

Now simply observe that the definition of conditional probability implies that

$$\begin{aligned}
&\Pr_{\mathbf{s}}(X_{t+\theta} \notin G_{k+1} \mid X_\theta \in G_k) \\
&= \Pr_{\mathbf{s}} \left(X_{t+\theta} \notin G_{k+1} \mid X_\theta \in G_k \cap SEL_{i^{(k)},\theta,t} \right) \Pr_{\mathbf{s}} \left(SEL_{i^{(k)},\theta,t} \mid X_\theta \in G_k \right) \\
&+ \Pr_{\mathbf{s}} \left(X_{t+\theta} \notin G_{k+1} \mid X_\theta \in G_k \cap \overline{SEL}_{i^{(k)},\theta,t} \right) \Pr_{\mathbf{s}} \left(\overline{SEL}_{i^{(k)},\theta,t} \mid X_\theta \in G_k \right) \\
&\le \Pr_{\mathbf{s}} \left(X_{t+\theta} \notin G_{k+1} \mid X_\theta \in G_k \cap SEL_{i^{(k)},\theta,t} \right) + \Pr_{\mathbf{s}} \left(\overline{SEL}_{i^{(k)},\theta,t} \mid X_\theta \in G_k \right),
\end{aligned}$$

where $\overline{SEL}_{i^{(k)},\theta,t}$ denotes the event that $SEL_{i^{(k)},\theta,t}$ does not occur. Since the probability that it occurs is at least $1 - \varepsilon$ regardless of the value of X_θ, then from (3) we achieve

$$\Pr_{\mathbf{s}}(X_{t+\theta} \notin G_{k+1} \mid X_\theta \in G_k) \le \eta t p + \varepsilon.$$

Lemma 2 *For any NBR-reducible game G, any imperfect best-response dynamics, any starting profile $\mathbf{s}$ and $k = 0, \ldots, \ell$, we have*

$$\Pr_{\mathbf{s}} (X_{kR} \in G_k) \ge 1 - k \cdot (\eta p R + \varepsilon).$$

Proof Observe that

$$\begin{aligned}
\Pr_{\mathbf{s}} (X_{kR} \notin G_k) &\le \Pr_{\mathbf{s}} \left(X_{kR} \notin G_k \mid X_{(k-1)R} \in G_{k-1} \right) + \Pr_{\mathbf{s}} \left(X_{(k-1)R} \notin G_{k-1} \right) \\
&\le \eta p R + \varepsilon + \Pr_{\mathbf{s}} \left(X_{(k-1)R} \notin G_{k-1} \right),
\end{aligned}$$

where the first inequality follows from the definition of conditional probabilities and the last one uses (2). Since $\Pr_{\mathbf{s}}(X_0 \notin G_0) = 0$ the lemma follows by iterating the argument. $\square$

Proof (of Theorem 3) Consider an interval of length $T = R \cdot \left\lceil \frac{\ln(2\ell/\delta)}{\ln(1/\varepsilon)} \right\rceil$. From Observation 1, the probability that all players are selected at least once in an interval of length T is $\frac{\delta}{2\ell}$. That is, every (R, ϵ)-fair schedule is also $(T, \delta/2\ell)$-fair. By applying Lemma 2 with $(R, \varepsilon) = (T, \delta/2\ell)$ we then have that for any NBR-reducible game G, any imperfect best-response dynamics and any starting profile $\mathbf{s}$

$$\Pr_{\mathbf{s}}(X_{\ell T} \in G_\ell) \ge 1 - \ell \cdot (\eta p T + \delta/2\ell).$$

The theorem then follows by taking $p \le \frac{\delta}{2} \cdot \frac{1}{\eta T \ell}$. $\square$

4 Incentive Compatibility Property

In this section we ask if the incentive compatibility property holds also in presence of noise, that is, if deviating from a p-imperfect best-response rule is not beneficial for the player. Note that adopting a p'-imperfect response rule, with $p' < p$, should be not considered a deviation, since this rule is also a p-imperfect response rule.

4.1 A Negative Result

In this section we show that the incentive compatibility condition is not robust against noise and mistakes. Indeed, the following theorem says that, even for arbitrarily small $p > 0$, there are games for which p-imperfect best response dynamics are no longer incentive compatible (though best-response dynamics are incentive compatible according to the main result of [19] – see Theorem 1 above).

Theorem 4 *For any $p > 0$, there exist a NBR-solvable game with clear outcome and a p-imperfect best-response dynamics whose response rule is not incentive compatible.*

Proof Consider the following game G with clear outcome (the gray profile)

	left	right
top	$L+2,1$	$1,0$
bottom	$0,0$	$0,L$

where L is a sufficiently large constant (to be specified below). We examine what happens when both players play according to a particular p-imperfect best response dynamics, namely the logit dynamics with parameter β (see Appendix A.2 for details). It turns out that the column player can improve her utility by playing *always* strategy "right" (thus deviating to a 1-imperfect best response).

We first use a known result by [7] to compute the expected utilities of logit dynamics for this game. The above game is a *potential game* and the potential function[4] Φ is given by the following table:

	left	right
top	$L+2$	$L+1$
bottom	0	L

For potential games, the logit dynamics converges to the following distribution π on the set of profiles:

$$\pi(\mathbf{s}) = \frac{e^{-\beta\Phi(\mathbf{s})}}{\sum_{\mathbf{s}'\in S}e^{-\beta\Phi(\mathbf{s}')}}.$$

---

[4]A game is a *potential game* if there exists a function Φ such that, for all i, for all $\mathbf{s}_{-i}$, and for all s_i, s_i', it holds that $\Phi(s_i, \mathbf{s}_{-i}) - \Phi(s_i', \mathbf{s}_{-i}) = u_i(s_i', \mathbf{s}_{-i}) - u_i(s_i, \mathbf{s}_{-i})$. Such function Φ is called a *potential function* of the game.

Hence, the expected utility of the column player when she plays according to the logit response rule is

$$\frac{1 \cdot e^{\beta(L+2)} + L \cdot e^{\beta L}}{1 + e^{\beta L} + e^{\beta(L+1)} + e^{\beta(L+2)}} < \frac{e^{2\beta} + L}{1 + e^{\beta} + e^{2\beta}}. \tag{4}$$

If instead the column player plays always strategy "right", then her expected payoff is determined by the logit dynamics on the corresponding subgame. The same potential argument above (note that Φ is also a potential for this subgame) says that the column player has expected payoff equal to

$$\frac{L \cdot e^{\beta L}}{e^{\beta L} + e^{\beta(L+1)}} = \frac{L}{1 + e^{\beta}}. \tag{5}$$

Since the right-hand side of (4) is smaller than (5) for $L \geq 1 + e^{\beta}$, the theorem follows from the observation that, in games with two strategies and integer payoffs, logit dynamics are p_{β}-imperfect with $p_{\beta} = \frac{1}{1+e^{\beta}}$ (see Section 2 and Appendix A.2 for details). □

4.2 A Positive Result

As done for convergence, we now investigate for sufficient conditions for incentive compatibility. We will assume that utilities are non-negative: note that there are a lot of response rules that are invariant with respect to the actual value of the utility function and thus, in these cases, this assumption is without loss of generality. Recall that we denote as $i^{(k)}$ and G_k the first occurrence of the player and the corresponding subgame in the elimination sequence given by the definition of game with clear outcome (Definition 7).

It turns out that we need a "quantitative" version of the definition of clear outcome, i.e., that whenever the player i has to eliminate a NBR her utility in the Nash equilibrium is sufficiently larger than the utility of any other profile in the subgame she is actually playing. Specifically, we have the following theorem.

Theorem 5 *For any NBR-solvable game G with clear outcome and any small $\delta > 0$, playing according to a p-imperfect rule is incentive compatible for player $i = i^{(k)}$ as long as $p \leq \frac{c}{\eta R \cdot \ell \ln \ell}$, for a suitable constant $c = c(\delta)$, the dynamics run for at least $r \cdot (R\ell \ln \ell)$, with $r = r(\delta)$ constant, and*

$$u_i(NE) \geq \frac{1}{1 - \delta} \left(\delta \cdot \max(u_i, G) + \max(u_i, G_k) \right),$$

where $u_i(NE)$ is the utility of i in the Nash equilibrium, $\max(u_i, G_k) = \max_{\mathbf{s} \in G_k} u_i(\mathbf{s})$ and $\max(u_i, G) = \max_{\mathbf{s} \in G} u_i(\mathbf{s})$.

We can summarize the intuition behind the proof of Theorem 5 as follows:

– If player i always updates according to the p-imperfect response rule, then the game will be in the Nash equilibrium with high probability and hence her expected utility almost coincides with the Nash equilibrium utility;

- Suppose, instead, player i does not update according to a p-imperfect response rule. Notice that the elimination of strategies up to G_k is not affected by what player i does. Therefore profiles of $G \setminus G_k$ will be played only with small probability (but i can gain the highest possible utility from these profiles), whereas the game will be in a profile of G_k with the remaining probability.

Let us now formalize this idea. We start with the following lemma.

Lemma 3 *For any NBR-reducible game G, any imperfect best-response dynamics, any starting profile* **s**, $k = 0, \ldots, \ell$ *and* $t \geq kR$, *we have*

$$\Pr_{\mathbf{s}} (X_t \in G_k) \geq 1 - \eta p \cdot (t - lkR) - k \cdot (\eta pR + \varepsilon),$$

where l is the largest integer such that $t \geq lkR$.

Proof We have

$$\Pr_{\mathbf{s}} (X_t \notin G_k) \leq \Pr_{\mathbf{s}} (X_t \notin G_k \mid X_{lkR} \in G_k) + \Pr_{\mathbf{s}} (X_{lkR} \notin G_k).$$

From Lemma 1 we have

$$\Pr_{\mathbf{s}} (X_t \notin G_k \mid X_{lkR} \in G_k) \leq \eta p \cdot (t - lkR).$$

For $\mathbf{s}' = X_{(l-1)kR}$ Lemma 2 implies

$$\Pr_{\mathbf{s}} (X_{lkR} \notin G_k) = \Pr_{\mathbf{s}'} (X_{kR} \notin G_k) \leq k \cdot (\eta pR + \varepsilon).$$

$\square$

Remark 3 Observe that Lemma 3 holds even if only players $i^{(1)}, \ldots, i^{(k)}$ are updating according to a p-imperfect response rule.

We then have the following corollary.

Corollary 1 *For any NBR-solvable game G, any imperfect best-response dynamics, any* $\delta > 0$ *and any starting profile* **s**, *if* $t \geq r \cdot (R \cdot \ell \ln \ell)$ *and* $p \leq \frac{c}{\eta R \cdot \ell \ln \ell}$, *for suitable constants* $r = r(\delta)$ *and* $c = c(\delta)$, *then*

$$\Pr_{\mathbf{s}}(X_t \in \hat{G}) \geq 1 - \delta.$$

Proof By using the same approach as in the proof of Theorem 3, we consider an interval T of length $R \cdot \left\lceil \frac{\ln(3\ell/\delta)}{\ln(1/\varepsilon)} \right\rceil$ and $p \leq \frac{\delta}{3} \cdot \frac{1}{\eta T \ell}$. From Observation 1, the schedule of G is $(T, \delta/3\ell)$-fair. By applying Lemma 3 with $k = \ell$ and $(R, \varepsilon) = (T, \delta/3\ell)$ we have that for any NBR-reducible game G, any imperfect best-response dynamics and any starting profile **s**, if $t \geq \ell T = r \cdot (R \cdot \ell \ln \ell)$, where $r = r(\delta) \leq 3 + \frac{\ln 1/\delta}{\ln 1/\varepsilon}$, then

$$\Pr_{\mathbf{s}}(X_t \in \hat{G}) \geq 1 - \frac{\delta}{3T\ell}(t - l\ell T) - \frac{2\delta}{3} \geq 1 - \delta,$$

where we used that, by definition of l in Lemma 3, $t - l\ell T \leq \ell T$.

$\square$

Proof (of Theorem 5) From Corollary 1, the expected utility of i, given that all players are playing according to the p-imperfect response rule, will be at least $(1 - \delta) \cdot u_i(NE)$.

Suppose now that i does not play a p-imperfect response rule. Similarly as done above, we let $T = R \cdot \left\lceil \frac{\ln(3k/\delta)}{\ln(1/\varepsilon)} \right\rceil$ and $p \leq \frac{\delta}{3} \cdot \frac{1}{\eta T \ell}$. Then, by applying Lemma 3 with $(R, \varepsilon) = (T, \delta/2k)$ we obtain

$$\Pr_{\mathbf{s}}(X_t \notin G_k) \leq \delta.$$

Hence, the expected utility of i will be at most $\delta \max(u_i, G) + \max(u_i, G_k)$ and the theorem follows. □

5 NBR-reducible Games

For general games it is not possible to prove convergence to a pure Nash equilibrium without making additional assumptions on the schedule and on the response rule. Such negative result applies also to NBR-reducible games, the natural extension of NBR-solvable ones. However, we shall see below that, for NBR-reducible games, several questions on the dynamics of a game G can be answered by studying the dynamics of the reduced game $\hat{G}$.

5.1 Impossibility Of General Results

At first sight NBR-solvable games may appear as a limited class of games. The class of NBR-reducible games is a natural generalization that can be applied to more settings. Unfortunately, we next show that no general result can be stated about the convergence of p-imperfect best-response dynamics. Specifically, we will show that the system behaves differently not only with respect to which schedule is adopted, but also with respect to how a p-imperfect response rule is implemented.

Same response rule and different schedules. Consider the classical 2-player coordination game:

	0	1
0	1,1	0,0
1	0,0	1,1

Obviously, the game above is not NBR-solvable but it may be the reduced game for a NBR-reducible game. We assume players update according to the best-response rule, but we consider two different schedules: the first one selects just one player randomly at each time step, whereas the second one updates all players at same time. Then, it is easy to see that the dynamics with the first schedule always converges in one of the two Nash equilibria of the game, (0,0) and (1,1), whereas with the second one it can cycle over profiles (1,0) and (0,1) and never converges (see, e.g., Alos-Ferrer and Netzer [1]).

Same schedule and different response rules. Consider the following 2-player game:

	0	1
0	0,0	0,1
1	0,1	1,0

The game has a unique Nash equilibrium, namely (1,0), but it is not NBR-solvable. We consider two different response rules: in both players update according to the best-response rule, but they differ in how handling multiple best-response strategies. The first response rule states that one of these strategies is selected randomly (as happens in logit response rule), whereas the second response rule choose the best response randomly only if the current strategy is not a best response. In these cases, the second response rule adopts a conservative approach and chooses the current strategy (this is exactly the way the mutation model handles multiple best responses). Let us pair these response rules with the schedule that selects just one player randomly at each time step. Then, it is easy to see that the dynamics with the second response rule always converges to the Nash equilibrium and never move from there, whereas with the first one it cycles infinitely over profiles $(1, 0)$, $(0, 0)$, $(0, 1)$ and (1,1) (see e.g. Alos-Ferrer and Netzer [1]).

5.2 Reductions Between Games

We will see now that some of the ideas developed in the previous sections about NBR-solvable games and their pure Nash equilibria can be extended to address questions about NBR-reducible games. In particular, we will see that for a wide class of questions about imperfect best-response dynamics for a NBR-reducible game G, an answer can be given simply by considering a restriction of these dynamics to the reduced game $\hat{G}$.

Before formally stating this fact, let us introduce some useful concepts.

The dynamics as a Markov chain. Recall that, in general, the dynamics is not a Markovian process (for instance, the schedule or the response rule of the players could depend on the history – the past strategy profiles). In order to state our results in full generality, we allow both the schedule and the response rule to depend on what we call the *status* of the dynamics, that is, on a set of information in addition to the current strategy profile.

We say that the dynamics is in a *status–profile pair* $(h, \mathbf{s})$ if h is the set of information currently available and $\mathbf{s}$ is the profile currently played. We denote with H the set of all status–profile pairs $(h, \mathbf{s})$ and with $\hat{H}$ only the ones with $\mathbf{s} \in \hat{G}$. Let X_t be the random variable that represents the status–profile pair $(h, \mathbf{s})$ after t steps of the imperfect best-response dynamics. Then, for every $(h, \mathbf{s}), (h', \mathbf{s}') \in H$ we set

$$P\big((h, \mathbf{s}), (h', \mathbf{s}')\big) = \Pr\Big(X_{i+1} = (h', \mathbf{s}') \mid X_i = (h, \mathbf{s})\Big) = \Pr\Big(X_1 = (h', \mathbf{s}') \mid X_0 = (h, \mathbf{s})\Big).$$

That is, P is the transition matrix of a Markov chain on state space H and it describes exactly the evolution of the dynamics. Note that we are not restricting the dynamics to be memoryless, since in the status we can save the history

of all previous iterations. For a set $A \subseteq H$ we also denote $P\big((h, \mathbf{s}), A\big) = \sum_{(h', \mathbf{s}') \in A} P\big((h, \mathbf{s}), (h', \mathbf{s}')\big)$.

Remark 4 Observe that, for any starting status-profile $(h, \mathbf{s}) \in H$, if we consider only the profile component of each X_t, this describes a particular imperfect best response dynamics (every h gives a different dynamics) starting at $\mathbf{s}$. Denoting by $X_t^{(h)}$ the random variable corresponding to this dynamics (now $X_t^{(h)}$ is a strategy profile), we can import prior bounds via the following useful inequality:

$$\Pr\big(X_t = (h', \mathbf{s}') \mid X_0 = (h, \mathbf{s})\big) \leq \sum_{(h'', \mathbf{s}') \in H} \Pr\big(X_t = (h'', \mathbf{s}') \mid X_0 = (h, \mathbf{s})\big)$$

$$= \Pr_{\mathbf{s}}\big(X_t^{(h)} = \mathbf{s}'\big).$$

Clearly, the response rule and the schedule of the imperfect best-response dynamics described by $X_t^{(h)}$ inherit the same parameters of the original dynamics X_t.

The restricted dynamics. As mentioned above, we will compare the original dynamics with a specific restriction on the subset $\hat{H}$ of status–profile pairs. Now we describe how this restriction is obtained. Henceforth, whenever we refer to the restricted dynamics, we use $\hat{X}_t$ and $\hat{P}$ in place of X_t and P. Then, the restricted dynamics is described by a Markov chain on state space H with transition matrix $\hat{P}$ such that for every $(h, \mathbf{s}), (h', \mathbf{s}') \in H$

$$\hat{P}\big((h, \mathbf{s}), (h', \mathbf{s}')\big) = \begin{cases} \frac{P((h,\mathbf{s}),(h',\mathbf{s}'))}{P((h,\mathbf{s}),\hat{H})}, & \text{if } (h, \mathbf{s}), (h', \mathbf{s}') \in \hat{H}; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the restricted dynamics is exactly the same as the original one except that the first never leaves the subgame $\hat{G}$, whereas in the latter, at each time step, there is probability at most p to leave this subgame. The following lemma quantifies this similarity, by showing that, for every $(h, \mathbf{s}) \in \hat{H}$, the *total variation distance* (TV)[5] between the original and the restricted dynamics starting from $(h, \mathbf{s})$ is small.

Lemma 4 *For any NBR-reducible game G, any imperfect best-response dynamics, every $(h, \mathbf{s}) \in \hat{H}$ and any $t \geq 0$,*

$$\big\| P^t\big((h, \mathbf{s}), \cdot\big) - \hat{P}^t\big((h, \mathbf{s}), \cdot\big) \big\| \leq \eta p t. \tag{6}$$

Proof First observe that, since $(h, \mathbf{s}) \in \hat{H}$, it holds that

$$\sum_{(h', \mathbf{s}') \in H \setminus \hat{H}} P((h, \mathbf{s}), (h', \mathbf{s}')) \leq \sum_{(\hat{h}, \hat{\mathbf{s}}) \in \hat{H}} \sum_{(h', \mathbf{s}') \in H \setminus \hat{H}} P((\hat{h}, \hat{\mathbf{s}}), (h', \mathbf{s}')) \tag{7}$$

$$\leq \Pr_{\mathbf{s}}\big(X_1^{(h)} \notin \hat{G} \mid X_0^{(h)} \in \hat{G}\big) \leq \eta p,$$

---

[5]See Appendix B for a review of the main properties of the total variation distance.

where the two inequalities follow from Remark 4 and Lemma 1, respectively. We next prove (6) by induction on t. The base case is $t = 1$ for which the set of status–profile pairs $(h', \mathbf{s}')$ such that $P((h, \mathbf{s}), (h', \mathbf{s}')) > \hat{P}((h, \mathbf{s}), (h', \mathbf{s}'))$ is exactly $H \setminus \hat{H}$ and thus

$$\left\| P\big((h, \mathbf{s}), \cdot\big) - \hat{P}\big((h, \mathbf{s}), \cdot\big) \right\| = \sum_{(h', \mathbf{s}') \in H \setminus \hat{H}} \Big(P((h, \mathbf{s}), (h', \mathbf{s}')) - \hat{P}((h, \mathbf{s}), (h', \mathbf{s}')) \Big)$$

$$= \sum_{(h', \mathbf{s}') \in H \setminus \hat{H}} P((h, \mathbf{s}), (h', \mathbf{s}')) \leq \eta p,$$

where the last inequality follows from (7). As for the inductive step, we have

$$\left\| P^t\big((h, \mathbf{s}), \cdot\big) - \hat{P}^t\big((h, \mathbf{s}), \cdot\big) \right\|$$

$$\leq \left\| P\big((h, \mathbf{s}), \cdot\big) P^{t-1} - \hat{P}\big((h, \mathbf{s}), \cdot\big) P^{t-1} \right\| + \left\| \hat{P}\big((h, \mathbf{s}), \cdot\big) P^{t-1} - \hat{P}\big((h, \mathbf{s}), \cdot\big) \hat{P}^{t-1} \right\|$$

$$\leq \left\| P\big((h, \mathbf{s}), \cdot\big) - \hat{P}\big((h, \mathbf{s}), \cdot\big) \right\| + \sup_{(h', \mathbf{s}') \in \hat{H}} \left\| P^{t-1}\big((h', \mathbf{s}'), \cdot\big) - \hat{P}^{t-1}\big((h', \mathbf{s}'), \cdot\big) \right\|$$

$$\leq \eta p + \eta p (t - 1) = \eta p t,$$

where the first and second inequality follows from the triangle inequality and monotonicity of the total variation distance, respectively, whereas the last inequality is by induction. $\qquad\square$

Status–profile events. We now describe the kind of questions about imperfect best-response mechanisms and NBR-reducible games for which a reduction can be beneficial. Roughly speaking, these are all questions about the occurrence (and the time needed for it) of events that can be described only by looking at status–profile pairs.

Specifically, a *status–profile set event* for an imperfect best-response dynamics is a set of status–profile pairs. A *status–profile distro event* for an imperfect best-response dynamics is a distribution on the status–profile pairs. More generally, we refer to *status–profile event* if we do not care whether it is a set or a distro event. Note that many equilibrium concepts can be described as status–profile events, like Nash equilibria, sink equilibria [12], correlated equilibria [6] or logit equilibria [4]: in any case we should simply list the set of states or the distribution over states at which we are interested in. Properties like "a profile that is visited for k times" or "a cycle of length k visited" are other examples of status–profile events. We remark that in these examples it is crucial that the equilibrium is defined on the status–profile pairs and not just on the profiles: indeed, the status can remember the history of the game and identify such events, whereas they are impossible to recognize if we only know the current profile.

For an NBR-reducible game G, a status–profile set event is *reducible* if the set of status–profile pairs that represent the event contains at least one profile from $\hat{G}$. A status–profile distro event is *reducible* if status–profile pairs on which is defined the distribution that represents the event contains only profiles of $\hat{G}$. It turns out that each one of the equilibria concepts described above is a reducible status-profile event: indeed, since all profiles not in $\hat{G}$ contain NBR strategies, they are not in the support

of any Nash, any sink and any correlated equilibrium; as for the logit equilibrium (that assigns non-zero probability to profiles not in $\hat{G}$) it is not difficult to show (see Appendix C) that the logit equilibrium of G is close to the logit equilibrium of $\hat{G}$.

A status–profile set event *occurs* if the imperfect best-response dynamics reaches a status–profile pair in the set of pairs that represent the event. Similarly, a status–profile distro event *occurs* if the distribution on the set of profiles generated by the dynamics is close to the one that represent the event. The *occurrence time* of a status–profile event is the first time step in which it occurs.

We are now in a position to state the main result on NBR-reducible games:

Theorem 6 *For any NBR-reducible game G and any small $\delta > 0$, if a reducible status–profile event for an imperfect best-response dynamics occurs in the restricted dynamics, then it occurs with probability at least $1 - \delta$. Moreover, let us denote with τ the occurrence time of the event E in the restricted dynamics. Then, E occurs in the original dynamics in $O(R \cdot \ell \ln \ell + \tau)$ steps with probability at least $1 - \delta$, whenever $p \leq \min\left\{\frac{c_1}{\eta R \cdot \ell \ln \ell}, \frac{c_2}{\eta \tau}\right\}$, for suitable constants $c_1 = c_1(\delta)$ and $c_2 = c_2(\delta)$.*

Proof We will show that the dynamics will be in $\hat{H}$ after $O(R \cdot \ell \ln \ell)$ with probability at least $1 - \delta/2$; moreover, if the dynamics is in $\hat{H}$ after t steps, then a reducible status–profile event occurs in further τ steps with probability at least $1 - \delta/2$. Hence, the probability that the event does not occurs in $O(R \cdot \ell \ln \ell + \tau)$ steps will be at most δ and the theorem follows.

Specifically, consider an interval T of length $R \cdot \left\lceil \frac{\ln(4\ell/\delta)}{\ln(1/\varepsilon)} \right\rceil$. From Remark 4 and by applying Lemma 2 with $k = \ell$, $(R, \varepsilon) = (T, \delta/4\ell)$ and $p \leq \frac{\delta}{4} \cdot \frac{1}{\eta T \ell}$ we have that for every $(h, \mathbf{s}) \in H$

$$\Pr\left(X_{\ell T} \notin \hat{H} \mid X_0 = (h, \mathbf{s})\right) \leq \Pr_{\mathbf{s}}\left(X_{\ell T}^{(h)} \notin \hat{G}\right) \leq \delta/2.$$

Finally, note that the probability that, for every $t > 0$, a reducible status–profile event occurs in $t + \tau$ steps given that after t steps it is in $(h', \mathbf{s}') \in \hat{H}$, is the same as if we assume the dynamics starts in $(h', \mathbf{s}')$, i.e., it is equivalent to the probability that the event occurs in τ steps from $(h', \mathbf{s}')$. If the event E is a distro event, i.e., the restricted dynamics converges after τ steps to a distribution π on the status–profile pairs, then, from (6), the distribution after τ steps of the original dynamics is π except for an amount of probability of at most $\eta p \tau$. On the other hand, if the event E is a set event, i.e. the restricted dynamics converges after τ steps to a set A of status–profile pairs, then, from (6) we have

$$\Pr\left(X_\tau \in A\right) \geq \Pr\left(\hat{X}_\tau \in A\right) - \mu p \tau = 1 - \mu p \tau,$$

and hence, after τ steps, the original dynamics is in A except with probability at most $\mu p \tau$. Then, by Lemma 4 and by taking $p \leq \frac{\delta}{2} \cdot \frac{1}{\eta \tau}$, the probability that the event occurs in τ steps for the original dynamics starting from $(h', \mathbf{s}')$ is at least $1 - \delta/2$. $\square$

Example 3 (*interference games*) The following is an instance of *interference games* [5] which model interference in wireless networks according to the "Signal to Interference plus Noise Ratio" (SINR) physical model. Formally, the game has the following utilities:

$$u_i(\mathbf{s}) = \begin{cases} v_i - s_i, & \text{if } s_i > \sum_{j \neq i} s_j; \\ -s_i, & \text{otherwise}; \end{cases}$$

where v_i is some value assigned to each player (representing the utility if the transmission succeeds), strategy s_i is the energy spent for transmitting, and the condition above indicates whether i's transmission is successful received or not. For two players with $v_1 = v_2 = 1 + \gamma$, $0 < \gamma < 1$, and $S_i = \{0, 1, 2, 3\}$ the payoffs are

	0	1	2	3
0	0,0	0,γ	0, $\gamma - 1$	0, $\gamma - 2$
1	γ,0	-1,-1	-1,$\gamma - 1$	-1,$\gamma - 2$
2	$\gamma - 1$,0	$\gamma - 1$,-1	-2,-2	-2,$\gamma - 2$
3	$\gamma - 2$,0	$\gamma - 2$,-1	$\gamma - 2$,-2	-3,-3

which is NBR-reducible to this subgame:

	0	1
0	0,0	0,γ
1	γ,0	-1,-1

Note that this is a potential game, with potential Φ given by the following table:

	0	1
0	0	$-\gamma$
1	$-\gamma$	$1 - \gamma$

The logit dynamics with parameter β on this subgame converges to the following distribution $\hat{\pi}$:

	0	1
0	$1/Z$	$e^{\beta\gamma}/Z$
1	$e^{\beta\gamma}/Z$	$e^{\beta(\gamma-1)}/Z$

where $Z = 1 + 2e^{\beta\gamma} + e^{\beta(\gamma-1)}$. This means that, for growing values of β, the dynamics "converges" to the two Nash equilibria of the game (the stationary probability of profiles (0,1) and (1,0) increase whereas the probability of the other profiles diminishes).

According to Theorem 6, the logit dynamics with parameter β on the original interference game can be analyzed by looking at the dynamics on the subgame only, provided p being small enough. (Note that in general both p and τ in Theorem 6 depend of β.) For instance, taking the event $E =$ "the dynamics is in one of the two PNE", τ is the hitting time of such PNE's and, for β large enough, the original dynamics converges to a PNE in $O(R + \tau)$ time steps.

The reader may have noticed that, in the previous example, the parameters are set so as to obtain a subgame which is a potential game. Indeed, the class of interference games [5] does not posses pure Nash equilibria in general. For instance, setting $v_1 = v_2 = 3$ and $S_i = \{0, 1, \ldots, 4\}$ would give an interference game which is NBR-reducible to a subgame containing a sink equilibrium [12] (but no pure Nash

equilibrium). Though the subgame is not a potential game, Theorem 6 suggests a natural way to tackle the problem by analyzing the sink equilibria of the subgame (e.g., the time for the restricted dynamics to reach such an equilibrium).

We conclude this section by observing that Theorem 6 is similar in spirit to the results by Candogan et al. [8]. Both results aim to link the behavior of dynamics for a game G to the behavior of the dynamics for a close game $\hat{G}$, for which the dynamics is well-understood. In [8] are considered games that are "close" to potential games and it has been showed that the behavior of best and better response dynamics, logit dynamics and fictitious play can be approximated by the equilibrium behavior of these dynamics on the close potential games. The extent at which this approximation is good depends on the distance between the games and on the number of profiles of these games (that is exponential on the number of players). Our results instead refer to generic imperfect response dynamics (and thus they hold for best response and logit dynamics but not for better response dynamics and fictitious play), but we consider only a very specific "closeness measure", namely that the game G is NBR-reducible to the game $\hat{G}$. On the other side, the game $\hat{G}$ can approximate G arbitrarily well, not only with respect to the equilibrium behavior, but also with respect to the transient behavior of the dynamics.

6 Conclusions

This work addresses how a (small) probability of selecting a non-best response by the players can affect the convergence and incentive compatibility properties of a class of dynamics and mechanisms in [19]. We have first shown that there are games for which convergence does not occur even for a probability p *exponentially* small in the number of players (Theorem 2). Our positive result (Theorem 3) says that convergence can be achieved if p is "sufficiently small" compared to certain parameters of the game and of the schedule (the number ℓ of rounds needed in the original dynamics without mistakes and the length R of a "probabilistic round" of the schedule – see Definition 3). Note that, in this case, the convergence time of the imperfect best-response dynamics is only moderately larger than the convergence time of the perfect best-response dynamics: When the latter is guaranteed to converge in $R\ell$ time steps (Theorem 1) the former converges in $O(R\ell \ln \ell)$ time steps with good probability (Theorem 3).

As for incentive compatibility, we showed that imperfect best-response dynamics require a stronger condition than the one sufficient for perfect best-response dynamics (cf. Theorems 1 and 4). The sufficient condition (Theorem 5) is essentially a quantitative version of the condition in [19] (clear outcome) which requires the payoffs at the equilibrium to be "sufficiently" larger than at other profiles.

Finally, we suggest a natural extension to games which are not NBR-solvable, and the elimination of NBR strategies yields only a subgame. In such cases, it is natural to consider the dynamics of the subgame as a good approximation of the dynamics of the original game. Theorem 6 gives a quantitative bound on the time to reach a certain "event" as the sum of two quantities: The time for the dynamics to converge to the subgame plus the time the dynamics restricted to the subgame takes to reach the

same event. We feel Theorem 6 might have several applications to games which are reducible to a subgame whose structure makes the analysis of the dynamics easier. For instance, if the subgame is a potential game, logit dynamics of the original game (even if this is not a potential game) can be accurately described by the logit dynamics of the subgame which is known to have an explicit simple stationary distribution [7].[6] Also, when the subgame $\hat{G}$ is *not* a potential game, it might be possible to combine our results with those in [8] in a two-steps analysis: Reduce the game G to a subgame $\hat{G}$ and then study this game by considering the "closest" potential game $\tilde{G}$. More in general, the advantage of reducing the analysis to a subgame $\hat{G}$ is that the resulting dynamics can be simpler to analyze using known techniques. For instance, for some potential games, it is possible to analyze variants of the logit dynamics that differ in the schedule of the players [1, 2].

Appendix A: Models for limited knowledge and bounded rationality

A.1 Mutation and mistakes models

The *mutation* and the *mistakes* models adopt the same response rule: at every time step, each selected player updates her strategy to the best response to what other players are currently doing except with probability ε. With such a probability, a *mutation* or *mistake* occurs, meaning that the selected player chooses a strategy uniformly at random. That is, suppose player i is selected at time step t and the current strategy profile is $\mathbf{s}^t$. We denote with $b_i(\mathbf{s}^t)$ the best response of i to profile $\mathbf{s}^t$ (if more than one best response exists and the current strategy x_i^t of i is a best response, then we set $b_i(\mathbf{s}^t) = x_i^t$, otherwise we choose one of the best responses uniformly at random). Then, a strategy $s_j \in S_i$ will be selected by i with probability

$$p_{ij} = \begin{cases} (1 - \varepsilon) + \varepsilon \cdot \frac{1}{|S_i|}, & \text{if } s_j = b_i(\mathbf{s}^t); \\ \varepsilon \cdot \frac{1}{|S_i|}, & \text{otherwise.} \end{cases}$$

The main difference between these models concerns the schedule: the mutation model assumes that at each time step every player is selected for update; the mistakes model assumes that at each time step only one player is selected uniformly at random for update.

---

[6]In the conference version of this work [9], we claimed that in a modified version of PageRank games [14], there exists a subgame which is a potential game and thus our results can be combined with [7] to obtain a good approximation of the logit dynamics for these games. Unfortunately, this claim was wrong and the logit dynamics for this subgame is in general not easy to analyze.

A.2 Logit dynamics

Logit dynamics is another kind of imperfect best response dynamics in which the probability of deviating from best response is determined by a parameter $\beta \geq 0$. The *logit dynamics* for a game G with parameter β runs as follows. At every time step

1. Select one player i uniformly at random;
2. Update the strategy of player i according to the following probability distribution:

$$\sigma_i(s_i, \mathbf{s}_{-i}) = \frac{e^{\beta u_i(s_i, \mathbf{s}_{-i})}}{Z(\mathbf{s}_{-i})},$$

where $Z(\mathbf{s}_{-i}) = \sum_{s_i'} e^{\beta u_i(s_i', \mathbf{s}_{-i})}$ and $\beta \geq 0$.

Remark 5 The parameter β is sometimes called the "inverse noise" parameter: for $\beta = 0$ the player chooses the strategy uniformly at random, while for $\beta \to \infty$ the player chooses a best response (if more than one best response is available, she selects one of these uniformly at random). Thus, the probability that i plays a best response is guaranteed to be least $1/|S_i|$. So, every logit dynamics is p-imperfect for some $p < 1$. Moreover, if the payoffs between a best response and a non-best response differ by at least γ, then the dynamics is p-imperfect with $p \leq (m-1)/(m-1+e^{\gamma\beta})$.

The above dynamics defines an ergodic Markov chain [7] over the set of all strategy profiles and with transition probabilities P given by

$$P(\mathbf{s}, \mathbf{s}') = \frac{1}{n} \cdot \begin{cases} \sigma_i(s_i, \mathbf{s}_{-i}), & \text{if } \mathbf{s}_{-i} = \mathbf{s}'_{-i} \text{ and } s_i \neq s_i'; \\ \sum_{i=1}^{n} \sigma_i(s_i, \mathbf{s}_{-i}), & \text{if } \mathbf{s} = \mathbf{s}'; \\ 0, & \text{otherwise.} \end{cases}$$

Since the Markov chain is ergodic, the chain converges to its (unique) *stationary distribution* π. That is, for any starting profile $\mathbf{s}$ and any profile $\mathbf{s}'$,

$$\lim_{t \to \infty} P^t(\mathbf{s}, \mathbf{s}') = \pi(\mathbf{s}'),$$

where $P^t(\mathbf{s}, \mathbf{s}')$ is the probability that the dynamics starting with profile $\mathbf{s}$ is in profile $\mathbf{s}'$ after t steps. Thus, the stationary distribution represents the equilibrium reached by the logit dynamics, also called the *logit equilibrium* of game G [4].

The stationary distribution of logit dynamics is fully understood for the class of so-called *potential games*. We recall that a game is a potential game if there exists a function Φ such that, for all i, for all $\mathbf{s}_{-i}$, and for all s_i, s_i', it holds that

$$\Phi(s_i, \mathbf{s}_{-i}) - \Phi(s_i', \mathbf{s}_{-i}) = u_i(s_i', \mathbf{s}_{-i}) - u_i(s_i, \mathbf{s}_{-i}).$$

Blume [7] showed that in this case, the stationary distribution of the corresponding logit dynamics is

$$\pi(\mathbf{s}) = \frac{e^{-\beta\Phi(\mathbf{s})}}{Z},$$

where $Z = \sum_{\mathbf{s}' \in S} e^{-\beta\Phi(\mathbf{s}')}$ is the normalizing constant.

Finally, the time to converge to the logit equilibrium is the so-called *mixing time* of the corresponding Markov chain [4], defined as follows. For any $\epsilon > 0$, consider

$$t_{mix}(\epsilon) := \min_{t \in \mathbb{N}} \max_{\mathbf{s} \in \Omega} \left\{ \left\| P^t(\mathbf{s}, \cdot) - \pi \right\| \leq \epsilon \right\},$$

where $\left\| P^t(\mathbf{s}, \cdot) - \pi \right\| = \frac{1}{2} \sum_{\mathbf{s}'' \in \Omega} |P^t(\mathbf{s}, \mathbf{s}'') - \pi(\mathbf{s}'')|$ is the total variation distance (see Appendix B for more details). This quantity measures the time needed for the chain (dynamics) to get close to stationary within an additive factor of ϵ. The *mixing time* of the chain is simply defined as $t_{mix} := t_{mix}(1/2)$, since it is well-known that $t_{mix}(\epsilon) \leq t_{mix} \cdot \log_2(1/\epsilon)$ for any ϵ.

Appendix B: Total variation distance

The *total variation distance* between distributions μ and $\hat{\mu}$ on an enumerable state space Ω is

$$\| \mu - \hat{\mu} \| := \frac{1}{2} \sum_{x \in \Omega} | \mu(x) - \hat{\mu}(x)| = \sum_{\substack{x \in \Omega \\ \mu(x) > \hat{\mu}(x)}} \mu(x) - \hat{\mu}(x).$$

Note that the total variation distance satisfies the usual triangle inequality of distance measures, i.e.,

$$\| \mu - \hat{\mu} \| \leq \| \mu - \mu' \| + \| \mu' - \hat{\mu} \|,$$

for any distributions μ and μ'. Moreover, the following monotonicity properties hold:

$$\| \mu P - \mu \hat{P} \| \leq \| \mu - \hat{\mu} \|, \tag{8}$$

$$\| \mu P - \mu \hat{P} \| \leq \sup_{x \in \Omega} \| P(x, \cdot) - \hat{P}(x, \cdot) \|, \tag{9}$$

$$\| \mu P - \hat{\mu} P \| \leq \sup_{x, y \in \Omega} \| P(x, \cdot) - P(y, \cdot) \|, \tag{10}$$

where P and $\hat{P}$ are stochastic matrices. Indeed, as for (8) we have

$$
\begin{aligned}
\| \mu P - \hat{\mu} P \| = \| (\mu - \hat{\mu}) P \| &= \frac{1}{2} \sum_{y \in \Omega} \left| \sum_{x \in \Omega} (\mu(x) - \hat{\mu}(x)) P(x, y) \right| \\
&\leq \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \hat{\mu}(x)| \sum_{y \in \Omega} P(x, y) \\
&= \| \mu - \hat{\mu} \|.
\end{aligned}
$$

As for (9) we observe that

$$\|\mu P - \mu \hat{P}\| = \|\mu(P - \hat{P})\| = \frac{1}{2} \sum_{y \in \Omega} \left| \sum_{x \in \Omega} \mu(x)(P(x, y) - \hat{P}(x, y)) \right|$$

$$\leq \sum_{y \in \Omega} \left(\frac{1}{2} \sum_{x \in \Omega} \mu(x) \left| P(x, y) - \hat{P}(x, y) \right| \right)$$

$$= \sum_{x \in \Omega} \mu(x) \left(\frac{1}{2} \sum_{y \in \Omega} \left| P(x, y) - \hat{P}(x, y) \right| \right)$$

$$\leq \sup_{x \in \Omega} \|P(x, \cdot) - \hat{P}(x, \cdot)\|.$$

Finally, for (10) we have

$$\|\mu P - \hat{\mu} P\| = \left\| \sum_{z \in \Omega} \mu(z) \sum_{w \in \Omega} \hat{\mu}(w) \left(P(z, \cdot) - P(w, \cdot) \right) \right\|$$

$$\leq \sum_{z \in \Omega} \mu(z) \sum_{w \in \Omega} \hat{\mu}(w) \|P(z, \cdot) - P(w, \cdot)\|$$

$$\leq \sup_{x, y \in \Omega} \|P(x, \cdot) - P(y, \cdot)\|.$$

Appendix C: Equilibria of logit dynamics and NBR-reducible games

In this section we specialize the approach described in Section 5.2 to the case of logit dynamics. First of all, the status "h" is immaterial in this case since the dynamics is by definition Markovian (i.e., the transition probabilities depend only on the current profile).

We next provide a sufficient condition for which the (equilibrium of) the logit dynamics of the subgame is a good approximation of the (equilibrium of) the logit dynamics of the original game. Let G be a game NBR-reducible to $\hat{G}$, and let π_β and $\hat{\pi}_\beta$ denote the stationary distributions of the logit dynamics with parameter β for G and $\hat{G}$, respectively. The following lemma says that π_β and $\hat{\pi}_\beta$ are close to each other (in total variation) if β is large enough.

Lemma 5 *Let $\hat{\tau}_\beta$ be the mixing time of the restricted chain given by the logit dynamics with parameter β, and let $p = p_\beta$ be the corresponding probability of selecting a NBR strategy. If $\lim_{\beta \to \infty}(p_\beta \hat{\tau}_\beta) = 0$, then for every $\delta > 0$, there exists a constant β_δ such that*

$$\|\pi_\beta - \hat{\pi}_\beta\| \leq \delta$$

for all $\beta \geq \beta_\delta$.

Proof Let $\tau = \hat{t}_{mix}^{(\beta)}(\delta/8)$ be the mixing time of the restricted chain. Consider first two copies of the chain starting in profiles $\hat{s}, \hat{s}' \in \hat{G}$ and bound the total variation

after τ time steps:

$$\begin{aligned}
\|P^\tau(\hat{\mathbf{s}}, \cdot) - P^\tau(\hat{\mathbf{s}}'\|, \cdot) &\leq \|P^\tau(\hat{\mathbf{s}}, \cdot) - \hat{P}^\tau(\hat{\mathbf{s}}\|, \cdot) + \|\hat{P}^\tau(\hat{\mathbf{s}}, \cdot) - \hat{\pi}\| \\
&\quad + \|\hat{\pi} - \hat{P}^\tau(\hat{\mathbf{s}}', \cdot)\| + \|\hat{P}^\tau(\hat{\mathbf{s}}', \cdot) - P^\tau(\hat{\mathbf{s}}', \cdot)\| \\
&\leq 4 \cdot \frac{\delta}{8} = \delta/2,
\end{aligned}$$

where the last inequality is due to Lemma 4 by taking β sufficiently large (note that $\eta p \tau = \eta p_\beta t_{mix}^{(\beta)}(\delta/8) \leq \eta p_\beta \hat{\tau}_\beta \ln(8/\delta)$, which tends to 0 as $\beta \to \infty$ by hypothesis). Consider an interval T of length $R \cdot \left\lceil \frac{\ln(8\ell/\delta)}{\ln(1/\varepsilon)} \right\rceil$. By applying Lemma 2 with $k = \ell$, $(R, \varepsilon) = (T, \delta/4\ell)$ and β sufficiently large we have that for every $\mathbf{s} \in G$

$$\Pr_{\mathbf{s}}(X_{\ell T} \notin \hat{G}) \leq \delta/8.$$

Let $t^\star = \ell T + \tau$ and $Q = P^{\ell T}$. Then, for every $\mathbf{s}, \mathbf{s}' \in G$

$$\begin{aligned}
\|\pi - P^{t^\star}(\mathbf{s}', \cdot)\| &\leq \|P^{t^\star}(\mathbf{s}, \cdot) - P^{t^\star}(\mathbf{s}', \cdot)\| = \|Q(\mathbf{s}, \cdot)P^\tau - Q(\mathbf{s}', \cdot)P^\tau\| \\
\text{(triangle inequality)} &\leq \|Q(\mathbf{s}, \cdot)P^\tau - \hat{Q}(\mathbf{s}, \cdot)P^\tau\| + \|\hat{Q}(\mathbf{s}, \cdot)P^\tau - \hat{Q}(\mathbf{s}', \cdot)P^\tau\| \\
&\quad + \|\hat{Q}(\mathbf{s}', \cdot)P^\tau - Q(\mathbf{s}', \cdot)P^\tau\|,
\end{aligned}$$

where, for every $\mathbf{s}, \mathbf{s}' \in G$, we set

$$\hat{Q}(\mathbf{s}, \mathbf{s}') = \begin{cases} \frac{Q(\mathbf{s}, \mathbf{s}')}{Q(\mathbf{s}, \hat{G})}, & \text{if } \mathbf{s}, \mathbf{s}' \in \hat{G}; \\ 0, & \text{otherwise.} \end{cases}$$

By (8) we obtain

$$\|Q(\mathbf{s}, \cdot)P^\tau - \hat{Q}(\mathbf{s}, \cdot)P^\tau\| \leq \|Q(\mathbf{s}, \cdot) - \hat{Q}(\mathbf{s}, \cdot)\| \leq \Pr_{\mathbf{s}}\left(X_{\ell T} \notin \hat{G}\right) \leq \delta/8.$$

By (10) we obtain

$$\|\hat{Q}(\mathbf{s}, \cdot)P^\tau - \hat{Q}(\mathbf{s}', \cdot)P^\tau\| \leq \max_{\hat{\mathbf{s}}, \hat{\mathbf{s}}' \in \hat{G}} \|P^\tau(\hat{\mathbf{s}}, \cdot) - P^\tau(\hat{\mathbf{s}}', \cdot)\| \leq \delta/2$$

and thus $\|\pi - P^{t^\star}(\mathbf{s}', \cdot)\| \leq 3\delta/4$. Finally, for every $\hat{\mathbf{s}} \in \hat{G}$, by triangle inequality

$$\begin{aligned}
\|\pi - \hat{\pi}\| &\leq \|\pi - P^{t^\star}(\hat{\mathbf{s}}, \cdot)\| + \left\|P^{t^\star}(\hat{\mathbf{s}}, \cdot) - \hat{P}^{t^\star}(\hat{\mathbf{s}}, \cdot)\right\| + \left\|\hat{P}^{t^\star}(\hat{\mathbf{s}}, \cdot) - \hat{\pi}\right\| \\
&\leq 3\delta/4 + \delta/8 + \delta/8 = \delta.
\end{aligned}$$

$\square$

References

1. Alós-Ferrer, C., Netzer, N.: The logit-response dynamics. Games Econ. Behav. **68**(2), 413–427 (2010)
2. Auletta, V., Ferraioli, D., Pasquale, F., Penna, P., Persiano, G.: Convergence to equilibrium of logit dynamics for strategic games. In: Proceedings of the 23rd Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), pp. 197–206. ACM (2011)
3. Auletta, V., Ferraioli, D., Pasquale, F., Penna, P., Persiano, G.: Logit dynamics with concurrent updates for local interaction games. In: Proceedings of the 21st Annual European Symposium on Algorithms (ESA), vol. 8125 of LNCS, pp. 73–84 (2013)

4. Auletta, V., Ferraioli, D., Pasquale, F., Persiano, G.: Mixing time and stationary expected social welfare of logit dynamics. Theory Comput. Syst. **53**(1), 3–40 (2013)
5. Auletta, V., Moscardelli, L., Penna, P., Persiano, G.: Interference games in wireless networks. In: Proceedings of the 4th International Workshop On Internet And Network Economics (WINE), vol. 5385 of LNCS, pp. 278–285 (2008)
6. Aumann, R.J.: Subjectivity and correlation in randomized strategies. J. Math. Econ. **1**(1), 67–96 (1974)
7. Blume, L.E.: The statistical mechanics of strategic interaction. Games Econ. Behav. **5**, 387–424 (1993)
8. Candogan, O., Ozdaglar, A., Parrilo, P.A.: Near-potential Games : Geometry and dynamics. ACM Trans. Econ. Comput. **1**(2), 11:1–11:32 (2013)
9. Ferraioli, D., Penna, P.: Imperfect best-response mechanisms. In: Proceedings of the 6th International Symposium on Algorithmic Game Theory (SAGT), pp. 243–254 (2013)
10. Gao, L., Rexford, J.: Stable internet routing without global coordination. IEEE/ACM Trans. Netw. **9**(6), 681–692 (2001)
11. Godfrey, B., Schapira, M., Zohar, A., Shenker, S.: Incentive compatibility and dynamics of congestion control. In: Proceedings of the International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS), pp. 95–106 (2010)
12. Goemans, M., Mirrokni, V., Vetta, A.: Sink equilibria and convergence. In: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 142–154. IEEE (2005)
13. Harsanyi, J.C., Selten, R.: A General Theory of Equilibrium Selection in Games. MIT Press, Cambridge (1988)
14. Hopcroft, J., Sheldon, D.: Network reputation games. Computing and Information Science Technical Reports, Cornell University. (2008). http://hdl.handle.net/1813/11579
15. Kandori, M., Mailath, G.J., Rob, R.: Learning, mutation, and long run equilibria in games. Econometrica **61**(1), 29–56 (1993)
16. Kandori, M., Rob, R.: Evolution of equilibria in the long run: a general theory and applications. J. Econ. Theory **65**(2), 383–414 (1995)
17. Levin, H., Schapira, M., Zohar, A.: Interdomain routing and games. SIAM J. Comput. **40**(6), 1892–1912 (2011)
18. Nisan, N., Schapira, M., Valiant, G., Zohar, A.: Best-response auctions. In: Proceedings of the 12th ACM Conference on Electronic Commerce (EC), pp. 351–360 (2011)
19. Nisan, N., Schapira, M., Valiant, G., Zohar, A.: Best-response mechanisms. In: Proceedings of the 2nd Symposium on Innovations in Computer Science (ICS), pp. 155–165 (2011)
20. Hobart Peyton Young: The evolution of conventions. Econometrica **61**(1), 57–84 (1993)