Imperfect best-response mechanisms

Diodato Ferraioli* Paolo Penna†

Abstract

Best-response mechanisms (Nisan, Schapira, Valiant, Zohar, 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".

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 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. [NSVZ11b]. 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. [NSVZ11b] 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. [NSVZ11b] 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

PSL,Universit[Pleaseinsert\PrerenderUnicode{Ãľ}intopreamble] Paris Dauphine, Lamsade and CNRS UMR 7243

†Institute of Theoretical Computer Science, ETH Zurich

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) in [NSVZ11b] 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 8) 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 [GR01]).

  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 in which convergence results *must consider* how players are scheduled. Indeed, we complement the negative result above with a general positive result (Theorem 10) saying that convergence can be guarantee whenever p is polynomially small in some parameters defining both the game and 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 13). 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 14) turns out to be a quantitative version of the one given in [NSVZ11b]. 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 [NSVZ11b].

Our focus is on the same class of (base) games of [NSVZ11b] 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 [GR01]) leads to games in this class and, ultimately, to the fact that BGP converges and is incentive compatible [LSZ11, NSVZ11b]. 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 [NSVZ11b] (and this work) to a natural generalization of their games. Intuitively speaking, these games guarantee only that best-response converge to a *sub-game*. In this case, the dynamics of the original game can be approximated by the dynamics of the sub-game (Theorem 17). Unfortunately, this "reduction" cannot be pushed further simply because the sub-game 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 sub-game are well-understood, then we can infer their behavior also on the original game. One applications of this approach is on PageRank games which turn out to be reducible to a sub-game for which the well-studied logit equilibrium has a closed formula (being the sub-game a potential game).

Related works. Convergence of best-response dynamics is a main topic in game theory. It relates to the so-called problem of equilibrium selection (how can the players converge to an equilibrium?). Noisy versions of such 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. [KMR93], and to the *mistakes model* by Young [You93], and Kandori and Rob [KR95]. A related model is the *logit dynamics* of Blume [Blu93] 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 effects the selected equilibrium is the main focus of a recent work by Alos-Ferrer and Netzer [AFN10]. 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.

Incentive compatibility of best-response dynamics provide a theoretical justification for several protocols and auctions widely adopted in practice. Levin et al. [LSZ11] proved convergence and incentive compatibility of the intricate BGP protocol in the current Internet (based on the mathematical model by Gao and Rexford [GR01] 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. [GSZS10] shows that certain variants of the current TCP protocol converge (the flow rate stabilizes) and is incentive compatible on arbitrary networks (this property assumes routers adopt specific queing policy). The so-called Generalized Second-Price auctions used in most of ad-auctions is another example of incentive compatible best-response mechanism as proved by Nisan et al.[NSVZ11a]. All of these problems (and others) and results have been unified by NIsan et al. [NSVZ11b] 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 $s = (s_1, \dots, s_n)$ is the corresponding strategy profile, with $u_i(s)$ being the payoff of player i. To stress the dependency of the utility $u_i(s)$ on the strategy s_i we adopt the standard notation (x, s_{-i}) to denote the vector $(s_1, \dots, s_{i-1}, x, s_{i+1}, \dots, s_n)$.

(Imperfect) Best-response dynamics. A *game dynamics* consists of a (possibly infinite) sequence of strategy profiles $s^0, s^1, \dots$, where s^0 is an arbitrarily chosen profile and the profile s^t is obtained from 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). In this work we focus on dynamics based on the following type 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 1 (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 [KMR93] or mistakes models [You93, KR95] (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 [Blu93] (see Appendix A.2 for a brief overview) are p-imperfect with

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

for all games in which the payoff between a non-best and a best-response differs by at least one[1] and each player has at most m strategies.

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 2 ((R,ε)-fair schedule)**.** *A schedule is (R,ε)-fair if there exists a nonnegative 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 $[a+1, a+R]$.

A scheduling of the players in round-robin fashion or concurrently corresponds to $(n,0)$-fair and $(1,0)$-fair schedules, respectively. Selecting a player at random at each time step is (R,ε)-fair with $R = \mathcal{O}(n \log n)$. Observe that if a schedule is (R,ε)-fair, then, for every $\delta > 0$, all players are selected at least once with probability at least $1-\delta$ in an interval of $R \cdot \left\lceil \frac{\log(1/\delta)}{\log(1/\varepsilon)} \right\rceil$ time steps (this holds because the probability $1-\varepsilon$ is guaranteed for any interval of R time steps). We also denote with η the maximum number of players selected for update in one step by the schedule. Note that $\eta \leq n$.

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. We highlight that we do not put any other constraint on the way the dynamics run. In particular we allow both the schedule and the response rule to depend on the *status* of the game, that is on a set of information other than the current strategy profile.

Convergence and incentive compatibility. We say that a game dynamics for a game G *converges* if it eventually converges to a 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.

Let us denote with X^t the random variable that represent 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 [NSVZ11b]. Nisan et al. [NSVZ11b] analyzed the convergence and incentive compatibility of the (perfect) best-response dynamics. Before stating their result, let us now recall some definitions.

Definition 3 (never best-response)**.** *A strategy s_i is a never best-response (NBR) for player i if, for every s_{-i}, there exists s'_i such that $u_i(s_i, s_{-i}) < u_i(s'_i, s_{-i})$.*[2]

[1]When the minimum difference is δ this extends easily by taking $\beta_\delta = \beta \cdot \delta$ in place of β.

[2]Nisan et al. [NSVZ11b] 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 $u_i(s_i, s_{-i}) = u_i(s'_i, s_{-i})$ and $s_i \prec_i s'_i$. However, such tie-breaking 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 [NSVZ11b].

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

Definition 4 (elimination sequence)**.** *An elimination sequence for a game G consists of a sequence of sub-games*

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

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 5 (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.*

For example, 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

(1)

Notice that strategy 2 is a NBR for both players. Hence, there exists an elimination sequence of length 2 that reduces above game in its upper-left 2×2 sub-game 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 sub-game 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 game.

While the convergence result of [NSVZ11b] holds for the class of NBR-solvable games, in order to guarantee incentive compatibility they introduce the following condition on the payoffs:

Definition 6 (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 sequence such that the following holds. If i appears the first time in this sequence at position k, then in the sub-game G_k the profile that maximizes the utility of player $i = i^{(k)}$ is the Nash equilibrium.*

Theorem 7 (main result of [NSVZ11b])**.** *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 [NSVZ11b] 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 to the Nash equilibrium even for values of p very small. Specifically we will prove the following theorem.

Theorem 8. *For every $0 < \delta < 1$, there exists a 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.*

The game. Consider the following game: there are n players with two strategies 0 and 1. For each player i utilities are defined 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., $u_i(\mathbf{1}_{1\cdots i-1}, 0, \mathbf{x}_{i+1\cdots n}) < u_i(\mathbf{1}_{1\cdots i-1}, 1, \mathbf{x}_{i+1\cdots n})$ for each strategy profile $\mathbf{x}$, otherwise i prefers to play 0, i.e. $u_i(\mathbf{x}_{-i}, 1) < u_i(\mathbf{x}_{-i}, 0)$ for any $\mathbf{x}$ such that $\mathbf{x}_{1\cdots i-1} \neq \mathbf{1}_{1\cdots i-1}$.

It is easy to check that 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 sub-game in which all players $1, \ldots, i-1$ have eliminated 0). The sub-game $\hat{G}$ consists of the unique PNE that is the profile $\mathbf{1} = (1, \ldots, 1)$.

The p-imperfect response rule. All players play the following p-imperfect response rule:

- Player i chooses strategy 0 with probability p if all players $j < i$ are playing strategy 1;
- Player i chooses strategy 0 with probability $1 - q$ if at least one player $j < i$ is playing strategy 0, where $0 \leq q \ll p$.

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.

A key observation about this schedule is in order.

Observation 9. *Between any two occurrences of player $i < n$ there is an occurrence of a player $j \geq i+1$.*

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 $\sigma_n(t)$ plays strategy 0 given that each player $j < \sigma_n(t)$ plays 1 (thus, a bad move occurs with probability p).

of Theorem 8. 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 $\Pr(X_t^n = 1) \leq \varepsilon$. 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 $\mathbf{x}$

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

Let us use $\Pr_{\mathbf{x}}(X_\tau^n = 1)$ as a shorthand for $\Pr(X_\tau^n = 1 \mid X_0 = \mathbf{x})$. 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

$$\Pr_{\mathbf{x}}(X_\tau^n = 1) = \Pr_{\mathbf{x}}\left(X_\tau^n = 1 \mid \overline{B}\right)\Pr\left(\overline{B}\right) + \sum_{t=1}^{\tau}\Pr_{\mathbf{x}}\left(X_\tau^n = 1 \mid B_t\right)\Pr\left(B_t\right).$$

Note that B_t has probability at most p and $\overline{B}$ has probability $(1-p)^\tau$. Obviously, $\Pr\left(X_\tau^n = 1 \mid \overline{B}\right) = 1$. Moreover, by Observation 9, 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 $1 - q$ 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 $1-q$. Hence,

$$\Pr\left(X_\tau^n \neq 1 \mid B_t\right) \geq (1-q)^n.$$

Then

$$\begin{aligned}\Pr\left(X_\tau^n = 1\right) &\leq (1-p)^\tau + \tau p(1-(1-q)^n) \\ &\leq \frac{1}{1+p\tau} + p\tau\frac{q}{1-q},\end{aligned}$$

where we repeatedly used that $1-x \leq e^{-x} \leq (1+x)^{-1}$.

The theorem follows by taking $p = \frac{1-\delta}{\delta \cdot 2^{n-1}}$ and q sufficiently small. □

Remark 1. *It is interesting that we can instantiate the abstract example above as follows. The game can be seen as an instance of the BGP game [LSZ11, NSVZ11b] with utilities:*

$$\begin{aligned}u_i(0, s_{-i}) &= 0 \\ u_i(1, s_{-i}) &= \begin{cases}1, & \text{if } s_1 = \cdots = s_{i-1} = 1; \\ -L, & \text{otherwise;}\end{cases}\end{aligned}$$

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

$$p = \frac{1}{1+e^\beta} \quad \text{and} \quad q = \frac{e^{-\beta L}}{e^\beta + e^{-\beta L}}.$$

Remark 2. *The proof of Theorem 8 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 of Theorem 8 shows that non-convergence of the imperfect best-response dynamics requires setting $p \approx \frac{1}{R}$ or greater. As a consequence, it may be possible to prove convergence to the equilibrium only by taking p being 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 10. *For any NBR-solvable game G and any small $\delta > 0$ an imperfect best-response dynamics converges to the Nash equilibrium of G in $\mathcal{O}(R \cdot \ell \log \ell)$ steps with probability at least $1-\delta$, whenever $p \leq \frac{c}{\eta R \cdot \ell \log \ell}$, for an suitably chosen constant $c = c(\delta)$.*

The following two lemmas represent the main tools in the proof of the theorem above. 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 $\Pr_{\mathbf{x},h}(E)$ the probability of the event E conditioned on the initial profile being $X_0 = \mathbf{x}$ and the initial status being h. Note that h may contain the history of iterations occurred before the dynamics were in X_0.

Lemma 11. *For any profile $\mathbf{x}$ and any initial status h, we have*

$$\Pr_{\mathbf{x},h}\left(X_{t+s} \in G_k \mid X_s \in G_k\right) \geq 1 - \eta p t, \tag{2}$$

$$\Pr_{\mathbf{x},h}\left(X_{R+s} \in G_{k+1} \mid X_s \in G_k\right) \geq 1 - \eta p R - \varepsilon. \tag{3}$$

Proof. Let the dynamics be in G_k at time s and observe that if the dynamics is not in G_k at time $t+s$, then in one of steps in the interval $[s+1, s+t]$ some selected player played a NBR. Since at every step at most η players are selected, (2) follows from the union bound.

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

$$\Pr_{h\mathbf{x},}\left(X_{t+s} \notin G_{k+1} \mid X_s \in G_k \cap SEL_{i^{(k)},s,t}\right) \leq \eta t p. \tag{4}$$

Now simply observe that

$$\begin{aligned}\Pr_{\mathbf{x},h}\left(X_{t+s} \notin G_{k+1} \mid X_s \in G_k\right) &\leq \Pr_{\mathbf{x},h}\left(X_{t+s} \notin G_{k+1} \mid X_s \in G_k \cap SEL_{i^{(k)},s,t}\right) \\ &\quad + 1 - \Pr(SEL_{i^{(k)},s,t}) \\ &\leq \eta t p + \left(1 - \Pr(SEL_{i^{(k)},s,t})\right),\end{aligned}$$

where the first inequality follows from the definition of conditional probabilities and the last one uses (4). Since $\Pr(SEL_{i^{(k)},s,R}) \geq 1-\varepsilon$ by definition of imperfect best-response dynamics, the lemma follows. □

Lemma 12. *For any profile* $\mathbf{x}$*, any initial status* h *and* $1 \leq k \leq \ell$*, we have*

$$\Pr_{\mathbf{x},h}\left(X_{kR} \in G_k\right) \geq 1 - k \cdot (\eta p R + \varepsilon).$$

Proof. Observe that

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

where the first inequality follows from to the definition of conditional probabilities and the last one uses (3). Since $\Pr_{\mathbf{x},h}\left(X_0 \notin G_0\right) = 0$ the lemma follows by iterating the argument. □

of Theorem 10. Consider an interval T of length $R \cdot \left\lceil \frac{\log(2\ell/\delta)}{\log(1/\varepsilon)} \right\rceil$. As discussed above, the probability that all players are selected at least once in an interval of length T is $\frac{\delta}{2\ell}$. The theorem follows by applying Lemma 12 with $k = \ell$, $(R, \varepsilon) = (T, \delta/2\ell)$ and $p \leq \frac{\delta}{2} \cdot \frac{1}{\eta T \ell}$. □

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.

A negative result. The following theorem shows that the incentive compatibility property is not resistant to the introduction of noise.

Theorem 13. *There is a NBR-solvable game with clear outcome and an 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`	$c+2, 1$	$1, 0$
`bottom`	$0, 0$	$0, c$

for some c that will be defined later and suppose to run the logit dynamics for G (we already noted that the logit dynamics is an example of imperfect best-response dynamics). We will show that the column player has a better expected payoff by playing always strategy `right`.

The above game is a *potential game* and the potential Φ is

	`left`	`right`
`top`	$c+2$	$c+1$
`bottom`	0	c

It is known (see, for example, [Blu93, AFPP10]) that in this case the logit dynamics converges to a distribution on the set of profiles such that the probability of a profile $\mathbf{x}$ is proportional to $e^{\beta\Phi(\mathbf{x})}$. Hence, the expected utility of the column player when she plays according to the logit response rule is

$$\frac{1 \cdot e^{\beta(c+2)} + c \cdot e^{\beta c}}{1 + e^{\beta c} + e^{\beta(c+1)} + e^{\beta(c+2)}} < \frac{e^{2\beta} + c}{1 + e^{\beta} + e^{2\beta}}. \tag{5}$$

If instead the column player always plays strategy `right`, then her expected payoff is determined by the logit dynamics on the corresponding sub-game and it is equal to

$$\frac{c \cdot e^{\beta c}}{e^{\beta c} + e^{\beta(c+1)}} = \frac{c}{1 + e^{\beta}}. \tag{6}$$

Since the right-hand side of (5) is smaller than (6) for $c \geq 1 + e^{\beta}$, the lemma follows. □

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 $i = i^{(k)}$ and G_k denote the occurrence of the player and the subgame in the definition of game with clear outcome (Definition 6).

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 sub-game she is actually playing. Specifically, we have the following theorem.

Theorem 14. *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 $p \leq \frac{c}{\eta R \cdot \ell \log \ell}$, for a suitable constant $c = c(\delta)$, the dynamics run for $\Omega(R \cdot \ell \log \ell)$ and*

$$u_i(NE) \geq \frac{1}{1 - 2\delta}\left(2\delta \cdot u_i^\star + u_i^k\right),$$

where $u_i(NE)$ is the utility of i in the Nash equilibrium, $u_i^k = \max_{\mathbf{x} \in G^{(k)}} u_i(\mathbf{x})$ and $u_i^\star = \max_{\mathbf{x} \in G} u_i(\mathbf{x})$.

We can summarize the intuition behind the proof of Theorem 14 as follows:

- If player i always updates according to the p-imperfect response rule, then the game will be in the Nash equilibrium for a lot of time steps 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 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 for a small number of times (but i can gain the highest possible utility from these profiles), whereas for the rest of the time the game will be in a profile of G_k.

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

Lemma 15. *For any profile $\mathbf{x}$, any initial status h, any $1 \leq k \leq \ell$ and any $t \geq kR$, we have*

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

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

Proof. We have

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

From Lemma 11 we have

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

Moreover, let $\mathbf{y} = X_{(l-1)kR}$ and let h' be the status that contains every information collected in the first $(l-1)kR$ steps of the dynamics. Then by Lemma 12 we have

$$\Pr_{\mathbf{x},h}(X_{lkR} \notin G_k) = \Pr_{\mathbf{y},h'}(X_{kR} \notin G_k) \leq k \cdot (\eta p R + \varepsilon). \quad \square$$

Remark 3. *Observe that Lemma 15 holds even if only players $i^{(1)}, \ldots, i^{(k)}$ are updating according to a p-imperfect response rule.*

of Theorem 14. Let us start by computing the expected utility of i, given that all players are playing according to the p-imperfect response rule. Let T and p as in the proof of Theorem 10. Then, by applying Lemma 15 with $k = \ell$ and $(R, \varepsilon) = (T, \delta/2\ell)$ we have for any $t = \Omega(R \cdot \ell \log \ell)$

$$\Pr_{\mathbf{x},h}\left(X_t \in \hat{G}\right) \geq 1 - 2\delta.$$

Hence, the expected utility of i will be at least $(1 - 2\delta) \cdot u_i(NE)$.

Suppose now that i does not play a p-imperfect response rule. Similarly as done above, we let $T = R \cdot \left\lceil \frac{\log(2k/\delta)}{\log(1/\varepsilon)} \right\rceil$ and then, by applying Lemma 15 with $(R, \varepsilon) = (T, \delta/2k)$ we obtain

$$\Pr_{\mathbf{x},h}\left(X_t \notin G_k\right) \leq 2\delta.$$

Hence, the expected utility of i will be at most $2\delta u_i^\star + u_i^k$ and the theorem follows. □

5 NBR-reducible Games

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 (we give one such application in Section 5.3 below). 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 behave 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 response rule, but we consider two different schedules: the first one select just one player randomly at each time step, whereas the second one update 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 [AFN10]).

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 an 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 happen 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 handle multiple best responses). Let us pair these response rules with the schedule that select 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 [AFN10]).

5.2 Reductions between games

The previous section highlights that no general convergence result can be given for NBR-reducible games. However, we will see 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 this dynamics to the reduced game $\hat{G}$.

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

The dynamics as a Markov chain. We say that the game is in a *status–profile pair* $(h, \mathbf{x})$ if h is the set of information currently available and $\mathbf{x}$ is the profile currently played. We denote with H the set of all status–profile pairs $(h, \mathbf{x})$ and with $\hat{H}$ only the ones with $\mathbf{x} \in \hat{G}$. Let X_t be the random variable that represents the status–profile pair $(h, \mathbf{x})$ of the game after t steps of the imperfect best-response dynamics. Then, for every $(h, \mathbf{x}), (z, \mathbf{y}) \in H$ we set

$$P\big((h, \mathbf{x}), (z, \mathbf{y})\big) = \Pr\Big(X_1 = (z, \mathbf{y}) \mid X_0 = (h, \mathbf{x})\Big).$$

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 $P\big((h, \mathbf{x}), A\big) = \sum_{(z,\mathbf{y}) \in A} P\big((h, \mathbf{x}), (z, \mathbf{y})\big)$.

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 $(h, \mathbf{x}), (z, \mathbf{y}) \in H$

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

Thus, the restricted dynamics is exactly the same as the original one except that the first never leaves the sub-game $\hat{G}$, whereas in the latter, at each time step, there is probability at most p to leave this sub-game. The following lemma quantifies this similarity, by showing that, for every $(h, \mathbf{x}) \in \hat{H}$, the *total variation distance* (TV)[3] between the original and the restricted dynamics starting from $(h, \mathbf{x})$ is small.

Lemma 16. *For every* $(h, \mathbf{x}) \in \hat{H}$,

$$\Big\| P^t\big((h, \mathbf{x}), \cdot\big) - \hat{P}^t\big((h, \mathbf{x}), \cdot\big) \Big\| \leq \eta p t. \tag{7}$$

Proof. The proof is by induction on t. The base case is $t = 1$ for which the set of status–profile pairs $(z, \mathbf{y})$ such that $P((h, \mathbf{x}), (z, \mathbf{y})) > \hat{P}((h, \mathbf{x}), (z, \mathbf{y}))$ is exactly $\overline{H} = H \setminus \hat{H}$ and hence

$$\begin{aligned} \Big\| P\big((h, \mathbf{x}), \cdot\big) - \hat{P}\big((h, \mathbf{x}), \cdot\big) \Big\| &= \sum_{(z,\mathbf{y}) \in \overline{H}} \Big(P((h, \mathbf{x}), (z, \mathbf{y})) - \hat{P}((h, \mathbf{x}), (z, \mathbf{y})) \Big) \\ &= P\big((h, \mathbf{x}), \overline{H}\big) \leq \eta p, \end{aligned}$$

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

where the last inequality follows from Lemma 11. Furthermore

$$
\begin{aligned}
\left\| P^t\big((h,\mathbf{x}),\cdot\big) - \hat{P}^t\big((h,\mathbf{x}),\cdot\big) \right\| &\leq \\
\text{(TV triangle inequality)} &\leq \left\| P\big((h,\mathbf{x}),\cdot\big)P^{t-1} - \hat{P}\big((h,\mathbf{x}),\cdot\big)P^{t-1} \right\| \\
&\quad + \left\| \hat{P}\big((h,\mathbf{x}),\cdot\big)P^{t-1} - \hat{P}\big((h,\mathbf{x}),\cdot\big)\hat{P}^{t-1} \right\| \\
\text{(TV monotonicity)} &\leq \left\| P\big((h,\mathbf{x}),\cdot\big) - \hat{P}\big((h,\mathbf{x}),\cdot\big) \right\| \\
&\quad + \sup_{(h',\mathbf{y})\in\hat{H}} \left\| P^{t-1}\big((h',\mathbf{y}),\cdot\big) - \hat{P}^{t-1}\big((h',\mathbf{y}),\cdot\big) \right\| \\
\text{(Lemma 11 and induction)} &\leq \eta p + \eta p(t-1) = \eta p t. \quad \square
\end{aligned}
$$

□

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 [GMV05], correlated equilibria [Aum74] or logit equilibria [AFPP10]: 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 some profile from $\hat{G}$. A status–profile distro event is *reducible* if status–profile pairs on which is defined the distribution that represent 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.

The Main Theorem.

Theorem 17. *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 $\mathcal{O}(R\cdot\ell\log\ell+\tau)$ steps with probability at least $1-\delta$, whenever $p \leq \min\left\{\frac{c_1}{\eta R\cdot\ell\log\ell}, \frac{c_2}{\eta\tau}\right\}$, 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 $\mathcal{O}(R\cdot\ell\log\ell)$ 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 $\mathcal{O}(R\cdot\ell\log\ell+\tau)$ steps will be at most δ and the theorem follows.

Specifically, consider an interval T of length $R \cdot \left\lceil \frac{\log(4\ell/\delta)}{\log(1/\varepsilon)} \right\rceil$. By applying Lemma 12 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 $(h, \mathbf{x}) \in H$

$$\Pr\left(X_{\ell T} \in \hat{H} \mid X_0 = (h, \mathbf{x})\right) \geq 1 - \delta/2.$$

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 $(h', \mathbf{y}) \in \hat{H}$, is the same as if we assume the dynamics starts in $(h', \mathbf{y})$, i.e., it is equivalent to the probability that the event occurs in τ steps from $(h', \mathbf{y})$. 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 (7), 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 side, 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 (7) 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 16 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 $(z, \mathbf{y})$ is at least $1 - \delta/2$. □

5.3 Application: PageRank games

PageRank [BP98] is the famous reputation system used by Google for ranking web pages based on the link structure of the world wide web. PageRank considers the web as a directed graph, whose vertices corresponds to web pages and an edge from node i to node j exists if the page that corresponds to i has a link to the page that corresponds to j. Then, PageRank executes an *α-random walk* on this directed graph, i.e., at each step with probability α it *jumps* to a vertex in the graph selected according to a distribution q and with probability $1 - \alpha$ it moves to a random neighbor of the currently visited node[4]. Finally, PageRank ranks web page according to the (unique) stationary distribution π of this random walk.

It is known that the PageRank approach can be affected by manipulation [BGS05, FC06, GBGMP06, GGM05a, GGM05b]: for example, link spammers and search engine optimizers can opportunely change the link structure of their pages in order to achieve an higher rank by PageRank. This undermine the assumption on which PageRank works, that is that the link structure reflects the real ranking of the web pages.

Hopcroft and Sheldon [HS08] proposed a game-theoretic framework to analyze the effect that the selfish behavior of agents can have on the link structure of the web. Specifically, in a *PageRank game* each node v is a player whose goal is to maximize her rank $\pi(v)$. The strategies of the players consist in subsets of players towards which they can link. The structure of the network is completely specified from these choices and, hence, also the ranking of the nodes.

However, in this setting, it looks extremely unrealistic to assume that players are able to efficiently compute the best response or to know exactly the current link structure of the network (note that the set of possible strategy is exponential on the size of the network). Thus, it is more appropriate to assume that players can be sometimes wrong and fail to play the best response. That is, it is more appropriate to analyze the game dynamics in this setting through an imperfect best-response dynamics.

The modified PageRank game. Indeed, we will adopt the logit dynamics for analyzing a slight variation of the game described above: if a node v has no outgoing edges, then the next edge is selected according to distribution q but it is conditioned to be different from v. That is, the only difference with the original version is that, whenever the random walk reaches a vertex v with no outgoing edges, the next vertex cannot be v itself. We refer to this game as a *modified PageRank game*.

To analyze the equilibrium behavior of this game at stationarity of the logit dynamics would be easy if it is a potential game (see Appendix A.2). However, it is easy to see this is not the case. Consider, indeed, three nodes, named 0, 1 and 2 with 1 that links to 2 and vice versa. Assume, moreover, the

[4]When a vertex has not outgoing edges, a vertex in the graph is selected according to distribution q (see [BP98] or [HS08] for details).

distribution q is uniform. We look at what happens in the four cases deriving from the existence or the non-existence of links between 0 and 1:

- If no link between 0 and 1 exists, then you can check that the stationary distribution of our modified PageRank assigns $\pi(0) = \frac{\alpha}{3}$ and $\pi(1) = \frac{3-\alpha}{6}$;
- If 1 links to 0, but 0 does not link to 1, then the stationary distribution assigns $\pi(0) = \frac{2}{3(3-\alpha)}$ and $\pi(1) = \frac{2(2-\alpha)}{3(3-\alpha)}$;
- If 0 links to 1, but 1 does not link to 0, then the stationary distribution assigns $\pi(0) = \frac{\alpha}{3}$ and $\pi(1) = \frac{3-2\alpha}{6-3\alpha}$;
- If 0 and 1 link each other, then the stationary distribution assigns $\pi(0) = \frac{3-\alpha}{6(2-\alpha)}$ and $\pi(1) = \frac{3-2\alpha}{3(2-\alpha)}$.

The reader can check that these ranking values implies that the game cannot be a potential game for any $\alpha < 1$.

A never best-response. However, Theorem 17 can be used for achieve an accurate description of the stationary distribution of this game. We start by observing that having no outgoing edges is an NBR (according to a suitable tie-breaking). Indeed, consider the matrix N such that $N(i,j)$ is the expected number of times that j is visited before a jump if the random walk starts from i. Consider the stopping time S defined as the time in which the first jump occurs. By denoting with X_t the node visited after t steps of the random walk, we observe that $\Pr_q(X_S \in A) = q(A)$ for each subset A of nodes. Moreover, we observe that the expected value of S is $\frac{1}{\alpha} < \infty$. Then, from Proposition 4 in [AF02], it turns out that

$$\pi(j) = \alpha \sum_i q(i)N(i,j) = \alpha(q^T N)(j).$$

Let us define the matrix H such that $H(i,j)$ is the probability that j is visited before a jump if the random walk starts from i. Observe that $N(i,j) = H(i,j) \cdot N(j,j)$. Moreover, in [HS07] it is proved that $H(i,j)$ does not depend on the links placed by j. Let H^+ be a vector such that $H^+(j)$ denotes the probability that the random walk starting from j visits again this node before a jump. Observe that $H^+(j) = \frac{1-\alpha}{n-1}\sum_k H(k,j)$ if j does not have any outgoing link and $H^+(j) = \frac{1-\alpha}{|D(j)|}\sum_{k\in D(j)} H(k,j)$ otherwise, where $D(j)$ denotes the set of nodes at which j links. Finally, observe that $N(j,j)$ can be seen as a geometric random variable counting the number of returns to j before a jump, where the probability of a jump before the next return is $1 - H^+(j)$. Thus, $N(i,j) = \frac{H(i,j)}{1-H^+(j)}$ and

$$\pi(j) = \alpha \cdot \frac{\sum_i q_i H(i,j)}{1 - H^+(j)}.$$

Thus, the only way in which j could influence her ranking is by maximizing $H^+(j)$. However, it is evident that it is always possible to define a subset of players $D^\star$ that achieves such a maximum. Then, the ranking of j when she links to other nodes is always as large as her ranking when she does not link anyone. Thus, the modified PageRank game is NBR-reducible to the game G' in which each player has at least one outgoing edge.

A potential function. We will finally show that G' is a potential game. Then, we can describe the logit equilibrium for this game and, by Theorem 17, this is sufficient for closely approximating the logit equilibrium of the full modified PageRank game (at least when the rationality parameter β is large enough), despite it is not a potential game.

Consider, indeed, a profile $\mathbf{x}$ of G' and let $D_{\mathbf{x}}(j)$ be the set of neighbor of j according to $\mathbf{x}$. Consider, moreover, the following function Φ that assigns at each profile $\mathbf{x}$ of G' the value

$$\Phi(\mathbf{x}) = -\alpha \cdot \sum_j \frac{\sum_i q(i)H(i,j)}{1 - \frac{1-\alpha}{|D_{\mathbf{x}}(j)|}\sum_{i\in D_{\mathbf{x}}(j)} H(i,j)}.$$

Then, it is immediate to see that the difference in player j ranking between the profile $\mathbf{x}$ and the profile $\mathbf{x}'$, that differs from $\mathbf{x}$ only in the set $D'_{\mathbf{x}}(j)$ of nodes at which j links, is exactly the same than the difference in the value of function Φ between the same pair of profiles. Thus, Φ is a potential function giving the aimed result.

References

[AF02] David Aldous and James Allen Fill. *Reversible Markov Chains and Random Walks on Graphs*. 2002. http://www.stat.berkeley.edu/ aldous/RWG/book.html.

[AFN10] Carlos Alós-Ferrer and Nick Netzer. The logit-response dynamics. *Games and Economic Behavior*, 68(2):413–427, 2010.

[AFPP10] Vincenzo Auletta, Diodato Ferraioli, Francesco Pasquale, and Giuseppe Persiano. Mixing time and stationary expected social welfare of logit dynamics. In *SAGT*, volume 6386 of *LNCS*, pages 54–65. Springer, 2010.

[Aum74] Robert J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96, 1974.

[BGS05] Monica Bianchini, Marco Gori, and Franco Scarselli. Inside pagerank. *ACM Trans. Internet Techn.*, 5(1):92–128, 2005.

[Blu93] Lawrence E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.

[BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, April 1998.

[FC06] Eric J. Friedman and Alice Cheng. Manipulability of PageRank under Sybil Strategies. In *Proc. of the 1st Workshop of Networked Systems (NetEcon)*, 2006.

[GBGMP06] Zoltán Gyöngyi, Pavel Berkhin, Hector Garcia-Molina, and Jan O. Pedersen. Link spam detection based on mass estimation. In *VLDB*, pages 439–450, 2006.

[GGM05a] Zoltán Gyöngyi and Hector Garcia-Molina. Link spam alliances. In *VLDB*, pages 517–528, 2005.

[GGM05b] Zoltán Gyöngyi and Hector Garcia-Molina. Web spam taxonomy. In *AIRWeb*, pages 39–47, 2005.

[GMV05] Michel Goemans, Vahab Mirrokni, and Adrian Vetta. Sink equilibria and convergence. In *Proc. of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 142–154. IEEE, 2005.

[GR01] Lixin Gao and Jennifer Rexford. Stable internet routing without global coordination. *IEEE/ACM Trans. Netw.*, 9(6):681–692, 2001.

[GSZS10] Brighten Godfrey, Michael Schapira, Aviv Zohar, and Scott Shenker. Incentive compatibility and dynamics of congestion control. In *SIGMETRICS*, pages 95–106, 2010.

[HS07] John Hopcroft and Daniel Sheldon. Manipulation-resistant reputations using hitting time. In *Algorithms and Models for the Web-Graph*, volume 4863 of *Lecture Notes in Computer Science*, pages 68–81. Springer Berlin Heidelberg, 2007.

[HS08] John Hopcroft and Daniel Sheldon. Network reputation games, 2008.

[KMR93] Michihiro Kandori, George J. Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):29–56, 1993.

[KR95] Michihiro Kandori and Rafael Rob. Evolution of equilibria in the long run: A general theory and applications. *Journal of Economic Theory*, 65(2):383 – 414, 1995.

[LSZ11] Hagay Levin, Michael Schapira, and Aviv Zohar. Interdomain routing and games. *SIAM J. Comput.*, 40(6):1892–1912, 2011.

[NSVZ11a] Noam Nisan, Michael Schapira, Gregory Valiant, and Aviv Zohar. Best-response auctions. In *Proc. of the 12th ACM Conference on Electronic Commerce (EC)*, pages 351–360, 2011.

[NSVZ11b] Noam Nisan, Michael Schapira, Gregory Valiant, and Aviv Zohar. Best-response mechanisms. In *Proc. of the 2nd Symp. on Innovations in Computer Science (ICS)*, pages 155–165, 2011.

[You93] H. Peyton Young. The evolution of conventions. *Econometrica*, 61(1):57–84, 1993.

A Models for limited knowledge and bounded rationality

A.1 Mutation and mistakes models

The *mutation* and the *mistakes* model adopt the same response rule: at every time step, each selected player updates her strategies 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 choose a strategy uniformly at random. That is, suppose player i is selected at time step t and the current strategy profile is $\mathbf{x}^t$. We denote with $b_i(\mathbf{x}^t)$ the best response of i to profile $\mathbf{x}^t$ (if more than one best response exists and the current strategy x_i^t of i is a best response, then we set $b_i(\mathbf{x}^t) = x_i^t$, otherwise we choose one of the best response uniformly at random). Then, a strategy $s_j \in S_i$ will be selected by i with probability

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

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.

A.2 Logit dynamics

The *logit dynamics* for a game G runs as follows: at every time step (i) Select one player $i \in [n]$ uniformly at random; (ii) Update the strategy of player i according to the *Boltzmann distribution* with parameter β over the set S_i of her strategies. That is, a strategy $s_i \in S_i$ will be selected with probability

$$\sigma_i(s_i \mid \mathbf{x}_{-i}) = \frac{1}{Z_i(\mathbf{x}_{-i})} \, e^{\beta u_i(\mathbf{x}_{-i}, s_i)}, \tag{8}$$

where u_i is the utility function of the player i, $\mathbf{x}_{-i} \in \{0,1\}^{n-1}$ is the profile of strategies played at the current time step by players different from i, $Z_i(\mathbf{x}_{-i}) = \sum_{z_i \in S_i} e^{\beta u_i(\mathbf{x}_{-i}, z_i)}$ is the normalizing factor, and $\beta \geq 0$. From (8), it is easy to see that for $\beta = 0$ player i selects her strategy uniformly at random, for $\beta > 0$ the probability is biased toward strategies promising higher payoffs, and for β that goes to ∞ player i chooses her best-response strategy (if more than one best-response is available, she chooses one of them uniformly at random).

The above dynamics defines a *Markov chain* $\{X_t\}_{t \geq 0}$ with the set of strategy profiles as state space, and where the probability $P(\mathbf{x}, \mathbf{y})$ of a transition from profile $\mathbf{x} = (x_1, \ldots, x_n)$ to profile $\mathbf{y} = (y_1, \ldots, y_n)$ is zero if $H(\mathbf{x}, \mathbf{y}) \geq 2$ and it is $\frac{1}{n}\sigma_i(y_i \mid \mathbf{x}_{-i})$ if the two profiles differ exactly at player i. More formally, we can define the logit dynamics as follows.

Definition 18 (logit dynamics [Blu93])**.** *Let G be a game and let $\beta \geq 0$. The logit dynamics for G is the Markov chain $\mathcal{M}_\beta = (\{X_t\}_{t \geq 0}, S, P)$ where S is the set of profiles of G and*

$$P(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \cdot \begin{cases} \sigma_i(y_i \mid \mathbf{x}_{-i}), & \textit{if } \mathbf{y}_{-i} = \mathbf{x}_{-i} \textit{ and } y_i \neq x_i; \\ \sum_{i=1}^{n} \sigma_i(y_i \mid \mathbf{x}_{-i}), & \textit{if } \mathbf{y} = \mathbf{x}; \\ 0, & \textit{otherwise;} \end{cases} \tag{9}$$

where $\sigma_i(y_i \mid \mathbf{x}_{-i})$ is defined in (8)*.*

The Markov chain defined by (9) is ergodic. Hence, from every initial profile $\mathbf{x}$ the distribution $P^t(\mathbf{x}, \cdot)$ over profiles after the chain has taken t steps starting from $\mathbf{x}$ will eventually converge to a *stationary distribution* π as t tends to infinity. As in [AFPP10], we call the stationary distribution π of the Markov chain defined by the logit dynamics on a game G, the *logit equilibrium* of G.

A game is a *potential game* if there exists a function Φ such that for every player i, every $\mathbf{x}_{-i}$ and $s, s' \in S_i$

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

For the class of potential games the stationary distribution is the well-known *Gibbs measure*.

Theorem 19 ([Blu93]). *If G is a potential game with potential function Φ, then the stationary distribution π of the Markov chain given by (9) is*

$$\pi(\mathbf{x}) = \frac{1}{Z} e^{-\beta\Phi(\mathbf{x})},$$

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

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 every distribution μ'. Moreover, the following monotonicity properties hold:

$$\|\mu P - \hat{\mu} P\| \leq \|\mu - \hat{\mu}\|, \tag{10}$$

$$\left\|\mu P - \mu \hat{P}\right\| \leq \sup_{x \in \Omega} \left\|P(x, \cdot) - \hat{P}(x, \cdot)\right\|, \tag{11}$$

$$\|\mu P - \hat{\mu} P\| \leq \sup_{x, y \in \Omega} \|P(x, \cdot) - P(y, \cdot)\|, \tag{12}$$

where P and $\hat{P}$ are stochastic matrices. Indeed, as for (10) we have

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

As for (11) we observing that

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

Finally, for (12) we have

$$\begin{aligned}
\|\mu P - \hat{\mu} P\| &= \left\| \sum_{z \in \Omega} \mu(z) \sum_{w \in \Omega} \hat{\mu}(w) \left(P(z, \cdot) - P(w, \cdot) \right) \right\| \\
&\leq \sum_{z \in \Omega} \mu(z) \sum_{w \in \Omega} \hat{\mu}(w) \|P(z, \cdot) - P(w, \cdot)\| \\
&\leq \sup_{x, y \in \Omega} \|P(x, \cdot) - P(y, \cdot)\|.
\end{aligned}$$

C Equilibria of logit dynamics and NBR-reducible games

Let G be a game NBR-reducible to $\hat{G}$. Let π be the stationary distributions of the logit dynamics for G and $\hat{\pi}$ be the stationary distribution of the restriction of this dynamics to $\hat{G}$. Then the following lemma holds.

Lemma 20. *For every $\delta > 0$,*

$$\|\pi - \hat{\pi}\| \leq \delta,$$

for β sufficiently large.

Proof. Let $\tau = \hat{t}_{\text{mix}}(\delta/8)$ be the mixing time of the restricted chain. Consider first two copies of the chain starting in profiles $\hat{\mathbf{x}}, \hat{\mathbf{y}} \in \hat{G}$ and bound the total variation after τ time steps:

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

where the last inequality is due to Lemma 16 by taking β sufficiently large. Consider an interval T of length $R \cdot \left\lceil \frac{\log(8\ell/\delta)}{\log(1/\varepsilon)} \right\rceil$. By applying Lemma 12 with $k = \ell$, $(R, \varepsilon) = (T, \delta/4\ell)$ and β sufficiently large we have that for every $\mathbf{x} \in G$

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

Let $t^\star = \ell T + \tau$ and $Q = P^{\ell T}$. Then, for every $\mathbf{x}, \mathbf{y} \in G$

$$\begin{aligned}\left\|\pi - P^{t^\star}(\mathbf{y}, \cdot)\right\| &\leq \left\|P^{t^\star}(\mathbf{x}, \cdot) - P^{t^\star}(\mathbf{y}, \cdot)\right\| = \|Q(\mathbf{x}, \cdot)P^\tau - Q(\mathbf{y}, \cdot)P^\tau\| \\ \text{(triangle inequality)} &\leq \left\|Q(\mathbf{x}, \cdot)P^\tau - \hat{Q}(\mathbf{x}, \cdot)P^\tau\right\| + \left\|\hat{Q}(\mathbf{x}, \cdot)P^\tau - \hat{Q}(\mathbf{y}, \cdot)P^\tau\right\| \\ &\quad + \left\|\hat{Q}(\mathbf{y}, \cdot)P^\tau - Q(\mathbf{y}, \cdot)P^\tau\right\|,\end{aligned}$$

where, for every $\mathbf{x}, \mathbf{y} \in G$, we set

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

By (10) we obtain

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

By (12) we obtain

$$\left\|\hat{Q}(\mathbf{x}, \cdot)P^\tau - \hat{Q}(\mathbf{y}, \cdot)P^\tau\right\| \leq \max_{\hat{\mathbf{x}}, \hat{\mathbf{y}} \in \hat{G}} \|P^\tau(\hat{\mathbf{x}}, \cdot) - P^\tau(\hat{\mathbf{y}}, \cdot)\| \leq \delta/2.$$

and hence $\left\|\pi - P^{t^\star}(\mathbf{y}, \cdot)\right\| \leq 3\delta/4$. Finally, for every $\hat{\mathbf{x}} \in \hat{G}$, by triangle inequality

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