Incentive-Based Control of Asynchronous Best-Response Dynamics on Binary Decision Networks
Riehl, James; Ramazi, Pouria; Cao, Ming

Published in:
IEEE Transactions on Control of Network Systems

DOI:
10.1109/TCNS.2018.2873166

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2019

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 05-11-2018
Incentive-Based Control of Asynchronous Best-Response Dynamics on Binary Decision Networks

James Riehl\textsuperscript{1}, Pouria Ramazi\textsuperscript{2} and Ming Cao\textsuperscript{3}

\textbf{Abstract}—Various populations of interacting decision-making agents can be modeled by asynchronous best-response dynamics, or equivalently, linear threshold dynamics. Building upon recent convergence results in the absence of control, we now consider how such a network can be efficiently driven to a desired equilibrium state by offering payoff incentives or rewards for using a particular strategy, either uniformly or targeted to individuals. We begin by showing that strategy changes are monotone following an increase in payoffs in coordination games, and that the resulting equilibrium is unique. Based on these results, for the case when a uniform incentive is offered to all agents, we show how to compute the optimal incentive using a binary search algorithm. When different incentives can be offered to each agent, we propose a new algorithm to select which agents should be targeted based on maximizing a ratio between the cascading effect of a strategy switch by each agent and the incentive required to cause the agent to switch. Simulations show that this algorithm computes near-optimal targeted incentives for a wide range of networks and payoff distributions in coordination games and can also be effective for anti-coordination games.

I. INTRODUCTION

Faced with the rapidly growing scale and complexity of networked multi-agent systems, in which agents often have different and possibly competing objectives, researchers across various disciplines are increasingly using tools from game theory to study convergence, stability, control, performance, and robustness of these systems in diverse contexts, e.g., potential games [1]–[5], stochastic games [6]–[8], matrix games [9], repeated games [10], [11], networked games [12], and others [13]–[19]. For investigating dynamics and control in large populations of interacting decision-making agents, evolutionary game theory has proven to be a particularly powerful tool [20]–[24]. The \textit{myopic best-response update rule}, in which agents choose the strategy that maximizes their total utility against the current strategies of their neighbors, is one of the simple yet intelligent mechanisms that evolutionary game theory postulates to understand the emergence of collective behaviors on networks of interacting individuals, and is thus perhaps the most widely studied dynamical regime in this domain [25]. The best-response rule can be thought of as a greedy optimization scheme, and perhaps unsurprisingly, social experiments have revealed that human decisions in certain game contexts are as much as 96% consistent with the prescriptions of this policy [26]. Moreover, for two-strategy matrix games, best response updates are equivalent to linear threshold dynamics, which are prevalent in wide-ranging fields including sociology [27], economics [28], and computational neuroscience [29].

To a large degree, such dynamics can be divided into two categories: \textit{coordination games}, in which individuals tend to adopt the action used by most of their neighbors, such as in the spread of social innovations and viral infections, and \textit{anti-coordination games}, in which individuals tend to adopt actions different from those used by a majority of neighbors, such as in traffic congestion and the division of labor [30]. We refer to agents whose payoffs correspond to the above games as \textit{coordinating} and \textit{anti-coordinating}, respectively. In either context, the agents may make their decisions simultaneously, resulting in a \textit{synchronous} update rule [31], or they may make decisions on independent time lines, resulting in an \textit{asynchronous} update rule [32], which is particularly suitable when the rewards and consequences of the decisions take place more frequently than the decisions themselves. Several studies have investigated convergence in best-response dynamics for coordination and anti-coordination games in homogeneous populations, that is, when the utility functions of the individuals are the same, both on well-mixed populations [33] and networks [34]–[36], and some others have studied the more general heterogeneous case [27], [31], [37], where each individual has a possibly unique utility function. In particular, we have recently shown that every network consisting of either all coordinating or all anti-coordinating agents who update asynchronously with best responses, in the absence of any control input, will eventually reach an equilibrium state [38].

Equipped with a better understanding of how such
networks evolve, we are now interested in the possibility of promoting more desirable global outcomes through the efficient use of payoff incentives. This research is motivated by applications such as marketing new technologies [39], stimulating socially or environmentally beneficial behaviors [40], or any other application that is well-modeled by networks of coordinating agents in networks of coordinating agents and in which individual decisions are subject to influence by rewards or incentives. Indeed, this is a fast growing research area in which several different approaches are possible, depending on what is considered as the control input. For example, under imitative dynamics, the goal in [41] is to find the minimum number of agents such that when these agents adopt a desired strategy, the rest of the agents in the network will follow. The input in this work is thus the strategies of the agents, but it leaves open the question of how to implement such strategy control. In the context of best-response dynamics, a natural mechanism for achieving strategy control is the use of payoff incentives. For instance, in [42], the payoffs of a stochastic snowdrift game are changed in order to shift the equilibrium to a more cooperative one. This type of mechanism is applicable to situations where a central regulating agency has the power to uniformly change the payoffs of all agents to encourage them to play a particular strategy. We refer to this control problem as uniform reward control where the goal is to lead individuals’ strategies to best responses by offering the minimum uniform incentive to play that strategy. On the other hand, if the central agency can offer different rewards to each agent, a more efficient control protocol may be possible. That is, by altering the payoffs of just some individuals, the population can be led to a desired equilibrium state [41], [43]. We refer to this problem as targeted reward control. In case the budget for offering such rewards is limited, which may often be the case, a typical goal would be to maximize the number of individuals playing the desired strategy subject to the budget constraint, and we refer to this problem as budgeted targeted reward control.

In this section, we describe a standard model for asynchronous best response dynamics for \(2 \times 2\) matrix games on networks. Let \(G = (V, E)\) denote a network in which the nodes \(V = \{1, \ldots, n\}\) correspond to agents and the edges \(E \subseteq V \times V\) represent 2-player games between neighboring agents. Each agent \(i \in V\) chooses strategies from a binary set \(\{A, B\}\) and receives a payoff upon completion of the game according to the matrix:

\[
\begin{pmatrix}
A & B \\
A & c_i \\
B & d_i
\end{pmatrix},
\quad a_i, b_i, c_i, d_i \in \mathbb{R}.
\]

The dynamics take place over a sequence of discrete times \(t = 0, 1, 2, \ldots\). Let \(x(t) := (x_1(t), \ldots, x_n(t))^\top\) be the state of the system, where \(x_i(t) \in \{A, B\}\) is the strategy of agent \(i\) at time \(t\), and denote the current number of agent \(i\)’s neighbors playing \(A\) and \(B\) at time \(t\) by \(n_i^A(t)\) and \(n_i^B(t)\). When there is no ambiguity, we may sometimes omit the time \(t\) for compactness of notation. The total payoffs to each agent \(i\) at time \(t\) are accumulated over all neighbors, and therefore equal to \(a_i n_i^A(t) + b_i n_i^B(t)\) when \(x_i(t) = A\), or \(c_i n_i^A(t) + d_i n_i^B(t)\) when \(x_i(t) = B\).

In asynchronous (myopic) best-response dynamics, at each time \(t\), one agent activates to revise its strategy at time \(t + 1\) to that which achieves the highest total...
payoff, i.e. is the best response, against the strategies of its neighbors at time $t$:

$$x_i(t+1) = \begin{cases} 
A, & \text{if } a_i n_i^A + b_i n_i^B > c_i n_i^A + d_i n_i^B \\
B, & \text{if } a_i n_i^A + b_i n_i^B < c_i n_i^A + d_i n_i^B \\
z_i, & \text{if } a_i n_i^A + b_i n_i^B = c_i n_i^A + d_i n_i^B 
\end{cases}$$

In the literature, the case in which strategies $A$ and $B$ result in equal payoffs is often either included in the $A$ or $B$ case, or set to $x_i(t)$ to indicate no change in strategy. For maximum generality, we allow for all three of these possibilities in our approach using the notation $z_i$, and we do not even require all agents to have the same $z_i$. However, to simplify the analysis, we assume that the $z_i$’s do not change over time.

It is convenient to rewrite these dynamics in terms of the number of neighbors playing each strategy. Let $\text{deg}_i$ denote the total number of neighbors of agent $i$. We can simplify the conditions above by using the fact that $n_i^B = \text{deg}_i - n_i^A$ and rearranging terms:

$$a_i n_i^A + b_i (\text{deg}_i - n_i^A) > c_i n_i^A + d_i (\text{deg}_i - n_i^A)$$

$$n_i^A (a_i - c_i + d_i - b_i) > \text{deg}_i (d_i - b_i)$$

$$\delta_i n_i^A > \gamma_i \text{deg}_i, \quad (1)$$

where $\delta_i := a_i - c_i + d_i - b_i$ and $\gamma_i := d_i - b_i$. The cases ‘$<$’ and ‘$>$’ can be handled similarly. First, consider the case when $\delta_i \neq 0$, and let $\tau_i := \frac{\gamma_i}{\delta_i}$ denote a threshold for agent $i$. Depending on the sign of $\delta_i$, we have two possible types of best-response update rules. If $\delta_i > 0$, the update rule is given by

$$x_i(t+1) = \begin{cases} 
A, & \text{if } n_i^A(t) > \tau_i \text{deg}_i \\
B, & \text{if } n_i^A(t) < \tau_i \text{deg}_i \\
z_i, & \text{if } n_i^A(t) = \tau_i \text{deg}_i 
\end{cases} \quad (2)$$

We call agents following such an update rule coordinating agents, because they seek to switch to strategy $A$ if a sufficient number of neighbors are using that strategy, and likewise for strategy $B$. On the other hand, we call agents for which $\delta_i < 0$ anti-coordinating agents, because if a sufficient number of neighbors are playing $A$, they will switch to $B$, and vice versa. The anti-coordination update rule is given by

$$x_i(t+1) = \begin{cases} 
A, & \text{if } n_i^A(t) < \tau_i \text{deg}_i \\
B, & \text{if } n_i^A(t) > \tau_i \text{deg}_i \\
z_i, & \text{if } n_i^A(t) = \tau_i \text{deg}_i 
\end{cases} \quad (3)$$

In the special case that $\delta_i = 0$, the result is a stubborn agent who either always plays $A$ or always plays $B$ depending on the sign of $\gamma_i$ and the value of $z_i$, and this agent can be considered as either coordinating or anti-coordinating with $\tau_i \in \{0, 1\}$, possibly with a different value of $z_i$.

Let $\Gamma := (G, \tau, \pm)$ denote a network game, which consists of the network $G$, a vector of agent thresholds $\tau = (\tau_1, \ldots, \tau_n)^T$, and either $+$ or $-$ corresponding to the cases of coordinating or anti-coordinating agents, respectively. The dynamics in (2) are in the form of the standard linear threshold model [27] and (3) can be considered as an anti-coordinating linear threshold model. An equilibrium state in the threshold model is a state in which the number of $A$-neighbors of each agent does not violate the threshold that would cause them to change strategies. For example, in a network of coordinating agents with $z_i = B$ for all $i$, this means that for each agent $i \in V$, $x_i = A$ implies $n_i^A > \tau_i \text{deg}_i$ and $x_i = B$ implies $n_i^A \leq \tau_i \text{deg}_i$. Note that this notion of equilibrium is equivalent to a pure strategy Nash equilibrium in the corresponding network game.

We emphasize that the dynamics (2) and (3) do not correspond to an engineering design, but rather to a model of individuals’ behaviors as part of collective phenomena. Therefore, except for the control input, which is limited to payoff increments, individual agent dynamics cannot be controlled. Instead, these payoff increments serve as incentives for the agents to change strategies on their own accord, which may then have a cascading effect as individual decisions depend on the actions of their neighbors. Ultimately, the collective of agents is the system to be controlled. Before presenting a specific approach to achieve this, we first investigate the transitional behavior of the network games after providing payoff incentives.

## III. UNIQUE EQUILIBRIUM CONVERGENCE OF COORDINATING NETWORK GAMES

Our approach for reward-based control of the dynamics (2) depends on some important convergence and monotonicity properties, for which we build upon our previous results in [38] for the case when no control is applied. The following theorem establishes convergence of asynchronous best-response dynamics on networks of coordinating agents, and requires only the weak assumption that each agent activates infinitely many times as time goes to infinity, stated formally as follows.

**Assumption 1.** For every agent $i \in V$ and every time $t \geq 0$, there exists a future time $t^1 > t$ such that agent $i$ is active at time $t^1$.

The results of this paper apply to any activation sequence satisfying the above assumption, where by activation sequence, we mean a sequence of agents
\{i^0, i^1, \ldots \} where \(i^t\) denotes the agent who activates at time \(t\).

Of course, it is not necessary that the sequence be known in advance; in practice, agents are likely to activate in random order.

**Theorem 1** (Theorem 2 in [38]). Every network of coordinating agents will reach an equilibrium state.

This theorem guarantees equilibrium convergence, leaving open the question of whether the equilibrium is unique. As the main theoretical result of this paper, we show that if the network starts from any equilibrium state, and the thresholds of some of the agents are decreased, the network reaches a new equilibrium state, which is unique in the sense that it does not depend on the sequence in which agents activate. Let \(\Gamma := (\mathcal{G}, \tau, +)\) denote a network game of coordinating agents such that \(x(0) = \bar{x}\), where \(\bar{x}\) is an equilibrium state, and let \(\epsilon := (\epsilon_1, \ldots, \epsilon_n)\) denote a vector of nonnegative real numbers \(\epsilon_i \in \mathbb{R}_{\geq 0}\) for each agent \(i \in \mathcal{V}\).

**Theorem 2.** In the network game \(\Gamma' := (\mathcal{G}, \tau', +)\) with modified thresholds \(\tau' := \tau - \epsilon\) and starting from an equilibrium state \(x(0) = \bar{x}\), there exists a time \(t^*\) and unique equilibrium state \(\bar{x}'\) such that \(x(t) = \bar{x}'\) for all \(t \geq t^*\).

For the proof, we first show that under the condition of Theorem 2, the number of agents playing \(A\) evolves monotonically: when the network is at equilibrium, a decrease in one or more thresholds can only result in agents switching from \(B\) to \(A\).

**Proposition 1.** In the network game \(\Gamma' := (\mathcal{G}, \tau', +)\) with modified thresholds \(\tau' := \tau - \epsilon\) and starting from an equilibrium state \(x(0) = \bar{x}\), no agent will switch from \(A\) to \(B\) at any time \(t \geq 0\).

**Proof:** The proof is done via contradiction. Assume the contrary and let \(t_1 > 0\) denote the first time that some agent \(i\) switches from \(A\) to \(B\). We know that the network was at equilibrium at time zero, so it follows from (2) that \(n_i^A(0) > \tau_i \text{deg}_i\). Since no thresholds are increased and node degrees are constant, the fact that agent \(i\) switched from \(A\) to \(B\) at time \(t_1\) means that the number of \(A\)-neighbors of agent \(i\) at time \(t_1 - 1\) must have been less than that at time 0, i.e., \(n_i^A(t_1 - 1) < n_i^A(0)\). Therefore, at least one of the neighbors of agent \(i\) must have switched from \(A\) to \(B\) at some time before \(t_1\), which contradicts how \(t_1\) is defined, completing the proof.

Next we show that after decreasing some of the thresholds in a network at equilibrium, any agents who switch from \(B\) to \(A\) under one activation sequence will do so under any activation sequence, although possibly at different times. Consider two activation sequences \(\mathcal{S}^1 := \{i^0, i^1, \ldots \}\) and \(\mathcal{S}^2 := \{j^0, j^1, \ldots \}\). Denote by \(x_i(t)\) the strategy of agent \(i\) at time \(t\) under the activation sequence \(\mathcal{S}^1\), and define \(x_i^2(t)\) similarly for \(\mathcal{S}^2\). Let \(t_0\) be the first time when agent \(j^0\) is active in \(\mathcal{S}^1\). Then define \(t_s\) as the first time after \(t_{s-1}\) that agent \(j^s\) is active in \(\mathcal{S}^1\), for \(s \in \{1, 2, \ldots \}\). The existence of \(t_s\) is guaranteed by Assumption 1.

**Lemma 1.** In the network game \(\Gamma' := (\mathcal{G}, \tau', +)\) with modified thresholds \(\tau' := \tau - \epsilon\) and starting from an equilibrium state \(x(0) = \bar{x}\), given any two activation sequences \(\mathcal{S}^1 = \{i^0, i^1, \ldots \}\) and \(\mathcal{S}^2 = \{j^0, j^1, \ldots \}\), the following holds for \(s \in \{0, 1, \ldots \}\):

\[
x_j^2(s + 1) = A \Rightarrow x_j^1(t_s + 1) = A.
\]

Intuitively, this lemma holds because \(\mathcal{S}^2\) is a subsequence of \(\mathcal{S}^1\) and Proposition 1 means that no agent will switch to \(B\) as a result of activations in \(\mathcal{S}^1\) that are not part of this subsequence. For a detailed proof by induction, see Appendix A.

We finally prove Theorem 2 by using Lemma 1 and Proposition 1.

**Proof of Theorem 2:** From Theorem 1, we know that the network will reach an equilibrium state under every activation sequence satisfying Assumption 1. So it remains to prove the uniqueness of the equilibrium for all activation sequences, which we do by contradiction. Assume that there exist two activation sequences \(\mathcal{S}^1 = \{i^0, i^1, \ldots \}\) and \(\mathcal{S}^2 = \{j^0, j^1, \ldots \}\) that drive the network to two distinct equilibrium states, implying the existence of an agent \(q\) whose strategy is different at the two equilibria, say \(B\) under the equilibrium of \(\mathcal{S}^1\) and \(A\) under the equilibrium of \(\mathcal{S}^2\). Hence, there exists some time \(\tau\) after which the strategy of agent \(q\) is \(A\) under \(\mathcal{S}^2\). So since each agent is active infinitely many times, there is some time \(s \geq \tau\) at which agent \(q\) is active and plays strategy \(A\) at time \(s + 1\) under \(\mathcal{S}^2\), i.e., \(x_q^2(s + 1) = A\). Then in view of (4) in Lemma 1, \(x_q^1(t_s + 1) = A\), that is the strategy of agent \(q\) becomes \(A\) at time \(t_s + 1\). On the other hand, according to Proposition 1, the strategy of agent \(q\) will not change after \(t_s + 1\), i.e., \(x_q^1(t) = A\) for all \(t \geq t_s + 1\). But this is in contradiction with the assumption that the strategy of agent \(q\) is \(B\) at the equilibrium state under \(\mathcal{S}^1\), completing the proof.

**IV. Control Through Payoff Incentives**

In this section we consider the use of payoff incentives to drive a network of agents who update asynchronously with best responses from any undesired equilibrium toward a desired equilibrium, in which all or at least
more agents play strategy $A$. Since these networks are guaranteed to converge [38], it is reasonable to assume that the network to be controlled has reached a steady state, and therefore the control problem becomes one of driving the network from one equilibrium to another, more desirable one.

A. Uniform Reward Control

Suppose a central regulating agency has the ability to provide a reward of $r_0 \geq 0$ to all agents who play strategy $A$. The resulting payoff matrix is given by

$$
\begin{pmatrix}
A & B \\
B & C
\end{pmatrix}
= \begin{pmatrix}
a_i + r_0 & b_i + r_0 \\
c_i & d_i
\end{pmatrix},
$$

for each agent $i \in \mathcal{V}$. The control objective in this case is the following.

Problem 1 (Uniform reward control). Given a network game $\Gamma = (\mathcal{G}, \tau, \pm)$ and initial strategies $x(0)$, find the infimum reward $r^*_0$ such that for every $r_0 > r^*_0$, $x_i(t)$ will reach $A$ for every agent $i \in \mathcal{V}$.

First, we observe that the solution to Problem 1 for networks of anti-coordinating agents is simply to choose $r^*_0$ such that the thresholds of all agents are greater than or equal to one. For networks of coordinating agents, we first investigate how the agents’ thresholds are affected by the reward. Let $\Delta \tau_i := \tau'_i - \tau_i$ denote the change in agent $i$’s threshold.

Proposition 2. If a coordinating agent $i$ receives a positive reward for playing $A$, then the corresponding threshold will not increase, i.e., $\Delta \tau_i \leq 0$.

Proof: First, we consider a non-stubborn coordinating agent, i.e., $\delta_i > 0$. The original threshold for such an agent is given by

$$
\tau_i = \frac{\gamma_i}{\delta_i} = \frac{d_i - b_i}{a_i - c_i + d_i - b_i}.
$$

After adding the reward, the new threshold is

$$
\tau'_i = \frac{d_i - b_i - r_0}{a_i - c_i + d_i - b_i} = \tau_i + \Delta \tau_i,
$$

where the change in threshold is given by

$$
\Delta \tau_i = \frac{-r_0}{\delta_i}.
$$

Hence, $\delta_i > 0$ implies $\Delta \tau_i \leq 0$. Next, we consider a stubborn coordinating agent, that is $\delta_i = 0$ and $\tau_i = 0$ if the agent is biased to $A$, and $\tau_i = 1$ if it is biased to $B$. Such an agent remains stubborn after adding any reward $r_0$. In particular, if the threshold of the agent is already 0, then the reward has no effect since the agent will still be biased to $A$. The threshold will also remain unchanged if it is originally 1, and the added reward is not enough to bias the agent to $A$. Otherwise, the reward changes the bias of the stubborn agent from $B$ to $A$, making the threshold change from 1 to 0. Therefore, the change in threshold of a stubborn agent $i$ is either 0 or $-1$, resulting in $\Delta \tau_i \leq 0$, which completes the proof.

To compute the value of $r^*_0$ for networks of coordinating agents, we take advantage of the following key properties of the dynamics: (i) the number of agents who converge to $A$ is monotone in the value of $r_0$ due to Propositions 1 and 2, and (ii) due to the unique equilibrium property established in Theorem 2, the effect of a reward can be evaluated by simulating the network game under any activation sequence. In other words, property (ii) means that since all activation sequences will result in the same equilibrium, we can choose a sequence consisting of only agents whose thresholds are violated, which will have a maximum length of $n$ before reaching equilibrium. We begin by generating a set $\mathcal{R}$ of candidate infimum rewards. Let $\hat{n}_i^A = [\tau_i, \text{deg}_i]$ denote the minimum number of $A$-playing neighbors of agent $i$ required for agent $i$ to either switch to or continue playing $A$. Then, we propose

$$
\mathcal{R} := \left\{ r \geq \gamma_{\max} \left| r = \frac{\delta_i(\hat{n}_i^A - j)}{\text{deg}_i}, i \in \mathcal{V}, j \in \{1, \ldots, \hat{n}_i^A\} \right. \right\},
$$

where

$$
\gamma_{\max} = \begin{cases} 
\max_{i \in \bar{B}} \gamma_i & \bar{B} \neq \emptyset \ 
0 & \bar{B} = \emptyset
\end{cases}
$$

and $\bar{B} = \{i \mid \delta_i = 0, x_i(0) = B\}$ is the set of stubborn agents biased to $B$. The set $\mathcal{R}$ is clearly finite, and indeed includes the optimal reward as shown in the following.

Proposition 3. For a network of coordinating agents, $r^*_0 \in \mathcal{R}$.

Proof: According to Proposition 2, $\Delta \tau_i \leq 0$ for all $i \in \mathcal{V}$. So in view of Theorem 2, after adding a reward $r_0 > r^*_0$, the network reaches a unique equilibrium where everyone plays $A$, at some time $t_f$. For stubborn agents, we know that if they initially play $A$, they will keep doing so, and hence do not require a reward. However, if a stubborn agent is initially playing $B$, then in view of (1), the necessary and sufficient condition on the reward $r_0$ to make a stubborn agent $i$ play $A$ is $r_0 \geq \gamma_i$. Hence, $r^*_0 \geq \gamma$, implying that $r^*_0$ must be greater than $\gamma_{\max}$. On the other hand, in view of the update rule (2), to have all non-stubborn agents $i$ play $A$, $r^*_0$ must make the new
thresholds $\tau_i'$ satisfy $n_i^A(t_f) \geq \tau_i'$ for all $i \in V$. Hence,

$$r_0^* = \inf \left\{ r \geq \gamma_{\max} \left| r \geq \frac{\delta_i(n_i^A(t_f) + 1)}{\deg_i} \quad \forall i \in V \right. \right\}.$$ 

By definition, $n_i^A(t_f) \geq \tau_i d_i + 1$ for all $i \in V$. Hence,

$$r_0^* = \inf \left\{ r \geq \gamma_{\max} \left| r \geq \frac{\delta_i(n_i^A - (n_i^A(t_f) + 1))}{\deg_i} \quad \forall i \in V \right. \right\}.$$ 

On the other hand, $n_i^A(t) \in \{0, 1, \ldots, \deg_i\}$ for all $t$ and $i \in V$, implying that

$$r_0^* \in \left\{ r \geq \gamma_{\max} \left| r = \frac{\delta_i(n_i^A(t) - j)}{\deg_i} \quad i \in V, j \in \{1, \ldots, \deg_i\} \right. \right\}$$

which completes the proof.

Let $v^R$ denote the vector containing the elements of $\mathcal{R}$ sorted from lowest to highest. Algorithm 1 uses the fact that convergence of the reward is monotone in the reward $r_0$ and performs a binary search to find the minimum candidate reward that results in all agents reaching strategy $A$. Let $S^0 := \{1, \ldots, n, 1, \ldots, n\}$ denote an arbitrarily chosen activation sequence, which satisfies Assumption 1, and let $t^*$ denote the index of the last entry of the first sequence of $n$ consecutive activations that occur without any change in strategy (i.e., when it is clear that an equilibrium state has been reached). In what follows, $1$ denotes the $n$-dimensional vector containing all ones.

1. $i^- := 1$
2. $i^+ := |\mathcal{R}|
3. while $i^+ - i^- > 1$ do
4. \hspace{1em} $r_0^* := v^R_1$, where $j := \left\lceil \frac{i^- + i^+}{2} \right\rceil$
5. \hspace{1em} $\Gamma' := (G, \tau, +)$
6. \hspace{1em} Evaluate $x(t^*)$ under $\Gamma'$ using $S^0$
7. \hspace{1em} $x := x(t^*)$
8. \hspace{1em} if $x_i = A$ for all $i \in V$ then
9. \hspace{1em} \hspace{1em} $i^- := j$
10. \hspace{1em} else
11. \hspace{1em} \hspace{1em} $i^+ := j$
12. end
13. end

**Proposition 4.** Algorithm 1 computes the reward $r_0^*$ that solves Problem 1 and terminates in $O(n \log |\mathcal{E}|)$ steps.

**Proof:** Since $r_0^* \in \mathcal{R}$ due to Proposition 3, the minimum $r_0 \in \mathcal{R}$ which results in all agents switching to $A$ is $r_0^*$. According to Theorem 2, if a given $r_0$ results in all agents switching to $A$ for one activation sequence, then it does for every activation sequence. Therefore, we can test any given $r_0$ by activating only those agents whose thresholds are violated. Since agents can only switch from $B$ to $A$ after a decrease in thresholds, such a simulation requires no more than $n$ activations. Due to Propositions 1 and 2, the number of agents switching to $A$ is monotone in $r_0$, which means we can perform a binary search on the ordered list $v^R$. Since the maximum number of elements in the set $\mathcal{R}$ is equal to the sum of the degrees of all nodes in the network which is equal to $2|\mathcal{E}|$, a binary search on $v^R$ will result in $O(\log |\mathcal{E}|)$ iterations of the loop in Algorithm 1. The algorithm performs one simulation per iteration, and therefore requires $O(n \log |\mathcal{E}|)$ operations in total.

**B. Targeted Reward Control**

If one has the ability to offer a different reward to each agent, it may be possible to achieve a desired outcome at a lower cost than with uniform rewards in networks of coordinating agents. This is because a small number of agents switching strategies can start a cascading effect in the network. Also, in a network with irregular topology and where the agents have different payoffs, some agents will generally require a smaller reward than others in order to adopt the desired strategy. Let $r := (r_1, \ldots, r_n)^T$ denote the vector of rewards offered to each agent, where $r_i$ is the reward to agent $i$. We now have the following payoff matrix for each agent $i \in V$:

$$A \quad B$$

$$A \begin{pmatrix} a_i + r_i & b_i + r_i \\ c_i & d_i \end{pmatrix}, \quad a_i, b_i, c_i, d_i \in \mathbb{R}, \quad r_i \in \mathbb{R}_{\geq 0}.$$ 

The targeted control objective is the following.

**Problem 2 (Targeted reward control).** Given a network game $\Gamma = (G, \tau, \pm)$ and initial strategies $x(0)$, find the targeted reward vector $r^*$ that minimizes $\sum_{i \in V} r_i^*$ such that if $r_i > r_i^*$ for each $i$, then $x_i(t)$ will converge to $A$ for every agent $i \in V$.

The solution to Problem 2 for networks of anti-coordinating agents is simply to set the threshold of every agent greater than or equal to one. Now consider a network of coordinating agents, which is at equilibrium.
at some time \( t_e \). Let \( \bar{r}_i \) denote the infimum reward required for an agent playing \( B \) in this network to switch to \( A \), which must satisfy the following according to (1):
\[
\delta_i n_i^A(t_e) = (\gamma_i - \bar{r}_i) \deg_i,
\]
\[
\Rightarrow \bar{r}_i = \gamma_i - \frac{\delta_i n_i^A(t_e)}{\deg_i}.
\]
(6)
The corresponding new threshold is \( \tau'_i = \tau_i + \Delta \tau_i \), where
\[
\Delta \tau_i = \begin{cases} 
-\frac{\bar{r}_i}{\delta_i} & \text{if } \delta_i \neq 0 \\
0 & \text{if } \delta_i = 0 \land \gamma_i \leq 0 \\
-1 & \text{if } \delta_i = 0 \land \gamma_i > 0
\end{cases}.
\]
In order to identify which agents should be offered incentives, we propose a potential function, which is a modification of the one used in [38] to prove convergence. Define the function \( \Phi(x(t)) = \sum_{i=1}^{n} \Phi_i(x_i(t)) \), where
\[
\Phi_i(x(t)) = \begin{cases} 
n_i^A(t) - \bar{n}_i^A(t) & \text{if } x_i(t) = A \\
n_i^A(t) - \bar{n}_i^A(t) - 1 & \text{if } x_i(t) = B
\end{cases}.
\]
(7)
This function has a unique maximum, which occurs when all agents play \( A \), and increases whenever an agent switches from \( B \) to \( A \).

To evaluate the resulting change in the potential function \( \Phi(x) \), we again use Theorem 2, which means that the network will reach a unique equilibrium and simulations are thus fast to compute using an activation sequence of length at most \( n \). Denote this unique equilibrium by \( \bar{x} \). The total change is then given by \( \Delta \Phi(\bar{x}) := \Phi(\bar{x}) - \Phi(x(0)) \). Let \( e_i \) denote the \( i \)th column of the \( n \times n \) identity matrix.

Algorithm 2 computes a set of agents and rewards such that when these rewards are offered to the corresponding agents, the network will eventually reach a state in which all agents play strategy \( A \), if there is no budget limit, and if there is a budget limit, it computes a set of rewards that satisfies this limit. It is a generic algorithm in the sense that the set of agents is computed iteratively, and the rule for selecting an agent at each iteration is the final piece that completes the algorithm. Since \( \bar{r}_i \) is an infimum reward, we add an arbitrarily small amount \( \epsilon \) to any nonzero reward \( r_i \) to ensure that the targeted agent will switch to \( A \).

The rule we propose for choosing an agent in line 4 of Algorithm 2 is to select the uncontrolled \( B \)-playing agent that maximizes the ratio \( \frac{\Delta \Phi(x)^\alpha}{r_i^\beta} \), where the exponents \( \alpha \geq 0 \) and \( \beta > 0 \) are degrees of freedom for the control designer, which we will explore further in Section V.

**Remark 1.** In the worst case, the computational complexity of Algorithm 2 will be \( O(nm) \), where \( m \) is the number of edges in the network, because simulating the network game takes \( O(m) \) computation steps, and the maximum number of iterations of the algorithm is \( O(n) \), which occurs when rewards are offered to every agent in the network.

### C. Budgeted Targeted Reward Control

It is quite likely that any agency that wishes to influence a network of agents through the use of rewards has a limited budget with which to do so. This leads to the following problem, which is perhaps of even greater practical importance than Problem 2.

**Problem 3** (Budgeted targeted reward control). Given a network game \( \Gamma = (G, \tau, \pm) \), initial strategy state \( x(0) \), and budget constraint \( \sum_{i \in V} r_i < \rho \), find the reward vector \( r \) that maximizes the number of agents in the network who reach \( A \).

Algorithm 2 is designed to approximate the solution to this problem as well, by incorporating the budget constraint in the definition of the set \( B \) of candidate nodes to target for each iteration. The only difference is that the algorithm will now terminate if no more agents can be incentivized to switch to \( A \) without violating the budget constraint \( \rho \).

### V. Simulations

In this section, we compare the performance of the proposed algorithm to some alternative approaches. Short descriptions of each algorithm are provided below. Each of these methods is applied iteratively, targeting agents until either the control objective is achieved or the budget limit is reached.
• **Iterative Random (rand):** target random agents in the network.

• **Iterative Degree-Based (deg):** target agents with maximum (minimum) degree for networks of coordinating (anti-coordinating) agents.

• **Iterative Potential Optimization (IPO):** target agents resulting in the maximum increase of the potential function ($\alpha = 1$, $\beta = 0$).

• **Iterative Reward Optimization (IRO):** target agents requiring minimum reward ($\alpha = 0$, $\beta = 1$).

• **Iterative Potential-to-Reward Optimization (IPRO):** target agents maximizing the potential-change-to-reward ratio ($\alpha > 0$, $\beta > 0$).

For each set of simulations, we generate geometric random networks by randomly distributing $n$ agents in the unit square and connecting all pairs of agents who lie within a distance $R$ of each other. We focus on the case when all agents are coordinating to align with our theoretical results, but we also include one simulation study on a network of anti-coordinating agents to show that the proposed algorithm can be applied to more general cases. In all simulations of the IPRO algorithm, we used $\alpha = 1$ and $\beta = 4$.

A. Uniform vs. Targeted Reward Control

First, we investigate the difference between uniform and targeted reward control to estimate the expected cost savings when individual agents can be targeted for rewards rather than offering a uniform reward to all agents. Figure 1 shows not only that targeted reward control offers a large cost savings over uniform rewards, but that the savings increases with network size.

![Fig. 1. Comparison of uniform and targeted reward control on geometric random networks for a range of sizes. For each size tested, 500 random networks were generated using a connection radius $R = \sqrt{(1 + \text{deg}_{\text{avg}})/\pi n}$, corresponding to a mean node degree of approximately $\text{deg}_{\text{avg}} = 10$. Thresholds $\tau_i$ for each agent are uniformly randomly distributed on the interval $[0, \frac{1}{2}]$, and the corresponding payoffs are $a_i = \frac{1-\tau_i}{\tau_i}$, $b_i = c_i = 0$, and $d_i = 1$.](image)

B. Targeted-Reward Control: Network Size

Next, we compare the performance of the proposed control algorithms to some alternative approaches for various sizes of networks of coordinating agents, using the same network and threshold setup as the previous section. Fig. 2 shows that the IPRO algorithm performs consistently better than the other proposed approaches across all network sizes, although the IRO method requires only slightly larger rewards on average than IPRO.

![Fig. 2. Algorithm performance comparison for different sizes of networks. The connection radius, threshold distribution, and payoffs are generated exactly as in the simulations for Fig. 1.](image)

C. Targeted-Reward Control: Network Connectivity

We now investigate how the connectivity of a network affects the reward needed to achieve consensus in strategy $A$. We consider geometric random networks of only 12 agents, which is small enough that we can compare against the true optimal solution computed using an exhaustive search algorithm. Fig. 3 shows that there appears to be a transition region in the required reward between sparsely and densely connected networks, and we see that the IPRO algorithm yields near-optimal results across the entire range, while the IRO algorithm also performs quite well for dense networks.

D. Targeted-Reward Control: Threshold Level

In this section, we investigate the performance of various algorithms as the thresholds of agents increase and thus become more costly to control. We again consider geometric random networks of only 12 agents and thresholds of no greater than 0.5 in order to compare against the optimal solution. Fig. 4 shows that the IPRO algorithm maintains the best performance across this range of threshold values, while the distance from optimality increases slightly as the mean threshold increases.
connection radius
mean reward (r)
rand
deg
IRO
IPO
IPRO
opt

Fig. 3. Algorithm performance comparison on sparsely to densely connected 12-node networks. 100 networks are tested for each connection range, and the threshold distribution and payoffs are generated exactly as in the simulations for Fig. 1.

mean threshold
mean reward (r)
rand
deg
IRO
IPO
IPRO
opt

Fig. 4. Algorithm performance comparison for various mean thresholds of coordinating agents. 500 12-node networks are tested for each mean threshold value $\tau_0$ and the connection radius $R$ is drawn uniformly at random from the interval $[0.3, 1]$. Agent thresholds are uniformly randomly distributed on the interval $\tau_0 \pm 0.1$.

E. Targeted-Reward Control: Threshold Variance

In the next set of simulations, we change the threshold variance to understand the effect of increasing heterogeneity on the performance of the algorithms. Fig. 5 shows that the IPRO algorithm again performs the best of the alternative algorithms. Moreover, as the threshold variance increases, its performance approaches that of the optimal solution.

F. Budgeted Targeted Reward Control

Finally, we consider the case when there is a limited budget from which to offer rewards. Figures 6 and 7 show the results for the cases of coordination and anti-coordination, respectively. In the coordination case, we see that IPRO achieves greater convergence to $A$ at lower costs when compared to the other approaches. Interestingly, the IPO algorithm also performs quite well for low-budget cases. However, there remains significant sub-optimality of all approaches in the low to middle range of reward budgets. Since budgeted targeted reward control is the only problem that has a nontrivial solution for anti-coordinating agents, we also compared the algorithms for an anti-coordinating case. Here, we observe that while IRO works best for small reward budgets, IPO performs best for larger reward budgets. This suggests setting the exponent $\alpha$ small for low budgets and large for high budgets while doing exactly the opposite for the exponent $\beta$.

Fig. 5. Algorithm performance comparison for different threshold variances $w$. 500 12-node networks are tested for each value of $w$ and the thresholds are uniformly randomly distributed in the interval $\frac{1}{2} \pm \frac{w}{2}$.

Fig. 6. Algorithm performance comparison for budgeted targeted reward control on networks of coordinating agents for a range of reward budgets. 500 networks were tested with 50-nodes each and a connection range $R = 0.2$. Thresholds are uniformly randomly distributed on the interval $0.5 \pm 0.1$.

Fig. 7. Algorithm performance comparison for budgeted targeted reward control on networks of anti-coordinating agents for a range of reward budgets. 500 networks were tested with 50-nodes each and a connection range $R = 0.2$. Thresholds are uniformly randomly distributed on the interval $0.5 \pm 0.1$. 
VI. CONCLUDING REMARKS

We have considered three problems related to the control of asynchronous best-response dynamics on networks through payoff incentives. Our proposed solutions are based on the following key theoretical results: (i) after offering rewards to some of the agents in a coordinating network which is at equilibrium, strategy switches occur only in one direction, and (ii) the network reaches a unique equilibrium state. When a central entity can offer a uniform reward to all agents, the minimum value of this reward can be computed using a binary search algorithm whose efficiency is made possible by these monotonicity and uniqueness results. If rewards can be targeted to individual agents, the desired convergence can be achieved at much lower cost; however, the problem becomes more complex to solve. To approximate the solution in this case, we proposed the IPRO algorithm, which iteratively selects the agent who, upon switching strategies, maximizes the ratio between the resulting change in potential and the cost of achieving such a switch, until desired convergence is achieved. A slight modification of this algorithm applies to the case when the budget from which to offer rewards is limited. In a simulation study on geometric random networks under various conditions, the algorithm performed significantly better than other algorithms based on threshold or degree, and in many cases came very close to the true optimal solution. Compelling directions for future work include making refinements to the IPRO algorithm, including prescriptions for the exponents \( \alpha \) and \( \beta \) under various conditions, and bounding the worst-case approximation error for various network structures and game dynamics.

REFERENCES

APPENDIX A

Proof of Lemma 1

Proof: The proof is via induction on \( s \). First the statement is shown for \( s = 0 \). Suppose \( x_{j^0}^0(1) = A \). If \( x_{j^0}^0(0) = A \), i.e., agent \( j^0 \)'s strategy was already \( A \) in the beginning, then in view of Proposition 1, this agent will not switch to \( B \) regardless of the activation sequence. Hence, \( x_{j^0}^s(t) = A \) for all \( t \geq 0 \), implying that (4) is in force. Next, assume that \( x_{j^0}^2(0) = B \). Then agent \( j^0 \) has switched strategies at \( t = 1 \) under \( S^2 \). Hence, in view of (2),

\[
\tau'_{j^0} = \tau'_{j^0} \geq \tau'_{j^0} \deg_{j^0}
\]

where \( \tau'_{j^0} \) denotes the (possibly new) threshold of agent \( i \) after decreasing some thresholds at time 0 and \( n_{i,t}^{A_{i,j}}(t) \) denotes the number of \( A \)-playing neighbors of agent \( i \) at time \( t \) under the activation sequence \( S^2 \). Similarly define \( n_{i,t}^{A_1}(t) \). Clearly

\[
n_{i,t}^{A_1}(0) = n_{i,t}^{A_2}(0).
\]

Due to Proposition 1, we also have \( n_{i,t}^{A_1}(0) = n_{i,t}^{A_2}(0) \). Hence, it follows from (9) that \( n_{i,t}^{A_1}(0) = n_{i,t}^{A_2}(0) \). Therefore, according to (8), \( n_{i,t}^{A_1}(0) \geq \tau'_{j^0} \deg_{j^0} \), implying that \( x_{j^0}^1(t_0 + 1) = A \), which proves (4) for \( s = 0 \). Now assume that (4) holds for \( s = 0, 1, \ldots, r - 1 \). Similar to the case of \( s = 0 \), the induction statement can be proven for \( s = r \): Suppose \( x_{j^r}^s(r + 1) = A \). If \( x_{j^r}^s(r) = A \), then according to Proposition 1, agent \( j^r \) will not switch to \( B \) regardless of the activation sequence. Hence, \( x_{j^r}^1(t) = A \) for all \( t \geq r \), implying that (4) is in force for \( s = r \). So assume that \( x_{j^r}^s(r) = B \). Then agent \( j^r \) switches strategies at \( t = r + 1 \) under \( S^2 \). Hence, in view of (2),

\[
n_{i,t}^{A_2}(r) \geq \tau'_{j^r} \deg_{j^r}.
\]

Since (4) holds for all \( s = 0, 1, \ldots, r - 1 \), and because of Proposition 1, we obtain

\[
n_{i,t}^{A_1}(t_{r-1} + 1) \geq n_{i,t}^{A_2}(r).
\]

On the other hand, in view of Proposition 1, since \( t_r \geq t_{r-1} + 1 \), we have \( n_{i,t}^{A_1}(t_r) \geq n_{i,t}^{A_2}(t_{r-1} + 1) \). So because of (11), we get \( n_{i,t}^{A_1}(t_r) \geq n_{i,t}^{A_2}(r) \). Therefore, according to (10), \( n_{i,t}^{A_1}(t_r) \geq \tau'_{j^r} \deg_{j^r} \), implying that \( x_{j^r}^1(t_r + 1) = A \), which proves (4) for \( s = r \), completing the proof. 

[38] P. Ramazi, I. Riehl, and M. Cao, “Networks of conforming or nonconforming individuals tend to reach satisfactory decisions,” in Decision and Control (CDC), 2012 IEEE 51st Annual Conference on . IEEE, 2012.