Exploration of misperceptions in hypergames

Bahman Gharesifard   Jorge Cortés

Abstract— This paper studies adversarial scenarios within the framework of hypergames, where rational players have misperceptions about the game they are involved in. We introduce the notion of inconsistent equilibrium to capture those equilibria of the hypergame that are not perceived by at least one player as an equilibrium in her own game. We identify a class of actions, termed exploratory, that a player can take from an inconsistent equilibrium to improve her payoff (even though such actions are perceived to be sanctioned by other players). We analyze the asymptotic convergence properties of the resulting dynamical system and characterize to what extent misperception can be decreased by the use of exploratory actions.

I. INTRODUCTION

This paper studies the issue of inconsistency in outcomes in adversarial scenarios, when players have subjective perceptions about the state of the game. We study this problem using a class of games of incomplete information, called hypergames. In hypergames, players can have mutually incompatible perceptions about their opponents' preferences. This class of games enjoys many properties, including the fact that, under suitable learning strategies, the repeated play converges to an equilibrium. However, an equilibrium of the overall hypergame may not be consistent with the perceptions of individual players. If the hypergame arrives at such inconsistent equilibria, then the players can correctly deduce that some misperception about each other exists among them. This paper introduces the notion of exploratory action, which correspond to actions that a player can take from an inconsistent equilibrium to improve her payoff (at the risk of being sanctioned by player's ensuing action). Our aim is to understand under what conditions such exploratory actions leads the players to correct errors in their misperception about others.

Literature review: In this paper, we consider games of incomplete information, where players are uncertain about some parameters involved, for example the payoff functions [1]. There is a large body of work on games; we refer the interested readers to [2] for a complete synthesis of games with incomplete information and to [3] for a summary of the literature. More specifically, we consider the framework of hypergames introduced in [4] and further developed in [5], [6], [7]. This class of incomplete information games is mostly used in the context of conflict analysis [8], [6] and is typically useful in scenarios where players are certain about their opponents' perceptions, while these certainties may be mutually inconsistent. Some well-known examples of the application of hypergame analysis include the Normandy

Bahman Gharesifard and Jorge Cortés are with the Department of Mechanical and Aerospace Engineering, University of California San Diego, {bgharesifard,cortes}@ucsd.edu

invasion and the Cuban missile crisis [6]. Other examples are wartime negotiation [9] and cybersecurity [10], where players play security strategies and the cost of risky actions is high.

The issue of inconsistency in equilibria in the framework of incomplete information games, where players have subjective belief about the state of the game, has been studied in [11]. In [12], [13], it is shown that the swap learning method is guaranteed not to increase the misperception of a player and converges to an equilibrium, possibly inconsistent with players' perception of equilibria. In [12], a self-blaming strategy, called the modified swap learning, is introduced. This strategy is guaranteed to converge to a consistent equilibrium, however, it can possibly increase the misperceptions. This is a consequence of the fact that, indeed, inconsistencies may entirely be due to opponent's misperception.

Statement of contributions: Our contributions are three-fold. The first contribution is the introduction of the notions of inconsistent equilibrium and exploratory action. An equilibrium of the hypergame is inconsistent if it is not perceived as an equilibrium by at least one of the players. An action from an inconsistent equilibrium is exploratory if it improves the payoff of the player but is perceived as not being free of sanctions by other players. We show that the digraph that contains both the actions perceived as sequentially rational by a player and the exploratory actions, termed exploratory H-digraph, is weakly acyclic. The second contribution establishes that, when players play sequentially, update their perceptions using swap update mechanism, compatible with their observations, and only one uses exploratory actions, the repeated play of the hypergame converges to an equilibrium which is either rational or consistent for the player. The third contribution concerns the setup where more than one player uses exploratory actions. In this case, we show that repeated play may not converge to an equilibrium, but a cycle. We introduce the notion of relative misperception, which allows us to precisely characterize those parts of the misperception that can be corrected.

Organization: Section II presents the basic notions on hypergames. Section III formalizes the problem under study. Section IV introduces the notion of exploratory action and analyzes the convergence properties of repeated play when such actions are allowed. Finally, we gather our conclusions and ideas for future work in Section V.

II. HYPERGAME THEORY

We consider games with inconsistent perceptions across players and model them as hypergames [5], [6], [12]. A 0-level hypergame is simply a *(finite) game*, i.e., a triplet

$\mathbf{G} = (V, \mathbf{S}_{\text{outcome}}, \mathbf{P})$, where V is a set of $n \in \mathbb{Z}_{\geq 1}$ players, $\mathbf{S}_{\text{outcome}} = S_1 \times \ldots \times S_n$ is the outcome set with finite cardinality $N = |\mathbf{S}_{\text{outcome}}| \in \mathbb{Z}_{\geq 1}$ and $\mathbf{P} = (P_1, \ldots, P_n)$, with $P_i = (x_1, \ldots, x_N)^T \in \mathbf{S}_{\text{P}}$ the preference vector of player v_i, $i \in \{1, \ldots, n\}$. Here, S_i is a finite set of actions available to player $v_i \in V$ and $\mathbf{S}_{\text{P}} \subset \mathbf{S}_{\text{outcome}}^N$ is the set of all elements in $\mathbf{S}_{\text{outcome}}^N$ with pairwise different entries. We denote by π_i the projection of $\mathbf{S}_{\text{outcome}}$ onto S_i.

A *n-person 1-level hypergame* is a set $H^1 = \{\mathbf{G}_1, \ldots, \mathbf{G}_n\}$, where $\mathbf{G}_i = (V, (\mathbf{S}_{\text{outcome}})_i, \mathbf{P}_i)$, for $i \in \{1, \ldots, n\}$, is the subjective finite game of player $v_i \in V$, and V is a set of n players; $(\mathbf{S}_{\text{outcome}})_i = S_{1i} \times \ldots \times S_{ni}$, with S_{ji} the finite set of strategies available to v_j, as perceived by v_i; $\mathbf{P}_i = (P_{1i}, \ldots, P_{ni})$, with P_{ji} the preference vector of v_j, as perceived by v_i. In a 1-level hypergame, each player $v_i \in V$ plays the game $\mathbf{G}_i$ with the perception that she is playing a game with complete information. Even though the notion of hypergame can be extended to higher levels [5], in this paper we focus on 1-level hypergames. This is the simplest class of games that allows for the explicit modeling of players' perceptions about their opponents' preferences and therefore exhibits all the intricacies associated with its analysis.

A. Stability and equilibria

We recall the notion of stability for 2-person 1-level hypergames. Let $H^1 = \{H_A^0, H_B^0\}$. Here, $H_A^0 = (\mathrm{P}_{AA}, \mathrm{P}_{BA})$ is the 0-level hypergame for player A, where P_{AA} and P_{BA} are, respectively, the preferences of player A and player B perceived by player A. The same convention holds for $H_B^0 = (\mathrm{P}_{AB}, \mathrm{P}_{BB})$. For simplicity, the 0-level hypergames have the same set of outcomes $\mathbf{S}_{\text{outcome}}$. We denote by $\succeq_{\mathrm{P}_{IJ}}$ the binary relation on $\mathbf{S}_{\text{outcome}}$ induced by P_{IJ}, where $I, J \in \{A, B\}$. For convenience, we let $\mathbf{S}_{\text{outcome}}|_{\pi_I(x)} = \{y \in \mathbf{S}_{\text{outcome}} \mid \pi_I(y) = \pi_I(x)\}$ and refer to it as a restricted outcome set. We also find it useful to use I' to denote the opponent of I in $\{A, B\}$. We assign $\mathrm{rank}(x, \mathrm{P}_{IJ}) \in \mathbb{R}_{>0}$ to each outcome $x \in \mathbf{S}_{\text{outcome}}$ such that $\mathrm{rank}(y, \mathrm{P}_{IJ}) > \mathrm{rank}(x, \mathrm{P}_{IJ})$ if and only if $x \succ_{\mathrm{P}_{IJ}} y$ (players prefer the outcomes with lower ranks). We use the set $\{1, \ldots, |\mathbf{S}_{\text{outcome}}|\}$ to rank the outcomes. The *misperception function* $\mathcal{L}_{BA} : \mathbf{S}_{\text{P}} \to \mathbb{R}_{\geq 0}$ of A about B's game is

$$\mathcal{L}_{BA}(P) = \sum_{i=1}^{N} |\mathrm{rank}(x_i, \mathrm{P}_{BB}) - \mathrm{rank}(x_i, \mathrm{P})| \quad (1)$$

Note that this function compares the rank of each outcome in the preference vector for B in H_A^0 to its rank in B's true preference vector in H_B^0. A similar misperception function $\mathcal{L}_{AB}$ can be defined.

Given two distinct outcomes $x, y \in \mathbf{S}_{\text{outcome}}$, y is an *improvement* from x for $I \in \{A, B\}$, perceived by $J \in \{A, B\}$ in H_J^0, if and only if $\pi_{I'}(y) = \pi_{I'}(x)$ and $y \succ_{\mathrm{P}_{IJ}} x$. An outcome $x \in \mathbf{S}_{\text{outcome}}$ is called *rational* for $I \in \{A, B\}$, as perceived by $J \in \{A, B\}$ in H_J^0, if there exists no improvement from x for I. The common notion of rationality in hypergames is the notion of sequential rationality [14], [6], [15]. An outcome $x \in \mathbf{S}_{\text{outcome}}$ is *sequentially rational* for

$I \in \{A, B\}$ with respect to H_J^0, $J \in \{A, B\}$, if and only if for each improvement y for I, perceived by J in H_J^0, there exists an improvement z for I', perceived by J in H_J^0, such that $x \succ_{\mathrm{P}_{IJ}} z$. Whenever this holds, we say that the improvement z from y for I' *sanctions* the improvement y from x for I in H_J^0. By definition, a rational outcome is also sequentially rational. An outcome $x \in \mathbf{S}_{\text{outcome}}$ is *unstable* for player I with respect to H_J^0 if it is not sequentially rational for player I, as perceived by player J and is an *equilibrium of H_J^0* if it is sequentially rational for both J and J', as perceived by player J. An outcome x is an *equilibrium of H^1* if it is sequentially rational for player A in H_A^0 and for player B in H_B^0. Note that x can be an equilibrium for H^1 and not an equilibrium of H_A^0. We denote by $\mathbf{Eq}(H_A^0)$, $\mathbf{Eq}(H_B^0)$, and $\mathbf{Eq}(H^1)$ the set of equilibria of H_A^0, H_B^0, and H^1, respectively.

B. H-digraphs

The notion of H-digraph encodes the stability information of hypergames. We begin by recalling some basic notions about directed graphs [16]. A *digraph* G is a pair (V, E), where V is a finite set, called the vertex set, and $E \subseteq V \times V$, called the edge set. Given $(u, v) \in E$, u is an *in-neighbor* of v and v is an *out-neighbor* of u. The set of in-neighbors and out-neighbors of v are denoted, respectively, by $\mathcal{N}^{\text{in}}(v)$ and $\mathcal{N}^{\text{out}}(v)$. The *in-degree* and *out-degree* of v are the number of in-neighbors and out-neighbors of v, respectively. A vertex is called a *sink* if its out-degree is zero and a *source* if its in-degree is zero. A *(directed) path* is an ordered sequence of vertices so that any two consecutive vertices are an edge of the digraph. A *cycle* in a digraph is a directed path that starts and ends at the same vertex and has no other repeated vertex. A digraph is called *acyclic* if it does not contain any cycle. A digraph is called *weakly acyclic* if there exists a path from any vertex to a sink. Note that a weakly acyclic digraph may contain cycles.

Formally, the *H-digraph* associated to H_A^0 is $\mathcal{G}_{H_A^0} = (\mathbf{S}_{\text{outcome}}, \mathcal{E}_{H_A^0})$, where there exists an edge $(x, y) \in \mathcal{E}_{H_A^0}$ if and only if either there exists an improvement y from x for A for which there is no sanction of B in H_A^0, or there exists an improvement y from x for B for which there is no sanction of A in H_A^0. One can similarly construct $\mathcal{G}_{H_B^0}$. By definition, an outcome x is sequentially rational for A (respectively for B) if and only if $\mathcal{N}^{\text{out}}(x) \cap \mathbf{S}_{\text{outcome}}|_{\pi_B(x)} = \emptyset$ (respectively $\mathcal{N}^{\text{out}}(x) \cap \mathbf{S}_{\text{outcome}}|_{\pi_A(x)} = \emptyset$). Moreover, an outcome belongs to $\mathbf{Eq}(H_A^0)$ if and only if its out-degree in the associated H-digraph is zero. H-digraphs enjoy some structural properties that play an important role in our discussion later. A path $\mathfrak{S} = (x_1, x_2, \ldots)$, $x_i \in \mathbf{S}_{\text{outcome}}$ for all $i \in \mathbb{Z}_{\geq 1}$, is *nondeteriorating* for H_A^0 if $(x_j, x_{j+1}) \in \mathcal{E}_{H_A^0}$, for all $j \in \mathbb{Z}_{\geq 1}$. A cycle $\mathfrak{S} = (x_1, x_2, \ldots, x_m, x_1)$, $m \in \mathbb{Z}_{\geq 1}$, is a *weak improvement cycle* for H_A^0 if it is nondeteriorating and $x_{j+1} \succ_{\mathrm{P}_{i(j)}} x_j$ for some $j \in \{1, \ldots, m-1\}$.

Theorem 2.1: (**0-level hypergames with two players contain no weak improvement cycle [13]):** For a 1-level hypergame $H^1 = \{H_A^0, H_B^0\}$ between players A and B, H_A^0 and H_A^0 contain no weak improvement cycle.

We denote the sequence of outcomes of the repeated play of a hypergame by a sequence $\sigma = (x_1, x_2, \ldots)$, where $x_1, x_2, \ldots \in \mathbf{S}_{\text{outcome}}$. We call σ a cycle when $\sigma = (x_1, x_2, \ldots x_1)$, where $x_i \neq x_1$ unless $i = 1$. Note that this notion of a cycle is different from the one for a digraph.

C. Learning

Suppose players A and B take actions that change the outcome from x to y. If A can perfectly observe B's action and believes that the opponent is rational, she concludes that player B prefers $(\pi_A(x), \pi_B(y))$ over x. Therefore, A can incorporate this information into her hypergame and update her perception about the preferences of player B.

1) Preference update mechanism: Let us describe the formal requirements that the method used for incorporating these observations should satisfy. Let P_{BA} be a preference vector of player B, perceived by player A. We define the *observation set* $\mathcal{O}_{BA}$ as the set of all binary relations observed by player A about player B. We say that the preference vector P_{BA} is compatible with an observation set $\mathcal{O}_{BA}$ if all the binary relations in $\mathcal{O}_{BA}$ hold with the order $\succeq_{P_{BA}}$. A *preference update mechanism* compatible with an observation set $\mathcal{O}_{BA}$ is a map $\Psi_{\mathcal{O}_{BA}} : \mathbf{S}_p \to \mathbf{S}_p$ such that $\Psi_{\mathcal{O}_{BA}}(P)$ is compatible with $\mathcal{O}_{BA}$ for $P \in \mathbf{S}_{\text{outcome}}$. Throughout this paper, when we say a player updates her preferences with some binary relation, we mean that this player adds this binary relation to her associated observation set and uses a preference update mechanism to generate a preference vector compatible with the observation set. We next describe one of such mechanisms, used in this paper.

2) Swap learning: We start with an algebraic construction. Let V be a set of cardinality N and let W be the subset of V^N with pairwise different elements. For $x_1, x_2 \in V$, let $\text{swap}_{x_1 \mapsto x_2} : W \to W$ be defined by

$$(\text{swap}_{x_1 \mapsto x_2}(v))_k = v_k \quad \text{if } v_k \neq x_1, x_2,$$

$$(\text{swap}_{x_1 \mapsto x_2}(v))_i = \begin{cases} v_j & \text{if } v_i = x_1, v_j = x_2 \text{ and } i < j, \\ v_i & \text{if } v_i = x_1, v_j = x_2 \text{ and } i > j, \end{cases}$$

$$(\text{swap}_{x_1 \mapsto x_2}(v))_j = \begin{cases} v_i & \text{if } v_i = x_1, v_j = x_2 \text{ and } i < j, \\ v_j & \text{if } v_i = x_1, v_j = x_2 \text{ and } i > j. \end{cases}$$

We refer to $\text{swap}_{x_1 \mapsto x_2}$ as the x_1 *to* x_2 *swap map*. The *swap learning maps* $\mathbf{Sw}_{x,y}^A : \mathbf{S}_P \to \mathbf{S}_P$ for player A is given by

$$\mathbf{Sw}_{x,y}^A(P) = \text{swap}_{x \mapsto (\pi_A(x), \pi_B(y))}(P).$$

One can show that if players are rational and perfectly observe each other's actions, then the misperception function (1) under swap learning does not increase [12].

Next, suppose players A and B sequentially take actions such that the hypergame outcomes are $\sigma = (x_1, \ldots, x_n)$. Let us denote by $\mathcal{O}_{BA}^{\sigma}$ and $\mathcal{O}_{AB}^{\sigma}$ the observation sets of players A and B, respectively, associated to σ. It is easy to observe that player A can make her preference vector compatible with $\mathcal{O}_{BA}^{\sigma}$ by executing a composition of swap updates, denote by $\mathbf{Sw}_{\mathcal{O}_{BA}^{\sigma}}^A$, each of which compatible with a binary relation in $\mathcal{O}_{BA}^{\sigma}$. We refer to this mechanism as

the swap update mechanism for A compatible with the observation set. Note that by its definition, $\mathbf{Sw}_{\mathcal{O}_{BA}^{\sigma}}^A$ also does not increase the misperception function. $\mathbf{Sw}_{\mathcal{O}_{AB}^{\sigma}}^B$ can be constructed similarly.

3) Modified swap learning: It is interesting to note that the preference vectors obtained using swap learning may be inconsistent with the stability properties of the outcomes as determined by the actions of other players. When faced with such situation, a player has to determine if this inconsistency is due to her own misperception or the misperception of the opponent. The modified swap learning method belongs to the first class and is defined as follows.

Consider a 1-level hypergame between players A and B. Suppose B takes an action that changes the outcome from $x_1 \in \mathbf{S}_{\text{outcome}}$ to $x_2 \in \mathbf{S}_{\text{outcome}}$, where $x_1 \succ_{P_{BA}} x_2$. The *modified swap learning map* $\mathbf{MSw}_{x_1, x_2}^A : \mathbf{S}_P \to \mathbf{S}_P$ is

- if x_1 perceived as unstable for B in $(P_{AA}, \mathbf{Sw}_{x_1, x_2}^A(P))$ and x_2 is perceived as sequentially rational for B in $P_{AA}, \mathbf{Sw}_{x_1, x_2}^A(P)$, then

$$\mathbf{MSw}_{x_1, x_2}^A(P) = \mathbf{Sw}_{x_1, x_2}^A(P),$$

- if x_1, x_2 are perceived as unstable for B in $(P_{AA}, \mathbf{Sw}_{x_1, x_2}^A(P))$, then

$$\mathbf{MSw}_{x_1, x_2}^A(P) = \mathbf{Sw}_{y, x_1}^A \circ \mathbf{Sw}_{x_1, x_2}^A(P),$$

where $y \in \mathbf{S}_{\text{outcome}}|_{\pi_A(x_2)}$ is the outcome with the highest rank, with respect to $\mathbf{Sw}_{x_1, x_2}^A(P)$, which is sequentially rational in $(P_{AA}, \mathbf{Sw}_{x_1, x_2}^A(P_{BA}))$.

- if x_1, x_2 are perceived as sequentially rational for B in $(P_{AA}, \mathbf{Sw}_{x_1, x_2}^A(P))$, then

$$\mathbf{MSw}_{x_1, x_2}^A(P) = \mathbf{Sw}_{x_1, z}^A \circ \mathbf{Sw}_{x_1, x_2}^A(P),$$

where $z \in \mathbf{S}_{\text{outcome}}|_{\pi_B(x_2)}$ is the outcome with the highest rank, with respect to $\mathbf{Sw}_{x_1, x_2}^A(P)$, which is an improvement from x_1 for player A in $(P_{AA}, \mathbf{Sw}_{x_1, x_2}^A(P_{BA}))$

Modified swap learning guarantees that the stability properties determined by the resulting perceptions are fully consistent with the actions of other players. However, this method may increase the value of the misperception function (1) if the misperception is due to the opponent.

III. PROBLEM STATEMENT

Consider a 1-level hypergame between A and B and suppose both players use the swap update mechanism compatible with their observation set about each other. Consider the resulting dynamical system on the perceptions,

$$P_{BA}(l + 1) = \mathbf{Sw}_{\mathcal{O}_{BA}(l)}^A(P_{BA}(l)), \quad (2a)$$

$$P_{AB}(l + 1) = \mathbf{Sw}_{\mathcal{O}_{AB}(l)}^A(P_{AB}(l)). \quad (2b)$$

Here, $\mathcal{O}_{BA}(l)$ and $\mathcal{O}_{AB}(l)$, respectively, denote the observation sets of players A and B at round $l \in \mathbb{Z}_{\geq 0}$ and $P_{BA}(0) = P_{BA}$ and $P_{AB}(0) = P_{AB}$ are, respectively, the initial A and B's perceptions about their opponent. Under the

evolution prescribed by (2), one can show that the repeated play of the hypergame converges to an equilibrium [13].

Note that, by definition, an equilibrium of a 1-level hypergame H^1 need not be an equilibrium of any of its associated 0-level hypergames, H_A^0 and H_B^0. Therefore, the equilibrium to which the hypergame converges to might not be consistent with the perception of individual players. For instance, A, according to her hypergame H_A^0 might expect B to move away from the equilibrium of H^1, even though B will actually not do so. This inconsistency makes players aware of the existence of some misperception that remains in their beliefs. Our objective in this paper is to study to what extent players can further explore the hypergame in order to arrive at fully consistent equilibria. We are particularly interested in methods that, unlike the modified swap learning strategy described in Section II-C.3, do not increase the misperception. Assuming players can afford to take some action within certain restricted class and move away from inconsistent equilibria, we are interested in answering the following questions

(i) under what conditions are these actions guaranteed to reduce the misperception about the opponent?

(ii) what is the asymptotic behavior of the resulting dynamical system on the perceptions? Do they converge to an equilibrium?

IV. EXPLORATORY ACTIONS

In this section, we introduce the notion of exploratory action and characterize its properties. We assume that players are playing sequentially, one after each other. We focus on player A's hypergame. All the notions can be established similarly for player B. In the rest of this paper, for simplicity of the statements, we assume that all the preorders are strict.

Let $\mathbf{Eq}_A^{\text{In-eq}}(H^1) = \mathbf{Eq}(H^1) \cap (\mathbf{S}_{\text{outcome}} \setminus \mathbf{Eq}(H_A^0))$ denote the set of equilibria of H^1 which are inconsistent with A's perceptions. By definition, these equilibria of H^1 are perceived by A as sequentially rational for her but unstable for B. Figure 1 illustrates this notion.
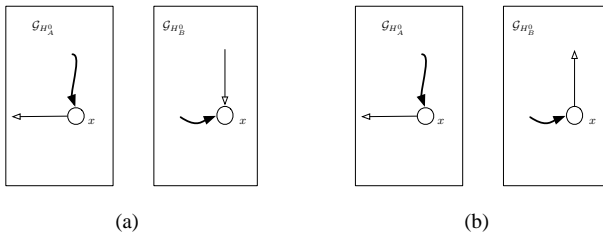

(a)   (b)

Fig. 1. The outcome x is an equilibrium of H^1 which is inconsistent with (a) A's perception (a) and (b) both A and B's perception. In each case, the arrows correspond in bold correspond to the actions for herself that the player perceives as sanction-free. The other arrows correspond to what the player perceives are sanction-free actions of her opponent.

We refer to the set of actions that players are allowed to take from inconsistent equilibria as *exploratory* and define them next.

Definition 4.1: (**Exploratory actions**): Consider the hypergame H^1 between players A and B. An action of player A from $x \in \mathbf{Eq}_A^{\text{In-eq}}(H^1)$ to $y \in \mathbf{S}_{\text{outcome}}|_{\pi_B(x)}$ is *exploratory* if $y \succ_{\mathrm{P}_{AA}} x$.

According to this definition, if an inconsistent equilibrium is rational for A, then this player has no exploratory action. Although an exploratory action leads to an outcome of the game that is more preferred by A, this player perceives at least a sanction of B against such improvement; thus exploratory actions, in general, can be costly. If such actions are permitted to A, then the H-digraph $\mathcal{G}_{H_A^0}$ has to be augmented by adding the set of edges

$$\mathcal{E}_A^{\text{exp}} = \{(x,y) \in \mathbf{S}_{\text{outcome}} \times \mathbf{S}_{\text{outcome}} \mid x \in \mathbf{Eq}_A^{\text{In-eq}}(H^1),$$
$$y \in \mathbf{S}_{\text{outcome}}|_{\pi_B(x)}, \ y \succ_{\mathrm{P}_{AA}} x\}.$$

We term the digraph $\widetilde{\mathcal{G}}_{H_A^0} = (\mathbf{S}_{\text{outcome}}, \mathcal{E}_{H_A^0} \cup \mathcal{E}_A^{\text{exp}})$ the *exploratory H-digraph* of A. The following result characterizes the basic properties of this digraph and is a consequence of Theorem 2.1.

Lemma 4.2: (**The exploratory H-digraph is weakly acyclic**): The digraph $\widetilde{\mathcal{G}}_{H_A^0}$ is weakly acyclic.

Note that the digraph $\widetilde{\mathcal{G}}_{H_A^0}$ is unknown to the player, unlike $\mathcal{G}_{H_A^0}$. This is because she does not know what the equilibria of H^1 are a priori.

A. Repeated play when only one player explores

Here, we analyze the convergence properties of the repeated play of the hypergame when only one player uses exploratory actions. The following result shows that, in this case, convergence to consistent equilibria for the player that uses exploratory actions is guaranteed.

Theorem 4.3: (**Convergence of repeated play of hypergames when only one player explores**): Consider a 1-level hypergame with two players A and B, where players play sequentially and update their perceptions using swap update mechanism compatible with their observation set. Suppose that only A uses exploratory actions. Then the repeated play of the hypergame converges to an equilibrium which is either consistent with A's perception or rational for A.

Proof: Since there are no exploratory actions from a rational outcome, the result is immediate if the play arrives at an outcome which is both an equilibrium of H^1 and rational for A. Assume this is not the case. Let us show that the repeated play of this hypergame will not converge to a cycle, i.e., players will not repeatedly play the actions given by a sequence $\sigma_{\mathcal{C}} = (x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}, x_i)$, $k \in \mathbb{Z}_{\geq 3}$. Without loss of generality, we assume that the action from x_{i+k} to x_i is taken by player B. Suppose otherwise, i.e., suppose that the repeated play of the hypergame starting at x_i converges to σ. Since the repeated play of this hypergame under learning and without using exploratory actions does not converge to a cycle [13], player A must have used an exploratory action. Without loss of generality, let us assume that the action that changes the outcome form x_i to x_{i+1} is exploratory. Since B is using the swap update mechanism

compatible with their observation set and does not explore, one can conclude that $x_{i+k} \succ_{P_{BB}} x_{i+1}$. However, since B has also observed the action of player A from x_i to x_{i+1}, taking the action that changes the outcome to x_{i+k} from x_i by B implies that $x_{i+1} \succ_{P_{BB}} x_{i+k}$, which is a contradiction.

Since the repeated play does not converge to a cycle and players use swap update mechanism compatible with their observation sets, unless the repeated play of the hypergame stops in an equilibrium of the evolved hypergame which is either perceived as rational or an equilibrium by A, eventually, for $x \in \mathbf{S}_{\text{outcome}}$ and $y \in \mathbf{S}_{\text{outcome}}|_{\pi_A(x)}$ we have $x \succ_{P_{BA}} y$ iff $x \succ_{P_{BB}} y$. Similarly, for $z \in \mathbf{S}_{\text{outcome}}|_{\pi_B(x)}$ we have $x \succ_{P_{AB}} z$ iff $x \succ_{P_{AA}} z$. Thus the repeated play of evolved hypergame is the same as the repeated play of the 0-level hypergame $\bar{H}^0 = (P_{AA}, P_{BB})$. Since the H-digraph associated to $\bar{H}^0$ is acyclic, Theorem 2.1, the repeated play converges to an equilibrium, as claimed. ∎

B. Repeated play when both players explore

The next natural step is to investigate the convergence of repeated play to an equilibrium when both players use exploratory actions. The following observation illustrates how cycles may exist in this case.

Remark 4.4: **(Cycles may exist when both players explore):** Consider a 1-level hypergame with players A and B. Assume that the sequence $\sigma = (x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k})$, where $k \in \mathbb{Z}_{\geq 3}$ and is odd, is the result of the first k rounds of play of this hypergame, where

- $x_i \in \mathbf{Eq}_A^{\text{In-eq}}(H^1)$ (A explores from this outcome);
- $x_{i+1} \in \mathbf{Eq}_B^{\text{In-eq}}(H^1)$ (B explores from this outcome);
- $x_{i+k} \in \mathbf{S}_{\text{outcome}}|_{\pi_A(x_i)}$.

Suppose that x_i is a sanction-free improvement from x_{i+k} for B in H_B^0. Note that this is not in contradiction with the actions taken by this player in σ. This is because the action taken by B from x_{i+1} to x_{i+2} is exploratory and thus, unlike in the proof of Theorem 4.3, one *cannot* conclude that $x_{i+k} \succ_{P_{BB}} x_{i+1}$. Thus executing σ followed by the action $\pi_B(x_i)$ of player B results in a cycle, which we denote by σ_C. Note that when one of the actions of B in σ is inconsistent with P_{BA}, A learns, and therefore, having taken the initial exploratory action is beneficial for A. However, it can be the case that the perceived preference vectors of players are such that no two outcomes are swapped during the execution of σ_C. •

Our observation in Remark 4.4 shows that if players persistently explore a cycle, the repeated play of the hypergame may never converge to an equilibrium. The source of this issue is that players cannot identify which actions of their opponent have been exploratory. In order to make this point precise, we introduce some useful notions. Let S_A denote the set of actions available to A. Given an action $a \in S_A$ and a preference vector P associated to a subset $\mathbf{U} \subseteq \mathbf{S}_{\text{outcome}}$, we assign $\text{rank}_a^{\mathbf{U}}(y, P) \in \mathbb{R}_{>0}$ to each outcome $y \in \mathbf{U}|_a$ such that

$$\text{rank}_a^{\mathbf{U}}(y, P) > \text{rank}_a^{\mathbf{U}}(z, P) \quad \text{iff} \quad y \succ_{P_{BA}} z$$

for all $y, z \in \mathbf{U}|_a$. We use the set $\{1, \ldots, |\mathbf{U}|_a|\}$ to rank the outcomes of $\mathbf{U}|_a$. The *relative misperception function* $\mathcal{R}_{BA}^{\mathbf{U}} : \mathbf{U}_P \to \mathbb{R}_{\geq 0}$ of A about B's game, where $\mathbf{U}_P \subseteq \mathbf{S}_P$ associated to $\mathbf{U}$, is defined then by

$$\mathcal{R}_{BA}^{\mathbf{U}}(P) = \sum_{\substack{a \in S_A \\ z \in \mathbf{U}|_a}} |\text{rank}_a^{\mathbf{U}}(z, P_{BB}) - \text{rank}_a^{\mathbf{U}}(z, P)|. \quad (3)$$

The relative misperception function measures how a player ranks the outcomes, on a subset of $\mathbf{S}_{\text{outcome}}$, which share the same action of the opponent. Similar to the misperception function, the relative misperception function is also monotonically decreasing. Clearly

$$\mathcal{L}_{BA}(P) = 0 \Rightarrow \mathcal{R}_{BA}^{\mathbf{S}_{\text{outcome}}}(P) = 0.$$

However, the reverse direction is not necessarily true.

Next, let σ_C be a cycle formed by the repeated play of the hypergame when both players use exploratory actions, see e.g., Remark 4.4. Given P, we let $P|_{\sigma_C} = (x_{i_1}, \ldots, x_{i_K})^T$, with $K = |\sigma_C| - 1$, denote the preference vector P restricted to σ_C, i.e.,

- x_{i_j} appears in σ_C, for all $j \in \{1, \ldots, K\}$, and
- $x_{i_j} \succ_{P|_{\sigma_C}} x_{i_k}$ iff $x_{i_j} \succ_P x_{i_k}$, $j, k \in \{1, \ldots, K\}$.

We have the following result.

Proposition 4.5: **(The relative misperception vanishes along the cycles of repeated play):** Consider a 1-level hypergame H^1 between A and B, where players play sequentially, update their perceptions using swap update mechanism compatible with their observation sets, and can take exploratory actions. If the repeated play of the hypergame starting from $x \in \mathbf{S}_{\text{outcome}}$ reaches x again, forming a cycle σ_C, then

$$\mathcal{R}_{BA}^{\mathbf{U}}(\mathbf{Sw}_{\mathcal{O}_{BA}^{\sigma_C}}^A(P_{BA})|_{\sigma_C}) = 0,$$
$$\mathcal{R}_{AB}^{\mathbf{U}}(\mathbf{Sw}_{\mathcal{O}_{AB}^{\sigma_C}}^B(P_{AB})|_{\sigma_C}) = 0,$$

where $\mathbf{U}$ is the subset of $\mathbf{S}_{\text{outcome}}$ associated to the outcomes that appear in σ_C.

Proof: We prove the first statement. The second statement follows similarly. Since players are using the swap update mechanism compatible with their observation sets, after traveling the cycle, for any outcomes $x, y \in \mathbf{U}$ such that $y \in \mathbf{S}_{\text{outcome}}|_{\pi_A(x)}$, $x \succ_{P_{BA}} y$ iff $x \succ_{P_{BB}} y$; thus, by definition of $\mathcal{R}_{BA}^{\mathbf{U}}$, the result follows. ∎

This result precisely captures the part of the misperception that can be reduced by using exploratory actions. The following result, in contrast, focuses on the misperception that may remain even after exploration.

Proposition 4.6: **(The remaining part of the misperception function):** Consider a 1-level hypergame H^1 between A and B, where players play sequentially, update their perceptions using swap update mechanism compatible with their observation sets and can take exploratory actions. Assume that

$$\mathcal{R}_{BA}^{\mathbf{S}_{\text{outcome}}}(P_{BA}) = \mathcal{R}_{AB}^{\mathbf{S}_{\text{outcome}}}(P_{AB}) = 0. \quad (4)$$

Suppose A takes an exploratory action from $x \in \mathbf{S}_{\text{outcome}}$. If the outcome of the hypergame comes back to x by an action of B in any future round of play, then A learns that B has also explored. Moreover, for each action of B from x, perceived as sanction-free by A, player A is indecisive between

  (i) executing modified swap learning, or
 (ii) labeling the action as exploratory.

    Proof: First, we show that had B not explored, repeated play will not return to the outcome x. Suppose otherwise, i.e. B has not explored and the repeated play of the hypergame is $\sigma_{\mathcal{C}} = (x_i, x_{i+1}, \ldots, x_{i+k}, x_i)$, where $x_i = x$, $k \in \mathbb{Z}_{\geq 3}$ and the action x_i to x_{i+1} is an exploratory action by A. Since B has not explored, all his actions are perceived by her as sanction-free. By assumption (4), one can conclude that $x_{i+k} \succ_{\mathrm{P}_{BB}} x_{i+1}$. Also, the action of B from x_{i+k} to x_i implies that $x_{i+1} \succ_{\mathrm{P}_{BB}} x_{i+k}$, a contradiction.

    To establish the other part, let us consider then an action of player B from $x \in \mathbf{S}_{\text{outcome}}$ to an outcome $y \in \mathbf{S}_{\text{outcome}}|_{\pi_A(x)}$. Note that, by the assumption on zero relative misperception in (4), this action is perceived as an improvement for B by A. Suppose A perceives x as unstable for B, i.e., A perceives that $w \succ_{\mathrm{P}_{BA}} x$, where w is an improvement from y for A. In other words, the preference vectors of A look like

$$\mathrm{P}_{AA} = (\cdots w \cdots y \cdots)^T,$$
$$\mathrm{P}_{BA} = (\cdots y \cdots w \cdots x \cdots)^T.$$

Since, again by the assumption on zero relative misperception in (4), B perceives correctly that $w \succ_{\mathrm{P}_{AB}} y$, two cases can happen:

  (i) x is, in fact, perceived as sequentially rational for B by B, i.e.,

$$\mathrm{P}_{AB} = (\cdots w \cdots y \cdots)^T,$$
$$\mathrm{P}_{BB} = (\cdots y \cdots x \cdots w \cdots)^T.$$

    Thus, the inconsistency is due to player A's misperception. This misperception can be removed by swapping x and w. This precisely corresponds to the operation $\mathbf{MSw}_{x,y}^A(\mathrm{P}) = \mathbf{Sw}_{x,w}^A \circ \mathbf{Sw}_{x,y}^A(\mathrm{P})$.

 (ii) x is perceived as unstable for B by B, and thus A's perception about B is correct and the action of B that changes the outcome x to y is exploratory.

This concludes the proof. ∎

    The result of Proposition 4.6 implies that when both players use exploratory actions, the repeated play of hypergame may fail to converge, unless players have some capability or procedure to distinguish what actions of their opponents are exploratory.

V. Conclusions

    This paper has considered the repeated play of 1-level hypergames with two players. We have defined the notion of inconsistent equilibrium to capture those equilibria of the hypergame for which at least one player expects the other to move away from. The existence of such equilibria signals the persistence of some misperception about the game the players are involved in. We have identified a class of actions, termed exploratory, that players can take from inconsistent equilibria in order to further decrease the misperception. We have shown that, if exploration is allowed and only one player does it, then the repeated play of the hypergame arrives at a consistent equilibrium or an outcome rational for the player. In contrast, when both players use exploratory actions, we have shown that the repeated play may finish in a cycle. The introduction of the notion of relative misperception function has allowed us to characterize the part of the misperception which is guaranteed to vanish along such cycles. Future work will consider reward-based mechanisms for exploration, where players obtain higher reward for exploratory actions which lead to learning. We also plan to study the impact of including the cost of exploratory actions, the influence of these actions on deception, and the collective exploration by groups of players.

References

[1] J. C. Harsanyi, "Bargaining in ignorance of the opponents' utility function," *Journal of Conflict Resolution*, vol. 6, no. 1, pp. 29–38, 1962.

[2] J. C. Harsanyi, "Games with incomplete information played by 'Bayesian' players, i-iii," *Management Science*, vol. 50, no. 12 Supplement, pp. 1804–1817, 2004.

[3] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[4] P. G. Bennett, "Towards a theory of hypergames," *OMEGA*, vol. 5, no. 6, pp. 749–751, 1977.

[5] N. M. Fraser and K. W. Hipel, *Conflict Analysis: Models and Resolutions*. Systems Science and Engineering, Oxford, UK: Elsevier, 1984.

[6] M. Wang, K. W. Hipel, and N. M. Fraser, "Solution concepts in hypergames," *Applied Mathematics and Computation*, vol. 34, pp. 147–171, 1989.

[7] R. Vane and P. Lehner, "Using hypergames to increase planned payoff and reduce risk," *Autonomous Agents and Multi-Agent Systems*, vol. 5, pp. 365–380, 2002.

[8] P. G. Bennett, "Hypergames, the development of an approach to modeling conflicts," *Futures*, vol. 12, pp. 489–507, 1980.

[9] B. L. Slantchev, "The principle of convergence in wartime negotiations," *American Political Science Review*, vol. 97, no. 4, pp. 621–632, 2003.

[10] J. T. House and G. Cybenko, "Hypergame theory applied to cyber attack and defense," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX*, vol. 7666, (Orlando, Florida), pp. 766604–766611, April 2010.

[11] D. Teneketzis and D. A. Castañón, "Informational aspects of class of subjective games of incomplete information: static case," *Journal of Optimization Theory & Applications*, vol. 54, no. 2, pp. 413–422, 1987.

[12] B. Gharesifard and J. Cortés, "Evolution of players' misperceptions in hypergames under perfect observations," *IEEE Transactions on Automatic Control*, vol. 57, no. 7, 2012. To appear.

[13] B. Gharesifard and J. Cortés, "Learning of equilibria and misperceptions in hypergames with perfect observations," in *American Control Conference*, (San Francisco, California), pp. 4045–4050, 2011.

[14] D. M. Kreps and R. Wilson, "Sequential equilibria," *Econometrica*, vol. 50, no. 4, pp. 863–894, 1982.

[15] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. SIAM, 2 ed., 1999.

[16] N. Biggs, *Algebraic Graph Theory*. Cambridge University Press, 2 ed., 1994.