Preservation of Misperceptions – Stability Analysis of Hypergames

Yasuo Sasaki, Kyoichi Kijima
Department of Value and Decision Science
Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552 JAPAN

ABSTRACT

The present paper tries to model how some kinds of misperceptions of agents are preserved in a decision making situation where multiple agents are involved. We use hypergame model which is a theoretical framework to deal with agents who may misperceive situations (Bennett et al., 1979). After each play of hypergame, agents may update their perceptions based on the result, that is, the structure of the hypergame may change. However, in some case, they may not, and the hypergame is 'stable', that is, their misperceptions are preserved. To discuss stability of hypergames, we newly define a solution concept what we call stable hyper Nash equilibrium. Using these ideas, we analyze the stability. To demonstrate change in perceptions of agents, we consider agent-based intrinsic motivation. Although we provide general foundation for discussion, we analyze a game called battle of sexes as an example case.

Keywords: hypergame theory, misperception, stable hyper Nash equilibrium, intrinsic motivation, battle of sexes.

1. Introduction

Hypergame theory is a theoretical framework to deal with agents who may misperceive a decision making situation (Bennett et al., 1979). In a hypergame situation, agents are assumed to perceive the situation subjectively, and the subjectivity may cause their misperceptions.

Hypergame theory was proposed to address a strict assumption of classical game theory, which generally requires complete information. It has been assumed in game theory that all agents fully understand the situation and thereby they all see the same game. The assumption has restricted the possibility of analysis of real world situation because imperfect knowledge or different perception affects decision making quite often. This perceptual problem has been pointed out in several ways since early time in the development of game theory. One of the most notable achievements is analysis in games with incomplete information by Harsanyi (1967). He provided a mathematical framework of games including subjective probability distributions by 'Bayesian' players. Although Bayesian games and hypergame theory share common philosophy in some aspects, we believe that suitable model for analysis depends on the situation.

In the present research in the framework of hypergame theory, we will conduct stability analysis of dynamically changing games. After agents play a hypergame, they may encounter outcomes that they have never expected because of their misperceptions. In those cases, intrinsic motivations to improve their own perceptions would arise from the surprises. These agent-based motivations may lead to learning of situations or communications among the agents and to changes of their perceptions, that is, agent-based motivation may bring about structural change of the hypergame. This process would never occur in classical game situations.

However, even if there exist misperceptions, some types of hypergames may remain unchanged and be stable in the sense that misperceptions of agents are preserved. Our main concern is with preservation of misperceptions, which has not been discussed rigorously so far, though Kaneko and Kline (2002) argue this problem epistemologically based on game theoretical analysis.

Our main purpose of the present paper is to formalize such stability of structure of hypergames and to give reasonable explanations to the stability from a viewpoint of agent-based intrinsic motivations for improving perceptions.

2. Hypergame Model and Its Solution Concepts

In this section, we introduce some formal definitions of models and concepts as a preparation for our study.

Before that, we will introduce an intuitive idea of noncooperative game and explain about our research motivation more clearly by using a well-known model.

Noncooperative game is a game which describes a decision making situation where the agents just seek their own payoff without communication. One of such typical models is the following "Battle of sexes". The story of this game is as follows:

> A boy and a girl, say i and j respectively, are going to have a date. Now, they have to decide where to go. Their options are to go to see a boxing game (B) and music concert (M). They, however, have to make a decision independently under some reason. The boy prefers boxing, while the girl prefers music concert. However they both prefer having a date together to going out by oneself.

The situation can be formulated in terms of matrix shown by Fig. 2.1, where the formal definitions will be given later.

$i \diagdown j$	B	M
B	2, 1	0, 0
M	0, 0	1, 2

Fig. 2.1 Battle of sexes

The matrix implies that both two agents completely understand the opponent's strategies and payoff. Although this is requirement of the classical game model, it is not always guaranteed in real world. For example, when it is not long since they got to know each other, they might not know about the opponent well. They might misunderstand the opponent's preference. If they play the game under misperceptions, they might encounter unexpected outcome and have intrinsic motivation to improve their perception. Or they might not encounter unexpected outcome in spite of the misperceptions and not have such a kind of motivations. We discuss this problem formally through the paper.

2.1 Noncooperative Game and Nash Equilibrium

Since the idea of hypergames are derived from game theory, especially from noncooperative game theory, we begin with formulating the game model and its central solution concept, Nash equilibrium.

A decision making situation in which multiple agents are involved and compete each other for their own utility can be formally modeled in terms of noncooperative game as definition 2.1.

Definition 2.1 (noncooperative games)

A noncooperative game is given by (N, S, u), where:

- $N = \{1, ..., n\}$ is a set of agents.
- $S = \times_{i \in N} S_i$ is a set of strategies, where S_i is a set of strategies of agent *i*. We call $s \in S$ an outcome.
- $u = (u_i)_{i \in N}$ is a profile of utility functions, where $u_i : S \to R$ is agent *i*'s utility function. Let $x, y \in S$, suppose $u_i(x) > u_i(y)$ iff agent i prefers outcome x to y, and $u_i(x) = u_i(y)$ iff agent *i* is indifferent between x and y.

The classical noncoperative games assume that every agent has common knowledge about the structure of the game, which means it supposes every agent knows all the components of the game as shown in definition 2.1 completely.

Nash equilibrium is a most well-known solution concept for noncooperative games and hypergame solutions concepts we use in our study are based on its idea. A Nash equilibrium is such an outcome of the game that every agent has no incentive to deviate from it as long as the others do not change their strategies.

Definition 2.2 (Nash equilibrium)

$s^* = (s_i^*, s_{-i}^*) \in S$ is a Nash equilibrium of a noncoopereative game $G = (N, S, u)$ iff $\forall i \in N$, $\forall s_i \in S_i$, $u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*)$.

In the above definition we call s_i* i's Nash strategy. For a noncooperative game G, let us denote the set of Nash equilibria by $N(G)$, while the set of Nash strategies of agent i by $N(G)_i$.

2.2 Hypergame Model

As mentioned in Introduction, hypergame theory provides a theoretical framework which contains critically different assumptions from those of noncooperative game theory with respect to perceptions of agents. Unlike in noncooperative game model, in a hypergame model, each agent is assumed to construct his/her 'subjective game' based on how he perceives the situation. Although hypergame theory has been examined in several ways (Inohara, 2002; Wang et al., 1988), we focus on the class of 'simple hypergames' because it provides the most basic but essential hypergame model.

Definition 2.3 (simple hypergame)
A simple hypergame H is given by $(N, (G^i)_{i\in N})$, where:

- $N = \{1, ..., n\}$ is a set of agents involved in the situation
- $G^i = (N^i, S^i, u^i)$ is the subjective game of agent i, where :
  - ➢ N^i is a set of agents perceived by agent i.
  - ➢ $S^i = \times_{j\in N^i} S^i_j$ is a set of strategies perceived by agent i, where S^i_j is a set of strategies of agent j perceived by agent i.
  - ➢ $u^i = (u^i_j)_{j\in N^i}$ is a profile of utility functions perceived by agent i, where $u^i_j : S^i \to R$ is agent j's utility function perceived by agent i.

In a simple hypergame, it is assumed that each agent perceives the situation subjectively in a form of a complete information game (subjective game). Agents consider their own perceptions are common knowledge respectively, so that they plays based on different perceptions but no one knows this.

The whole game, the list of subjective games of all agents, is called a hypergame H, that is, $H = (G^1, ..., G^n)$, and nobody except 'God' can know H. Note that the superfix in Definition 2.3 indicates an agent who perceives the component. In the subsequent discussions, we use the word 'hypergame' in the sense of simple hypergame. Although we can consider mixed strategies in a similar way as classical game theory does (Sasaki et al., 2007), we deal only with pure strategies and ordinal utility in the present paper.

Next, we define such an 'imaginary' game that we would obtain by assuming that every misperception has been eliminated and every agent perceives his/her own components, strategies and payoff correctly. We call such a game a base game.

Definition 2.4 (base game)
$BG = (N, S, u)$ is a base game which is a noncooperative game generated from a hypergame $H = (N, (G^i)_{i\in N})$ satisfying the following conditions:

- $S_i = S^i_i$
- $u_i = u^i_i$

A base game is a noncooperative game where each agent's strategy set and payoff are ones in his/her own subjective game in the hypergame. We should notice that base game is not well-defined as a noncooperative game if the agent cannot define his/her payoff on outcomes including the other's strategy that he/she dose not recognize when there exist such strategies. In the subsequent analysis, we focus only on hypergames in which the agents may misperceive only the others' payoff. We call this class of hypergames 'perturbed hypergame'. We can always define base game from perturbes hypergame.

Definition 2.5 (perturbed hypergame)
A hypergame $H = (N, (G^i)_{i \in N})$ is called a perturbed hypergame of base game $G = (N, S, u)$, iff we have $\forall i, j \in N,\ N^i = N,\ S^i_j = S_j$.

In a perturbed hypergame, all agents correctly perceive the set of agents involved in the situation and every agent's strategy set, while they may misperceive the other's payoff

2.3 Solution Concepts in Hyperames

We introduce two types of solution concepts of hypergames: hyper Nash equilibrium and stable hyper Nash equilibrium, where the latter is original of this paper.

Definition 2.6 (hyper Nash equilibrium)
$(s^i_i{*})_{i \in N} \in \times_{j \in N} S^j_j$ is called a hyper Nash equilibrium of a hypergame H iff $\forall i \in N$, $s^i_i{*} \in N(G^i)_i$.

Hyper Nash equilibrium is defined as a profile of such strategies that each agent plays according to his/her Nash strategy in his/her own subjective game (Kijima, 1996). Agent *i*'s strategy in hyper Nash equilibrium, $s^i_i{*}$, is called *i*'s hyper Nash strategy. Let us denote the set of hyper Nash equilibria in hypergame H by $HN(H)$. The following lemma is straightforward from the definition but useful for the calculation.

Lemma 2.1 (set of hyper Nash equilibrium)
In a hypergame H, we have $HN(H) = \times_{i \in N} N(G^i)_i$.

If we assume Nash strategies as decision making discipline of agents, rational outcomes of a hypergame are necessarily hyper Nash equilibrium, so that it is well-suited to predict the outcome of an one-shot hypergame. However, in a hyper Nash equilibrium, it may occur that strategies the others have chosen are different from what the agent expects. For example, let $i, j \in N$ $(i \neq j)$, then *j*'s hyper Nash strategy, $s^j_j{*} \in N(G^j)_j$, might not be included in any Nash equilibria of *i*'s subjective game. In such cases, the structure of the hypergame would change, because agents would 'learn' something from the outcome. They may try to update their perceptions to have more 'correct' understanding or to communicate each other to resolve the misperceptions. Learning of situations is necessarily one of main interest in hypergame study (Takahashi et al., 1999). We cannot find this nature of equilibrium in Nash equilibrium in noncooperative games, where expectations about equilibrium by agents are consistent with each other.

Now we introduce another solution concept, stable hyper Nash equilibrium, to treat changes of equilibria.

Definition 2.7 (stable hyper Nash equilibrium)
$(s_i^i{**})_{i\in N} \in \times_{j\in N} S_j^j$ is called a stable hyper Nash equilibrium of a hypergame H iff $^\forall k \in N$, $(s_i^i{**})_{i\in N} \in N(G^k)$.

Stable Nash equilibrium is as such a profile of strategies that is Nash equilibrium in every subjective game. Let us denote the set of stable hyper Nash equilibria of a hypergame H by $SHN(H)$. It is given by the intersection of the sets of Nash equilibria perceived by each agent in similar way as Lemma 2.1.

Lemma 2.2 (set of stable hyper Nash equilibrium)
In a hypergame H, $SHN(H) = \cap_{i\in N} N(G^i)$.

At stable hyper Nash equilibrium, strategies the others have chosen are necessarily consistent with an agent's anticipation so that the agent has no incentive to update his/her perception or to communicate with other agents for resolving misperceptions. This solution concept not only can predict which outcome will occur in an one-shot hypergame similar to hyper Nash equilibrium. But, in contrast to hyper Nash equilibrium, stable hyper Nash equilibrium is once accomplished, it is also consistently 'stable' even when the situation is repeated, as long as there is no change in agents' decision making discipline or perceptions about the game (We will discuss this issue in the next section).

2.4 Example

Using a hypergame shown below (Fig. 2.2) as an example, we illustrate the models and concepts introduced so far.

$i \diagdown j$	X	Y
A	4, 2	1, 3
B	3, 1	2, 4

G^i: i's subjective game

$i \diagdown j$	X	Y
A	2, 4	3, 3
B	1, 1	4, 2

G^j: j's subjective game

Fig. 2.2 Hypergame H

Let us consider a hypergame $H = (N, (G^i)_{i \in N})$, where $N = \{i, j\}$ and G^i and G^j are *i*'s and *j*'s subjective game, respectively. There $S_i^i = S_i^j = \{A, B\}$ and $S_j^i = S_j^j = \{X, Y\}$. Values in the matrix express their ordinal utility. By definition 2.4, the base game BG is generated from H as Fig. 2.3. It is the game which would be obtained if all misperceptions in H disappear.

$i \diagdown j$	X	Y
A	4, 4	1, 3
B	3, 1	2, 2

Fig. 2.3 Base game BG

Equilibria in this hypergame are given as follows:

$N(G^i) = \{(B, Y)\}$;
$N(G^j) = \{(A, X), (B, Y)\}$;
$HN(H) = N(G^i)_i \times N(G^j)_j = \{(B, X), (B, Y)\}$;
$SHN(H) = N(G^i) \cap N(G^j) = \{(B, Y)\}$;
$N(BG) = \{(A, X), (B, Y)\}$.

An outcome (B, X) is a hyper Nash equilibrium but not a stable hyper Nash equilibrium, thus according to our definition, it is not stable, while the other outcome (B, Y) is stable (see 2.3).

2.5 Relationships Among the Solution Concepts

In this section, we examine global relationships among the solution concepts.

First, we have the following proposition with respect to relationship between the two solution concepts in hypergames.

Proposition 2.3 (relation between two equilibria in hypergames)
In a hypergame H, we have $SHN(H) \subseteq HN(H)$.

Proof:
It is obvious from definition 2.6 and 2.7.

□

Proposition 2.3 says that if there exists a stable hyper Nash equilibrium, it is necessarily a hyper Nash equilibrium. In other words, stable hyper Nash equilibrium is more strict solution concept than hyper Nash equilibrium.

Next, we have the following theorem with respect to relationship between stable hyper Nash equilibrium and Nash equilibrium of the base game.

Theorem 2.4 (stable hyper Nash equilibrium and Nash equilibrium of base game)
For a given hypergame H, let BG be base game which derives from H. Then we have $SHN(H) \subseteq N(BG)$.
Proof:
Let $(s_i^i **)_{i \in N} = (s_1^1 **, ..., s_n^n **) \in \times_{j \in N} S_j^j$ be a stable hyper Nash equilibrium.
By definition 2.7,
$^\forall k \in N$, $(s_i^i **)_{i \in N} \in N(G^k)$
$\Rightarrow {}^\forall k \in N$, in G^k, $^\forall s_k^k \in S_k^k$, $u_k(s_k^k **, s_{-k}^k **) \geq u_k(s_k^k, s_{-k}^k **)$
$\Leftrightarrow {}^\forall k \in N$, in BG, $^\forall s_k \in S_k$, $u_k(s_k **, s_{-k} **) \geq u_k(s_k, s_{-k} **)$, where $s_k ** = s_k^k **$.
$\Leftrightarrow (s_i^i **)_{i \in N} \in N(BG)$
On the other hand, $(s_i^i **)_{i \in N} \in N(BG) \Rightarrow (s_i^i **)_{i \in N} \in SHN(H)$ does not always hold.

□

Theorem 2.4 claims that if there exists a stable hyper Nash equilibrium, then it is necessarily a Nash equilibrium of the base game. It implies that an outcome which is not Nash equilibrium in the base game cannot be stable hyper Nash equilibrium in the hypergame.

Furthermore, we have the following theorem from these results.

Theorem 2.5 (the nonexistence condition of stable hyper Nash equilibrium)
For a given hypergame H, let BG be base game which derives from H. If $HN(H) \cap N(BG) = \phi$, then $SHN(H) = \phi$.

Proof:
It is direct from proposition 2.3 and theorem 2.4.

□

Theorem 2.5 shows that when there exist hyper Nash equilibria in a hypergame, if all of them are not Nash equilibrium in the base game, there does not exist stable hyper Nash equilibrium. An intuitive interpretation of the theorem is that when we anticipate all of outcomes which seem to happen actually (hyper Nash equilibrium) would not happen if all the misperceptions are eliminated, those outcomes are necessary unstable.

Hence the relationships among the solution concepts in a hypergame H can be depicted by Fig. 2.4. 'All outcomes' in the figure means $\times_{i \in N} S_i^i$. We can easily see that the results with respect to the example in 2.4 of course satisfy these relationships.

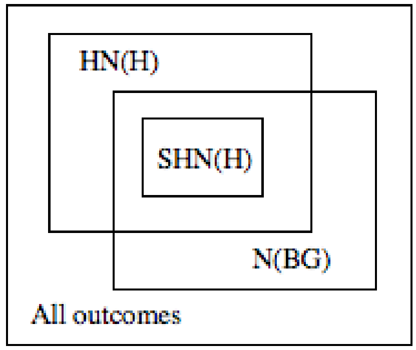

Fig. 2.4 Relationships among the solution concepts

3. Stability Analysis in Battle of Sexes and Preservation of Misperceptions

In this section, by using battle of sexes game as an example, we will illustrate how the structure of hypergames may change and how, in some cases, misperceptions of agents may be preserved in the context introduced in the previous chapter.

3.1 Preparation for analysis

For preparing for our further analysis, we first categorize how the agents' perceptions change into two types, A and B, as follows:

Change of type A is change independent of the agent's intrinsic motivation, while Change of type B is change generated by the agent's intrinsic motivation.

Change of type A may occur even if the agent does not seek it intentionally, while change of type B cannot occur unless he/she seeks it, that is, unless he/she has 'intrinsic motivation' to update his/her perception.

Whether an agent has such an intrinsic motivation or not depends on (a) if the outcome is as he/she expected or not and, if so, (b) if he/she accepts the outcome completely or not (Fig. 3.1). The former is simply related whether or not the outcome is Nash equilibrium in his/her perception. On the other hand, the latter is an 'emotional' issue of agents as will be argued later.

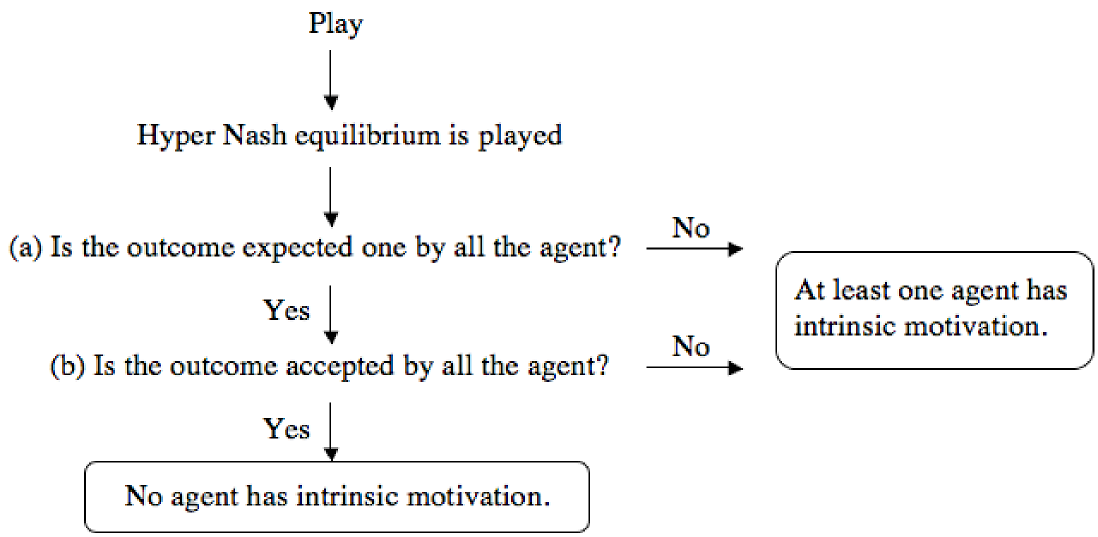

Fig. 3.1 Intrinsic motivation

Change of agents' perceptions is based on some information they can get. We restrict here such information only to that with respect to the opponent's payoff.

Next, we categorize hypergame situations in two cases according to whether there exists stable hyper Nash equilibrium or not. We call these two types of hypergame situations 'unstable hypergame' and 'stable hypergame', respectively as follows:

Definition 3.1 (stable and unstable hypergame)
We say a hypergame H unstable hypergame if we have $SHN(H) = \phi$, while we call a hypergame H stable hypergame if we have $SHN(H) \neq \phi$.

In Fig. 3.1, if the answer to the question (a) is No, the situation is an unstable hypergame, while if it is Yes, the situation is a stable hypergame.

3.2 Stability analysis of battle of sexes

By using the ideas introduced so far, we conduct stability analysis, that is, we analyze in what situation, what type of hypergame is stable or unstable and explain why misperceptions may be preserved in some cases. Although here the analysis is conducted for case of battle of sexes between two agents, the discussion can be generalized.

First, when both the agents perceive the situation correctly and know the opponent's information completely, we can express the situation of the battle of sexes as base game shown by Fig. 3.2. It is the same as $G = (N, S, u)$ of Fig. 2.1, where $N = \{i, j\}$, $S_i = S_j = \{B, M\}$, u_i and u_j are as in the matrix.

$i \diagdown j$	B	M
B	2, 1	0, 0
M	0, 0	1, 2

Fig. 3.2 G: **Battle of sexes (base game)**

The game has two Nash equilibria in the range of pure strategies, i.e. $N(G) = \{(B, B), (M, M)\}$. Although to predict which of the results of the one-shot game will really happen has been a difficult coordination problem in game theory, it is out of our interest of the present paper. Instead we regard this game as a base game and analyze how and why the situation may or may not change, when first date is its perturbed hypergame.

<u>Case1: Unstable hypergame</u>

Now suppose both the agents misunderstand the opponent's preference each other as follows:

> i: "j is not interested in boxing at all.";
> j: "i is not interested in music concert at all."

Then the situation can be formulated in terms of hypergame model by Fig. 3.3. We set $H = (G^i, G^j)$.

$i \diagdown j$	B	M
B	2, 0	0, 1
M	0, 0	1, 2

G^i: i's subjective game

$i \diagdown j$	B	M
B	2, 1	1, 0
M	0, 0	0, 2

G^j: j's subjective game

Fig. 3.3 Battle of sexes (unstable hypergame)

Since Nash equilibrium in each subjective game is given by $N(G^i) = \{(M, M)\}$ and $N(G^j) = \{(B, B)\}$, the unique hyper Nash equilibrium is (M, B). However, there does not exist stable hyper Nash equilibrium in this hypergame. The fact is consistent with theorem 2.5.

When the agents play this hypergame, they both would encounter unexpected outcome. Then it necessarily leads to emergence of intrinsic motivation for improvement of perception in each agent as shown in Fig. 3.1. That is, change of type B would always occur. For example, if it is possible to communicate each other directly, they may try to do so and to find out the opponent's 'real' preference. Even if they cannot have such a direct communication, they would *learn* something from the result and update their perceptions. In this case, i expected that j would surely choose music concert, while she chooses boxing actually. Thus i would reconstruct his perception for the next date taking into account possibility of j's preference for boxing.

Furthermore, we can also consider change of type A in agents' perceptions in this case. For example, one of j's friends may tell i that she prefers having a date to going to a music concert alone, and if he believes the information, he would update his perception about j's payoff. This kind of communication occurs often independent of whether or not i seeks it, that is, independent of his intrinsic motivation.

All of such *interactions or/and learning* by agents may result in change of the structure of the hypergame, so that H may become another hypergame H' ($\neq H$) including different misperceptions from the previous play. If they could communicate each other thoroughly and reach mutual understanding, it may become a normal noncooperative game same as the base game which does not include misperceptions at all (Fig. 3.2). In some cases, agent i may get to perceive the possibility that agent j may misperceive how i perceives the situation. Then, the 'hierarchy of perception' emerges. This kind of hypergame is formulated as n-level hypergames (Wang et al., 1988). In other cases, an agent may get to take into account more than one possibility with respect to what is the 'real' situation. Then, the situation can be regarded as a kind of Bayesian game (Harsanyi, 1967). Anyway when they have second date (Play 2), they would play different situation from the first date (Play1) (See Fig. 3.4).

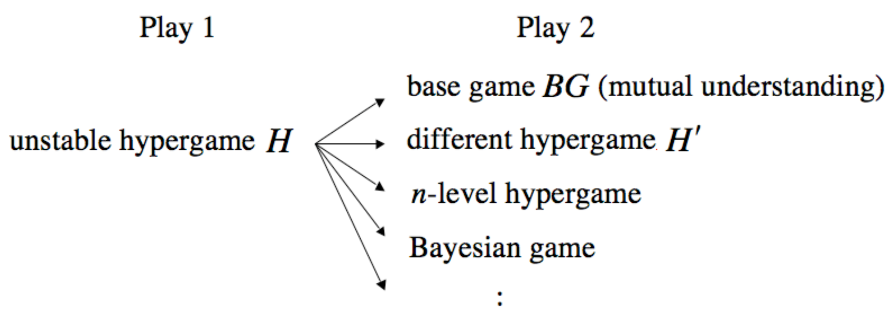

Fig. 3.4 Change of battle of sexes (starting from unstable hyperegame)

<u>Case2: Stable hypergame</u>

Next, suppose both the agents misunderstand the opponent's preference each other as follows:

> i: "j is happiest when she enjoy my hobby together.";
> j: "i is not interested in music concert at all."

i's perception is different from the previous case, while *j*'s perception is same. Then the situation can be formulated in terms of hypergame model as Fig. 3.5. We set $H = (G^i, G^j)$.

$i \setminus j$	B	M
B	2, 2	0, 0
M	0, 0	1, 1

G^i: *i*'s subjective game

$i \setminus j$	B	M
B	2, 1	1, 0
M	0, 0	0, 2

G^j: *j*'s subjective game

Fig. 3.5 Battle of sexes (stable hypergame)

Since Nash equilibrium in each subjective game is $N(G^i) = \{(B,B),(M,M)\}$ while $N(G^j) = \{(B,B)\}$, this hypergame has two hyper Nash equilibria, (B,B) and (M,M). Therefore it can be predicted the result of this hypergame as either of these two outcomes. In contrast to Case 2, this hypegame has a stable hyper Nash equilibrium, which is (B,B). It is quite natural assumption that when there exist multiple Nash equilibria, any agents do not choose such a strategy that leads to Pareto dominated one. According to the assumption, i should choose B, and (B,B), stable hyper Nash equilibrium, is the unique predictable outcome.

In this case, the outcome is expected one for both. Thus intrinsic motivation for improvement of perception does not emerge in neither of the both agents directly from the outcome. This is a crucial difference from unstable hypergame.

However, even if the outcome is expected one for both, the motivation may emerge in an agent in some specific cases. It is because the agent cannot accept the outcome due to some reasons. Then a kind of 'emotion' like frustration arises, and it may provoke motivation to correct perception. This would lead to change in agents' perceptions of type B. For example, *i* may possibly find out that *j* does not look so happy though he has expected (B,B) is the best for her as well and ask her why. Or, seeing *i* does not appreciate at all that *j* comes to the date sacrificing herself to enjoy her own hobby, *j* may become so stressful that she confesses her actual preference to *i*. These processes can change agents' perceptions.

As similar to unstable hypergame, we can also consider the possibility of change of type A in this case. For example, one of *j*'s friends may tell *i* that she prefers music concert to boxing in fact. This information may come independent of his intrinsic motivation.

Considering all of such update of perceptions, possibility of change of the hypergame is similar with unstable hypergame (Refer to Fig. 3.4).

However, as stated above, agents have intrinsic motivation for improvement of perception only in some specific cases where they have emotions to do so, while they always have such motivation in unstable hypergame. Now let us consider the following two assumptions:

(i) There is no information coming to an agent unless he/she seeks it.
(ii) Both agents accept the outcome with no frustration.

Then, since neither of change in perceptions of type A nor B occurs, H would be consistently stable. When they have the second date (Play 2), they would play the same hypergame and choose same strategies as the first date (Play 1). Likewise they will play the same in their third date (Play 3), and forever. In this way, the misperceptions of agents are preserved through all the dates (Fig. 3.6).

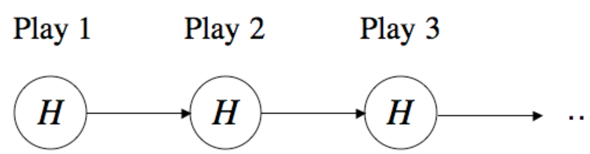

Fig. 3.6 Repeated battle of sexes (stable hypergame)

The two assumptions dose not influence on the changing process of unstable hypergame (Fig. 3.4) because agents always have intrinsic motivation for learning even under the assumptions.

Before closing this section, we would like to point out one thing regarding misperceptions based on our discussion so far. Misperception is often problematic. In the case of the example in section 2.4, elimination of all misperceptions leads to Pareto optimal equilibrium, (A, X), while the current stable hyper Nash equilibrium, (B, Y), is not Pareto optimal. To achieve this, for example, intervention by a third party is important. Otherwise agents may remain to stay at the not-better outcome for both.

However, misperception is not always problematic, either. Suppose the case of Fig. 3.6. The boy and girl, i and j, may be happy and continue to be along with each other. On the other hand, if they know the opponent's actual preference correctly, they may become to have trouble with each other to decide where they have a date.

4. Conclusions

We examined how perceptions of agents are changed and why some sorts of misperceptions are preserved in terms of hypergame model mainly using the game of battle of sexes as an example. We defined formally the hypergame model and introduced a new solution concept, stable hyper Nash equilibrium, which is essential for analyzing the stability of hypergames. Then we clarified their entire relationship. Based on these theoretical foundations, we conducted intensive stability analysis. To argue change in perception of agents, we focused on agent-based intrinsic motivation.

To discuss more about the changing factors of perceptions we stated in section 3 like intrinsic motivation, emotion, communication and learning is one of our main future works.

REFERENCES

Bennett, P. G. (1980), "Hypergames: Developing a Model of Conflict," *Futures*, 6, 489-507.

Bennett, P. G. and M. R. Dando (1979), "Complex Strategic Analysis: A Hypregame Study of the Fall of France," *Journal of the Operational Research Society*, 30(1), 23-32.

Harsanyi, J. C. (1967), "Games with Incomplete Information Played by Bayesian players," *Management Science*, 14, 159-182, 320-334, 486-502.

Inohara, T. (2002), *Emotions and Perceptions (Irrational Strategies in Competition and Society II)*, Keiso-Shobo, Tokyo (in Japanese)

Kaneko, M. and J. J. Kline (2002), "False Beliefs and Game Theory: Implications from the Japanese Comic Story Konnyaku Mondo", Discussion Paper Series No.970, Institute of Policy and Planning Sciences, University of Tsukuba.

Kijima, K. (1996), "An Intelligent Poly-agent Learning Model and Its Application", *Information and Systems Engineering*, 2, 47-61.

Sasaki, Y., N. Kobayashi and K. Kijima (2007), "Mixed Extension of Hypergames and Its Application to Inspection Games," *ISSS Proceedings of the 51st Annual Conference,* 777.

Takahashi, S., N. Hinago, T. Inohara and B. Nakano (1999), "Evolutionary approach to three-person hypergame situation," *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, 4, 254-259

Wang, M., K. W. Hipel and N. M. Fraser (1988), "Modeling Misperceptions in Games," *Behavioral Science*, 33, 207-223.

Wang, M., K. W. Hipel and N. M. Fraser (1989), "Solution Concepts in Hypergames," *Applied Mathmatics and Computation*, 34(3), 147-171.