Functional dynamic by intention recognition in iterated games

Yuma Fujimoto and Kunihiko Kaneko

Department of Basic Science, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

1 Author to whom any correspondence should be addressed.

E-mail: yfujimoto@complex.c.u-tokyo.ac.jp and kaneko@complex.c.u-tokyo.ac.jp

Keywords: cognitive game, functional dynamics, equilibrium, cooperation, exploitation

Abstract

Intention recognition is an important characteristic of intelligent agents. In their interactions with others, they try to read others’ intentions and make an image of others to choose their actions accordingly. While the way in which players choose their actions depending on such intentions has been investigated in game theory, how dynamic changes in intentions by mutually reading others’ intentions are incorporated into game theory has not been explored. We present a novel formulation of game theory in which players read others’ intentions and change their own through an iterated game. Here, intention is given as a function of the other’s intention, whereas the functional dynamics reach equilibria in which both players’ intentions are optimized. These cover a classical Nash and Stackelberg equilibria but we extend them in this study: Novel equilibria exist depending on the degree of mutual recognition. Moreover, the degree to which each player recognizes the other can also differ. This formulation is applied to resource competition, duopoly, and prisoner’s dilemma games. For example, in the resource competition game with player-dependent capacity on gaining the resource, the superior player’s recognition leads to the exploitation of the other, while the inferior player’s recognition leads to cooperation through which both players’ payoffs increase.

1. Introduction

How each individual decides his/her own behavior is a long-standing problem in nature. Each agent freely behaves and receives a reward as a result. This situation is generally formulated as a ‘normal form game’ in which the ‘player,’ ‘action,’ and ‘payoff’ are given [1]. In a standard game played only once, each player needs to decide what action he/she chooses as his/her own strategy. An optimal strategy is to choose the best rewarded action depending on others’ actions, which results in the Nash equilibrium [2, 3]. However, while this is the optimal strategy in a one-shot game, it is common that the game is repeatedly played in reality. In a repeated game, each player can refer to a data set on the history of actions played in the past and use this for the next game or a later one. The Nash equilibrium was not originally introduced to deal with such a situation. Instead, one possible equilibrium is given by the ‘folk theorem’ [4–6], where deviation from the Nash equilibrium can happen; however, the theorem only provides the requirements for the achieved equilibrium and cannot specify which equilibrium is really achieved. To determine an optimal strategy, which is now nothing given as in a one-shot game, we must therefore adopt a concrete learning process through which each player improves his/her own strategy against that of the other player.

An important characteristic of an intelligent agent (i.e. a human) is to recognize and make an image of others by using the history of the other’s actions. Such an agent assumes that the other intentionally changes the next action in response to the agent’s own action. For example, a descriptive and predictive model for a person’s cognitive behavior has been proposed [7, 8], as is based on the experiments of a repeated beauty contest game.
In this model, each person is given a cognitive level. The level 0 person chooses an action with no recognition (randomly), while the level \( k (> 0) \) person best responds to the image of others at the \((k − 1)\)th level (or lower). Recently, such cognitive levels have been uncovered in neural economics [11]. As another example, it has been reported that information on others’ true intentions increases the performance in the game when the actions of others are transferred with disturbances by noise [12–15]. Following this experimental result, the evolution of the ability to recognize others’ intentions is theoretically discussed [16–18]. Although the above studies sufficiently justify the existence of humans’ ability to recognize others and benefit from that, a theory for the dynamic coevolution of images between agents has been underdeveloped. In [19], a player builds and deconstructs the other’s image in prisoner’s dilemma games has been studied by using a recurrent network, whereas several studies use such recognition of intention for decision-making [20] or as means of compassion [21]. How the equilibrium of actions is shaped and how it deviates from the Nash equilibrium, however, are not analyzed. Here, we develop a theoretical framework with mutual learning that shapes the other’s internal intentions, generally applicable to any games, without resorting to specific learning algorithms.

Once every agent has an image of the other and best responds to it, the sequential actions are given. Therefore, how an agent constructs the other’s image itself is now a strategy in repeated games, which is represented as a function, as shown later. Initially, both agents best respond to each other without referring to the other’s strategy. Hence, both best-response functions as strategies achieve nothing but the Nash equilibrium [2, 3]. Then, before considering the dynamics of a pair of such strategy functions, we see an extreme case that an agent one-sidedly reads the other’s strategy function. In this case, the Stackelberg equilibrium [22] is achieved, which is defined as an equilibrium in an ‘extensive form game’ [23] under perfect information.

Following these introductory results, we study the dynamics of strategy functions, which represent that agents mutually recognize the other’s intention. With repeated games, each agent accurately reads the other’s strategy function and optimizes his/her own one based on it. This dynamics reaches an equilibrium when there is no additional advantage for the further recognition of the other’s strategy. At this point, a ‘functional equilibrium’ is achieved between both players’ strategies instead of the original Nash equilibrium.

Note that our formulation can be applied to general games. Here, applications to resource competition, duopoly, and probabilistic prisoner’s dilemma games are provided as examples. In the former case, it is found that learning by an inferior agent increases the payoffs of both players, while that by a superior agent enhances exploitation and decreases the payoff of the other.

The rest of the paper is organized as follows. In section 2, we briefly review the classic Nash equilibrium in terms of best response. This section is intended as an introduction to the following sections, and the function for recognition is not yet introduced. One-sided recognition, in which the function is introduced only for one player, is studied in section 3. In section 4, the recognition of both players is studied through the introduction of dynamics for the two functions. This leads to a set of novel equilibria in the strategy of the game. In these dynamics, one parameter specifies the degree of learning. In section 5, the dynamics of this parameter are introduced to optimize the payoff, as with learn-to-learn in machine learning. For each of sections 3–5, we apply the formulation of the function dynamics to three renowned examples (resource-competition, duopoly, and the prisoner’s dilemma game) and discuss how the influence of recognition and learning depend on the characteristics of each game. Up to and including section 5, function dynamics are based on perfect recognition of the other player’s function at one time. This, of course, is an idealization. In reality, players make errors in recognizing the other’s action. In section 6, we discuss how estimates from sampling many repeated games under uncertainty lead to the ideal case in the earlier function dynamics. Section 7 discusses the relevance of recognition and learning to social behavior, based on the function dynamics.

2. Nash equilibrium

We consider a two-player game in which players are denoted by \( i \in \{1, 2\} \). In addition, each player \( i \)’s action and its payoff are represented by \( x_i \) and \( u_i(x_1, x_2) \), respectively, which are continuous variables. A player tries to receive a higher payoff by optimizing his/her action depending on the other’s action. Now, each player has an intention on which action he/she chooses depending on the other’s action. The intentions vary depending on how the player imagines the other’s action. Thus, the intention of player 1 is given as strategy function \( f_1(x_2) \), which represents that action \( x_1 \) is chosen when player 2 takes action \( x_2 \). Player 2’s strategy function is similarly defined as \( f_2(x_1) \). Then, assuming that each player’s action follows his/her own strategy, the equilibrium set of actions, denoted by \((x_1^{eq}, x_2^{eq})\), is given by the crossing point of both players’ functions. In other words, we get

\[
\begin{align*}
x_1^{eq} &= f_1(x_2^{eq}), \\
x_2^{eq} &= f_2(x_1^{eq}).
\end{align*}
\]
In this section, we consider a situation in which both players have no recognition of the other’s intention. In this case, each player simply maximizes his/her own payoff without referring to the other’s strategy. To be consistent with the standard terminology, this strategy is called the ‘best-response’ [3], as denoted by \( f_1^B(x_1) \) (B-response) for player 1. According to this definition, \( f_1^B(x_1) \) satisfies

\[
f_1^B(x_1) := \arg\max_{x_2} u_1(x_1, x_2).
\]  

Equation (2) simply means that player 1’s strategy function is given by maximizing the payoff under the assumption that the other’s action is constant independent of his/her own action. At this point, note that the strategy function given by B-response depends on the other’s action \( x_2 \).

Player 2’s B-response is given in the same way. Thus, when both players make B-responses, the equilibrium, denoted by \((x_1^{BB}, x_2^{BB})\), is nothing but the Nash equilibrium from its definition. In the present paper, however, we call it the BB equilibrium, where the left index indicates the player’s strategy to the other’s strategy given by the right index, because the same equilibrium set of actions can be achieved by different sets of functions. In this study, which pair of functions results in the equilibrium action is important; hence, we need to specify not only the equilibrium point but the pair of functions to achieve it. At the BB equilibrium, each player’s payoff is defined as \( u_i^{BB} := u_i(x_i^{BB}, x_2^{BB}) \).

### 3. Definition of the learning response and one-sided recognition

Next, we define another type of intention where a player perfectly recognizes the other’s intention. Then, each player optimizes his/her action based on the information on the other’s strategy function. This strategy is termed the ‘learning response’ (L-response), denoted by \( f_1^L(x_1) \), which is the response to the function of \( f_2(x_1) \). Hence, it follows that

\[
f_1^L(x_1) := \arg\max_{x_2} u_1(x_1, f_2(x_1)).
\]  

An obvious difference between the L- and B-responses lies in the form of the recognized player’s action. Recall that in the B-response, 1’s strategy is given under the image that the other’s action is independent of his/her own action (see equation (2)). On the contrary, in the L-response, 1’s strategy is given by learning that the other’s action depends on his/her own action (see equation (3)). Therefore, the L-response is independent of \( x_2 \), while the B-response depends on \( x_2 \).

We now consider a situation in which player 1 one-sidedly recognizes 2’s intention. In this case, player 1 (2) makes the L- (B-) response. The crossing of these functions is defined as the LB equilibrium \((x_1^{LB}, x_2^{LB})\), which is given by

\[
\begin{align*}
x_1^{LB} &= f_1^L(x_2^{LB}), \\
x_2^{LB} &= f_2^B(x_1^{LB}).
\end{align*}
\]  

Then, player i’s payoff is defined as \( u_i^{LB} := u_i(x_i^{LB}, x_2^{LB}) \). In the same way, the BL equilibrium is defined as the crossing point between the B-response of player 1 and the L-response of player 2. In the duopoly game to be discussed later, the LB (BL) equilibrium is known as the ‘Stackelberg equilibrium’ [22], while in general games, it belongs to ‘sub-game perfect equilibria’ [24, 25]. Here, we use the term the Stackelberg equilibrium in any games. Therefore, one-sided recognition means a transition from the Nash equilibrium to the Stackelberg equilibrium.

We now study some of the general properties of such one-sided recognition. First, a player does not lose any benefit by learning the other’s B-response; in other words, \( u_1^{LB} \geq u_1^{BB} \) holds. This is easily proven as

\[
\begin{align*}
u_1^{LB} &= \max_{x_1} u_1(x_1, f_2^B(x_1)) \\
&\geq u_1(x_1^{BB}, f_2^B(x_1^{BB})) \\
&= u_1^{BB}.
\end{align*}
\]  

This inequality is understood as follows: as the player adopting B-response chooses the strategy depending on the other, the player can take advantage of the other’s strategy and shift the equilibrium point (i.e. \( x_1^{LB} \) or \( x_2^{BL} \)) one-sidedly, in order to get more payoff. Note that the Zero Determinant strategy by Press and Dyson [26] in prisoner’s dilemma game, adopts a similar strategy, as the optimization strategy of one player itself is taken advantage by the other to increase the payoff.

Second, we obtain a necessary and sufficient condition for a recognizing player to increase his/her own payoff. When player 1 makes the L-response, 1 refers to 2’s strategy, which represents how 2’s action changes depending on 1’s action. The difference between LB and BB equilibria lies in whether the gradient of agent 2’s action \( \partial f_2^B(x_1) / \partial x_1 \) is considered or not. Considering the case where the LB and BB equilibria are achieved \([x_{min}, x_{max}]\), the possible range of players’ actions (i.e. at \( x_{min} < x < x_{max} \)), the condition for the disagreement
between LB and BB is given by

\[
\begin{align*}
\frac{\partial u_1}{\partial x_2} &\bigg|_{BB} \leq 0, \\
\frac{\partial^2 u_2}{\partial x_1 \partial x_2} &\bigg|_{BB} \geq 0,
\end{align*}
\]

as shown in the supplementary data available online at stacks.iop.org/NJP/21/023025/medial for the detailed calculation. The condition for player 2’s L-response is obtained in the same way. This condition will be referenced in the later examples.

The above result is interpreted by the relationship between both players’ strategies. For the fixed 2’s strategy \( f_2(x_1) \), 1’s strategy \( f_1(x_2) \) enables him/herself to realize benefit \( u_1^{eq} \). Therefore, the condition that 1’s strategy is optimal and is not changed by the other’s strategy is given by

\[
u_1^{eq} = \max_{x_2} u_1(x_1, f_2(x_2)).
\]

We only consider the situation for one-way recognition in the present section. However, the same recognition equation can also be introduced for player 2, as will be discussed in depth in section 4. Then, the set of strategy functions satisfies the equilibrium for such a mutual recognition, which is defined as ‘function equilibrium’.

As illustrated in the following two examples, the function equilibrium is not always satisfied at the BB equilibrium in general because the function of the other player imagined by one player does not agree with the real function of the other. In the B-response, the player imagines that the other’s strategy function is constant, and he/she chooses his/her strategy accordingly. When both players make B-responses, however, the function of each player is no longer constant in contrast to the assumption for the B-response. Therefore, both players still gain an advantage by learning the other’s strategy function.

On the contrary, there is no such disagreement at the LB or BL equilibria, where the L-response player imagines that the other’s action can be changeable depending on the learning side’s action, and as a result, the real strategy function is made to be constant. Thus, B- and L-response functions are consistent with each other. Then, there are no more advantages of learning the other’s strategy for both players, and the function equilibrium is satisfied.

Next, we consider whether the learned side increases or decreases his/her payoff. Let us consider the ‘competitive’ case in which an increase in \( x_i \) leads to disutility for the other as is given by \( \partial u_1 / \partial x_2, \partial u_2 / \partial x_1 < 0 \).

Indeed, a few nontrivial games satisfy such a relation, as discussed in these two examples. In this case, if player 1 is more competitive owing to recognition \( x_1^{LB} > x_1^{BB} \), the following relationship is satisfied:

\[
\begin{align*}
u_2^{LB} &= u_2(x_1^{LB}, x_2^{LB}) \\
&< u_2(x_1^{BB}, x_2^{LB}) \\
&\leq \max_{x_2} u_2(x_1^{BB}, x_2) \\
&= u_2^{BB}.
\end{align*}
\]

Then, the learned player is proven to receive a payoff below the BB equilibrium.

On the contrary, if player 1 is less competitive \( x_1^{LB} < x_1^{BB} \), we get \( u_2^{BB} < u_2^{LB} \) in the same way. In addition, we can deal with another case, for example, a cooperative situation \( \partial u_1 / \partial x_2, \partial u_2 / \partial x_1 > 0 \) by inverting each player’s action \( x_1 \rightarrow -x_1, x_2 \rightarrow -x_2 \). For example, the public goods game belongs to this type, which is originated in [27] (see [28–30] for recent examples).

Here, in contrast to earlier studies, we consider not only the equilibrium set of actions but also the functions of the players to achieve it, based on the recognition of the other’s intention. Accordingly, the function equilibrium that deviates from the Nash equilibrium is introduced. We explicitly calculate the BB, LB, and BL equilibria in specific examples.

3.1. Example 1: resource competition game

As an example of the BL and LB responses formulated above, we consider a ‘resource competition’ game. In this game, both players \( i \in \{1, 2\} \) pay cost \( x_i \geq 0 \) to compete for a restricted resource with the total amount of unit one. Each player’s reward, defined as the distributed resource, is proportional to both the paid cost \( x_i \) and the efficiency to get resource per cost \( r_i \). As the available resource is fixed to one, each player’s reward is given by \( r_i x_i / (\sum_j r_j x_j) \). Then, each player’s payoff \( u_i \) is defined as the difference between the reward and cost \( x_i \), so that
In this case, we get $<$.

Here, recall equation (2). The condition of one-way learning is given as long as the abilities of two agents are not equal $r = 1$. Without loss of generality, $r_2$ is set at 1, and we take $r_1 \equiv r \leq 1$. When $r = 1$, the abilities of the two players are identical, while $r > 1$ means that player 1 is superior to 2. This game is a continuous version of the hawk-dove game [31, 32]. In addition, this continuous game was recently applied to hierarchical competition [33].

By applying equation (2) to this resource competition game, we get B-response function. Since $u_i$ is an upward-convex function, the condition of local maximum for $x_i$ is given by

$$
\frac{\partial u_i}{\partial x_i} \bigg|_{x_i=x_i^B} = 0
$$

$$\Leftrightarrow f_i^B(x_i) = \sqrt{x_i^2/r} - x_i/r.
$$

Recalling that the range of paid cost is limited to be positive, we get each player’s B-response as

$$f_1^B(x_2) = \max(\sqrt{x_2^2/r} - x_2/r, 0),$$

$$f_2^B(x_1) = \max(\sqrt{x_1^2/r} - x_1/r, 0).$$

From equation (1), we get the set of actions at the BB (Nash) equilibrium as the crossing of the strategy functions.

Next, we consider the case in which only player 1 learns 2’s B-response. Player 1’s L-response is given (see figure 1(B)) by

$$f_1^L(x_2) = \begin{cases} r/4 & (1 \leq r < 2) \\ 1/r & (2 \leq r) \end{cases}$$

Here, recall equation (6), the condition in which the equilibrium actions are changed by one-way recognition. This condition is satisfied as long as the abilities of two agents are not equal $r = 1$. In fact, for $r = 1.5$, where LB and BB disagree (compare figures 1(B) with (A)).

We now study how these two are different (see the supplementary data for the detailed calculation). From figure 1, we get $x_1^{LB} > x_1^{BB}, x_2^{LB} < x_2^{BB}, u_1^{LB} > u_1^{BB}$, and $u_2^{LB} < u_2^{BB}$. These equations indicate that owing to the superior player’s one-way learning, he/she increases his/her cost but increases his/her payoff, while the other player decreases his/her payoff while decreasing his/her cost.

On the contrary, when player 2 one-sidedly learns 1’s B-response, 2’s L-response is given (see figure 1(C)) by

$$f_2^L(x_1) = 1/(4r).$$

In this case, we get $x_1^{BL} < x_1^{BB}, x_2^{BL} < x_2^{BB}$, and $u_1^{BL} > u_1^{BB}, u_2^{BL} > u_2^{BB}$ as shown in figure 1 (see the supplementary data for the detailed calculation). Hence, both the players decrease their costs and increase their payoffs owing to the one-way learning of the inferior player in contrast to that of the superior player.

As explained earlier, the LB and BL correspond to the classical Stackelberg equilibria [22, 24, 25]. Here, however, note again that we focus not only on the crossing equilibrium but also on a pair of strategy functions.

Figure 1. The BB (left), LB (center), and BL (right) equilibria in the case of $r = 1.5$ in a resource competition game. For all three figures, the X-axis (Y-axis) indicates player 1’s (2’s) action, denoted by $x_1$ ($x_2$). The blue (green) line indicates 1’s (2’s) intention $f_1(x_2)$ ($f_2(x_1)$) at the equilibrium. The cyan (magenta) dots indicate the Nash (Stackelberg) equilibria, respectively.
achieved in the equilibrium. Thus, we use the term ‘LB’ or ‘BL’ equilibrium, instead of the Stackelberg equilibrium here.

The superior player’s one-way learning results in exploitation by gaining more benefit by increasing its own cost, while the inferior player’s learning results in cooperation by decreasing its own cost. This is interpreted as follows. First, the cost a player should pay depends on the other’s cost. A player would not need to pay so much when the other’s cost is too small because the player would monopolize most resources by paying not so much cost. On the contrary, if the other’s cost is too large, the player would not pay much cost either, because one should pay too much cost to obtain more resources. Therefore, a player’s optimal cost is maximal when the other pays a moderate cost (see figure 1(A)). Second, in the BB equilibrium, both players pay a moderate cost. Hence, no matter whether the learning player is superior or not, the player has to repress the other’s cost to gain more benefits. How to repress the other’s cost, however, depends on whether the player is superior or not. The superior player increases his/her cost and forces the inferior one to give up competition. Therefore, the former exploits the latter by learning the other’s strategy. On the contrary, to gain a higher payoff, the inferior player decreases his/her cost and relaxes the competition. Therefore, the learning player cooperates with the learned one.

How the learning and learned players’ payoffs change depends on the type of game. Below, we discuss an alternative example, namely a duopoly game, in which competition always persists to the point that an increase in the payoff of the learning player always decreases the other’s payoff.

3.2. Example 2: duopoly game

In a duopoly game, two companies $i \in \{1, 2\}$, which separately supply products, compete for a limited market. The more products supplied in this limited market, the cheaper their prices are. Here, player $i$’s action $x_i$ is the number of products he/she supplies. We assume that the price is $max(0, p - x_1 - x_2)$, where $p$ represents the maximal price. In addition, player $i$’s cost of supplying products is assumed to be $c_i$. Thus, player $i$’s payoff is given by

$$u_i(x_i, x_2) = x_i(max(0, p - x_1 - x_2) - c_i),$$

$$u_2(x_1, x_2) = x_2(max(0, p - x_1 - x_2) - c_2).$$  \hspace{1cm} (14)

Here, we assume $c_1 \leq c_2$ without loss of generality. In other words, player 1 is superior to 2.

In this duopoly game, equation (6) is always satisfied regardless of $p$, $c_1$, and $c_2$. Examples of the strategy functions in the BB, LB, and BL equilibria are plotted in figures (A)–(C), respectively. Accordingly, from the crossing points, we obtain $u_1^{LB} > u_1^{BB}$ and $u_2^{LB} < u_2^{BB}$. These equations indicate that the superior company’s learning results in the exploitation of the learned one, as in the resource competition game. On the contrary, we get $u_1^{BL} < u_1^{BB}$ and $u_2^{BL} > u_2^{BB}$. In contrast to the resource competition game, the inferior company’s learning also results in exploitation. (See the supplementary data for the detailed calculation.)

Although both the resource competition and the duopoly games are categorized as competitive ($\partial u_i / \partial x_i, \partial u_i / \partial x_2 < 0$), the change in the learned player’s payoff differs between the two. Every player’s learning brings a disbenefit on the other in duopoly game, while only the strong side’s learning does in resource competition game. As has already been explained, this difference depends on whether the learning side is more or less competitive according to one-way learning (see equation (8)). In spite of the difference in the learning effect between the two games, the following dynamical process of learning itself, however, can be applied to both examples similarly. Hence, we discuss this example only in the Supplementary Data, and summarize the result in the later table 1 in section 7.
Table 1. Summary of the effect of recognition in resource-competition (RC), duopoly, and prisoners’ dilemma (PD). In the resource-competition game, the result depends on whether the recognizing agent has an advantage (ad.) or not (disad.), so we put the result in a separate column. In section 3, we considered one-way recognition, which leads to the change of equilibrium from BB (Nash) to LB (Stackelberg). The left part of table shows the change in cost and payoff for the recognizing and recognized side between LB and BB, as obtained in section 3. The right part represents the increase in cost and payoff to the player as a result of increasing his/her learning rate $\epsilon_i$, while the next row represents those of the other player, as obtained in section 4.

|       | Recognizing | Recognized_i | Recognized_other |
|-------|-------------|--------------|-----------------|
|       | Cost | Payoff | Cost | Payoff | Cost | Payoff | Cost | Payoff |
| RC (ad.) | $x_{LB} > x_{BB}$ | $u_{LB} > u_{BB}$ | $x_{LB} < x_{BB}$ | $u_{LB} < u_{BB}$ | + | + | - | - |
| RC (disad.) | $x_{LB} < x_{BB}$ | $u_{LB} < u_{BB}$ | $x_{LB} > x_{BB}$ | $u_{LB} > u_{BB}$ | - | + | - | + |
| Duopoly | $x_{LB} > x_{BB}$ | $u_{LB} > u_{BB}$ | $x_{LB} < x_{BB}$ | $u_{LB} < u_{BB}$ | + | - | - | + |
| PD | $x_{LB} = x_{BB}$ | $u_{LB} = u_{BB}$ | $x_{LB} = x_{BB}$ | $u_{LB} = u_{BB}$ | 0 | 0 | 0 | 0 |

3.3. Example 3: probabilistic prisoner’s dilemma game

The third example is prisoner’s dilemma game with probabilistic choice of actions. In contrast to examples 1 and 2, in this example BB lies on the boundary of the range of actions $[X_{\min}, X_{\max}]$ players can take (see the Supplementary Data for the detailed calculation). Therefore, $x_{LB} = x_{1LB} = x_{1BB}$ holds, that is, the learning discussed above changes neither the action nor strategy function. Hence, we do not discuss this example further.

4. Functional dynamics of strategies

So far, we first considered the BB equilibrium in which both players have no recognition of the other’s intention. Then, we introduced the LB and BL equilibria in which a player one-sidedly recognizes the intention of the other. When one recognizes the other but not vice versa, the recognizing player has an advantage. However, such one-way recognition rarely appears because both of the players usually try to recognize each other’s intention. Another problem in the L-response is that one player knows the other player’s strategy function rather than the dynamic systems of state variables of a finite dimension. Therefore, we do not focus on this example.

To represent such a gradual recognizing process, we assume that each player $i$ learns the other’s strategy at a rate of $\epsilon_i$. In this case, each player’s strategy function $f_i(x_i)$ changes depending on the other’s one as

$$f_i(x_i)[t + 1] = \arg\max_{x_i} u_i(x_i, x_i, \epsilon f_i(x_i)[t] + (1 - \epsilon_i)x_i),$$

$$f_j(x_j)[t + 1] = \arg\max_{x_j} u_j(\epsilon f_j(x_j)[t] + (1 - \epsilon_j)x_j, x_i).$$ (15)

For $\epsilon_i = 0$, $f_i$ in the one-shot game corresponds to the B-response (see equation (2)); for $\epsilon_i = 1$, $f_i$ corresponds to the L-response (see equation (3)). In addition, when at least one player makes the B-response, both players’ strategies in the equilibrium are given as fixed functions, as already mentioned. In the present case with $\epsilon_1, \epsilon_2 > 0$, however, it is necessary to consider the functional dynamics, where both players change their strategy functions by learning the other’s strategy function. Therefore, we add the time variable $t$.

Equation (15) represents the functional dynamics [34–36], where the change in time depends on the function rather than the dynamic systems of state variables of a finite dimension (for example, in dynamical-systems game [37]). Hence, we need to solve the dynamics of infinite dimensions.

We now analyze the equilibrium state of equation (15). In the following, we assume that there exist a pair of fixed-point functions as an equilibrium state of the functional dynamics, which is denoted by $f_i^eq(x_i), f_j^eq(x_j)$ satisfying

$$f_i^eq(x_i) = \arg\max_{x_i} u_i(x_i, \epsilon f_i^eq(x_i) + (1 - \epsilon_i)x_i),$$

$$f_j^eq(x_j) = \arg\max_{x_j} u_j(\epsilon f_j^eq(x_j) + (1 - \epsilon_j)x_j, x_i).$$ (16)

As demonstrated later numerically, fixed-point functions are reached in various games. To study the behavior near the equilibrium, we derive a crossing point of the equilibrium functions and its neighborhood. By assuming the continuity of the functions around the crossing point, we expand the equilibrium functions as

$$f_i^eq(x_i) = x_i^eq + a_i^eq(x_i - x_i^eq),$$

$$f_j^eq(x_j) = x_j^eq + a_j^eq(x_j - x_j^eq).$$ (17)
By substituting equations (17) into (16), we get the first-order term as

\[
\frac{\partial}{\partial x_1} (u_1(x_1, x_2)|_{x_2=x_2^e}) \bigg|_{eq^*} = 0,
\]

\[
\frac{\partial}{\partial x_2} (u_2(x_1, x_2)|_{x_1=x_1^e}) \bigg|_{eq^*} = 0.
\]

Then, we also get the second-order term as

\[
\frac{\partial}{\partial x_1} \left( \frac{\partial}{\partial x_1} (u_1(x_1, x_2)|_{x_2=x_2^e}) \bigg|_{x_1=f_1^*(x_2)} \right) \bigg|_{eq^*} = 0,
\]

\[
\frac{\partial}{\partial x_2} \left( \frac{\partial}{\partial x_2} (u_2(x_1, x_2)|_{x_1=x_1^e}) \bigg|_{x_2=f_2^*(x_1)} \right) \bigg|_{eq^*} = 0,
\]

where \(x_1 = \epsilon_2 f_1^*(x_2) + (1 - \epsilon_2) x_1 \) and \(x_2 = \epsilon_1 f_2^*(x_1) + (1 - \epsilon_1) x_2\). Here, equation (18) indicates that the crossing point \((x_1^{eq^*}, x_2^{eq^*})\) satisfies the optimization condition for the other’s function, while equation (19) is a consequence of the fixed-point functions, indicating that the set of equilibrium functions \((f_1^*(x_2), f_2^*(x_1))\) satisfies the optimization condition in the neighborhood of the crossing point. To compute the player’s equilibrium payoff given by \((x_1^{eq^*}, x_2^{eq^*})\), the above calculation of \(f_1^*\) and \(f_2^*\) is thus sufficient.

As an extreme case, we consider \(\epsilon_i = 1\), in which both players perfectly recognize the other’s intention. The fixed point in this case is the LL equilibrium. Here, both players make L-responses, which are constant for the other’s action. Therefore, from equation (3), the achieved actions at the LL equilibrium correspond to those at BB (i.e., the Nash equilibrium), namely \(x_i^{LL} = x_i^{BB}\) (see the upper right of figure 3). Note that the equilibrium points are identical, whereas they are different in the functional dynamics. Indeed, from equations (18) and (19), we can confirm that \(x_i^{eq^*}\) is equal for LL and BB, while \(a_i^{eq^*}\) is not (see the Supplementary Data for the detailed calculation). Owing to this inequality in \(a_i^{eq^*}\), the function equilibrium holds in LL, but not in BB: In the functional dynamics with \(\epsilon_i = 1\), both players’ strategy functions are constant, while for \(\epsilon_i = 0\), they are not.

Figure 3. Equilibrium state of mutual learning in a resource competition game. For all nine figures, the X-axis (Y-axis) indicates player 1’s (2’s) action, denoted by \(x_1\) (\(x_2\)). The blue (green) line indicates 1’s (2’s) strategy function and the solid (broken) line indicates the simulated (analytical) solution. The yellow circle is the crossing point and the red circles are the Nash and Stackelberg equilibria to allow for a comparison of positions. The left, center, and right figures are respectively the cases of \(\epsilon_1 = 0, 0.5, 1\), while the upper, center, and lower figures are respectively the cases of \(\epsilon_2 = 0, 0.5, 1\).
It may be disappointing that the classical Nash equilibrium is achieved as the LL equilibrium in the game between the learning players. As explained later, however, this LL equilibrium is rarely achieved; indeed, in many cases, novel equilibria are achieved according to the functional dynamics. Below, we discuss some specific examples for the functional dynamics.

4.1. Example 1: resource competition game

We again consider the resource competition game. From equations (18) and (19), we get the set of equilibrium actions \( (x_1^{eq}, x_2^{eq}) \) and the set of equilibrium gradients around them \( (a_1^*, a_2^*) \) as given by

\[
\begin{align*}
    x_1^{eq} &= \frac{1}{r + \epsilon_i a_2^*} \left( \sqrt{r(x_2^{eq} - \epsilon_i a_2^* x_1^{eq})} - (x_1^{eq} - \epsilon_i a_2^* x_1^{eq}) \right), \\
    x_2^{eq} &= \frac{1}{1 + \epsilon_2 a_1^*} \left( \sqrt{r(x_1^{eq} - \epsilon_2 a_1^* x_2^{eq})} - (x_2^{eq} - \epsilon_2 a_1^* x_2^{eq}) \right), \\
    a_1^* &= \frac{1}{r + \epsilon_i a_2^*} \left\{ \frac{r(1 - \epsilon_i)}{2 \sqrt{r(x_1^{eq} - \epsilon_i a_1^* x_1^{eq})}} - (1 - \epsilon_i) \right\}, \\
    a_2^* &= \frac{1}{1 + \epsilon_2 a_1^*} \left\{ \frac{r(1 - \epsilon_2)}{2 \sqrt{r(x_1^{eq} - \epsilon_2 a_2^* x_2^{eq})}} - (1 - \epsilon_2) \right\}.
\end{align*}
\]

We now simulate equation (15) and compare the simulation results with the calculation, confirming that both players’ strategy functions immediately converge to fixed ones (see figure 4). The crossing points \( (x_1^{eq}, x_2^{eq}) \) of these functions agree well with the above analytic estimation and the converged strategy functions in the neighborhood of the crossing points are well estimated by equation (17) with the above values \( a_1^* \) and \( a_2^* \). In addition, the action compared with the other’s action is calculated from fixed-point function \( f \), as shown in figure 3. This indicates that the more (less) each player learns the other’s strategy, the less (more) dependent the strategy function is on the other’s action.

5. Dynamics of the degree of learning

Thus far, the learning degree \( \epsilon_i \) has been given and fixed. Thus, for each player, the case with \( \epsilon_i = 1 \) would be the better one for receiving a higher payoff. Each player, however, can change the degree to which he/she learns the other’s strategy. Initially, each player may not care about the other, and he/she learns the other’s strategy more through the repeated game. In the following, we consider this temporal evolution in the degree of learning, \( \epsilon_1, \epsilon_2 \). Here, assuming that the other’s strategy function is fixed, each player tries to increase his/her payoff by changing his/her learning degree. Therefore, the dynamics of both players’ learning degrees are given by
where $S_1, S_2$ is the speed with which each player optimizes the intensity of recognition. In the following, we simulate these dynamics for the introduced example and examine what equilibrium is reached.

5.1. Example 1: resource competition game

Figure 5 shows the dynamics of $\epsilon_1$ and $\epsilon_2$ for various sets of learning speeds $(S_1, S_2)$, while this temporal evolution in the payoff of each player according to equation (21) is shown in figure 6.

First, the initial BB (Nash) equilibrium is unstable compared with the learning dynamics. In other words, both players are motivated to learn the other’s strategy and to change their strategies accordingly because the payoff has a nonzero gradient at $(x_1, x_2)$ around the BB equilibrium point.

Second, as one player’s learning is superior, the other’s learning is repressed. During the evolution of learning, one player learns the other’s gradient of the strategy function, and his/her strategy approaches a constant function (L-response). Each player tries to gain more benefit by repressing the other’s cost. Therefore, the superior player 1 increases his/her cost as a result of his/her own learning, while the inferior player 2 decreases his/her cost.

Third, the intermediate state between the LB/BL and LL equilibria is finally achieved depending on the evolutionary speed relationship. Not only the edges (LB, BL, and LL equilibria), but also the intermediate states of LB (BL) and LL ($\epsilon_1, \epsilon_2$ is between 0 and 1) are achieved, where every player succeeds in optimizing his/her strategy function for the other’s one. In other words, the function equilibrium is achieved as a result of these dynamics.
6. Functional dynamics as a result of repeated games under uncertainty

So far, we have assumed that each agent updates his/her function to represent the action with which he/she responds to the other’s action by reading the other’s function. This seems too demanding for the recognition capability of the agents, who must recognize not only the other’s action but also the whole function. Further, no uncertainty is introduced in the recognition. In this section, we show that the original function dynamics can be derived by carrying out repeated games under uncertainty for each step and estimating the other’s function.

For a single function update, a large enough number of games must be repeatedly played. Then, the other’s function is estimated from samples of repeated actions. Here, two time-scales of dynamics are assumed: faster dynamics for each action under an actual function and slower dynamics to update each function based on a collection of actions.

First, each agent’s response function \( f^A \) is fixed within the faster dynamics time scale. Then, the dynamics of actions \( x_i(T_f) \) are given by

\[
\begin{align*}
x_i(T_f + 1) & = f^A_i(x_i(T_f)) + \eta_i(T_f), \\
x_i(T_f + 1) & = f^A_i(x_i(T_f)) + \eta_2(T_f).
\end{align*}
\]

(22)

Here, the noise term \( \eta_i(T_f) \) is introduced. This represents inevitable errors in decision-making or recognition and is often introduced in physical analysis for game theory. The above actions are repeated over \( T_f \) times for a single update of functions.

To update the function, each agent estimates the other’s function \( f^A \). This estimated function \( f^E \) is obtained by using linear regression from the data of \( x_i(T_f) \) over \( T_f \) times. Then \( f^E \) is given by

\[
\begin{align*}
f^E_i(x_i) & = a_i(x_i - x_i^{eq}), \\
f^E_2(x_i) & = a_2(x_i - x_i^{eq}),
\end{align*}
\]

where \( a_i \) and \( x_i^{eq} \) are estimated by minimizing errors

\[
\begin{align*}
E_1([x_i(T_f)]) & = \sum_{T_f=0}^{T_f \max} (f^E_i(x_i[T_f]) - x_i[T_f + 1])^2, \\
E_2([x_i(T_f)]) & = \sum_{T_f=0}^{T_f \max} (f^E_2(x_i[T_f]) - x_i[T_f + 1])^2.
\end{align*}
\]

(24)

Note that estimated function \( f^E \) and the actual function \( f^A \) do not agree due to the noise terms. This disagreement corresponds to misunderstanding between the two agents. On the other hand, noise is essential to get data not only at the equilibrium point but also around it, which is necessary for the linear-regression estimate.

On the basis of the other’s estimated function, equation (15) is replaced by updating actual functions as

\[
\begin{align*}
f^A_i(x_i) & = \arg\max_{x_i} \{ u_i(x_i, \epsilon_1 f^E_1(x_i) + (1 - \epsilon_1)x_i) \}, \\
f^A_2(x_i) & = \arg\max_{x_i} \{ u_2(\epsilon_2 f^E_2(x_i) + (1 - \epsilon_2)x_i, x_i) \}.
\end{align*}
\]

(25)

Recall that as \( T_f \) increases, the number of samples for estimation, the other’s function will be more accurately estimated. Therefore, equation (25) shows some deviation due to noise, which decreases with the increase of \( T_f \). With the increase, equation (25) converges to the original function dynamics. Figure 7 shows that the present function dynamics of \( f^A(x) \) in the linear regime around the equilibrium point converges to the original function dynamics studied in earlier sections as \( T_f \) increases.

Note also that the dynamics of learning degrees (equation (21)) can be replaced by \( f^E_1 \). Then, the dynamics approach the theoretical result shown by figure 5, with the increase of \( T_f \).

7. Summary and discussion

In this study, we introduce a new formulation for the mutual recognition of intention, which is represented as functional dynamics. In the formulation, every player can read the other’s strategy function \( f(x) \), which determines the action to be chosen for the other’s action \( x \).

As a result, we proved that both players can increase their payoffs according to their learning. The more a player learns the other’s strategy, the less his/her action depends on the other’s action (i.e., the function approaches a constant function). Since such a constant function does not provide any motivation to learn, the process of mutual learning stops when one player perfectly learns the other’s strategy function.
The resultant function equilibrium includes and extends two kinds of well-known equilibria (Nash and Stackelberg), which are achieved in contrasting situations. In the extreme case in which one player one-sidedly learns the other, the Stackelberg equilibrium is achieved. On the contrary, when the evolutionary speeds of the learning of both players are of comparative order, the Nash equilibrium is achieved again. In contrast to previous studies, our study discusses such equilibria as the crossing points of the strategy functions. Furthermore, depending on the evolutionary speeds of learning, the intermediate equilibria between the Nash and Stackelberg equilibria are achieved. This finding may lead to understanding how the leader–follower relationship is formed in game theory according to intention recognition.

Furthermore, we also confirm that each player’s payoff changes according to intention recognition in the resource competition game. The faster the superior player learns the other’s strategy function, the more that player exploits the other, where the learning (learned) player’s payoff increases (decreases). On the contrary, the faster the inferior person learns, the more the players cooperate, meaning that both players’ payoffs increase. In the learning process, one player’s positive (negative) gradient of his/her strategy function leads to a decrease (increase) in the other’s competitiveness.

The theoretical formulation of recognition and learning presented here is applied to resource-competition, duopoly, and prisoner’s dilemma games, as summarized in table 1.

In our formulation, we assume that each player’s learning speed is independent of its accuracy. This assumption results in a monotonic advantage for the increase in learning speed, at least in the duopoly game. In reality, however, there is a trade-off between the accuracy of reading and speed of evolution, which provides another disadvantage for the fast evolution owing to incomplete information. Indeed, some previous studies show that such incomplete information on the other’s action leads to disutility [41].

This study investigated the game and function dynamics between two people. The present formulation can be extended to games of more than two players, straightforwardly. With mutual recognition, the same equilibria and function exist as a solution, where each of the two players achieves the same relationship studied here. However, there is an important difference. A third person can influence the game and the recognition between the first two players. Since each agent’s response is given by functions with more than two variables, novel equilibria that cannot be characterized by Nash and Stackelberg exist, as do oscillatory solutions that appear in some cases. For example, in loop recognition, where each of agents 1, 2, and 3 one-sidedly recognize 2 and 3, 1’s response is not covered by the extensive form game [23]. The function dynamics here can be used in future to study cognitive relationships in the network of agents, as is often important in the study of sociophysics [42–44].

Acknowledgments

The authors would like to thank E Akiyama, T S Hatakeyama and N Takeuchi for useful discussions. This research was partially supported by Hitachi The University of Tokyo Laboratory.

References

[1] Von Neumann J and Morgenstern O 2007 Theory of Games and Economic Behavior (Commemorative Edition) (Princeton, NJ: Princeton University Press)
[2] Nash J 1950 Equilibrium points in n-person games Proc. Natl Acad. Sci. 36 48–9
[3] Nash J 1951 Non-cooperative games Ann. Math. 54 286–95
[4] Aumann R J and Shapley L S 1994 Long-term competition: a game-theoretic analysis *Essays in Game Theory* (New York: Springer) pp 1–15
[5] Rubinstein A 1980 Strong perfect equilibrium in supergames *Int. J. Game Theory* 9 1–12
[6] Fudenberg D and Maskin E 2009 The folk theorem in repeated games with discounting or with incomplete information *A Long-Run Collaboration On Long-Run Games* (London: Imperial College Press) pp 209–30
[7] Nagel R 1995 Unraveling in guessing games: an experimental study *Am. Econ. Rev.* 85 1313–26
[8] Camerer C F, Ho T H and Chong J K 2004 A cognitive hierarchy model of games *Q. J. Econ.* 119 861–98
[9] Keynes J M 2018 *The General Theory of Interest, Employment and Money* (Berlin: Springer) ch 12 (https://doi.org/10.1007/978-3-319-70344-2)
[10] Moulin H 1986 *Game Theory for Social Sciences* (New York: University Press)
[11] Coricelli G and Nagel R 2009 Neural correlates of depth of strategic reasoning in medial prefrontal cortex *Proc. Natl Acad. Sci.* 106 9163–8
[12] Charness G and Levine D I 2007 Intention and stochastic outcomes: an experimental study *Econ. J.* 117 1051–72
[13] Falk A, Fehr E and Fischbacher U 2008 Testing theories of fairness: intentions matter *Games Econ. Behav.* 62 287–303
[14] Fischbacher U and Utitkal V 2013 On the acceptance of apologies *Games Econ. Behav.* 82 592–608
[15] Rand D G, Fudenberg D and Dreber A 2015 It’s the thought that counts: the role of intentions in noisy repeated games *J. Econ. Behav. Organ.* 116 861–99
[16] Anh H T, Pereira L M and Santos F C 2011 Intention recognition promotes the emergence of cooperation *Adapt. Behav.* 19 264–79
[17] Anh H T, Santos F C, Lenaerts T and Pereira L M 2015 Synergy between intention recognition and commitments in cooperation dilemmas *Sci. Rep.* 5 9312
[18] Nakamura M and Ohtsuki H 2016 Optimal decision rules in repeated games where players infer an opponent’s mind via simplified belief calculation *Games* 7 19
[19] Taiji M and Ikegami T 1999 Dynamics of internal models in game players *Physica D* 134 253–66
[20] Wang W X, Ren J, Chen X and Wang B H 2006 Memory-based snowdrift game on networks *Phys. Rev. E* 74 056113
[21] Li Y, Zhang J and Perc M 2018 Effects of compassion on the evolution of cooperation in spatial social dilemmas *Appl. Math. Comput.* 320 437–45
[22] von Stackelberg H 2011 *Market Structure and Equilibrium* 1st edn (Berlin: Springer) (Transl. English) (https://doi.org/10.1007/978-3-642-12586-7)
[23] Kuhn H W 1950 Extensive games *Proc. Natl Acad. Sci.* 36 570–6
[24] Selten R 1965 Spieltheoretische behandlung eines otigopolmodells mit nachfragegetriggert *J. Institutiona Theor. Econ.* 121 301–24
[25] Selten R 1975 Reexamination of the perfectness concept for equilibrium points in extensive games *Int. J. Game Theory* 4 25–55
[26] Press W H and Dyson F J 2012 Iterated Prisoner’s dilemma contains strategies that dominate any evolutionary opponent *Proc. Natl Acad. Sci.* 109 10409–13
[27] Hardin G 2009 The tragedy of the commons *J. Nat. Resour. Policy Res.* 1 243–53
[28] He N, Chen X and Szolnoki A 2019 Central governance based on monitoring and reporting solves the collective-risk social dilemma *Appl. Math. Comput.* 347 334–41
[29] Wang Q, He N and Chen X 2018 Replicator dynamics for public goods game with resource allocation in large populations *Appl. Math. Comput.* 328 162–70
[30] Chen X and Szolnoki A 2018 Punishment and inspection for governing the commons in a feedback-evolving game *PLoS Comput. Biol.* 14 e1006347
[31] Smith J M and Price G R 1973 The logic of animal conflict *Nature* 246 15–8
[32] Smith J M 1988 *Evolution and the Theory of Games* (United States: Springer)
[33] Fujimoto Y, Sagawa T and Kaneko K 2017 Hierarchical prisoner’s dilemma in hierarchical game for resource competition *New J. Phys.* 19 073008
[34] Kataoka N and Kaneko K 2000 Functional dynamics. I: Articulation process *Physica D* 138 225–50
[35] Kataoka N and Kaneko K 2001 Functional dynamics. II: Syntactic structure *Physica D* 149 174–96
[36] Kataoka N and Kaneko K 2003 Dynamical networks in function dynamics *Physica D* 181 235–51
[37] Sato Y, Akiyama E and Farmer J D 2002 Chaos in learning a simple two-person game *Proc. Natl Acad. Sci.* 99 4748–51
[38] Perc M et al 2017 Statistical physics of human cooperation *Phys. Rep.* 687 1–51
[39] Wang Z et al 2016 Statistical physics of vaccination *Phys. Rep.* 664 1–113
[40] Perc M 2016 Phase transitions in models of human cooperation *Phys. Lett. A* 380 2803–8
[41] Wu J and Axelrod R 1995 How to cope with noise in the iterated prisoner’s dilemma *J. Conflict Resolution* 39 183–9
[42] Szolnoki A and Perc M 2018 Evolutionary dynamics of cooperation in neutral populations *New J. Phys.* 20 013031
[43] Battiston F, Perc M and Latora V 2017 Determinants of public cooperation in multiplex networks *New J. Phys.* 19 073017
[44] Wang Z, Szolnoki A and Perc M 2014 Self-organization towards optimally interdependent networks by means of coevolution *New J. Phys.* 16 033041