IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 24, 2020, accepted October 3, 2020, date of publication October 7, 2020, date of current version October 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029502

Artificial Empathy: A New Perspective for Analyzing and Designing Multi-Agent Systems

JIZE CHEN, (Graduate Student Member, IEEE), DALI ZHANG, ZHENSHEN QU, (Member, IEEE), AND CHANGHONG WANG, (Senior Member, IEEE)
Harbin Institute of Technology, Harbin 150001, China

Corresponding author: Changhong Wang (cwang@hit.edu.cn)

This work was supported by the Heilongjiang Touyan Project.

ABSTRACT Learning from biological mechanisms is an essential method of devising interaction rules among agents. Inspired by neuropsychology, we introduce empathy, a vital ability of higher animals' affective systems, to assist us in analyzing and designing multi-agent systems. In this paper, we abstract the process of empathy as an optimization problem and establish a reasonable model of empathy by optimizing the corresponding free energy. Variable temperatures on the integrated utility associated with empathy provide agents with several different modes, including collectivity, equality, oligopoly, and monopoly. Therefore, we can change the agent's mode artificially according to the task requirement and examine agents' evolution from the perspective of continuous changes on temperature. Then we present a bandit algorithm called Empathy-based Interactive Learner (EIL), by which agents can enable affective utility evaluation and adaptive learning procedure in multi-agent systems. We test EIL's performance in four games, including the iterated prisoners' dilemma, the ultimatum game and its multi-player variant, and the survival game. The results showed that EIL could significantly improve cooperation, promote altruism behaviors, and stimulate the sense of fairness in the equal mode, whereas increasing the trend of self-interest gradually in the process of switching to other modes. To sum up, our model illustrates that empathy can act as a virtual drive underlying cooperation and competition. This provides novel methods and insights in regulating behaviors in multi-agent systems, as well as artificial subjects in psychology and behavioral economics experiments.

INDEX TERMS Empathy, multi-agent system, bandit algorithm, cooperation and competition.

I. INTRODUCTION

"Agents" are individuals with information and capabilities of information processing. "Multi-agent system" is a system composed of agents following specific information flow. If we want to make agents' final information consistent by changing the mechanism of information processing, the "consensus problem" arises. In general, given the strong coupling, high complexity, and potential chaos of multi-body dynamics, most research on multi-agent collaboration can only be carried out according to bionics. For this reason, many consensus algorithms were proposed under the conjecture of "nearest neighbor rules" [1]–[3]. These algorithms can coordinate multiple agents with given dynamic characteristics and partially observed ability into the same state. As most traditional applications of a multi-agent system (such as robots formation, multi-arm robotic assemblies, and multi-satellite formation) require agents' states converge to "consistency" for a cooperative purpose, the theory of consensus control has been fully developed in the past two decades [4]–[7]. However, there are two questions worth further consideration. Firstly, is the intuitive imitation of biological synergy the most effective collaboration, and is there a more fundamental cooperative mechanism? Secondly, general collaboration in biology is reflected not only in the consistency of state but also in multi-dimensional interactions of cooperating and competing, as well as the exploration of unknown environments with more complex forms of expression.

For the first question, researchers explore it in two ways: extrinsically and intrinsically. Rules and laws can act as extrinsic drives [8]. Extrinsic regulation can also be achieved by establishing a reputation system, where other agents in the environment can rate each other based on performance [9]. On the contrary, intrinsic regulation is usually achieved by specific hard-wiring strategies or modeling analog of intrinsic properties, which is an internal drive, feedback,

The associate editor coordinating the review of this manuscript and approving it for publication was Jonghoon Kim.

and self-discipline. Some previous studies focused on modeling emotions or affective functions such as guilt and forgiveness or modeling social fairness to achieve prosocial behaviors [10]–[13]. These studies have successfully enhanced cooperation in certain games to some extent. However, there are no results of eliciting other facets of behaviors such as helping, altruism, and a sense of fairness, let alone the self-adaption of multiple behaviors. For the second question, significant results were achieved in the fields of machine learning and economics. The former, represented by multi-agent reinforcement learning [14] and adversarial bandit learning [15], focuses on the distributed learning methods the agent can carry out to achieve the Nash equilibrium in an unknown environment. The latter is represented by mechanism design theory [16] and selective incentive theory [17]. The main purpose of these theories is to design an incentive mechanism to make Nash equilibrium meet the specific task. However, the fact that adjustment on the structure of the environment is unrealistic for agents in the unknown environment further emphasizes the demands for improving the internal mechanism of decision-making.

We devote to exploring a more general distributed method for agents to promote appropriate interaction in the broader sense of collaboration including cooperation and competition. We also wish to provide a new method to explain the forming of different individual behaviors (self-interest or altruism) and social structures (differentiation or equality). Therefore, inspired by social neuropsychology, we introduce ''empathy'' as an underlying paradigm in multi-agent interaction. Empathy is a psychological term which refers to the ability to bring others' emotion into oneself [18]. This ability widely exists in mammals and plays a critical role in regulating the relationship among individuals [19], [20]. The strength of empathy affects how we perceive things and how we make decisions. Higher empathy often means a tendency to cooperate, whereas lower empathy promotes competition. Since empathy works almost the lifeline of a living being, it makes sense to integrate empathy for non-living agents to solve problems involving complicated relationships.

The discovery of mirror neurons once made academics think that the secret of empathy was revealed [21]. However, subsequent research and experimental results do not support this once-and-for-all hypothesis. The exact mechanism of empathy, which involves multiple regions of the brain is still too complex to understand. In order to introduce the concept of empathy into multi-agent systems, previous studies have simplified most of the neurological connotations of empathy, focusing on the influence of empathy as a linear information-sharing model on the dynamics of multi-agent networks. The work in [22] established a framework for moral learning in which the process of merging the associated individual's utility is similar to the process of empathy. This work aims to design a moral identity, represented as the weight vector of some representative relational characteristic indicators, and conversely inferred based on its behavior. Another work in [23] established an individual utility model based on empathy. It mainly discusses how to make group decisions to maximize social welfare in the known network of empathy weights. The work in [24] also aimed to maximize social welfare, but the difference is that the proposed algorithm is designed in a distributed learning form. As a whole, most articles do not study how the empathy model maps individual characteristics to the degree of empathy and remain focus on how to get rid of the economic dilemma and maximize social welfare under the action of altruism. We believe an in-depth discussion of a more general model is significant. With the analysis of potential parameters' impact on individual behavior patterns and the resulting social patterns (not only the maximization of social welfare but also the equalization and equity of society), we can design the adaptive structure of the interaction algorithm more clearly.

For the above consideration, we conducted further research on artificial empathy. The main contributions of this work are in three aspects: (1) an analytical method is proposed to establish the empathy model; (2) four special cases, including the collective mode, the equal mode, the oligarchic mode, and the monopolistic mode, are considered and demonstrated in detail to facilitate the analysis of the dynamic characteristics of this model; (3) an adversarial bandit algorithm integrating the proposed model, called Empathy-based Interactive Learner (EIL), is designed to make agents act in a way that takes into account the feelings of others.

The reminder of this paper is organized as follows. Section II introduces the mathematical tools and algorithm framework used in this work, as well as the biological basis of empathy and the progress of artificial empathy. Section III formally describes the empathy problem as an optimization problem by comparing the process of empathy with anti-clustering and gives a solution of this problem by minimizing the corresponding system's free energy, thus obtaining an empathy model and corresponding utility model. Section IV introduces this empathy model into an adversarial bandit framework and theoretically prove that, with some general hypothesis of egocentric drive, individuals have various working patterns, such as monopolism, oligarchy, equalitarianism, and collectivism, under different model parameters (temperatures). Algorithm EIL is shown at the end of this section and proved to be adaptive and robust to some extent. Section V shows empirical results of the proposed algorithm. In the prisoner's dilemma and standard ultimatum game, EIL-agents under the equal mode achieved self-consistency on cooperation, equality in rewards, and adaptability in dynamic environments. In the tests of the multi-player ultimatum game and the survival game, different temperatures on the process of empathy lead to diversity on individual characteristics. Section VI gives conclusions and the future work.

II. BACKGROUND

A. GRAPH THEORY

A multi-agent system containing n agents can be represented by a weighted directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{A})$, where

$\mathcal{V} = \{v_1, \cdots, v_n\}$ is a set of vertices, here representing the agents, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of the graph's directed edges. If there is a path from agent v_i to v_j, we call there exists a directed edge $e_{ij} = (v_i, v_j) \in \mathcal{E}$. The graph is called strongly-connected if there is a directed edge e_{ij} for any $v_i, v_j \in \mathcal{V}(\mathcal{G}), v_i \neq v_j$. $\boldsymbol{A} = [A_{ij}]$ is the adjacency matrix of $\mathcal{G}$ with nonnegative A_{ij} [1]. A_{ij} is positive if e_{ij} exists, otherwise $A_{ij} = 0$. The degree of agent v_i is then defined as $D_i = \sum_{v_j \in \mathcal{V}(\mathcal{G})} A_{ij}$. If some properties of agents, such as position, voltage, income and temperature, need to be represented, we can combine the graph $\mathcal{G}$ with agents' value $\boldsymbol{x} = (X_1, \cdots, X_n)^T, X \in \mathbb{R}$ and get a algebraic graph $\mathcal{G}_x = (\mathcal{G}, \boldsymbol{x})$ [25]. Moreover, the set of agent v_i's neighbors can be denoted as $\mathcal{N}_i = \{v_j \in \mathcal{V}(\mathcal{G}) : (v_i, v_j) \in \mathcal{E}(\mathcal{G})\}$.

B. EXTENDED STRUCTURE OF ADVERSARIAL BANDIT

The adversarial bandit problem is closely related to the problem of learning to play in an unknown n-agent finite game, where the game is played repeatedly by n agents [15]. In the traditional $K-$armed adversarial bandits, there is an arbitrary sequence of reward vectors $\boldsymbol{r} = (R_1, \cdots, R_m)$ where $R_t \in [0, 1]^K$ for each $t \leq m$. In each round, the agent chooses an action $a_t \in [K]$ and observes the reward R_t. The agent's purpose is to maximize the total reward $S_m = \sum_{t=1}^{m} R_t$.

In order to endow the feedback of the environment with more psychological guidance, an extended structure of adversarial bandits was proposed, as shown in Fig. 1. In this structure, the actual stimulus acting on the decision system is the integrated utility U, which contains the effect of emotional factors. As a result, the motivation for an agent's decision-making is derived from the maximization of $S_m = \sum_{t=1}^{m} U_t$.

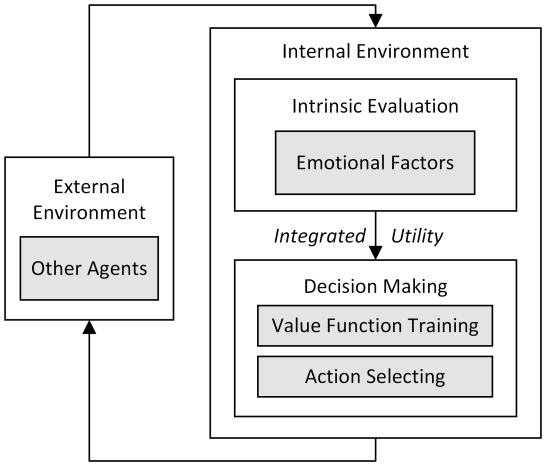


FIGURE 1. Schematic representation of dynamic interaction between the external environment and internal environment based on [26].

C. DEFINITION AND MECHANISMS OF EMPATHY

Empathy was first defined as a state of feeling another's emotions in a genuine way, later corrected by psychologist Rogers, to an experiential process of putting oneself on hold and entering another's world [27]. This process allows the therapist to understand the counselor better and guide his or her recovery. As a result, many therapists also see empathy as a professional skill. Besides, due to the need to observe and understand human's complex psychological structure, the humanistic theory emphasizes the complexity of the process of human empathy, as well as the strong but sensitive characteristics. In recent years, with the development of neurobiological science, the research on empathy has been carried out in a more reductionist way with the consensus that "empathy is the perceptual ability leading to pre-rationality and animism," and empathy has been divided into two categories, namely "affective empathy" and "cognitive empathy" [28], [29].

"Affective empathy" refers to the alternative sharing process of other people's emotions, which has the spontaneous nature of unconsciousness. "Cognitive empathy" refers to the process of an individual's understanding of others' emotions, which has subjective and dynamic attributes. The early humanistic definition of empathy emphasizes the latter more. From the neural mechanism perspective, the nervous system involved in empathy mainly includes the core emotional system, the mirror nervon system, and the theory of mind system. "Affective empathy" is thought to be driven by the mirror nervous system, which has been found in the brains of monkeys and later humans, combined with the core emotional system [30], [31]. The researchers believe that this neural system stores code for specific behavioral patterns that allow individuals not only to perform necessary actions autonomously but also to see others perform the same actions without thinking about it [32]. "Cognitive empathy" is a form of empathy based on higher-level emotional information such as inference from sensory information, prior knowledge, and perspective-taking [33]. The stimulus material may not have direct emotion, and it requires individuals to conduct cognitive processing and inference on the stimulus content and its background information, which can only be completed with the participation of the theoretical system of mind, which is, to a large extent a unique function of human beings. Therefore, Walter [34] proposed a loop model of empathy and explained the mechanism of empathy at the system level for the first time. According to the model, as shown in Fig. 2, empathy can be evoked through both bottom-up emotional signals (affective empathy) and top-down information containing content and context (cognitive empathy).

It needs to be clarified that there is no unified way to distinguish the different levels of empathy, which inevitably leads to the existence of multiple interpretations of "empathy" in different contexts and makes empathy not only a direct mechanism at the bottom but also a manifestation of behavior at the top. For now, unraveling the ultimate mystery of empathy requires more exploration of the biological mechanisms involved in brain cognition. To avoid ambiguity in this article, we define "empathy" as a general term for this type of phenomenon, without subdividing it. The derived representation of empathy will then be described separately, such as "the

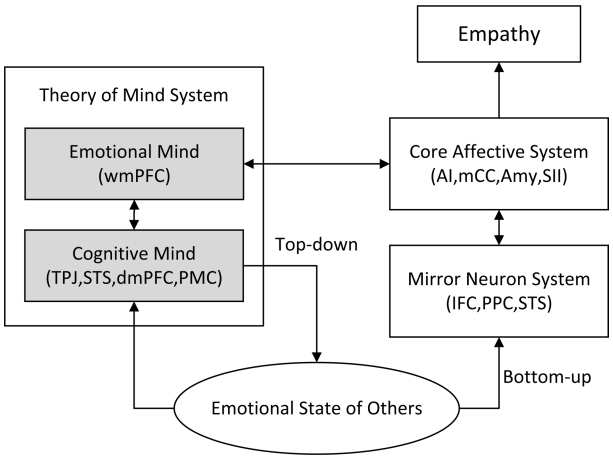


FIGURE 2. Brain circle for empathy. TPJ: temporo-parietal junction; STS: superior temporal sulcus; dmPFC: dorsomedial prefrontal cortex; PMC: posteromedial cortex; vmPFC: ventromedial prefrontal cortex; AI: anterior insula; mCC: midcingulate cortex; AMY: amygdala; SII: secondary somatosensory cortex; IFG: inferior frontal gyrus; PPC: posterior parietal cortex.

mechanism of empathy'', ''the process of empathy'', and ''the effect of empathy''.

D. ARTIFICIAL EMPATHY

The process of empathy in the biological sense requires the linkage of multiple neural regions of the brain, and the realization mechanism is relatively complex. It is difficult to give a complete and accurate description from a mathematical point of view. However, this does not mean that empathy cannot be reproduced and used in other disciplines. For example, in the multi-agent domain, researchers have developed artificial empathy models that greatly simplify the implication of empathy in the biological sense [23]. According to the previous papers, we can divide this simplified empathy model into two layers. The first layer contributes a formal linear structure of utility affected by empathy, such as the local form

$$U_i = p_{i1}X_1 + p_{i2}X_2 + \cdots + p_{in}X_n \tag{1}$$

where $p_{ij}, j \in [n]$ is the degree of empathy from agent i to agent j, X_i is agent i's characteristic value. Another form is designed for global effect as

$$\boldsymbol{u} = \lim_{t\to\infty} \boldsymbol{P}^t \boldsymbol{x} \tag{2}$$

where $\boldsymbol{P} = [p_{ij}]$ can viewed as the transition probability matrix with limitation of $\sum_{j=1}^{n} p_{ij} = 1, \forall i \in [n]$, $\boldsymbol{u} = (U_1, \cdots, U_n)$ and $\boldsymbol{x} = (X_1, \cdots, X_n)$ are vectors specifying the utility and characteristic of each agent. This layer reveals a fusion structure oriented to information or state. The agent incorporates into his utility the characteristics of others he touches. The local model captures scenarios where an agent is concerned about its neighbors' direct preference, whereas the global model is suitable for scenarios where an agent needs to be more cautious about its neighbors' integrated utility. A similar definition can be found in [22]. Some artificial empathy research only focuses on the first layer's dynamics and assumes that the degree of empathy is known. The second layer contributes details of how agents' characteristics determine the degree of empathy and make it possible to carry out an adaptive design for the decision algorithm. Given $p_{ij} = f(\boldsymbol{x})$, works in [24], [35] take into account the factors of changing empathy, such as social comparison and companion impression, and define $f(\cdot)$ in an incremental form. There are also some works that give the explicit function directly. For example, [36] makes an analogy between the process of empathy and the process of conditioned reflex, and establishes a unified model of empathy and counter-empathy.

On the whole, previous modeling approaches ignore the vast majority of ''cognitive empathy'' and is closer to ''affective empathy'' triggered from the bottom up. The simplification of empathy is partially reasonable because the linear model can effectively promote prosocial behaviors such as collaboration and mutual assistance, the same as the effect of empathy in the real community. However, the study of a universal second-level model of artificial empathy is still blank, which makes it impossible to further explore the evolutionary dynamics and game dynamics of multi-agent networks with different characteristics under the environment of distributed empathetic interaction.

III. AN ANALYTICAL MODEL OF EMPATHY

In this section, we will give a clear description of the problem of empathy and establish a bridge from agents' characteristics to the degree of empathy. Although the mechanism of empathy can be extremely complex, we hold the point that the fuzziness of internal mechanism does not directly affect the establishment of an effective model, which means ''effective'' without ''cause''. Suppose further exploration of the internal mechanism is needed in the absence of breakthrough on the measurement technology and analytical tools. In that case, one possible scientific methodology is ''analogy'' that can guide us in a top-down way.

A. DESCRIPTION OF EMPATHETIC PROBLEM

We make an analogy between the process of empathy and the inverse process of clustering. Clustering refers to the process of classifying individuals according to their similarity to the specific cluster centers. On the contrary, empathy can be regarded as the process in which the agent takes itself as the local center and moves outwards to its neighbors. The corresponding information flow can be seen in Fig. 3. Therefore, similar to the definition of the clustering problem, the purpose of the empathy problem is to find the best probability distribution of transference from a given local central point to its neighbors according to a certain standard.

Suppose there is a multi-agent system, represented by weighed directed graph $\mathcal{G}_x$. For a given local central agent $v_o \in \mathcal{V}(\mathcal{G}_x)$ and its neighbor $v_i \in \mathcal{N}_o$, the probability of

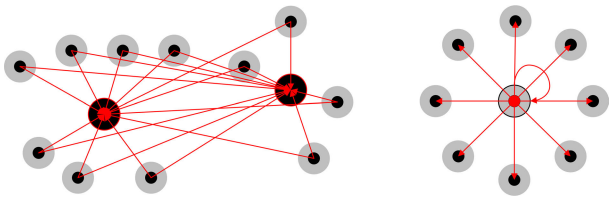


FIGURE 3. The information flow of clustering (left) and empathizing (right).

empathy from v_o to v_i is p_{oi}, and satisfies

$$\sum_{v_i \in \mathcal{N}_o} p_{oi} = 1 \quad \forall p_{oi} \in [0, 1] \tag{3}$$

It is important to note that when dealing with empathy problem, each agent in the default graph has an edge connected to itself, that is

$$v_o \in \mathcal{N}_o \quad \forall v_o \in \mathcal{V}(\mathcal{G}_x) \tag{4}$$

and the set of neighbors removed themselves is called deleted neighborhood, indicated as $\mathcal{N}_o$. There are circuits connected to the self, which means the individual's emotional pathway back to itself always persists.

On this basis, the process of empathy for v_o can be further expressed as an optimization problem, that is, the probability distribution of empathy is to be solved to minimize the cost function

$$E_o = \sum_{v_i \in \mathcal{N}_o} p_{oi} d_{oi} \tag{5}$$

where d_{oi} represents a measure of distance from v_o to v_i. Then, from the optimization perspective, a basic purpose of empathy can be explained as that the agent needs to experience neighbors' feelings with the minimum cost.

B. SOLVING METHOD OF EMPATHY PROBLEM

In the process of solving a specific empathy problem, as we have described above in (4), the agent needs to consider the weight from itself to itself. In this case, if $d_{oi} = 0$, the minimization process will directly lead to the extreme selfishness state $p_{oo} = 1$, that is to say, simply minimizing the cost function in (5) maybe not the spontaneous purpose of empathy. For this reason, drawing on the theory of chemical thermodynamics, we introduce the Gibbs free energy

$$F_o = E_o - T_o H_o \tag{6}$$

where E_o is the cost function defined in (5), which is also called the internal energy of the neighborhood system of v_o and T_o is the absolute temperature of this system. H_o is the empathy entropy of v_o's neighborhood system, defined as

$$H_o = -\sum_{v_i \in \mathcal{N}_o} p_{oi} \log_b p_{oi} \tag{7}$$

where b is the base of the logarithm, satisfying $b > 1$. According to the laws of thermodynamics, in the isothermal, isobaric process of a closed system, if no other non-expansion work is done, the system's spontaneous process will always go in the direction of the reduction of free energy [37]. We assume that v_o's neighborhood system satisfies this condition, namely closed (no particle exchange), isothermal (the average kinetic energy of the particle remains unchanged), isobaric (particle density remains unchanged), and no other work is done. Therefore, we can replace the purpose of empathy by minimizing the free energy of the v_o's neighborhood system, and this process can be expressed as

$$\begin{aligned} &\min\ F_o = E_o - T_o H_o \\ &s.t.\ \sum_{v_i \in \mathcal{N}_o} p_{oi} - 1 = 0 \end{aligned} \tag{8}$$

To solve the conditional optimal problem, we first construct the Lagrangian function

$$L_o = F_o + \phi \left(\sum_{v_i \in \mathcal{N}_o} p_{oi} - 1 \right) \tag{9}$$

Suppose that $\mathcal{N}_o = \{v_{n_1}, \cdots, v_{n_q}\}$, then

$$\begin{cases} L'_{p_{on_1}} = d_{on_1} + T_o \left(\log_b p_{on_1} + \dfrac{1}{\ln b} \right) + \phi = 0 \\ \qquad\qquad \vdots \\ L'_{p_{on_q}} = d_{on_q} + T_o \left(\log_b p_{on_q} + \dfrac{1}{\ln b} \right) + \phi = 0 \\ L'_{\phi} = \sum_{v_i \in \mathcal{N}_o} p_{oi} - 1 = 0 \end{cases} \tag{10}$$

According to (10), the probability distribution of empathy can be expressed as

$$p_{oi} = \frac{b^{\left(-\frac{d_{oi}}{T_o}\right)}}{\sum_{v_j \in \mathcal{N}_o} b^{\left(-\frac{d_{oj}}{T_o}\right)}} \quad \forall v_j \in \mathcal{N}_o \tag{11}$$

If b is the base of the natural logarithm, (11) is called the Gibbs distribution. We believe that the condition that the probability of empathy conforms to the distribution as (11) is necessary for the agent to be in an equilibrium state at a specific temperature. Then, another critical question is how to choose the temperature in the process of empathy.

According to the free energy function in (6), when the temperature is high, the proportion of empathy entropy increases, and the decrease of free energy tends to increase entropy, and the transference probability distribution tends to be uniform. When the temperature is low, the proportion of internal energy in the free energy increases, and the decrease of free energy tends to decrease the internal energy; that is, it tends to transference to the nearest agent with the maximum probability.

Therefore, the selection of temperature is similar to the exploration and exploitation problems faced by recommendation systems in machine learning. Generally speaking, in the face of such a problem, a reasonable balance weight will be given according to experience, or the idea of simulated

annealing to solve the problem of transition from exploration to exploitation. Of course, the speed of annealing still needs to be empirically set. In the case of empathy, we seek a more natural basis for temperature selection. In other words, we need to consider whether there are other constraints on the temperature in the neighborhood system of an equilibrium state with minimum free energy. For this reason, we will talk about this issue with the learning procedure in the next section.

IV. LEARNING WITH EMPATHY

In this section, we will define some important concepts about empathy first to make it convenient to introduce empathy into the learning process. Considering the subjective initiative of the learning process, we believe the setting of agents' temperatures can be purposeful. The detailed analysis of temperature will be shown in the following part. Then, to give a general application paradigm of artificial empathy, we will demonstrate an empathy-based learning method in bandit environments.

A. BASIC CONCEPTS ABOUT EMPATHY

Definition 1 (Empathetic Transfer Matrix): In a multi-agent system $\mathcal{G}_x$, if $\boldsymbol{A} = [a_{i,j}]$ is the adjacency matrix with $a_{ij} = a_{ji} = b^{\left(-\frac{d_{ij}}{T}\right)}$ and $\boldsymbol{D}$ is the a diagonal matrix formed with the degrees of agents, we can calculate the empathetic transfer matrix with following expression

$$\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{A} \tag{12}$$

and we call the element p_{ij} of $\boldsymbol{P}$ empathetic coefficient from agent v_i to v_i for convenience.Under this definition, the rows of $\boldsymbol{P}$ are all equal to 1, which means empathetic transfer matrix is a stochastic matrix applied to the theory of Markov random walks.

Definition 2 (Empathetic Utility): We define the expectation of agent's emotional characteristics under the probability distribution of the empathetic coefficient as the empathetic utility. If $\boldsymbol{x} = \{X_1, \cdots, X_n\}$ denotes the emotional characteristics of agents, the empathetic utility of v_o can be calculated as

$$E_o = \sum_{v_i \in \mathcal{V}(\mathcal{G}_x)} p_{oi} X_i \tag{13}$$

Obviously, we use the local form to define the empathetic utility, which is due to the focus on direct effects of empathy.

Definition 3 (Empathy Entropy): As mentioned in the previous section, we define the degree of randomness in which an agent empathizes with its neighbors as empathy entropy. For agent v_o, its empathy entropy can be expressed as

$$H_o = -\sum_{v_i \in \mathcal{V}(\mathcal{G}_x)} p_{oi} \log_b p_{oi} \tag{14}$$

The higher empathy entropy means the higher uncertainty of the empathetic object, as well as the greater similarity within the group.

Definition 4 (Integrated Utility): We call the utility directly acting on the decision system as integrated utility. For agent v_o, its integrated utility can be expressed as

$$U_o = E_{o,T_1} H_{o,T_2} \tag{15}$$

Integrated utility expressed in this combined form is based on two inherent motivations. The first is related to selfish attributes, and the second is related to the motive force of entropy increase. What needs to be explained here is that although we have used the idea of increasing entropy to solve the basic structure of empathy in the previous section, this does not conflict with the entropy increase behavior of the decision-making layer. It should also be noted that the temperatures of empathetic utility and empathetic entropy could be different.

B. BASIC MODES OF AGENTS

The learning process has subjective initiative, which not only has the basic characteristic of promoting the individual to maximize some goal, but also has the characteristic of variable intensity. Intuitively, temperatures on empathy reflect the latter characteristic which determine the behavior mode of an agent. Here, we introduce four special cases of the temperature setting, including the collective mode, the equal mode, the oligarchic mode, and the monopolistic mode.

Definition 5 (Collective Mode): We call agents in the collective mode when the temperatures on empathy satisfy $T_2 \to \infty$. In this mode, the integrated utility of any agent v_o can be simplified as

$$U_o^{col} = \lim_{T_2 \to \infty} E_{o,T_1} H_{o,T_2} = E_{o,T_1} \cdot \log_b n \tag{16}$$

where n is the total number of agents in the multi-agent system $\mathcal{G}_x$.

Definition 6 (Equal Mode): We call agents in the equal mode when the temperatures on empathy satisfy $T_1 \to \infty$. In this mode, the integrated utility of agent v_o can be simplified as

$$U_o^{equ} = \lim_{T_1 \to \infty} E_{o,T_1} H_{o,T_2} = \frac{\chi}{n} \cdot H_{o,T_2} \tag{17}$$

where $\chi = \sum_{v_i \in \mathcal{N}_o} X_i$ is the total characteristic of v_o's neighborhood system.

Definition 7 (Oligarchic Mode): We call agents in the oligarchic mode when the temperatures on empathy satisfy $T_2 \to 0$. In this mode, the integrated utility of agent v_o can be simplified as

$$U_o^{oli} = \lim_{T_2 \to 0} E_{o,T_1} H_{o,T_2} = E_{o,T_1} \cdot \log_b k \tag{18}$$

where k is the numbers of agents with the same income as v_o (including v_o).

Definition 8 (Monopolistic Mode): We call agents in the monopolistic mode when the temperatures on empathy satisfy $T_1 \to 0$. In this mode, the integrated utility of agent v_o can be simplified as

$$U_o^{mon} = \lim_{T_1 \to 0} E_{o,T_1} H_{o,T_2} = X_o \cdot H_{o,T_2} \tag{19}$$

Theorem 1: In a strongly-connected n-agent system $\mathcal{G}_x$ with non-negative characteristics $\boldsymbol{x}$, for agents in the collective mode, the correlation between collective interests and individual integrated utility increases as T_1 increases, and the conflict of interests between collective and individual is eliminated if and only if $T_1 \to \infty$.

Proof of Theorem 1: We set $\boldsymbol{x} = (X_1, X_2, \cdots, X_n)^T$ as a set of bases. Then agent v_o's integrated utility and collective interests can be expressed as

$$\begin{cases} U_o^{col} = \log_b n\,(p_{o1}, p_{o2}, \cdots, p_{on})\,\boldsymbol{x} = C_n \boldsymbol{p}_o \boldsymbol{x} \\ U_{all}^{col} = (1, 1, \cdots, 1)\,\boldsymbol{x} = \boldsymbol{e}\boldsymbol{x} \end{cases} \tag{20}$$

which means that $C_n\boldsymbol{p}_o$ and $\boldsymbol{e}$ represent the mapping of U_o^{col} and U_{all}^{col} to $\boldsymbol{x}$. We use the direction cosine as the measure of correlation, calculated as

$$\begin{aligned} \lambda_{o,all} &= cos\left(C_n\boldsymbol{p}_o, \boldsymbol{e}\right) = \frac{\sum_{i=1}^n p_{oi}}{\sqrt{\sum_{i=1}^n p_{oi}^2} \cdot \sqrt{n}} \\ &= \frac{1}{\sqrt{n \sum_{i=1}^n \left(\frac{b^{-\frac{d_{io}}{T_1}}}{\sum_{j=1}^n b^{-\frac{d_{oj}}{T_1}}} \right)^2}} \end{aligned} \tag{21}$$

The derivative of $\lambda_{o,all}$ with respect to T_1 is

$$\begin{aligned} \frac{\mathrm{d}\lambda_{o,all}}{\mathrm{d}T_1} &= -M \sum_{i=1}^n \sum_{j=1}^n \left(\frac{d_{oi} - d_{oj}}{T_1^2} \right) b^{-\frac{2d_{oi}+d_{oj}}{T_1}} \\ &= -\frac{M}{T_1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{d_{oi}}{T_1} - \frac{d_{oj}}{T_1} \right) \left(b^{-\frac{d_{oi}}{T_1}} - b^{-\frac{d_{oj}}{T_1}} \right) \\ &= -\frac{M}{T_1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n f\left(\frac{d_{oi}}{T_1}, \frac{d_{oj}}{T_1} \right) \end{aligned} \tag{22}$$

where

$$M = \frac{\ln b}{\sqrt{n}} \left(\sum_{i=1}^n b^{-\frac{2d_{oi}}{T_1}} \right)^{-\frac{3}{2}} \tag{23}$$

Known that

$$f\,(x, y) = (x - y)\left(b^{-x} - b^{-y}\right) \le 0, \quad \forall x, y \in \mathbb{R} \tag{24}$$

We get

$$\frac{\mathrm{d}\lambda_{o,all}}{\mathrm{d}T_1} \ge 0, \quad \forall T_1 > 0, \forall d_{oi}, d_{oj} \in \mathbb{R} \tag{25}$$

which means $\lambda_{o,all}$ increases with respect to T_1 and the correlation $Cr_{o,all} = 1$ is independent of d_{oi} and d_{oj} only if $T_1 \to \infty$.

Theorem 2: In a strongly-connected n-agent system $\mathcal{G}_x$ with non-negative characteristics $\boldsymbol{x}$ and total characteristics χ, if every agent stays in the equal mode and the distribution of characteristics is unlimited, the optimal distribution of rewards obtained by maximizing the integrated utility of any agent $v_i \in \mathcal{V}\,(\mathcal{G}_x)$ is consistent as $\boldsymbol{r}^* = (\chi/n, \cdots, \chi/n)$.

Proof of Theorem 2: The maximum feasible integrated utility for any $v_o \in \mathcal{V}\,(\mathcal{G}_x)$ in the equal mode is

$$U_o^{equ^*} = \frac{\chi}{n} \max_{\boldsymbol{x} \in \mathcal{M}} H_{o,T_2} \tag{26}$$

where

$$\mathcal{M} = \left\{ \left(x_1^2, \cdots, x_n^2\right) \,\middle|\, \sum_{s=1}^n x_s^2 = \chi \right\} \tag{27}$$

According to the principle of maximum entropy, H_{o,T_2} is maximized if and only if the distribution of p_{oi,T_2} is uniformly distributed [38]. Then we get

$$\frac{b^{\left(-\frac{d_{o1}}{T_2}\right)}}{\sum_{j=1}^n b^{\left(-\frac{d_{oj}}{T_2}\right)}} = \cdots = \frac{b^{\left(-\frac{d_{on}}{T_2}\right)}}{\sum_{j=1}^n b^{\left(-\frac{d_{oj}}{T_2}\right)}} \tag{28}$$

which means

$$\begin{aligned} d_{oi} = d_{oo} = 0 \quad \forall v_i \in \mathcal{V}\,(\mathcal{G}_x) \\ s.t. \sum_{s=1}^n X_s = \chi \end{aligned} \tag{29}$$

As a result, when U_o is at its maximum, the only solution to the distribution of characteristics is

$$X_i = \frac{\chi}{n} \quad \forall v_i \in \mathcal{V}\,(\mathcal{G}_x) \tag{30}$$

Theorem 3: In a strongly-connected n-agent system $\mathcal{G}_x$ with non-negative characteristics $\boldsymbol{x}$ and total characteristics χ, the maximum feasible integrated utility of an agent in the oligarchic mode can be achieved only if χ is equally distributed to $1 < k \le n$ agents (including this one). In particular, if $T_1 \to 0$, $k = 3$ and if $T_1 \to \infty$, $k = n$.

Proof of Theorem 3: For any agent $v_o \in \mathcal{V}\,(\mathcal{G}_x)$, assuming that the set of agents with the same characteristics as v_o is $\mathcal{N}_{o,\{k\}}$. According to (18), the integrated utility is greater than zero only if $k > 1$ and then the maximum feasible integrated utility of v_o satisfies

$$\begin{aligned} U_{o,\{k\}}^{oli^*} &= \max_{\boldsymbol{x} \in \mathcal{M}} \left(E_{o,T_1} \cdot \log_b k \right) \\ &= \log_b k \cdot \max_{\boldsymbol{x} \in \mathcal{M}} \sum_{i=1}^n \frac{b^{\left(-\frac{d_{oi}}{T_1}\right)}}{\sum_{j=1}^n b^{\left(-\frac{d_{oj}}{T_1}\right)}} X_i \end{aligned} \tag{31}$$

where

$$\mathcal{M} = \left\{ \left(x_1^2, \cdots, x_n^2\right) \,\middle|\, \sum_{s=1}^n x_s^2 = \chi \right\} \tag{32}$$

Known that, $\forall v_p \in \mathcal{N}_{o,\{k\}}$ and $\forall v_q \in \mathcal{N}_{o,\{k\}}$

$$\frac{b^{\left(-\frac{d_{op}}{T_1}\right)}}{\sum_{j=1}^n b^{\left(-\frac{d_{oj}}{T_1}\right)}} > \frac{b^{\left(-\frac{d_{oq}}{T_1}\right)}}{\sum_{j=1}^n b^{\left(-\frac{d_{oj}}{T_1}\right)}} \tag{33}$$

which means

$$U_{o,\{k\}}^{oli^*} = \log_b k \cdot \sum_{i=1}^n \frac{b^{\left(-\frac{d_{oi}}{T_1}\right)} \frac{\chi}{k} \mathbb{I}_{\mathcal{N}_{o,\{k\}}}(v_i)}{\sum_{j=1}^n b^{\left(-\frac{d_{oj}}{T_1}\right)}}$$

$$= \chi \cdot \frac{\log_b k}{k + (n-k)\, b^{-\frac{d_{o0}}{T_1}}} \tag{34}$$

where $\mathbb{I}_{\mathcal{N}_{o,\{k\}}} : \mathcal{V}(\mathcal{G}_x) \to \{0, 1\}$ is the indicator function of $\mathcal{N}_{o,\{k\}}$. Particularly, when $T_1 \to 0$,

$$\begin{aligned} U^{oli^*}_{o,T_2\to 0} &= \max_{1<k\le n} U^{oli^*}_{o,T_2\to 0,\{k\}} = \chi \cdot \max_{1\le k\le n} \frac{\log_b k}{k} \\ &= \chi \cdot \frac{\log_b k}{k}\bigg|_{k=3} = \chi \cdot \frac{\log_b 3}{3} \end{aligned} \tag{35}$$

In addition, when $T_1 \to \infty$,

$$\begin{aligned} U^{oli^*}_{o,T_2\to \infty} &= \max_{1<k\le n} U^{oli^*}_{o,T_2\to \infty,\{k\}} = \chi \cdot \max_{1\le k\le n} \frac{\log_b k}{n} \\ &= \chi \cdot \frac{\log_b k}{n}\bigg|_{k=n} = \chi \cdot \frac{\log_b n}{n} \end{aligned} \tag{36}$$

Then we can confirm that, of all possible distributions, dividing the total characteristics equally among k agents can maximize one's integrated utility, but also results in minimal gains for those agents outside the k ones when the absolute temperature $T_2 \to 0$. Note that the size of k depends on χ and T_1, and the particular cases that $T_1 \to 0$ or $T_1 \to \infty$ can lead to $k = 3$ or $k = n$ respectively.

Theorem 4: In a strongly-connected n-agent system $\mathcal{G}_x$ with non-negative characteristics $\boldsymbol{x}$ and total characteristics χ, if an agent stays in the monopolistic mode, the maximum feasible integrated utility of this one can be achieved only when $T_2 \to \infty$ and the agent monopolizes χ.

Proof of Theorem 4: According to the definition of empathy entropy, we can calculate the derivative of v_o's empathy entropy function concerning temperature as

$$\begin{aligned} \frac{\mathrm{d}H_{o,T}}{\mathrm{d}T} &= -\sum_{i=1}^{n} \frac{\sum_{j=1}^{n} \frac{d_{oi}-d_{oj}}{T^2} b^{\left(-\frac{d_{oi}+d_{oj}}{T}\right)}}{\left(\sum_{j=1}^{n} b^{\left(-\frac{d_{oj}}{T}\right)}\right)^2} \log_b \frac{b^{\left(-\frac{d_{oi}}{T}+1\right)}}{\sum_{j=1}^{n} b^{\frac{d_{oj}}{T}}} \\ &= \sum_{i\ne j} \frac{b^{-\frac{d_{oi}+d_{oj}}{T}}}{T} \left(\frac{d_{oi}-d_{oj}}{\sum_{j=1}^{n} b^{\left(-\frac{d_{oj}}{T}\right)}} \right)^2 \ge 0 \end{aligned} \tag{37}$$

which means the empathy entropy function is monotone increasing with respect to temperature and

$$\max_{T_2>0} H_{o,T_2} = \lim_{T_2\to\infty} H_{o,T_2} = \log_b n \tag{38}$$

Note that this maximum of empathy entropy is independent of X_o, which leads to

$$\begin{aligned} U^{mon^*}_{o} &= \max_{\substack{\boldsymbol{x}\in\mathcal{M} \\ T_2\ge 0}} \left(X_o \cdot H_{o,T_2}\right) = \max_{\boldsymbol{x}\in\mathcal{M}} \left(X_o \cdot \log_b n\right) \\ &= X_o \cdot \log_b n\big|_{X_o=\chi} = \chi \cdot \log_b n \end{aligned} \tag{39}$$

where

$$\mathcal{M} = \left\{ \left(x_1^2, \cdots, x_n^2\right) \,\middle|\, \sum_{s=1}^{n} x_s^2 = \chi \right\} \tag{40}$$

The transformation relationship of the four modes is shown in Fig. 4. We can see that different temperatures on the process of empathy can lead to very different individual characteristics. For a cooperative task, setting temperature as Theorem 1 shows can unify agents' purposes and facilitate the convergence of collaboration. For a competitive task, Theorem 2 and 3 provide possible settings to make the competitive state eventually tend to complete monopoly or oligopoly with a balance of power. Moreover, if an agent has a dynamic internal environment with changing temperature, its characteristics can be transformed from one kind to another.

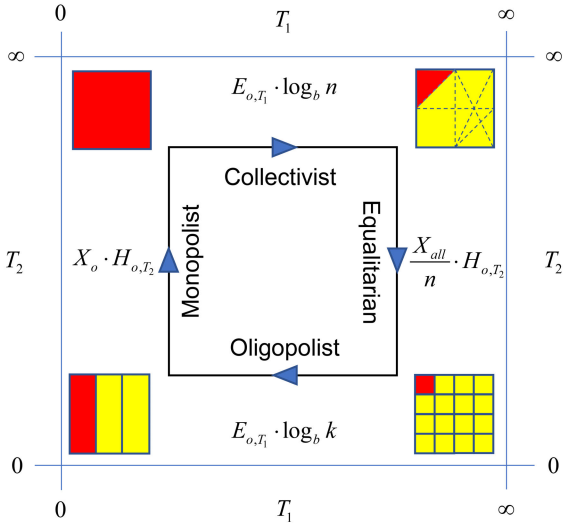


FIGURE 4. Changing modes with respect to T_1 and T_2. In the collective mode, as the value of T_2 increases, the correlation between individual interests and collective interests continues to increase, eventually reaching full overlap. Thereafter, if the value of T_1 decreases, the individual transitions to the equal mode, which strengthens the emphasis on the average degree of distribution. Furthermore, T_1 is reduced to 0, and the individual enters the oligopoly mode. At this time, as T_1 decreases, the individual tends to oligopoly resources, and ultimately, the individual benefits can be maximized in a situation like a tripod. If T_2 is increased further, individuals can enter the monopoly mode. Larger T_2 will cause individuals to take complete monopoly as their ultimate goal. Finally, by increasing T_2, the monopoly mode returns to the collective mode, forming a closed loop.

C. ALGORITHM FOR INTERACTIVE ENVIRONMENT

The core of a learning algorithm with empathy is to solve how empathy affects an agent's decision in an interactive environment. In this paper, we choose a simplified pattern that the process of empathy and decision-making are decoupled from each other. In other words, the temperature on the process of empathy is fixed, and the empathy system is only responsible for providing utility values to the decision system. Therefore, we can use the extended structure of adversarial bandit, shown in Fig. 1, to design our algorithm.

Next, in addition to the definition of bandit in the background, we still need to introduce a few related concepts to facilitate the algorithm's explanation.

Definition 9 (Income Entropy): For the multi-agent system $\mathcal{G}_x$ in a bandit environment, if the agent's memory capacity of recorded integrated utility is L, and l_o is the number of different values in the memory, we can calculate the income entropy of agent $v_o \in \mathcal{V}(\mathcal{G}_x)$ by

$$\hat{H}_o = -\sum_{i=1}^{l_o} c_{oi} \log_b c_{oi} \tag{41}$$

where c_{oi} represents the percentage of ith value in the memory space.

Definition 10 (Policy Entropy): For the multi-agent system $\mathcal{G}_x$ in a bandit environment, the policy entropy of agent $v_o \in \mathcal{V}(\mathcal{G}_x)$ represents the uncertainty of its decision on behavior. If there are K_o kinds of actions can be chosen by agent v_o, we can calculate the policy entropy of agent $v_o \in \mathcal{V}(\mathcal{G}_x)$ by

$$\widetilde{H}_o = -\sum_{a\in[K_o]} \pi_{o,a} \log_b \pi_{o,a} \tag{42}$$

where $\pi_{o,a}$ represents the probability that agent v_o chooses action a.

Assumption 1: Agents' emotional characteristics $\boldsymbol{x}$ is simplified as agents' corresponding rewards $\boldsymbol{r}$ in a bandit environment.

According to assumption 1, we can calculate agents' utility with rewards from the environment. Furthermore, to increase the algorithm's dynamic response-ability, we use a value function $Q_{o,a}$ of rolling update to evaluate the quality of agents' decisions. The update form of $Q_{o,a}$ is

$$Q_{o,a,t} = Q_{o,a,t-1} + \alpha \left(U_{o,t} - Q_{o,a,t-1}\right) \tag{43}$$

where α represents the learning rate. Based on these, we present a policy for agent v_o as

$$\pi_{o,a,t} = \frac{e^{(\mu_{o,t} Q_{o,a,t})}}{\sum_{a'\in[K_o]} e^{(\mu_{o,t} Q_{o,a',t})}} \tag{44}$$

where μ_o is the coldness (the reciprocal of temperature) of v_o's decision system. The structure like (44) is also called soft-max policy, which is similar to Gibbs distribution. The key to this policy lies in how to design an appropriate adjustment scheme for coldness. A traditional method is deterministic annealing, by which an agent can find the optimal solution in a static environment. If we want to apply the algorithm in a dynamic and adversarial environment, the update strategy's adaptivity for coldness should be improved.

The introduction of income entropy in 41 and policy entropy in 42 can provide multiple references for the algorithm's adaptive structure. Accordingly, we present an update strategy as

$$\mu_{o,t} = \max\left\{\min\left\{\hat{\mu}_{o,t}, \mu_{max}\right\}, \mu_{min}\right\} \tag{45}$$

with

$$\hat{\mu}_{o,t} = \beta \cdot \mu_{o,t-1} \left(\frac{\widetilde{H}_{o,t}}{\ln K_o}\right)^{\frac{\gamma_{o,t}}{\widetilde{H}_{o,t}}} \tag{46}$$

where β is a gain parameter, μ_{max} and μ_{min} is the upper bound and lower bound of μ, γ_o is a filtering parameter connected with the variation of income entropy and can be described as

$$\gamma_{o,t} = \min_{i\in[3]} \left(\hat{H}_{o,t-(i-1)w} - \hat{H}_{o,t-iw}\right) \tag{47}$$

where w is the sliding window's size on the variation of income entropy, which is also related to the pulse width of the external change that the agent wants to filter out. Empirically, full exploration followed with rapid convergence of μ can be achieved if β is slightly greater than 1.

Now we can present the complete structure of Empathy-based Interactive Learner (EIL) in Algorithm 1 and the information flow of each module defined above is shown in Fig. 5.

Algorithm 1 Empathy-Based Interactive Learner for Agent v_o

Require: base of logarithm b, learning rate α, gain parameter β, memory capacity L, window size w, coldness boulds $[\mu_{min}, \mu_{max}]$, initial value $Q_{o,0}$, temperature T_1 and T_2

Ensure: π_i

```
if not end of training then
  if not at end of each episode then
    Choose action a with the probability in (44) and observe the emotional characteristic X_{i,t} of each neighbor v_i ∈ N_o;
    Calculate the corresponding empathetic utility E_{o,T_1,t} with (13) and empathy entropy H_{o,T_2,t} with (14);
    Calculate integrated utility U_{o,t} with (15), income entropy Ĥ_{o,t} with (41) and policy entropy H̃_{o,t} with (42);
    Update value functiton Q^A_{o,t} with (43);
    Adjust coldness parameter μ_{o,t} with (45);
  end if
end if
```

Property 1: In the discrete-time case, agents with EIL have abilities to filter out the external pulse change on integrated utility with the width $d_p \le w$ and sense the ones with $d_p > w$ if $w \in \left[1, \frac{L-2}{3}\right)$ and $\hat{H}_{o,t} = 0, \forall t \in [t_p - w, t_p)$, where t_p is the time of rising edge of the external pulse change.

Proof of Property 1: For any agent v_o, the statement of "filter out" can be translated as "$\forall t \in [t_p, t_p + L)$, $\gamma_{o,t} \le 0$" and the statement of "sense" can be translated as "$\exists t \in [t_p, t_p + L)$ makes $\gamma_{o,t} > 0$".

When $t_p \le t < t_p + w$,

$$\gamma_{o,t} \le \hat{H}_{o,t-w} - \hat{H}_{o,t-2w} = 0 - \hat{H}_{o,t-2w} \le 0 \tag{48}$$

When $t_p + w \le t < t_p + 2w$,

$$\gamma_{o,t} \le \hat{H}_{o,t-2w} - \hat{H}_{o,t-3w} = 0 - \hat{H}_{o,t-3w} \le 0 \tag{49}$$

When $t_p + 2w \le t < t_p + L$,
Case 1 ($d_p \le w$):

$$\gamma_{o,t} \le \hat{H}_{o,t} - \hat{H}_{o,t-w} \le 0 \tag{50}$$

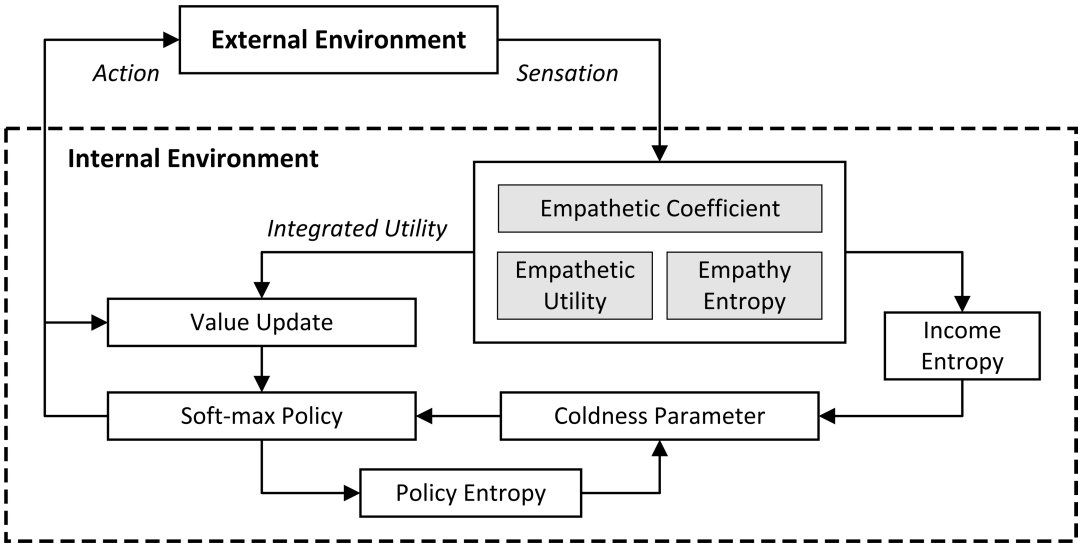


FIGURE 5. The information flow of EIL.

Therefore, if $d_p \leq w$ and $\hat{H}_{o,t} = 0, \forall t \in [t_p - w, t_p)$, we get $\gamma_{o,t} \leq 0, \forall t \in [t_p, t_p + L)$.

Case 2 ($d_p > w$ and $t = t_p + 2w$):

$$\begin{aligned}\Delta_2 &= \hat{H}_{o,t-2w} - \hat{H}_{o,t-3w} = \hat{H}_{o,t_p} - \hat{H}_{o,t_p-w} \\ &= -\frac{1}{L}\ln\frac{1}{L} - \frac{L-1}{L}\ln\frac{L-1}{L} - 0 > 0 \end{aligned} \tag{51}$$

$$\begin{aligned}\Delta_1 &= \hat{H}_{o,t-w} - \hat{H}_{o,t-2w} = \hat{H}_{o,t_p+w} - \hat{H}_{o,t_p} \\ &= -\frac{w+1}{L}\ln\frac{w+1}{L} - \frac{L-w-1}{L}\ln\frac{L-w-1}{L} \\ &\quad + \frac{1}{L}\ln\frac{1}{L} + \frac{L-1}{L}\ln\frac{L-1}{L} > 0 \end{aligned} \tag{52}$$

If $d_p \leq 2w$,

$$\begin{aligned}\Delta_0 &= \hat{H}_{o,t} - \hat{H}_{o,t-w} = \hat{H}_{o,t_p+2w} - \hat{H}_{o,t_p+w} \\ &= -\frac{d_p+1}{L}\ln\frac{d_p+1}{L} - \frac{L-d_p-1}{L}\ln\frac{L-d_p-1}{L} \\ &\quad + \frac{w+1}{L}\ln\frac{w+1}{L} + \frac{L-w-1}{L}\ln\frac{L-w-1}{L} \end{aligned} \tag{53}$$

Known that $w \in \left[1, \frac{L-2}{3}\right)$, we get $\frac{w+1}{L} < \frac{d_p+1}{L} \leq \frac{2w+1}{L} < \frac{L-w-1}{L}$. This means

$$\hat{H}_{o,t} - \hat{H}_{o,t-w} > 0 \tag{54}$$

If $d_p > 2w$,

$$\begin{aligned}\Delta_0 &= \hat{H}_{o,t} - \hat{H}_{o,t-w} = \hat{H}_{o,t_p+2w} - \hat{H}_{o,t_p+w} \\ &= -\frac{2w+1}{L}\ln\frac{2w+1}{L} - \frac{L-2w-1}{L}\ln\frac{L-2w-1}{L} \\ &\quad + \frac{w+1}{L}\ln\frac{w+1}{L} + \frac{L-w-1}{L}\ln\frac{L-w-1}{L} > 0 \end{aligned} \tag{55}$$

Thus,

$$\gamma_{o,t_p+2w} = \min\{\Delta_0, \Delta_1, \Delta_2\} > 0 \tag{56}$$

Therefore, we can confirm that if $d_p > w + 1$, $w \in \left[1, \frac{L-2}{3}\right)$ and $\hat{H}_{o,t-w} = 0, \forall t \in [t_p - w, t_p), \exists t \in [t_p, t_p + L)$ makes $\gamma_{o,t} > 0$.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present several experimental results of different games. Each game is designed specifically to test the performance of our algorithm. The parameters of EIL shared in the experiment are ($b = \mathrm{e}, \alpha = 0.001, L = 500, w = 100, \mu_{min} = 0.0001, \mu_{max} = 200, Q_0 = 0$) and as a further display for the significance of agents' different modes, we choose the equal mode with ($T_1 = 10^4, T_2 = 1$) for the first two experiments and changing modes with different temperature pairs (T_1, T_2) for the last two. We use the Euclidean distance to compute the empathetic coefficient, and all the experimental results were recorded in the average forms after 50 episodes of training.

A. ITERATED PRISONER'S DILEMMA

Iterated prisoner's dilemma is the iterated form of prisoner's dilemma, which is used to explain why it is difficult to maintain cooperation even when the cooperation is beneficial to both agents. In iterated prisoner's dilemma, agents interact in asynchronous mode with the canonical payoff matrix defined as

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} R, R & S, T \\ T, S & P, P \end{bmatrix} \end{array}$$

where 0, 1 stand for action *cooperate* and action *defect*, the payoff relationship $T > R > P > S$ is required because $R > P$ implies that mutual cooperation is superior to mutual defection, whereas $T > R$ and $P > S$ imply that defection is the dominant strategy for both agents. Moreover, $2R > T + S$ is required in order to ensure that mutual cooperation

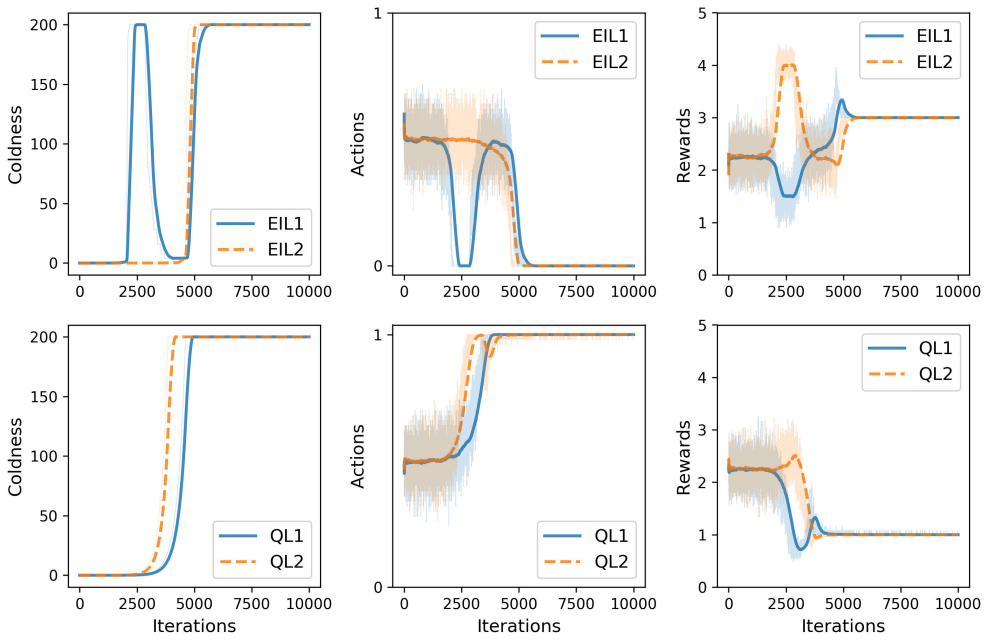


FIGURE 6. Coldness, actions, and rewards of the test between two EIL-agents (upper) and the test between two QL-agents (lower). Parameters: $\beta_{EIL1} = 1.005$, $\beta_{EIL2} = 1.002$, $\beta_{QL1} = 1.003$, $\beta_{QL2} = 1.0036$. State: equal mode. Environment: iterated prisoner's dilemma.

TABLE 1. Agent types considered in IPD.

Agent Type	Agent Characteristics
EIL	Act to maximize of (43)
QL	Act to maximize one's own reward
Mstep	Act with increasing width of pulse change on *cooperation*

is preferred over the alternation between cooperation and defection [39]. In this paper, we set $T = 5$, $R = 3$, $P = 1$, $S = 0$.

For two rational agents, this game's strategies will converge to pure-strategy Nash equilibrium, more precisely *defect*(1) even though the *cooperate*(0) is beneficial to both agents. As a contrast, this experiment's primary purpose is to verify EIL-agents' ability to solve the dilemma problem and the ability to adapt to the dynamic environment. So we simulated these games using agents with different strategies, including EIL, Q-learning (QL), and multi-step (Mstep), as shown in Table. 1. For the sake of comparison, the QL-agents tested here used the soft-max strategy in (44) of non-adaptive form.

Our first test is between two EIL-agents. The corresponding results depicted in Fig. 6 (upper) showed that EIL-agents with different gain parameters tended to choose *cooperate* progressively, which is consistent with the self-compatibility that cooperative state should be promoted if both agents use EIL strategy. By contrast, in the controlled test shown in Fig. 6 (lower), two rational QL-agents converged to a state of defection for the reason that betraying each other is the only Nash equilibrium of external payoffs. Note that, compared with the update speed on QL-agents' coldness, EIL-agents switched to exploitation after full exploration more rapidly. This property reduces the complexity of multi-agent interaction to some extent and ensures maximum utility acquisition through the sensitivity of income entropy and the on-going update of integrated utility.

In another test of IPD, we combined an EIL-agent and a QL-agent as a group. According to the results in Fig. 7 (upper), the EIL-agent showed sufficient goodwill in an uncertain environment. Once the opponent changed to an antagonistic state, the EIL-agent was always equipped to protect the safe payoffs (including the passive change of action caused by the increase of income entropy or the active change of action caused by the decrease of integrated utility). To further research the sensitivity of EIL-agents to the environment, we designed a test between an EIL-agent and a Mstep-agent. The Mstep-agent could change its actions in the form of several pulses, including the widths of 20, 40, 80, 160 and 320. The relevant results were depicted in Fig. 7 (lower). We can see EIL-agent with the window size of 100 filtered out the pulses with width less than 100 and sensed the ones wider than 100. The result is consistent with theorem 3 and proves that the interference of pulse noises can be effectively reduced by setting certain window size w of EIL-agents.

B. ULTIMATUM GAME

The ultimatum game is a classic non-zero game with two participants. It is the golden standard to examine fairness in behavioral economics [40]. In this game, the total amount of

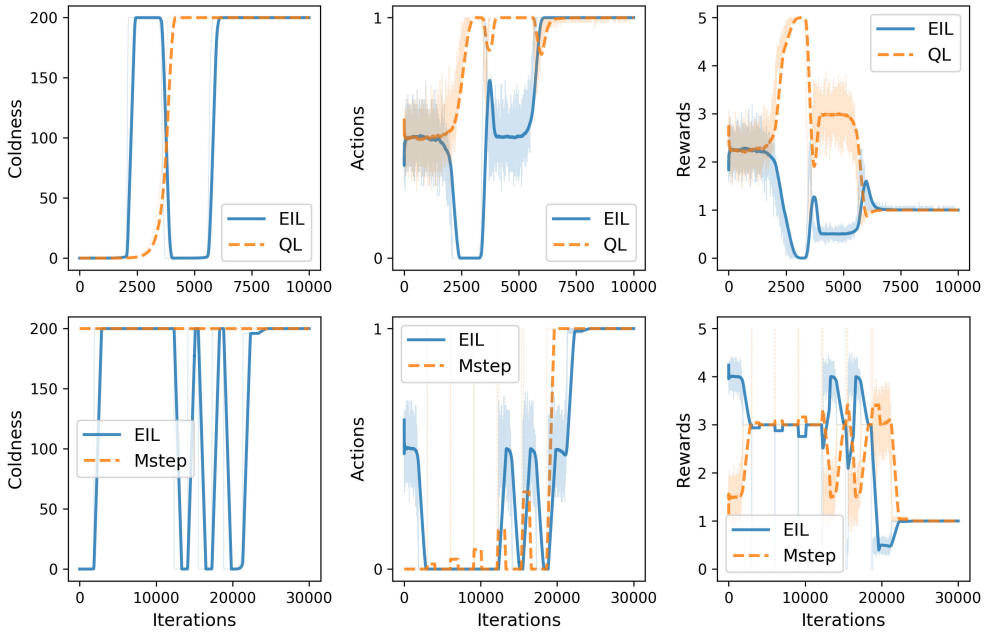


FIGURE 7. Coldness, actions, and rewards of the test between EIL-agent and QL-agent (upper) and the test between EIL-agent and Mstep-agent (lower). Parameters: $\beta_{EIL} = 1.005$, $\beta_{QL} = 1.0036$, $d_{Mstep} = \{d_{p1}, d_{p2}, d_{p3}, d_{p4}, d_{p5}\} = \{20, 40, 80, 160, 320\}$. State: equal mode. Environment: iterated prisoner's dilemma.

resources is fixed. One proposer proposes a plan to allocate resources to the participants, and the respondent has veto power. The canonical payoff matrix of the ultimatum game is defined as

$$\begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} P_r & \dots & P_* \quad \dots \quad P_0 \end{array} \\ \begin{bmatrix} 0, r & r-*, * & r, 0 \\ 0, 0 & 0, 0 & 0, 0 \end{bmatrix} \end{array}$$

where r stands for the total resources, P_* is the proposal that $*$ for the partner and $(r - *)$ for itself, 0, 1 stand for partner's action *accept* or *reject*. In this paper, the total resources r is 10, and the tick size of r_p (minimum change of proposal) is 1. For the proposer, the rational strategy is to offer minimum resources to the respondent, and the respondent is suggested to accept this instead of getting nothing [41]. In this experiment, what concerns us most is whether the unfair allocation pattern will be broken if both proposer and respondent learn with EIL in the equal mode.

We simulated this experiment with two EIL-agents and two QL-agents, respectively. As Fig. 8 showed, test between EIL-agents with different β converged to a absolute fairness $(5 - 5)$ successfully whereas two QL-agents converged to $(9 - 1)$. According to Theorem 2, this result happened because the uniform allocation conforms to the demand of maximizing the integrated utility of each EIL-agent in equal mode. On the view of game theory, we can say that the Nash equilibrium $(9 - 1)$ of external payoffs is transferred to Nash equilibrium $(5 - 5)$ of integrated utility by introducing empathy into the internal evaluation of behaviors. Proposer's proactive pursuit of fairness with a sacrifice on potential income can also be seen as an expression of altruism, which proved beneficial to the survival of the community [42].

To further study the allocation rules as agents' modes change, we extended the ultimatum game to four players. The proposer had the right to specify an allocation scheme, and each partner could express the will to accept or reject the distribution ratio. Similar to the rule of the original ultimatum game, the allocation can be carried out as planned only with the consent of all the partners; otherwise, no one will be rewarded. Therefore, the payoff matrix of multi-player ultimatum game can be defined as

$$\begin{array}{c} \\ (0,0,0) \\ else \end{array} \begin{array}{c} \begin{array}{c} P_{\{r_1, r_2, r_3, r-r_1-r_2-r_3\}} \end{array} \\ \begin{bmatrix} r_1, r_2, r_3, r - r_1 - r_2 - r_3 \\ 0, 0, 0, 0 \end{bmatrix} \end{array}$$

where r_1 is the resources for the proposer itself and r is the total resources to be allocated.

We first calculated the maximum feasible integrated utility of proposer under this circumstance with a mass of different combinations of (T_1, T_2) and recorded the maximum feasible integrated utility and the corresponding rewards in Fig. 9(b) and 9(d). As can be seen from the figure, egalitarianism was strengthened along the abscissa's positive direction, and the monopoly was strengthened along the ordinate's positive direction. We further simulated EIL-agents for the regions of $T_2 = 0.2$ (Fig. 9(a)) and the regions of $T_1 = 0.2$ (Fig. 9(c)) respectively. The proposal of the proposer in Fig. 9(a) switched from the equal allocation with

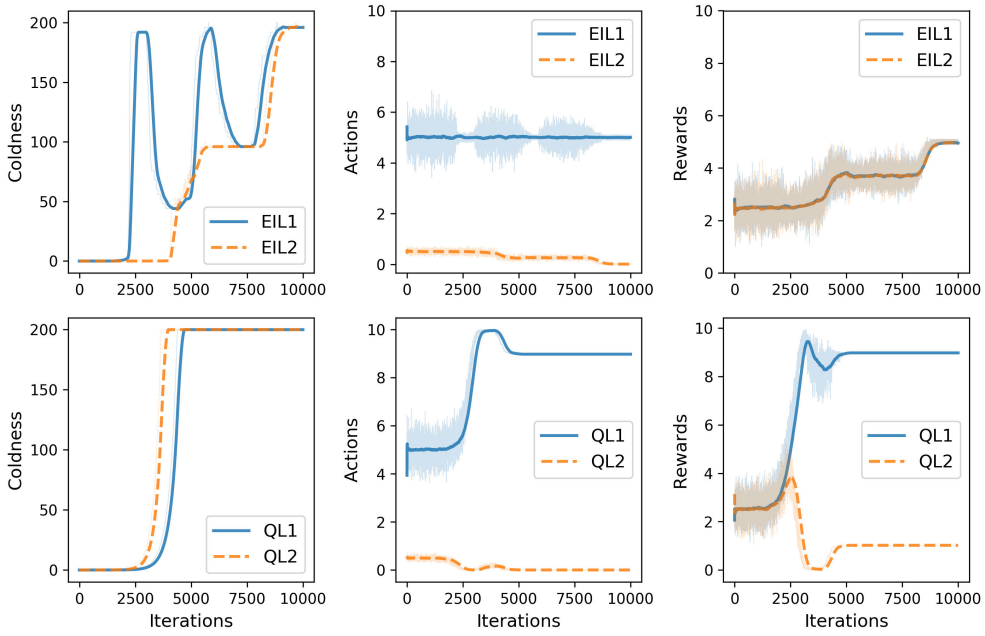


FIGURE 8. Coldness, actions, and rewards of the test between two EIL-agents (upper) and the test between two QL-agents (lower). Parameters: $\beta_{EIL1} = 1.005$, $\beta_{EIL2} = 1.002$, $\beta_{QL1} = 1.0036$, $\beta_{QL2} = 1.003$. State: equal mode. Environment: ultimatum game.

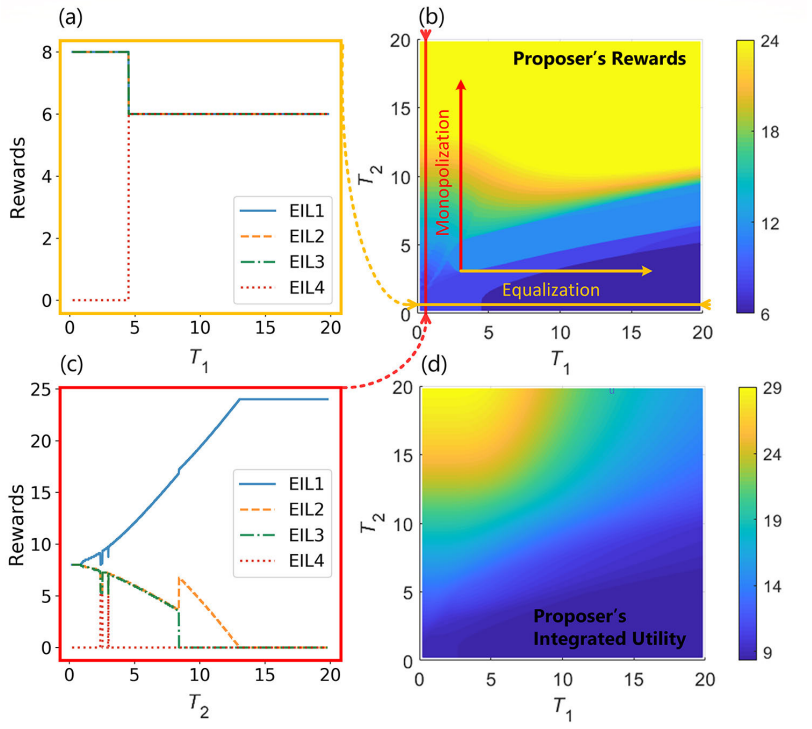


FIGURE 9. (a) Convergent rewards of each test between four EIL-agents when $T_1 = 0.2$ and T_2 changed from 0.2 to 19.8. (c) Convergent rewards of each test between four EIL-agents when $T_2 = 0.2$ and T_1 changed from 0.2 to 19.8. (b)(d) Theoretical value of the proposer's rewards and the corresponding maximum feasible integrated utility under different (T_1, T_2) pairs. Parameters: $\beta_{EIL1} = \beta_{EIL2} = \beta_{EIL3} = \beta_{EIL4} = 1.005$. Environment: 4-players ultimatum game.

3 members to the equal allocation with 4 members at a position of nearly 5 degrees of T_1, realizing the switch from oligopoly to equality as Theorem 3 describes. On the contrary, the proposer in Fig. 9(c) raised its proportion of resources as T_2 increased, which is consistent with the individual changes described in Theorem 2.

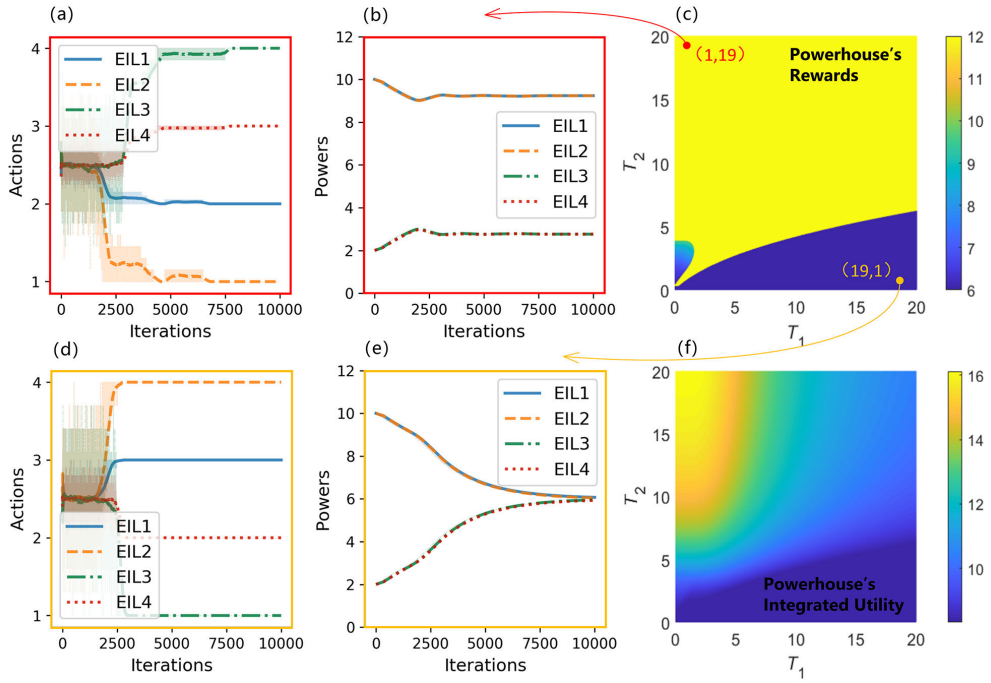


FIGURE 10. (a)(b) Actions and powers of the test between two strong agents {EIL1,EIL2} and two weak agents {EIL3,EIL4} when $T_1 = 1$ and $T_2 = 19$. (d)(e) Actions and powers of the test between two strong agents {EIL1,EIL2} and two weak agents {EIL3,EIL4} when $T_1 = 19$ and $T_2 = 1$. (c)(f) Theoretical value of the maximum feasible integrated utility and corresponding rewards of the powerhouse EIL1 or EIL2 under different (T_1, T_2) pairs. Parameters: $\beta_{EIL1} = \beta_{EIL2} = 1.005$, $\beta_{EIL3} = \beta_{EIL4} = 1.004$. Environment: survival game.

C. SURVIVAL GAME

Survival game is a positive-sum stochastic game with multiple participants. We designed this game to simulate the primitive hunting environment. In this game, agents can choose to hunt alone or team up with someone. Generally, the rational agents' will tend to cooperate with others because of the larger payoff of hunting in the team. It seems that introducing empathy on this basis makes no sense. However, things will change if we slightly complicated the game setting by giving agents different powers as follows.

1) Agent types:
   - $\{v_1, v_2\}$ – stands for the strong agents;
   - $\{v_3, v_4\}$ – stands for the weak agents.
2) Agent actions:
   - $a_i \in \{1, 2, 3, 4\}$ – stands for the action of v_i;
   - $a_i = i$ – stands for hunting alone;
   - $a_i = j \left(a_j \neq i, j \neq i\right)$ – stands for hunting alone;
   - $a_i = j \left(a_j = i, j \neq i\right)$ – stands for forming a team with v_j.
3) Reward for each agent:

$$R_i = \frac{\sum_{j \in \mathcal{C}_i} P_j}{\sum_{k \in [4]} \sum_{j \in \mathcal{C}_k} P_j} \cdot R$$

where $\mathcal{C}_i$ is the set of all teammates of v_i (including itself) and $C_i = \{v_i\}$ if v_i hunts alone, R is the total resources available, P_i is the power of agent i and it updates in the following form

$$P_{i,t} = P_{i,t-1} + \delta \left(R_{i,t} - P_{i,t-1}\right)$$

According to this setting, one team can accommodate up to two members, making it possible to form two different teams among the four agents. We can also regard the survival game as the bundling rewards version of the ultimatum game. The experiment was simulated to test EIL-agent's performance under different (T_1, T_2) pairs. In this test, the initial power of strong agents was $P_{1,0} = P_{2,0} = 10$ and the initial power of weak agents was $P_{3,0} = P_{4,0} = 2$. It should be noted that, in order to demonstrate the repeatability of this experiment, we only recorded the strong-weak combined teams of $\{v_1, v_3\}$, and $\{v_2, v_4\}$.

Fig. 10(f) and 10(c) depicted the theoretical value of the maximum feasible integrated utility and corresponding rewards of the strong EIL-agents under different (T_1, T_2) pairs. Therefore, within the dark blue area in the lower right corner of Fig. 10(c), EIL-agents with any difference of initial value would tend to combine the strong with the weak, thus making the powers converge to equality. For a further manifest of this area, we showed a test under the temperature pair of (19, 1) in Fig. 10(d) and 10(e). It is obvious that EIL-agents finally formed teams that combined the weak agents and the strong agents after a period of exploration. Go back and look at the yellow part of Fig. 10(c); the uniform distribution is no longer the optimal distribution in this area, which means that it is possible to maintain the difference in powers of

agents when the initial powers are unequal. Similarly, we further showed a sampled test under the temperature pair of (1, 19) depicted in Fig. 10(a) and Fig. 10(b). In this sample, EIL-agents formed teams of strong-by-strong and weak-by-weak with no mixture.

From the perspective of cooperation, it is positive that agents with different powers within a specific range can collaborate to reduce the gap between rich and poor. We can regard this tendency as a primordial drive of help from the strong to the weak. This phenomenon is also consistent with the assumption that generalized empathy towards similar creatures is beneficial to the community's stability by forming a virtuous circle for the symbiotic community in the early stage. Conversely, when the stability margin is exceeded, agents will enter the competitive state, such as monopoly and oligopoly. This will lead to differentiation within the multi-agent system.

VI. CONCLUSION

Empathy, a kind of spontaneous resonance on others' emotions, has positive implications for the continuation of populations and social harmony. In view of the critical role of empathy in living nature, we present a learning method called Empathy-based Interactive Learner and believe that the introduction of empathy in learning procedure is of profound significance for analyzing and designing multi-agent systems for the purpose of stability or evolution.

Previous works inspired by psychology has already illustrated that introducing affective functions such as guilt and forgiveness could enhance cooperation in a narrow sense. However, as for an autonomous multi-agent system as complex as human society, competition other than cooperation is also imperative to guarantee the smooth functioning of the system. More broadly, diverse collaboration needs to accommodate both stability under cooperation and evolution under competition. By modeling empathy as a state sharing function using thermodynamics, similar emotional states are induced among agents, enabling each agent to feel what others feel. In this case, using the learning method EIL with the optimization goal of empathetic utility multiplied by empathy entropy, agents can change modes by adjusting the temperature on the process of empathy. Under the specific temperature of the equal mode, when conflicted interest occurs as in the prisoner dilemma game, empathetic agents can consider their opponents and fully express goodwill on the premise of protecting their safe utility, whereas rational homo economics only maximize one's external reward. Besides, empathetic agents in the equal mode can feel others' sorrow and provide targeted help, which assures each individual's well-being and secures society's stable function as a whole. For agents in other modes, such as monopoly and oligopoly, the agent's awareness of competition will be highlighted. We believe that the unequal distribution of resources means the occurrence of evolution to some extent. Furthermore, in terms of adaptability, agents equipped with EIL can sense the trends of environmental fluctuation by monitoring the changing of income entropy, which provides EIL-agents with the ability to adapt to a dynamic environment.

The behaviors usually mentioned in psychology and behavioral economics, such as cooperation, altruism, fairness, and help, are considered as the different external manifestations under the same internal mechanism of empathy in our paper. However, the behavior in a more complex form is also dominated by the environment. In future work, we will try to explain the evolutionary patterns of the behavior under different external constraints of the environment, thereby further exploring the mechanism of complex collaboration and forming a relatively systematic collaboration theory, including the design of the collaboration model, analysis of collaboration index and optimization of collaboration algorithm.

REFERENCES

[1] W. Ren, R. W. Beard, and E. M. Atkins, "A survey of consensus problems in multi-agent coordination," in *Proc. ACC*, Portland, OR, USA, 2005, pp. 1859–1864.

[2] R. Olfati-Saber, "Flocking for multi-agent dynamic systems: Algorithms and theory," *IEEE Trans. Autom. Control*, vol. 51, no. 3, pp. 401–420, Mar. 2006, doi: 10.1109/TAC.2005.864190.

[3] S. Yong-Zheng and R. Jiong, "Consensus problems of multi-agent systems with noise perturbation," *Chin. Phys. B*, vol. 17, no. 11, pp. 4137–4141, Nov. 2008, doi: 10.1088/1674-1056/17/11/029.

[4] M. Nourian, P. E. Caines, R. P. Malhame, and M. Huang, "Nash, social and centralized solutions to consensus problems via mean field control theory," *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 639–653, Mar. 2013, doi: 10.1109/TAC.2012.2215399.

[5] J. Zhou, Q. Hu, and M. I. Friswell, "Decentralized finite time attitude synchronization control of satellite formation flying," *J. Guid., Control, Dyn.*, vol. 36, no. 1, pp. 185–195, Jan. 2013, doi: 10.2514/1.56740.

[6] Q. Wang, H. Gao, F. Alsaadi, and T. Hayat, "An overview of consensus problems in constrained multi-agent coordination," *Syst. Sci. Control Eng.*, vol. 2, no. 1, pp. 275–284, Mar. 2014, doi: 10.1080/21642583.2014.897658.

[7] J. A. Marvell, R. Bostelman, and J. Falco, "Multi-robot assembly strategies and metrics," *ACM Comput. Surv.*, vol. 51, no. 1, pp. 14–45, Jan. 2018, doi: 10.1145/3150225.

[8] J. Goldsmith and E. Burton, "Why teaching ethics to AI practitioners is important," *Proc. AAAI* San Francisco, CA, USA, 2017, pp. 4836–4840.

[9] D. C. Parkes and M. P. Wellman, "Economic reasoning and artificial intelligence," *Science*, vol. 349, no. 6245, pp. 267–272, Jul. 2015, doi: 10.1126/science.aaa8403.

[10] L. A. Martinez-Vaquero, T. A. Han, L. M. Pereira, and T. Lenaerts, "Apology and forgiveness evolve to resolve failures in cooperative agreements," *Sci. Rep.*, vol. 5, no. 1, p. 10639, Jun. 2015, doi: 10.1038/srep10639.

[11] L. M. Pereira, "Social manifestation of guilt leads to stable cooperation in multi-agent systems," *Proc. AAMAS*, Richland, SC, USA, 2017, pp. 1422–1430.

[12] L. M. Pereira, "Why is it so hard to say sorry? Evolution of apology with commitments in the iterated Prisoner's Dilemma," *Proc. IJCAI*, Beijing, China, 2013, pp. 177–183.

[13] C. Feng, Z. Li, X. Feng, L. Wang, T. Tian, and Y.-J. Luo, "Social hierarchy modulates neural responses of empathy for pain," *Social Cognit. Affect. Neurosci.*, vol. 11, no. 3, pp. 485–495, Oct. 2015, doi: 10.1093/scan/nsv135.

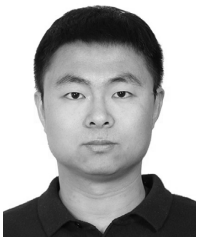
[14] L. Buşoniu, R. Babuška, and B. D. Schutter, "Multi-agent reinforcement learning: An overview," in *Innovations in Multi-Agent Systems and Applications—1*. Berlin, Germany: Springer, 2010, pp. 183–221.

[15] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM J. Comput.*, vol. 32, no. 1, pp. 48–77, Jan. 2002, doi: 10.1137/S0097539701398375.

[16] D. Bergemann and J. Välimäki, "Dynamic mechanism design: An introduction," *SSRN Electron. J.*, pp. 235–274, 2017, doi: 10.2139/ssrn.3024528.

[17] C. P. Cell, "Selective incentives versus ideological commitment: The motivation for membership in wisconsin farm organizations," *Amer. J. Agricult. Econ.*, vol. 62, no. 3, pp. 517–524, Aug. 1980, doi: 10.2307/1240207.

[18] L. T. Rameson and M. D. Lieberman, "Empathy: A social cognitive neuroscience approach," *Social Personality Psychol. Compass*, vol. 3, no. 1, pp. 94–110, Jan. 2009, doi: 10.1111/j.1751-9004.2008.00154.x.
[19] T. Singer and C. Lamm, "The social neuroscience of empathy," *Ann. New York Acad. Sci.*, vol. 1156, no. 1, pp. 81–96, Mar. 2009, doi: 10.1111/j.1749-6632.2009.04418.x.
[20] M. K. Davis, *Empathy: A Social Psychological Approach*. London, U.K.: Macmillan, 1985.
[21] V. Gazzola, L. Aziz-Zadeh, and C. Keysers, "Empathy and the somatotopic auditory mirror system in humans," *Current Biol.*, vol. 16, no. 18, pp. 1824–1829, Sep. 2006, doi: 10.1016/j.cub.2006.07.072.
[22] M. Kleiman-Weiner, R. Saxe, and J. B. Tenenbaum, "Learning a commonsense moral theory," *Cognition*, vol. 167, pp. 107–123, Oct. 2017, doi: 10.1016/j.cognition.2017.03.005.
[23] A. Salehi-Abari, C. Boutilier, and K. Larson, "Empathetic decision making in social networks," *Artif. Intell.*, vol. 275, pp. 174–203, Oct. 2019, doi: 10.1016/j.artint.2019.05.004.
[24] C. Zhang, X. Li, J. Hao, S. Chen, K. Tuyls, W. Xue, and Z. Feng, "SA-IGA: A multiagent reinforcement learning method towards socially optimal outcomes," *Auto. Agents Multi-Agent Syst.*, vol. 33, no. 4, pp. 403–429, May 2019, doi: 10.1007/s10458-019-09411-3.
[25] C. Godsil and G. F. Royle, *Algebraic Graph Theory*. New York, NY, USA: Springer, 2013.
[26] T. M. Moerland, J. Broekens, and C. M. Jonker, "Emotion in reinforcement learning agents and robots: A survey," *Mach. Learn.*, vol. 107, no. 2, pp. 443–480, Aug. 2017, doi: 10.1007/s10994-017-5666-0.
[27] C. R. Rogers, "Empathic: An unappreciated way of being," *Counseling Psychologist*, vol. 5, no. 2, pp. 2–10, Jun. 1975, doi: 10.1177/001100007500500202.
[28] N. Eisenberg and N. D. Eggum, "Empathic responding: Sympathy and personal distress," in *The Social Neuroscience of Empathy*, vol. 6. Cambridge, MA, USA: MIT Press, 2009, pp. 71–83.
[29] Y. Fan, N. W. Duncan, M. de Greck, and G. Northoff, "Is there a core neural network in empathy? An fMRI based quantitative meta-analysis," *Neurosci. Biobehavioral Rev.*, vol. 35, no. 3, pp. 903–911, Jan. 2011, doi: 10.1016/j.neubiorev.2010.10.009.
[30] G. Rizzolatti, "Premotor cortex and the recognition of motor actions," *Cogn. Brain Res.*, vol. 3, no. 2, pp. 131–141, Mar. 1996, doi: 10.1016/0926-6410(95)00038-0.
[31] P. Molenberghs, R. Cunnington, and J. B. Mattingley, "Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies," *Neurosci. Biobehavioral Rev.*, vol. 36, no. 1, pp. 341–349, Jan. 2012, doi: 10.1016/j.neubiorev.2011.07.004.
[32] J. A. C. J. Bastiaansen, M. Thioux, and C. Keysers, "Evidence for mirror systems in emotions," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 364, no. 1528, pp. 2391–2404, Aug. 2009, doi: 10.1098/rstb.2009.0058.
[33] C. D. Frith and T. Singer, "The role of social cognition in decision making," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 363, no. 1511, pp. 3875–3886, Oct. 2008, doi: 10.1098/rstb.2008.0156.
[34] H. Walter, "Social cognitive neuroscience of empathy: Concepts, circuits, and genes," *Emotion Rev.*, vol. 4, no. 1, pp. 9–17, Jan. 2012, doi: 10.1177/1754073911421379.
[35] J. Chen, "Promoting constructive interaction and moral behaviors using adaptive empathetic learning," in *Proc. ICIRA*, Shenyang, China, 2019, pp. 3–14.
[36] J. Chen and C. Wang, "Reaching Cooperation using Emerging Empathy and Counter-empathy," in *Proc. AAMAS*, Montreal, QC, Canada, 2019, pp. 746–753.
[37] S. R. de Groot, P. Mazur, and S. Choi, "Non-equilibrium thermodynamics," *Phys. Today*, vol. 16, no. 5, pp. 70–71, May 1963, doi: 10.1063/1.3050930.
[38] J. H. Justice, *Maximum Entropy and Bayesian Methods in Applied Statistics*. Cambridge, U.K.: Cambridge Univ. Press, 1984.
[39] R. Axelrod, "The evolution of strategies in the iterated prisoner's dilemma," in *The Dynamics of Norms*. Cambridge, U.K.: Cambridge Univ. Press, 1987, pp. 1–16.
[40] S. Debove, N. Baumard, and J.-B. André, "Models of the evolution of fairness in the ultimatum game: A review and classification," *Evol. Human Behav.*, vol. 37, no. 3, pp. 245–254, May 2016, doi: 10.1016/j.evolhumbehav.2016.01.001.
[41] M. van 't Wout, R. S. Kahn, A. G. Sanfey, and A. Aleman, "Affective state and decision-making in the ultimatum game," *Exp. Brain Res.*, vol. 169, no. 4, pp. 564–568, Feb. 2006, doi: 10.1007/s00221-006-0346-5.
[42] D. S. Wilson and E. O. Wilson, "Rethinking the theoretical foundation of sociobiology," *Quart. Rev. Biol.*, vol. 82, no. 4, pp. 327–348, Dec. 2007, doi: 10.1086/522809.



JIZE CHEN (Graduate Student Member, IEEE) was born in 1992. He received the B.S. and M.S. degrees from the Harbin Institute of Technology, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include multi-agent learning and intelligent control.

DALI ZHANG was born in 1991. He received the B.S. and M.S. degrees from the Harbin Institute of Technology, in 2013 and 2015, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include swarm algorithms and trajectory planning.

ZHENSHEN QU (Member, IEEE) was born in 1973. He received the B.S., M.S., and Ph.D. degrees in control science and engineering from the Harbin Institute of Technology (HIT), in 1995, 1998, and 2003, respectively. He is currently an Associate Professor of control science and engineering with the HIT. His current research interests include signal and image processing, autonomous vehicle systems, and visual servo.

CHANGHONG WANG (Senior Member, IEEE) was born in 1961. He received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology (HIT), in 1983, 1986, and 1991, respectively. He is currently a Professor and Ph.D. Student Supervisor with the HIT. His research interests include inertial navigation, precise servo control systems, and robust control.