

Received December 29, 2019, accepted January 25, 2020, date of publication February 3, 2020, date of current version February 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2971278

Evaluation Mechanism of Collective Intelligence for Heterogeneous Agents Group

ANNA DAI¹, ZHIFENG ZHAO², RONGPENG LI¹, HONGGANG ZHANG¹,
AND YUGENG ZHOU³

¹College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

²Zhejiang Lab, Hangzhou 311121, China

³Zhejiang Wanfeng Technology Development Company, Ltd., Zhejiang 312500, China

Corresponding author: Rongpeng Li (lirongpeng@zju.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1301003, in part by the National Natural Science Foundation of China under Grant 61701439 and Grant 61731002, in part by the Zhejiang Key Research and Development Plan under Grant 2019C01002 and Grant 2019C03131, in part by the Project sponsored by the Zhejiang Lab under Grant 2019LC0AB01, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY20F010016.

ABSTRACT Collective intelligence is manifested when multiple agents coherently work in observation, interaction, decision-making and action. In this paper, we define and quantify the intelligence level of heterogeneous agents group organized in a flat structure without explicit leadership by the improved *Anytime Universal Intelligence Test (AUIT)*, an extension of the existing evaluation of homogeneous agents group. The relationship of intelligence level with agents' composition, group size, spatial complexity and testing time is analyzed. The intelligence level of heterogeneous agents groups is compared with that for the homogeneous groups as well, so as to demonstrate the effects of heterogeneity on collective intelligence. Our work will contribute to understand the essence of collective intelligence more deeply and reveal the effect of various key factors on group intelligence level.

INDEX TERMS Collective intelligence, heterogeneous agents group, intelligence level, intelligence test.

I. INTRODUCTION

Collective or group is a very common organizational structure of intelligent creatures. Collective intelligence means that group of individuals behaves collectively in an intelligent manner [1]. Numerous kinds of tasks are increasingly accomplished by groups, making it ever more important to understand the determinants of group performance [2]. And the quantitative analysis of collective intelligence is an important part of studying and understanding group performance. A human group's performance on a wide variety of tasks can be explained by a general collective intelligence factor [2]. The theory of collective intelligence is helpful for understanding many aspects of group performance, bringing benefits to scientific research and practical applications [3]. Notably, the terminology collective, which involves hierarchies or human beings, is treated differently from the term group [4], [5]. However, our paper primarily focuses on a decentralized group structure [6], consisting of artificial

machines or simple agents [7]. Such a decentralized group structure, which is widely applied to the field of artificial machines like UAV (unmanned aerial vehicle), often has a flat structure without explicit leadership. So, in this paper, we assume the terminologies "collective" and "group" interchangeably to both refer to the aggregation of agents without leadership [8].

Collective intelligence only occurs when there are interactions between agents in the group [9]. The agents' ability to observe/perceive and sense the environment is a fundamental characteristic of agent-based systems [10]. Based on what they observe in the environment, agents in groups interact with each other in a direct (e.g., human talking) or indirect (e.g., ants leaving pheromones) way. Then the agents can carry out actions, and achieve their predefined goals. The process above illustrates a fundamental framework of multi-agent groups interacting within the environment. In this regard, heterogeneous groups are the aggregations of two or more interactive agents of different behaviors, while homogeneous ones are those agents with the same behaviors [11]. Heterogeneous agents have been widely used

The associate editor coordinating the review of this manuscript and approving it for publication was Chee Keong Kwoh.

to study real-world problems regarding cybersecurity [12] and economy [13], [14].

The intelligence level of heterogeneous groups cannot be achieved directly by a homogeneous group model due to the changes of group behavior [15]. In this paper we develop a mechanism of quantifying the intelligence level of heterogeneous group by the improved Anytime Universal Intelligence Test (AUIT). Our simulation results demonstrate that the intelligence level of heterogeneous collectives is higher than same size homogeneous collectives in most cases. And the composition/heterogeneity of heterogeneous collectives also has an important impact on the intelligence level.

The remainder of the paper is organized as follows. Possible ways of evaluating the intelligence of groups and state-of-the-art work is presented in Section II. Section III formulates the system model and evaluation mechanism. The experiment settings and parameters are in Section IV. We present our experiment results in Section V along with some discussion and analysis of quantitative results. In Section VI, we briefly draw conclusion and introduce the future directions.

II. BACKGROUND

There are several authoritative methods to quantify the intelligence of isolated agent, but it’s hard to quantify the intelligence of groups. David L. Dowe [16] proposed an additional computational requirement on intelligence, the ability of expression, as an extension to the Turing Test. C-Test is a test for comprehending ability, equally applicable to both humans and machines, which was presented by J Hernandez-Orallo [17]. Kannan and Parker [18] proposed an effective metric for the evaluation of learning capability towards understanding system level fault-tolerance. Schreiner [19] presented a study related to creating standard measures for systems that can be considered intelligent, which is realized by the US National Institute of Standards and Technology (NIST). Javier Insa-Cabrera [20], [21] analysed the influence after including agents with different degrees of intelligence and identified the components that should be considered when measuring social intelligence in multi-agent systems. Presented by Fox and Martin [22], an agent benchmark model is developed as a basis for analyzing and comparing multiple agent systems with cognitive capabilities. In the research of Anthon and Jannett [23], the intelligence is gauged in terms of the task intelligent costs. Hibbard [24] proposed a metric for intelligence measuring based on a hierarchy of increasingly complex environment sets. And an agent’s intelligence is measured as the ordinal of the most difficult set of environments it can pass. Chmait et al. [25] proposed a metric considered “universal” and appropriate to empirically measure the intelligence level of different agents or groups. J Hernández-Orallo [26] presented a way to estimate the difficulty and discriminating power of any task instance. A measure for machine intelligence was proposed by Legg and Hutter [27], who mathematically formulated essential features about human intelligence to produce a general measure of intelligence for

arbitrary machines. Peter M. Kraff [28] conceived general collective intelligence as measuring group performance on classes of cognitive tasks. X. Zhu [29] compared the optimization performance of three different group intelligence algorithms that were run on a support vector machine (SVM). Albert B. Kao and Iain D. Couzin [30] modeled modularity within animal groups and examined how it affected the amount of information represented in collective decisions, as well as the accuracy of those decisions. N.T. Nguyen [31], [32] presented an approach to calculate the collective knowledge state with the collective elements’ knowledge states, involving the distances from the collective knowledge state to the collective elements on a quasi-Euclidean space. J. Korczak [33] evaluated the collective intelligence in trading decisions by calculating the linear combination of several indicators related to the performance of market transaction process. Nader Chmait [25], [34], [35] provided an information-theoretic solution, and quantified and analyzed the impact of communication and observation abilities on the intelligence of homogeneous multi-agent system. They considered a series of factors hindering and influencing the effectiveness of interactive cognitive systems [11], [36], [37]. Finally, we summarize the possibly used taxonomy for evaluation criteria and calculation methods in Table 1.

However, there is no authoritative or universal way of collective intelligence evaluation for agent based groups at present. Besides, the current researches on group intelligence evaluation have not differentiated between homogeneous and heterogeneous groups. The impact of heterogeneity on collective intelligence has not been fully discussed. Considering the above situation, the improved AUIT model and method, together with the evaluation analysis of heterogeneous groups, are brought forward in this paper.

TABLE 1. The possible evaluation criteria and calculation methods of collective intelligence.

Criteria	Task-oriented	Cost [23] Complexity [24] Accuracy [19], [33] ...
	Ability-oriented	Learn [18] Comprehend [17] Cognize [22] ...
Calculation methods	Statistical [38]	Arithmetic average Geometric average Harmonic average Median/Mode Distances on a quasi-Euclidean space [31], [32] ...
	Structural	Weighted average [33] Probability(abstract modeling) [39] ...
	Psychometrical	Voting Theoretically semi-computable [40] ...

III. MODEL AND METHOD

A. THE SYSTEM MODEL

The Anytime Universal Intelligence Test (AUIT) [41] is a method to evaluate the intelligence level of homogeneous

multi-agent groups. The test simulates agents working in a finite environment and calculate the rewards corresponding to their actions. The average rewards over all agents are considered as the intelligence level of this group. The model works in a toroidal grid space, named *The Λ^* (Lambda Star) Environment*.

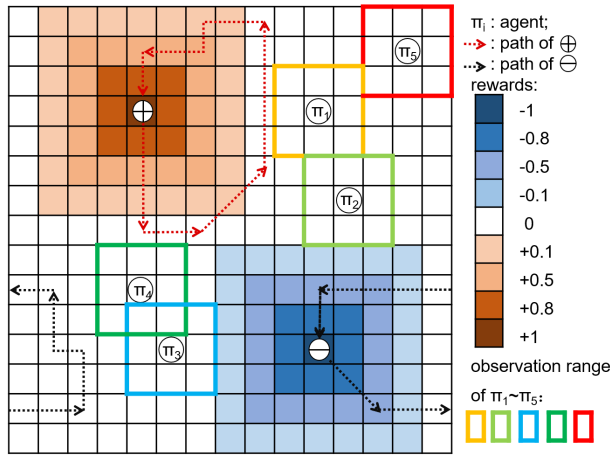


FIGURE 1. Test in a 15×15 environment with 5 agents and 2 special objects.

To evaluate the intelligence level of heterogeneous groups, we extend the AUIT model as shown in Fig. 1. The environment is a toroidal grid space (periodic boundaries) which means that moving off one border makes you appear on the facing one. In this test environment, there are objects from finite set $\Omega = \{\pi_1, \pi_2, \dots, \pi_x, \oplus, \ominus\}$ which contains working agents ($\Pi \subseteq \Omega, \Pi = \{\pi_1, \pi_2, \dots, \pi_x\}$) and two moving special objects, Good (\oplus) and Evil (\ominus). The two special objects travel with measurable complexity movement patterns in the environment. Once the size of the environment is determined, each cell is marked with an ordered index. The movement patterns are repeating series of cell index which illustrate where the special objects will be in one test episode. And the complexity of the series is considered to be the complexity of movement patterns [11]. Each element in Ω can work as a finite set of move actions $A = \{\text{left}, \text{right}, \text{up}, \text{down}, \text{up} - \text{left}, \text{up} - \text{right}, \text{down} - \text{left}, \text{down} - \text{right}, \text{stay}\}$. Reward is defined as a function of the distance of the evaluated agent to objects \oplus and \ominus [11]. In the test environment, given an agent π_j , its reward r_j^i at some test iteration i is a real number, calculated as follows, in which r_{j+}^i represents the reward caused by \oplus , and r_{j-}^i represents the reward caused by \ominus :

$$r_{j+}^i = \begin{cases} +1 & d_{\pi_j, \oplus} = 0, \\ +0.8 & d_{\pi_j, \oplus} = 1, \\ +0.5 & d_{\pi_j, \oplus} = 2, \\ +0.1 & d_{\pi_j, \oplus} = 3, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$$r_{j-}^i = \begin{cases} -1 & d_{\pi_j, \ominus} = 0, \\ -0.8 & d_{\pi_j, \ominus} = 1, \\ -0.5 & d_{\pi_j, \ominus} = 2, \\ -0.1 & d_{\pi_j, \ominus} = 3, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

$$r_j^i = r_{j+}^i + r_{j-}^i \quad (3)$$

where $d_{\pi_j, \oplus}$ and $d_{\pi_j, \ominus}$ means the (toroidal) chessboard distance [42] between π_j and \oplus or \ominus respectively. For example, in a 10-by-10 grid-world, the distance from cell (2, 1) to (2, 10) is 1. The reward of agent is the combination of the effects of two distances. The snapshot of agent rewards map is shown in Fig. 2.

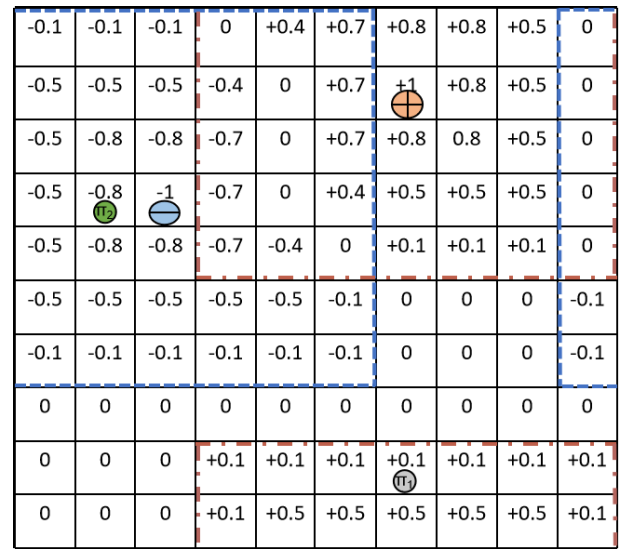


FIGURE 2. The snapshot of reward map.

Besides, the two special objects act as the moving targets in the evaluation, and the agents' work is to chase Good (\oplus) and keep away from Evil (\ominus). With these settings, a test episode consist of a series of ϑ iterations working as Algorithm 1. In Algorithm 1, an observation means the reward information of π_j^i 's observation range (1 Moore neighbour cells) [43]. The evaluation result is the average reward of each agent over each iteration [11], shown in Algorithm 1.

In environment μ , there are two types of complexity. One is task complexity $K(\mu)$ [11], which corresponds to the difficulty of the task, and is mathematically represented by the Kolmogorov complexity [44] of the two special objects' movement patterns. The other one is the search space complexity or environmental complexity $H(\mu)$ [11], represented by Shannon entropy [45] of the environment, which stands for the uncertainty of μ and corresponds to the size of the environment. As for an environment μ with a size of $m \times n$, S_μ is a possible environment status. N is the set of S_μ . And the environment complexity is calculated as follows,

$$|N| = \frac{(m \times n)!}{(m \times n - 2)!} \quad (4)$$

Algorithm 1 Evaluation Algorithm

Input: Π (set of n evaluated heterogeneous agents), special objects (\oplus and \ominus), A (set of actions), environment size $m \times m$, iteration number ϑ

Output: The evaluation of an n -agent group's intelligence

- 1: Agents from $\Pi \subseteq \Omega$ and the two special objects \oplus and \ominus are randomly distributed in the m -by- m toroidal grid-world // Initialize
- 2: **for** $i \leftarrow 1$ to ϑ **do**
- 3: **for** $j \leftarrow 1$ to n **do**
- 4: The environment sends an observation to π_j^i
 // Observation
- 5: **end for**
- 6: **for** $j \leftarrow 1$ to n **do**
- 7: π_j^i interacts with other agents about the observation and takes an action from A // Action
- 8: **end for**
- 9: The two special objects \oplus and \ominus perform the next action in their movement pattern and renew the rewards distribution in the environment
- 10: **for** $j \leftarrow 1$ to n **do**
- 11: The environment returns a reward r_j^i to π_j^i according to its distance to the special objects // Reward
- 12: **end for**
- 13: **end for**
- 14: Return $\tilde{R}_{\Pi, \mu, \vartheta} \leftarrow \frac{\sum_{j=1}^n \sum_{i=1}^{\vartheta} r_j^i}{n \times \vartheta}$

$$p(S_\mu) = \frac{1}{|N|} \quad (5)$$

$$H(\mu) = - \sum_{s_\mu \in N} p(s_\mu) \log_2 p(s_\mu) = \log_2 |N| \quad (6)$$

The movement patterns are randomly generated before each test. To evaluate the collectives over the same task complexity, we do the simulation with the special objects following the same movement pattern while the other settings (total number of agents, environment size, test and iteration times) remain unchanged. Furthermore, we increase the search space complexity by enlarging the size of the environment.

B. THE EVALUATION METHOD

In the heterogeneous group evaluation, we take several types of agents into consideration. They include Local Search Agent, Oracle Agent, and Random Agent. Agents are able to move on the same cell, while the special objects are not allowed to be in the same cell.

- a. Local Search Agent: This kind of agent will choose the cell of highest reward in its observation range to be the target cell in one iteration. If the rewards of cells in the observation range are all equal, it will randomly choose one.
- b. Oracle Agent: It knows the movement pattern of the Good special object and can get close to it in the fastest way.

- c. Random Agent: It randomly chooses one neighbor cell as its next moving target cell.

The synergy among the agents in the group is crucial to the performance [46]. The evaluation also considers different group communication methods, such as Talking, Stigmergy, Imitation.

- a. Talking (direct communication): Agents exchange observation information with each other in communication range, then choose the highest reward cell as their moving target.
- b. Stigmergy [47] (indirect communication): Agents get their own observation exactly. Each of them gets others' observations with random fake rewards.
- c. Imitation: Agents take the same action as other agent in their observation range. When there are more than one agent in range, they randomly choose one to follow.

The extended AUIT environment is like a real map, while the agent is like UAV. UAVs gather for search and avoidance missions in a decentralized way. They act on the observation of the environment and the information gathered from others. That constitutes a practical application scenario for the evaluation mechanism. In particular, different agents get information at each episode according to the communication methods they take. Agents with direct communication will get all agents' exact observations, while agents with indirect communication will get others' observations with bias. And agents with the imitation capability will get the action information of other agents in their observation range. After that, agents choose and perform actions. Then they get rewards from the environment. At the end of a test episode, the average rewards for each agent are considered as the intelligence level of the group.

For instance, considering a scenario where there is a heterogeneous group consisting of 8 local search agents using stigmergy and 2 local search agents using talking, these ten agents, as well as two extra special objects, are randomly scattered in the environment at the beginning of one episode. Then in each iteration, the ten agents individually get their own observation which contains the true rewards of cells in their own observation range. Afterwards, each local search agent using talking gets all the other agents' true observations while every local search agent using stigmergy gets the other agents' observations added with their fake rewards. Next, each agent chooses the cell with the highest reward to be its target cell according to the information it observes and receives. And each agent takes a selected action to get close to its target cell. Then, the two special objects change their positions (i.e., the action performed by objects) following their own pre-defined movement patterns and renew the rewards distribution in the environment. Subsequently, the environment returns the reward of the cell where the agent stays as the final reward it gets for this iteration. In a word, each episode includes several iterations and every agent/object only takes one action in each iteration. At the end of one episode, the average of each agent's reward in each iteration is the evaluation result in this episode. The similar procedure

applies to the following episodes and we take the final average of all episodes' result as the collective intelligence level of the heterogeneous group.

In the field of organizational behavior, the calculation of collective intelligence possibly need to take leadership and decision making into consideration [48]. The evaluation result could be much more complex than the average result of each agent. However, the mechanism we use here applies to the criterion of accuracy mentioned in Table 1 under the assumption that the collectives for the study now are decentralized. In this case, to assess the overall accuracy of the task, we consider the average rewards to be the final performance of the collectives. When the test environment and agent behaviors get more complicated in the future, the calculation method of the collective intelligence should be accordingly extended. For example, the weighted average or target reaching probability under abstract modeling could be exploited for the hierarchical collective cases.

IV. NUMERICAL EXPERIMENT PARAMETERS

Our experiments try to figure out the impact of agent composition, group size, environmental complexity and evaluation time on the intelligence level of heterogeneous groups. And we also compare the intelligence of heterogeneous group with a homogeneous group with the same size to understand the impact of heterogeneity. The agents and communication methods in our simulation, as well as the corresponding symbols are listed in Table 2.

TABLE 2. Symbols and Expression.

Symbol	Meaning
<i>SL</i>	Local search agents using indirect communication (Stigmergy Local Search Agent).
<i>TL</i>	Local search agents using direct communication (Talking Local Search Agent)
<i>IL</i>	Local search agents follow any other type of agent in its observation range (Imitative Local Search Agent), or just work as a Local Search agent exchanging no information if no agent in observation range.
<i>O</i>	Oracle agents. They offer their observations according to other agents' communication method.
<i>R</i>	Random agents. They offer their observations according to other agents' communication method.
$\&$	" <i>Ax&By</i> " represents a heterogeneous group consists of <i>x</i> type <i>A</i> agents and <i>y</i> type <i>B</i> agents ($x, y \in N$)

e.g., "*SL10*" represents a homogeneous group consists of 10 *SLs*. "*SL9&O1*" represents a heterogeneous group consists of 9 *SLs* and 1 *O*.

In the evaluation of heterogeneous groups, we mainly carry out the simulation in a 20×20 environment, corresponding to $H(\mu) = 17.2$ bits. And each test lasts 20 iterations. As for the fake rewards for agents using stigmergy, we firstly get the true observation of an agent π_j , and we set the fake rewards to be the average result of the maximum and minimum of the observed rewards. After adding the fake rewards to the true observation, we get a biased observation and when needed it will be sent to those agents who use stigmergy. In our simulation, the observation range is set to be 1, which means

that an agent can observe the cell where it is and one circle around it, 9 cells in total. Currently, agents using talking or stigmergy are able to get the information they need from all the other agents, indicating that the communication range is not limited, which can be optimized in the future. All agents have the same speed as each takes one action from the finite action set *A* to move from a cell to an adjacent one or stay still in every iteration. In order to figure out the impact of environmental complexity, time and group size, $H(\mu)$ varies from 13.2 bits to 19.6 bits as shown in Table 3, number of iterations varies from 10 to 500, and the number of agents varies from 10 to 60 (the ratio of the components remains the same). At the beginning of each test iteration, all the agents are randomly scattered in the environment. To reduce the influence of random initial positions, we use the average result of ten test episodes with the same settings.

TABLE 3. Environment Complexity.

$m \times n$	10×10	15×15	20×20	25×25	30×30
$H(\mu)$	13.2 bits	15.6 bits	17.2 bits	18.5 bits	19.6 bits

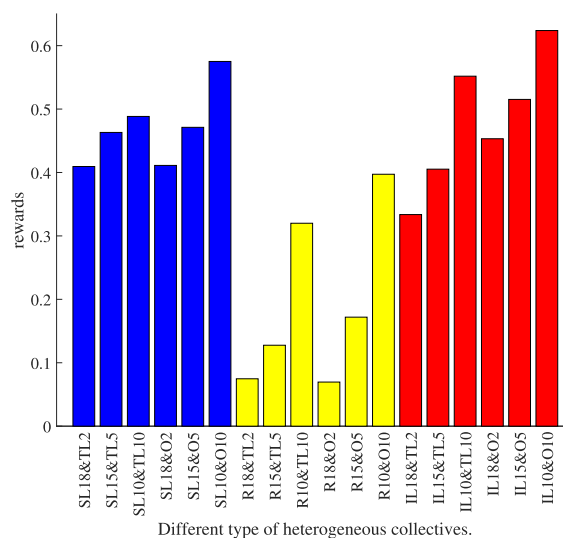


FIGURE 3. Evaluation of heterogeneous collectives.

V. NUMERICAL EXPERIMENTS AND RESULTS

A. EVALUATION OF HETEROGENEOUS AGENTS GROUP

1) AGENTS COMPOSITION

In heterogeneous groups, agents of different decision strategies or communication methods work cooperatively. We combine different agents to get heterogeneous groups and evaluate their intelligence level, and the corresponding results are illustrated in Fig. 3. It can be observed from Fig. 3 that the intelligence level of heterogeneous groups is mainly determined by the intelligence level of the components. If the group size is fixed, the group shows higher intelligence level as the heterogeneity gets stronger. For example, the first three columns in the left of Fig. 3 show the increase of group intelligence level as the proportion of *SL* and *TL* getting closer.

This indicates that heterogeneity can indeed help improve the group intelligence.

2) IMPACT OF AGENT NUMBER

Just as the saying goes “many hands make light the work [49]”. We enlarge the group size step by step to observe the changes in their quantified collective intelligence level, and the result is shown in Fig. 4. In Fig. 4, the expression “nbr” means the total number of agents, and the ratio of one type agents to the other types in the collectives remains 9:1 (e.g., the second column of the first cluster named “SL&TL” with a slash pattern represents a collective consisting of 18 SLs and 2 TLs). In the test environment, more agents means more information can be observed, leading to a better reward. But with the increase in the number of agents, the result will come to an upper limit. Since the amount of information in the environment is limited, the increase in agent number makes the available information to be mined quickly and the group performance will reach a ceiling in finite evaluation iteration times.

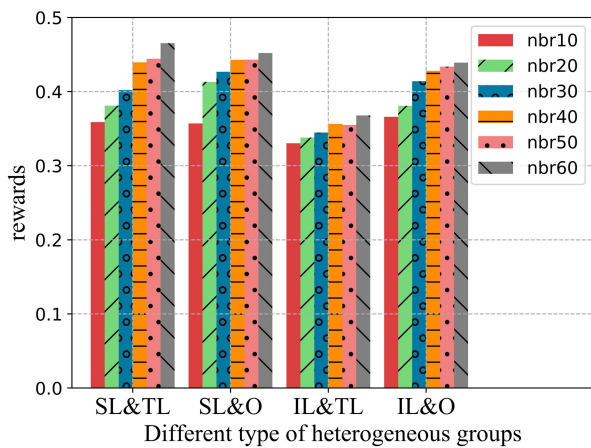


FIGURE 4. Evaluation of heterogeneous collectives of different agent numbers.

3) IMPACT OF ENVIRONMENT COMPLEXITY

We gradually extend the size of the environment to increase the search space complexity, and get results in Fig. 5. Fig. 5 illustrates that heterogeneous collective intelligence will decrease with the increase of the environment complexity. The reason lies in that in the environment with a large space, it is challenging to meet \oplus and elude \ominus within a finite numbers of iterations, which results in the degradation of heterogeneous group performance. In other words, when the environment is too large for any agent to sense or learn the position of the special objects, all agents in the group may behave like random and aimless walks, leading to the rewards they get closing to 0.

4) IMPACT OF EVALUATION TIME

We extend the evaluation time by increasing the number of iterations for each test, and we get Fig. 6. It can

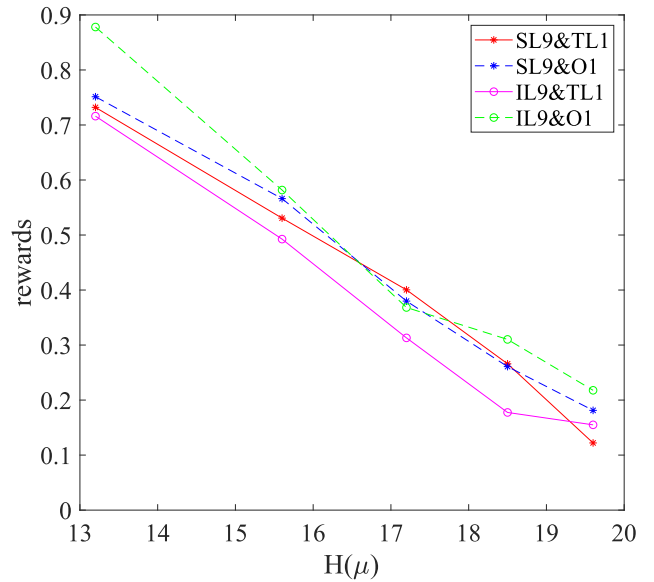


FIGURE 5. Evaluation of heterogeneous collectives with different environment complexities.

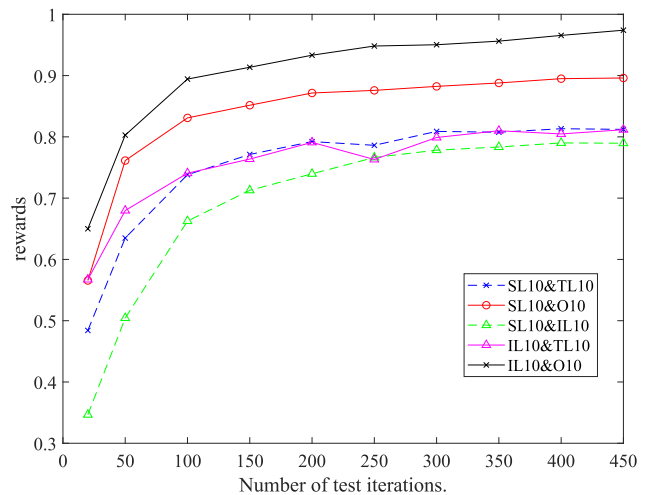


FIGURE 6. Evaluation of heterogeneous collectives with different evaluation time.

be observed from Fig. 6 that collectives demonstrate an increase in the intelligence level. As time goes longer, agents get more chance to seek and follow \oplus as well as staying away from \ominus . And the evaluation results gradually come to an upper limit with sufficient iterations. In Fig. 6, the gap between *IL10&TL10* and *IL10&O10* is larger than that between *SL10&TL10* and *SL10&O10*. That means the intelligence level of indirect communicating heterogeneous groups is more stable than that of imitative heterogeneous groups.

B. COMPARISON BETWEEN HOMOGENEOUS GROUP AND HETEROGENEOUS ONE

1) AGENTS COMPOSITION

We evaluate the intelligence level of heterogeneous groups and homogeneous groups of the same size. As a comparison,

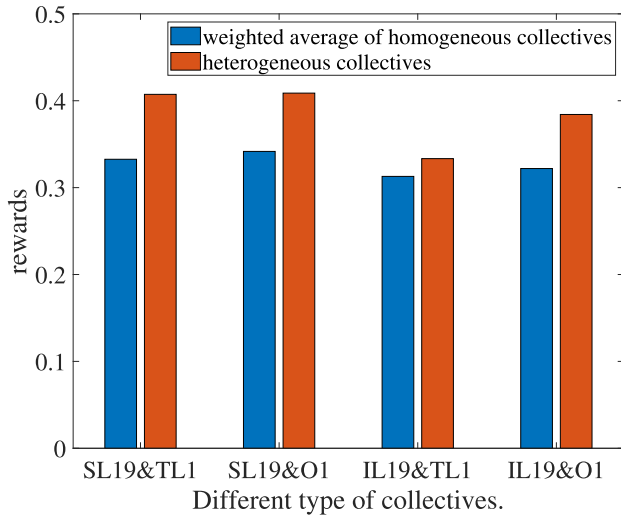


FIGURE 7. Comparison of weighted average of homogeneous collectives and heterogeneous collectives.

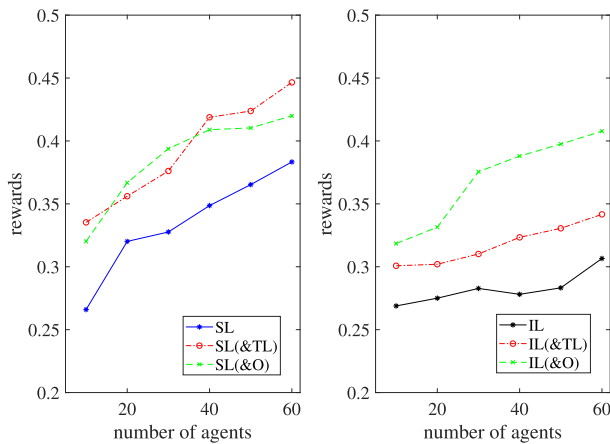


FIGURE 8. Comparison between homogeneous groups and the less intelligent part of agents in heterogeneous groups with agent number increasing.

we calculate the weighted average of homogeneous groups that consist of the same types and numbers of agents with the heterogeneous groups. For example, the weighted average of homogeneous collectives “SL19&TL1” is calculated as

$$(19 \times SL_{homo} + 1 \times TL_{homo})/20 \quad (7)$$

where SL_{homo} is the intelligence level of 20-agent homogeneous SL group, and TL_{homo} is the intelligence level of 20-agent homogeneous TL group. And that’s how we get Fig. 7. It can be observed from Fig. 7 that the improvement of the heterogeneous group intelligence level benefits from the better performance of most agents rather than the minor more intelligent ones. Fig. 7 implies that heterogeneity does have a positive impact on group intelligence level.

Besides, we compare homogeneous groups and the less intelligent part of agents in heterogeneous groups with the number of agents increasing, showing the results in Fig. 8.

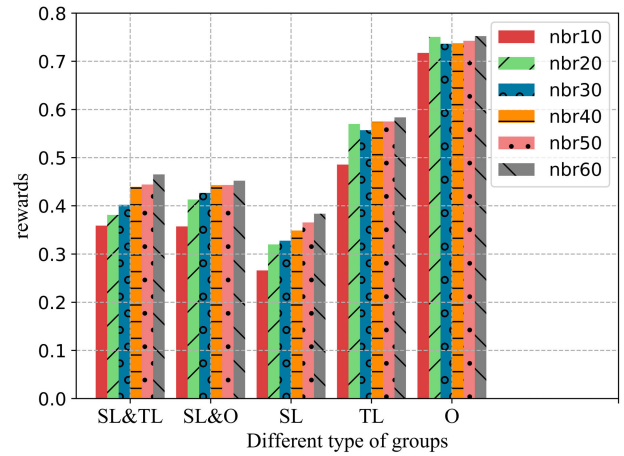


FIGURE 9. Evaluation of groups of various agent numbers.

In Fig. 8, the symbol “SL(&TL)” means the quantified result of SL in the heterogeneous collectives which also contain TL. Moreover, the ratio of SL or IL to the other types of agents remains 9:1. It can be observed from Fig. 8 that the homogeneous part in heterogeneous collectives apparently shows the improvement caused by heterogeneity. In the left part of Fig. 8, the improvement obtained from introducing TL is similar to that benefited from O. However, the right part of Fig. 8 shows that working together with O can make better improvement for the performance of IL, since they are highly dependent on other types of agents.

2) IMPACT OF AGENT NUMBER

We evaluate the intelligence level of homogeneous and heterogeneous groups with the number of agents increasing, showing the results in Fig. 9. The group performance improves with the agents’ number increasing in both heterogeneous and homogeneous cases. In Fig. 9, homogeneous group “TL” is apparently less intelligent than “O”. However, in heterogeneous groups, “SL&TL” and “SL&O” get similar performance, indicating that for multi-agents with indirect communications, the improvement caused by heterogeneity may have unclear relationship with the original performance of the wiser agents. Indirectly communicating groups achieve a similar intelligence level, even if the agents with different intelligent levels are added to the groups.

3) IMPACT OF ENVIRONMENT COMPLEXITY

Homogeneous groups and heterogeneous groups are evaluated with the increase of environment complexity, and the results are shown in Fig. 10. In Fig. 10, both homogeneous groups and heterogeneous groups show intelligence level decrease when environment complexity increases. The decrease rate of heterogeneous group intelligence level is between that of the individual components. The impact of environment complexity on heterogeneous group intelligence level is mostly determined by the components. But,

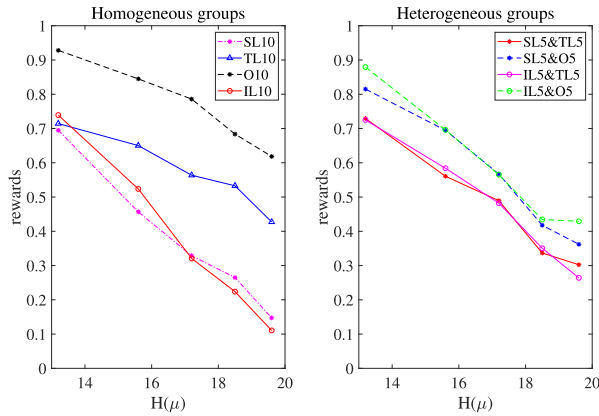


FIGURE 10. Evaluation of homogeneous and heterogeneous groups with different environment complexities.

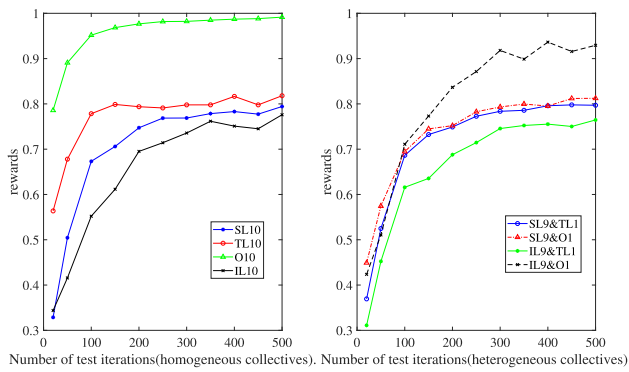


FIGURE 11. Evaluation of homogeneous and heterogeneous groups of different evaluation time.

the heterogeneity can make the group performance more stable.

4) IMPACT OF EVALUATION TIME

Fig. 11 shows the variation of the reward with respect to the iteration index. It can be observed from Fig. 11 that homogeneous and heterogeneous collectives show similar performance along with the increasing of iterations. Most heterogeneous groups' performance rises slower than that of homogeneous groups. But "IL9&O1" outperforms "TL10" and "SL9&O1" may work as well as "TL10" when number of iterations is sufficient. The performance of heterogeneous collectives which contain few high-performance agents (i.e., SL9&TL1) is very close to that of the homogeneous collectives containing only high-performance agents (i.e., TL10). When time expanding, we can expect heterogeneous groups mainly consist of low-performance agents to get high intelligence level. That is of great significance in reality since high-performance agents, like TL and O, are often energy-intensive or even impractical.

VI. CONCLUSION AND FUTURE WORK

To evaluate the intelligence level of heterogeneous group, we improve the AUIT model and the related evaluation method. We evaluate the intelligence level of different heterogeneous groups and study the impact of agent

composition together with communication methods, group size, environment complexity and evaluation time. Experiment results prove that (a) Heterogeneity can improve the group intelligence level; (b) More agents and longer test time can also lead to better group performance; (c) The intelligence level improvement of heterogeneous groups that mainly adopt indirect communication is quite stable, while groups most made of imitative agents are more likely to be affected by external conditions such as the space size and evaluating time.

In the future, to make the simulation closer to the actual situation, especially for the indirect communication method, the generation of fake rewards should have something to do with the distances between agents. And effective ways of controlling the task complexity should be added to the evaluation mechanism. To expand our work, we consider enriching the agent types, for example, incorporating reinforcement learning agents, and adopting different position initiation strategies in simulation. More complex heterogeneous collectives consisting of more types of agents can be taken into consideration. Finding a proper way to define the costs of groups working on a task is helpful for the utility analysis. There is substantial practical value in formulating a computational framework to optimally tune the best agent group for a given mission. The framework tends to optimize the group intelligence level by adjusting agent numbers, types and communication methods and constraints about cost. As groups get larger and agents become more intelligent, the organizational behavior can play a vital role in the collective performance [50]. In this regard, it is quite valuable to take account of different organizational structures when the evaluation mechanism gets more comprehensive and applicable for more complex agents. Accordingly, the calculation method of collective intelligence should be improved as well. We leave such a meaningful research direction as future works.

REFERENCES

- [1] T. W. Malone and M. S. Bernstein, *Handbook of Collective Intelligence*. Cambridge, MA, USA: MIT Press, 2015.
- [2] A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, and T. W. Malone, "Evidence for a collective intelligence factor in the performance of human groups," *Science*, vol. 330, no. 6004, pp. 686–688, Oct. 2010.
- [3] A. W. Woolley, I. Aggarwal, and T. W. Malone, "Collective intelligence and group performance," *Current Directions Psychol. Sci.*, vol. 24, no. 6, pp. 420–424, Dec. 2015.
- [4] F. J. Yammarino and F. Dansereau, "A new kind of organizational behavior," in *Multi-level Issues in Organizational Behavior and Processes*. Bingley, U.K.: Emerald Group Publishing Limited, 2009, pp. 13–60.
- [5] F. Dansereau, J. A. Alutto, and F. J. Yammarino, *Theory Testing in Organizational Behavior: The Variant Approach*. Upper Saddle River, NJ, USA: Prentice-Hall, 1984.
- [6] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Auton. Agents Multi-Agent Syst.*, vol. 11, no. 3, pp. 387–434, Nov. 2005.
- [7] D. H. Wolpert and K. Tumer, "An introduction to collective intelligence," 1999, *arXiv:cs/9908014*. [Online]. Available: <https://arxiv.org/abs/cs/9908014>
- [8] M. J. Mataric, "From local interactions to collective intelligence," in *The Biology and Technology of Intelligent Autonomous Agents*. Berlin, Germany: Springer, 1995, pp. 275–295.

- [9] D. Weschler, "Concept of collective intelligence," *Amer. Psychol.*, vol. 26, no. 10, pp. 904–907, 1971.
- [10] S. Franklin and A. Graesser, "Is it an agent, or just a program?: A taxonomy for autonomous agents," in *Proc. Int. Workshop Agents Theories*, 1997, pp. 27–35.
- [11] N. Chmait, Y.-F. Li, D. L. Dowe, and D. G. Green, "A dynamic intelligence test framework for evaluating AI agents," in *Proc. Workshop Eval. Gen.-Purpose AI, EGPAI*, 2016, pp. 1–8.
- [12] J. Grossklags, N. Christin, and J. Chuang, "Security and insurance management in networks with heterogeneous agents," in *Proc. 9th ACM Conf. Electron. Commerce (EC)*, 2008, pp. 160–169.
- [13] R. M. Burlando and F. Guala, "Heterogeneous agents in public goods experiments," *Exp. Econ.*, vol. 8, no. 1, pp. 35–54, Apr. 2005.
- [14] H. Dawid, P. Harting, S. Van Der Hoog, and M. Neugart, "Macroeconomics with heterogeneous agent models: Fostering transparency, reproducibility and replication," *J. Evol. Econ.*, vol. 29, no. 1, pp. 467–538, Mar. 2019, doi: [10.1007/s00191-018-0594-0](https://doi.org/10.1007/s00191-018-0594-0).
- [15] S. Magg and R. T. Boekhorst, "Pattern formation in homogeneous and heterogeneous swarms: Differences between versatile and specialized agents," in *Proc. IEEE Symp. Artif. Life (ALIFE)*, Apr. 2007, pp. 311–316.
- [16] D. Dowe and A. Hajek, "A non-behavioural, computational extension to the turing test," in *Proc. Int. Conf. Comput. Intell. Multimedia Appl.* Singapore: World Scientific Publishing, 1998, pp. 101–106.
- [17] I. Arel and S. Livingston, "Beyond the turing test," *Computer*, vol. 42, no. 3, pp. 90–91, Mar. 2009.
- [18] B. Kannan and L. E. Parker, "Metrics for quantifying system performance in intelligent, fault-tolerant multi-robot teams," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 951–958.
- [19] K. Schreiner, "Measuring IS: Toward a US standard," *IEEE Intell. Syst.*, vol. 15, no. 5, pp. 19–21, Sep. 2000.
- [20] J. Insa-Cabrera, J.-L. Benacloch-Ayuso, and J. Hernandez-Orallo, "On the influence of intelligence in (social) intelligence testing environments," 2012, *arXiv:1202.0837*. [Online]. Available: <https://arxiv.org/abs/1202.0837>
- [21] J. Insa-Cabrera and J. Hernández-Orallo, "Instrumental properties of social testbeds," in *Proc. Int. Conf. Artif. Gen. Intell.* Cham, Switzerland: Springer, 2015, pp. 101–110.
- [22] J. Fox, M. Beveridge, and D. Glasspool, "Understanding intelligent agents: Analysis and synthesis," *AI Commun.*, vol. 16, no. 3, pp. 139–152, 2003.
- [23] A. Anthony and T. C. Jannett, "Measuring machine intelligence of an agent-based distributed sensor network system," in *Advances and Innovations in Systems, Computing Sciences and Software Engineering*. Dordrecht, The Netherlands: Springer, 2007, pp. 531–535.
- [24] B. Hibbard, "Measuring agent intelligence via hierarchies of environments," in *Proc. Int. Conf. Artif. Gen. Intell.*, 2011, pp. 303–308.
- [25] N. Chmait, "Understanding and measuring collective intelligence across different cognitive systems: An information-theoretic approach," in *Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI)*, Melbourne, Australia, 2017, pp. 5171–5172.
- [26] J. Hernández-Orallo, "On environment difficulty and discriminating power," *Auton. Agents Multi-Agent Syst.*, vol. 29, no. 3, pp. 402–454, May 2015.
- [27] S. Legg and M. Hutter, "A formal measure of machine intelligence," 2006, *arXiv:cs/0605024*. [Online]. Available: <https://arxiv.org/abs/cs/0605024>
- [28] P. M. Krafft, "A simple computational theory of general collective intelligence," *Topics Cogn. Sci.*, vol. 11, no. 2, pp. 374–392, Apr. 2019.
- [29] X. Zhu, N. Li, and Y. Pan, "Optimization performance comparison of three different group intelligence algorithms on a SVM for hyperspectral imagery classification," *Remote Sens.*, vol. 11, no. 6, p. 734, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/6/734>
- [30] A. B. Kao and I. D. Couzin, "Modular structure within groups causes information loss but can improve decision accuracy," *Philos. Trans. Roy. Soc. B*, vol. 374, no. 1774, Jun. 2019, Art. no. 20180378.
- [31] N. T. Nguyen, "Inconsistency of knowledge and collective intelligence," *Cybern. Syst., Int. J.*, vol. 39, no. 6, pp. 542–562, 2008.
- [32] N. T. Nguyen, "A method for improving the quality of collective knowledge," in *Proc. Asian Conf. Intell. Inf. Database Syst.* Cham, Switzerland: Springer, 2015, pp. 75–84.
- [33] J. Korczak, M. Hernes, and M. Bac, "Collective intelligence supporting trading decisions on FOREX market," in *Proc. Int. Conf. Comput. Collective Intell.* Cham, Switzerland: Springer, 2017, pp. 113–122.
- [34] N. Chmait, D. L. Dowe, D. G. Green, and Y. F. Li, "Observation, communication and intelligence in agent-based systems," in *Proc. Int. Conf. Artif. Gen. Intell.*, 2015, pp. 50–59.
- [35] N. Chmait, D. L. Dowe, D. G. Green, Y.-F. Li, and J. Insa-Cabrera, "Measuring universal intelligence in agent-based systems using the anytime intelligence test," FIT, Clayton, Monash Univ., Melbourne, VIC, Australia, Tech. Rep. 2015/279, 2015.
- [36] N. Chmait, D. L. Dowe, Y.-F. Li, D. G. Green, and J. Insa-Cabrera, "Factors of collective intelligence: How smart are agent collectives?" in *Proc. ECAI*, 2016, pp. 542–550.
- [37] N. Chmait, D. L. Dowe, Y.-F. Li, and D. G. Green, "An information-theoretic predictive model for the accuracy of AI agents adapted from psychometrics," in *Proc. Int. Conf. Artif. Gen. Intell.* Cham, Switzerland: Springer, 2017, pp. 225–236.
- [38] J. A. Fadul, "Collective learning: Applying distributed cognition for collective intelligence," *Int. J. Learn., Annu. Rev.*, vol. 16, no. 4, pp. 211–220, 2009.
- [39] T. Szuba, "A formal definition of the phenomenon of collective intelligence and its IQ measure," *Future Gener. Comput. Syst.*, vol. 17, no. 4, pp. 489–500, 2001.
- [40] D. L. Dowe and J. Hernández-Orallo, "How universal can an intelligence test be?" *Adapt. Behav.*, vol. 22, no. 1, pp. 51–69, Feb. 2014.
- [41] J. Hernández-Orallo and D. L. Dowe, "Measuring universal intelligence: Towards an anytime intelligence test," *Artif. Intell.*, vol. 174, no. 18, pp. 1508–1539, Dec. 2010.
- [42] Y.-H. Lee and S.-J. Horng, "The chessboard distance transform and the medial axis transform are interchangeable," in *Proc. 10th Int. Conf. Parallel Process. (IPPS)*, Dec. 1996, pp. 424–428.
- [43] L. Gray, "A mathematician looks at wolfram's new kind of science," *Notices-Amer. Math. Soc.*, vol. 50, no. 2, pp. 200–211, 2003.
- [44] P. M. Vitanyi and M. Li, *An Introduction to Kolmogorov Complexity and Its Applications*, vol. 34, no. 10. Heidelberg, Germany: Springer, 1997.
- [45] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.
- [46] S. Liemhetcharat and M. Veloso, "Weighted synergy graphs for effective team formation with heterogeneous ad hoc agents," *Artif. Intell.*, vol. 208, pp. 41–65, Mar. 2014.
- [47] P.-P. Grassé, "La reconstruction du nid et les coordinations interindividuelles chez *Bellicositermes natalensis* et *Cubitermes* sp. La théorie de la stigmergie: Essai d'interprétation du comportement des termites constructeurs," *Ins. Soc.*, vol. 6, no. 1, pp. 41–80, Mar. 1959.
- [48] K. A. Mchugh, F. J. Yammarino, S. D. Dionne, A. Serban, H. Sayama, and S. Chatterjee, "Collective decision making, leadership, and collective intelligence: Tests with agent-based simulations and a Field study," *Leadership Quart.*, vol. 27, no. 2, pp. 218–241, Apr. 2016.
- [49] B. Latané, K. Williams, and S. Harkins, "Many hands make light the work: The causes and consequences of social loafing," *J. Personality Social Psychol.*, vol. 37, no. 6, pp. 822–832, Jun. 2006.
- [50] F. J. Yammarino, S. D. Dionne, J. U. Chun, and F. Dansereau, "Leadership and levels of analysis: A state-of-the-science review," *Leadership Quart.*, vol. 16, no. 6, pp. 879–919, Dec. 2005.



ANNA DAI received the B.S. degree from Zhejiang University, Hangzhou, China, in 2018. She is currently pursuing the master's degree with the College of Information Science and Electronic Engineering, Zhejiang University. Her research interests include collective intelligence, intelligence evaluation, and machine learning.



ZHIFENG ZHAO received the bachelor's degree in computer science, the master's degree in communication and information system, and the Ph.D. degree in communication and information system from the PLA University of Science and Technology, Nanjing, China, in 1996, 1999, and 2002, respectively.

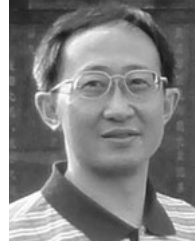
From 2002 to 2004, he was a Postdoctoral Researcher with Zhejiang University, China, where his works were focused on multimedia next-generation networks (NGNs) and soft-switch technology for energy efficiency. From 2005 to 2006, he was a Senior Researcher with the PLA University of Science and Technology, where he performed research and development on advanced energy-efficient wireless router, ad hoc network simulator, and cognitive mesh networking test bed. He is currently the Director of the Research Development Department, Zhejiang Lab, China. He is also with Zhejiang University. His research interests include cognitive radio, wireless multihop networks (Ad Hoc, Mesh, and WSN), wireless multimedia networks, and green communications.

Dr. Zhao is the Symposium Co-Chair of the China Com 2009 and 2010. He is the Technical Program Committee (TPC) Co-Chair of the IEEE 10th IEEE International Symposium on Communication and Information Technology (ISCIT) 2010.



RONGPENG LI received the B.E. degree from Xidian University, Xi'an, China, in 2010, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2015, both as "Excellent Graduates." From 2015 to 2016, he was a Research Engineer with the Wireless Communication Laboratory, Huawei Technologies Co. Ltd., Shanghai, China. He returned to academia, in 2016, initially as a Postdoctoral Researcher with the College of Computer Science and Technologies, Zhejiang

University, which is sponsored by the National Postdoctoral Program for Innovative Talents. He is currently an Assistant Professor with the College of Information Science and Electronic Engineering, Zhejiang University. His research interests currently focus on reinforcement learning, data mining, and all broad-sense network problems (e.g., resource management and security). He has authored/coauthored several articles in the related fields. He serves as an Editor for *China Communications*.



HONGGANG ZHANG was the International Chair Professor of excellence with the Université Européenne de Bretagne and Supélec, France. He is currently a Full Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. He is also an Honorary Visiting Professor with the University of York, U.K. He was a coauthor and an Editor of two books *Cognitive Communications-Distributed Artificial Intelligence (DAI)*, *Regulatory Policy and Economics, Implementation* (John Wiley & Sons) and *Green Communications: Theoretical Fundamentals, Algorithms and Applications* (CRC Press), respectively. He served as the Chair for the Technical Committee on Cognitive Networks of the IEEE Communications Society, from 2011 to 2012. He is also active in the research on cognitive radio and green communications. He was a leading Guest Editor of the *IEEE Communications Magazine* Special Issues on Green Communications and a Series-Editor of the *IEEE Communications Magazine* for its Green Communications and Computing Networks Series. He is the Associate Editor-in-Chief of *China Communications*.



YUGENG ZHOU received the B.S. and M.S. degrees from University of Science and Technology of China, in June 2010 and June 2007, respectively. Since October 2010, he has been an Engineer with Zhejiang Wanfeng Technology Development Company, Ltd., Zhejiang, China. He is currently an Engineer with the Institute of Wanfeng Jinyuan Holding Group Company, Ltd., Shaoxing. His research interests currently focus on control system for industrial robot, multirobot cooperative work and intelligent manufacturing. He has authored/coauthored several articles in the related fields.

• • •