



HuGoS: a virtual environment for studying collective human behavior from a swarm intelligence perspective

Nicolas Coucke^{1,2} · Mary Katherine Heinrich¹ · Axel Cleeremans² · Marco Dorigo¹

Received: 22 December 2020 / Accepted: 29 June 2021 / Published online: 3 August 2021
© The Author(s) 2021

Abstract

Swarm intelligence studies self-organized collective behavior resulting from interactions between individuals, typically in animals and artificial agents. Some studies from cognitive science have also demonstrated self-organization mechanisms in humans, often in pairs. Further research into the topic of human swarm intelligence could provide a better understanding of new behaviors and larger human collectives. This requires studies with multiple human participants in controlled experiments in a wide variety of scenarios, where a rich scope of possible interactions can be isolated and captured. In this paper, we present HuGoS—‘Humans Go Swarming’—a multi-user virtual environment implemented using the Unity game development platform, as a comprehensive tool for experimentation in human swarm intelligence. We demonstrate the functionality of HuGoS with naïve participants in a browser-based implementation, in a coordination task involving collective decision-making, messaging and signaling, and stigmergy. By making HuGoS available as open-source software, we hope to facilitate further research in the field of human swarm intelligence.

Keywords Swarm intelligence · Human behavior · Collective behavior · Virtual environment · Unity 3D · Human dynamics · Social psychology · Self-organization · Cognitive science · Human–swarm interaction · Complex systems · Swarm robotics

✉ Nicolas Coucke
nicolas.coucke@ulb.ac.be

Mary Katherine Heinrich
mary.katherine.heinrich@ulb.ac.be

Axel Cleeremans
axcleer@ulb.ac.be

Marco Dorigo
mdorigo@ulb.ac.be

¹ IRIDIA, Université Libre de Bruxelles, Brussels, Belgium

² CO3, Center for Research in Cognition and Neurosciences, Université Libre de Bruxelles, Brussels, Belgium

1 Introduction

Many collective human behaviors arise from simple interactions between individuals, akin to behaviors traditionally studied within the field of swarm intelligence (Krause et al. 2010). For example, models of human crowds have shown that there are many similarities between the collective behavior of human crowds and that of large groups of animals (Moussaïd et al. 2009). However, the study of most collective human behaviors—such as those seen in team dynamics (O’Bryan et al. 2020), online social networks (Lepri et al. 2016), or collective problem solving (Quinn and Bederson 2011)—is challenging, because the behaviors often result from complex interactions between individuals that may be difficult to capture or quantify (Ellwart 2011).

To study the cognitive mechanisms underpinning complex human interactions, research areas such as human social cognition (Frith and Frith 2012) or joint action (i.e., collective action limited to few individuals, as in Vesper et al. 2017) study the cognitive processes that are used when two individuals coordinate their actions on a joint task. These research areas typically study dyads, where two individuals interpret and coordinate with the other’s actions, and each has a distinct and essential role in the task (Vesper et al. 2017). These studies perform fine-grained analyses of individual mechanisms of coordination. On the other end of the spectrum, studies looking at large groups of individuals either consider cases such as virtual crowds, in which individuals usually have minimal interactions (Lorenz et al. 2011), or make abstractions from the complex interactions that occur between the individuals (Navajas et al. 2018). To conduct a complete analysis of collective human behavior and study swarm intelligence topics, we require comprehensive experimental data describing individual behavior, interactions, collective behavior, and relationships to task and environment features.

Virtual environments have been proposed as useful tools for studying human behavior in controlled experiments (Bailenson et al. 2004). A virtual environment allows the experimenter to analyze and manipulate all information available to participants as well as actions executed by participants, which would be prohibitively cumbersome during in-person experiments. In this paper, we present HuGoS—‘Humans Go Swarming’—a multi-user virtual environment built to support experiments in human collective behavior. This paper is an extension of a previous conference paper (Coucke et al. 2020), in which we proposed the idea of HuGoS and presented an initial proof-of-concept. In this paper, we contribute a fully developed prototype of HuGoS that is open-source¹ and ready to be used by the research community. We also demonstrate the functionalities of HuGoS by running case studies with anonymous naïve participants and assess the performance, advantages, and limitations of HuGoS as a tool for potential experimenters. In HuGoS, human participants interact via avatars in a controlled experimental setup. HuGoS supports a wide variety of interactions among participants, and between participants and the environment. HuGoS also captures detailed data about each of these interactions, and the participants, objects, and properties involved. By enabling these interactions, HuGoS supports the study of complex forms of human self-organization such as self-organized hierarchy or emerging patterns of communication.

To enable direct comparative studies between human groups and artificial swarms, HuGoS supports autonomous avatars controlled by, e.g., a finite state machine or an AI

¹ HuGoS is available on GitHub at <https://github.com/NicolasCoucke/HuGoS-code-MAIN>.

algorithm, in addition to human-controlled avatars. This comparison could bring new approaches to commonly studied problems in swarm intelligence, such as the best-of- n problem. Comparative studies could also focus on mechanisms that are not commonly studied in swarm intelligence. For example, self-organized leadership and hierarchy are infrequently studied in artificial swarms but are typical for human groups, and have recently been described as a key development for future swarm robotics applications (Dorigo et al. 2020).

This paper is organized as follows. In Sect. 2, we discuss existing virtual environments used for studying collective human behavior. In Sect. 3, we give a general description of the design of HuGoS and the scope of experimentation that it targets, with further details provided in Appendices 1 and 2. In Sect. 4, we describe the methodology used to demonstrate the functionalities of HuGoS via online experiments with anonymously recruited participants. The experiments are organized into three case studies in a coordination task: (1) basic collective decision-making, (2) additional messaging and signaling in collective decision-making, and (3) stigmergic coordination. In Sect. 5, we describe the results of the case studies and also assess the advantages and limitations of HuGoS as a tool for experimenters, according to the following criteria: (1) connection and latency, (2) participant responses to questionnaires, and (3) tool performance, usability, and flexibility. Based on these results, Sect. 6 discusses the suitability of HuGoS for studying human swarm intelligence, and Sect. 7 concludes the paper by summarizing the main contributions.

2 Related work

A number of multi-user virtual environments have been developed for scientific experiments with multiple humans participants. In this section, we discuss environments designed for experiments relevant to the following: (1) solving external problems, (2) ‘embodied’ collective behavior, (3) large-large-scale social networks, and (4) joint action. Lastly, we also discuss the potential of existing video games and robot simulators for studying swarm intelligence in humans.

The first category of environments harnesses the collective intelligence of multiple participants to solve difficult problems. Many of these environments are aimed at solving computationally intensive problems (Barrington et al. 2011; Cooper et al. 2010; Eberhart et al. 2015; Kirschenbaum and Palmer 2015; Lin et al. 2014; Jensen et al. 2020). The UNUM platform (Rosenberg et al. 2016; Rosenberg 2015), also referred to as Swarm AI®, lets participants collaboratively explore a decision space. Each player controls a ‘magnet’ that exerts influence on a ‘puck.’ The participants can make a collective decision by moving the puck to one of several locations that are labeled with an answer to a question asked by the experimenter. While HuGoS is also used to study humans when they are solving complex problems, the platform is mainly developed to study participants’ behavior during problem solving, rather than the outcome.

The second category of virtual environments is geared toward studying real-time physical coordination between individuals. For example, Unity has been adopted to study crowd behaviors by supporting human-like avatars that have a first-person view of the environment (Moussaïd et al. 2016; Zhao et al. 2018, 2020). In these environments, participants’ user inputs are accurately transferred to an avatar, which creates an almost ecological (representative of real-life) setting (Thrash et al. 2015). Other environments (e.g., the HoneyComb game for human crowd movement in Boos et al. 2019) support the study

of leadership—specifically, the impact of better informed individuals on implicit leadership (Boos et al. 2014). These environments are promising, but lack features for setting up complex tasks in a dynamic environment. They also have a limited range of possible interactions and relations between participants that could be instrumental in solving complex tasks. For example, higher-order mechanisms for coordination such as hierarchies would require explicit leadership links between participants, reminiscent of the ‘follower’ functionalities in online trading networks (Krafft et al. 2016).

The third category of environments takes an approach that is suited to study this type of advanced coordination. Participants are embedded in a network and make explicit decisions based on the information they see about their neighbors in the network. These approaches are useful for studying the impact of network ties on collective behavior (e.g., in experiments on game theory). A commonly investigated game theory scenario is the public good game (PGG), where people can behave as contributors (cooperators) or free-riders (defectors). In the VIAPPL software,² each participant is represented by a 2D dot that is embedded in a (often partially observable) network of other players who are also represented by dots. This software allows one to study the influence of social psychological factors, such as a shared identity, on behavior in a PGG (Titlestad et al. 2019). The Breadboard software³ is similar but is more suitable for large-scale online experiments studying the influence of changing network structures on behavior (Rand et al. 2011; Fowler and Christakis 2010). It also allows for some of the nodes to be replaced by autonomous agents (Shirado and Christakis 2017). Similar tools such as *Empirica* (Almaatouq et al. 2020) have been used to study collective intelligence in large networked groups. These platforms facilitate the study of several important modulators of human behavior and collective intelligence; however, they leave out real-time embodied interaction between individuals. With HuGoS, we wish to integrate aspects of both by implementing an embodied approach with dynamic interactions that also supports the development of, e.g., interaction networks between participants.

The fourth category of virtual environments is used in disciplines such as social neuroscience, experimental semiotics, and joint action. These environments are geared toward studying cognitive mechanisms that underlie human collective action, but at the level of the individual or of the dyad. These virtual environments consider simple tasks for two or three participants. For example, Stolk et al. (2013) used a simple game where precisely two participants have their own perspectives on a joint playing field. Participants could only communicate through the movements of their avatars. In a controlled fMRI experiment, this game was instrumental in elucidating the neuro-cognitive mechanisms underlying coordination based on mutual understanding. Similar approaches even study how simple languages emerge when participants are involved in a coordination game (Selten and Warglien 2007; Scott-Phillips et al. 2009). Other approaches allow unlimited text messages between participants and analyze the relationship between message content and task execution (Nölle et al. 2020). A typical feature of human collective action is the accumulation and usage of cultural practices. These dynamics can be studied on a short time scale with, e.g., a simple video game (Derex and Boyd 2015). We do not propose that our environment can study all these mechanisms with the same level of detail. Rather, we equip our environment with sufficient features to study the influence of many of these mechanisms, such as the development of signaling conventions, on collective behavior.

² <http://viappl.org/>.

³ <https://breadboard.yale.edu/>.

Many existing video games such as Minecraft can also be used and modified to conduct controlled experiments (Nebel et al. 2016). For example, Minecraft is well suited to studying a joint construction task. Data collected during these tasks can be used to study how the interdependence of sub-tasks impacts collaboration, performance, and learning among the participants (Nebel et al. 2017). Other approaches have studied how participants achieve coordinated play on tasks that are specific to video games such as World of Warcraft, using audio and video recordings of players' screens (Williams and Kirschner 2012). Games like World of Warcraft have also shown the potential to study collective behavior under external events such as pandemics (Lofgren and Fefferman 2007). Within the scope of existing video games, there is emerging research that builds computational models of players, characterizing the relationships between player inputs and outputs during a game. Players have a large range of possible actions at each particular moment, such that modeling the rich interactions between players and game is considered “a holy grail of game design and development” (Yannakakis and Togelius 2018). By building on technology originally purposed for video game design, our approach includes many of the perks inherent to multi-player video games, such as an immersive user experience. Beyond this, our approach supports the modeling of player behavior better than most first-person video games. Because of the uncluttered nature of our game environments, player modeling will be a more feasible undertaking than in existing video games, where specific interactions and influences are much more difficult to isolate.

Comparing human collective behavior with existing swarm robotics approaches requires simulated robots and player-controlled avatars to operate in the same environment. Tools such as ARGoS (Pinciroli et al. 2012), ROS (Quigley et al. 2009), and Webots (Michel 2004) have been used to support simulations of multi-robot systems. A few studies using these tools have looked at human–swarm interaction, e.g., using support from ROS (Walker et al. 2014), or simulated in Webots (Vasile et al. 2011). In these setups, human participants gave high-level commands to parts of the robot swarm and were not able to directly control any robot. Tavakoli et al. (2016) provide human participants with the information that a robot would typically have, and let participants control a 2D avatar in a typical robot scenario. We aim to expand on these approaches by enabling not only scenarios used to study robots, but also those suitable to the study of human behavior. Specifically, we aim to provide a user experience that is intuitive and motivating to the participants, and to support the study of coordination patterns that humans would display in real-world scenarios.

We aim to expand on the scope of these tools by enabling human participants to control avatars in scenarios that are otherwise identical to those used to study robots. Lastly, existing video games that have been designed for multiple human players are increasingly being used to train and study AI agents (e.g., OpenAI et al. 2019; Jaderberg et al. 2019). Future research in this area could compare the collective behavior of human players with that of AI-trained artificial agents in video games. Since our environment can, next to player-controlled avatars, include autonomous agents and accurate robot models, we support the comparison of human collective behavior to that of robots or AI agents, as well as the study of hybrid human–robot swarms.

3 HuGoS: ‘Humans Go Swarming’

We designed HuGoS as an environment to facilitate a wide range of experiments on the topic of human swarm intelligence. In this section, we identify the experimental scope that HuGoS targets and then describe the architecture and features of HuGoS that enable these experiments.

3.1 Experimentation scope for human swarm intelligence

We take the scenarios explored in existing studies of robot and artificial swarms as starting points for studying human swarm intelligence. By studying both human and artificial swarms in the same experiment setup, one can more easily compare their performance and transfer behaviors between them. This allows studies of collective human behavior to build on results found in artificial swarms and can also generate new approaches for swarm robotics that are inspired by humans. With this in mind, there are several classes of experiments that HuGoS should support to facilitate comprehensive study of human swarm intelligence. One class of experiments would study physical coordination between individuals. This class includes behaviors such as aggregation, pattern formation, and self-assembly (e.g., Rubenstein et al. 2014). In HuGoS, this would require an avatar controlled by each participant, which other participants can observe. Another class of experiments studies behaviors that involve observation of environmental features. In best-of- n collective decision-making, a swarm might choose the best of several options based on observations of the environment and move to the corresponding location (e.g., Valentini et al. 2017). In cooperative navigation, agents might extract and share information to find the shortest path in an environment (e.g., Ducatelle et al. 2013). In HuGoS, these experiments would require environments populated with observable and changeable features. A third class of experiments studies agents that actively modify the environment. For example, agents might use coordination via stigmergy by leaving a virtual pheromone trail (e.g., Hunt et al. 2019). In tasks such as collective construction, agents might pick up and move construction blocks (e.g., Werfel et al. 2014). In HuGoS, this would require that some objects can be modified or manipulated using avatar controls.

Several HuGoS capabilities are required across all classes of experiments. Each class involves various methods of direct and indirect communication between players. Studies of direct communication might include, for instance, the impact of simple signals and messages on group performance. HuGoS is therefore equipped with simple signaling between avatars (such as placing a crown above the avatar), and the exchanging of short text messages. It is also important that an experimenter can place express limitations on communication. For instance, during indirect communication via observation, an experimenter might limit a player’s view to include only its avatar’s immediate neighbors.

Any class might also include studies of explicit coordination approaches, such as self-organized hierarchical control structures (Mathews et al. 2017; Zhu et al. 2020; Zhang et al. 2021) or task allocation (Labella et al. 2004). By analyzing—or imposing—communication network structures, HuGoS can facilitate the study of coordination mechanisms. Each class could also include comparison or collaboration between human and artificial agents. This requires HuGoS to support autonomous agents with avatars that may be indistinguishable from human players. For direct comparison between humans and robots following the same approach, HuGoS should also support robot models as avatars. In our

Fig. 1 HuGoS can be used to define agent-based models that both provide insights into human behavior and lead to swarm robotics applications

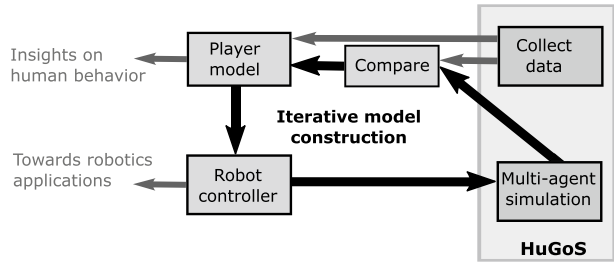
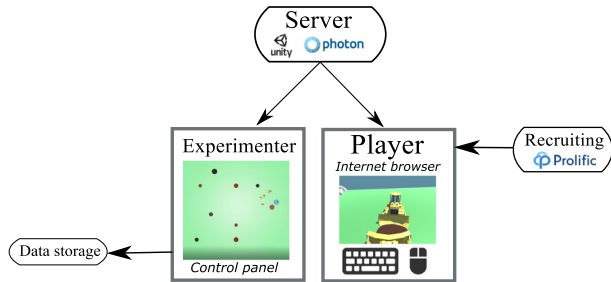


Fig. 2 Illustration of the HuGoS infrastructure. Both experimenter and participant start their HuGoS instance on their local client device and are connected via a server. Participants can be recruited using external services. Data are stored either in an online repository or on the experimenter’s local device



initial presentation of HuGoS (Coucke et al. 2020), we have integrated a 3D model of an e-puck robot (Mondada et al. 2009), such that each instance of the robot runs its control independently in 3D space with a 3D physics engine, roughly similar to the setup of the ARGoS multi-robot simulator (Pinciroli et al. 2012). Therefore, it should be feasible to replicate state-of-the-art swarm robotics studies conducted in a robot simulator such as ARGoS, for the purpose of one-to-one comparison with human behaviors. Furthermore, autonomous avatars following simple behavioral rules might be used to verify conclusions drawn from experiments with human players. If results from human experiments indicate that certain behavioral rules lead to certain group dynamics, the hypothesis can be further investigated by encoding those behavioral rules into artificial agents in the same setup as the human participants (i.e., the same information, capabilities, environment, and task), see Fig. 1. With this approach, studying human behavior in tasks relevant to robots could inspire new algorithms for the control of robot swarms.

Beyond experiment classes that relate to artificial swarms, HuGoS may also serve as a new tool to study topics in cognition, psychology, and social psychology. For instance, the study of human behavior in real-world scenarios, such as collective self-organization in emergencies (Drury 2018), could be supported by further study in a game environment. The graphical capabilities of Unity could be used to create valid simulations of real-world scenarios in HuGoS.

3.2 Features of HuGoS

HuGoS is built in Unity, a game development platform that supports networked games with multiple players, as well as autonomous agents (Juliani et al. 2018). In this section, we summarize some features and capabilities of HuGoS (for a more detailed description, see “Appendix 1”). Figure 2 shows a simplified architecture of the HuGoS platform. Both

the experimenter and participants run their own instance of the platform on their workstations and connect to a server. The experimenter can change setup options before and during an experiment (see Sect. 3.4). During each experiment trial, the experimenter sees an overview of the environment. Players (i.e., recruited participants) observe the environment through a first-person or third-person view of their avatar, which they control with their keyboard and mouse. Depending on the nature of the experiment, the environment can be populated with many *game objects* that are either static or controlled via *behavior scripts*. These scripts program game objects to change according to player behavior (e.g., *lava spills* in Sect. 4). An experimenter can also use the scripts to program avatars to act as autonomous agents, for instance, to perform the same tasks as human participants or to interact with human-controlled avatars.

Players can interact and communicate in several ways. They can communicate indirectly by each individual observing the motion of other avatars. They can also send explicit signals by changing avatar appearance or sending text messages in the HuGoS chat system. This direct and indirect communication is recorded during each experiment and can be analyzed to study communication network dynamics over time. Communication networks between players can also be explicitly manipulated by the experimenter, to study the effect of different configurations on collective behavior (see Fig. 17 in “Appendix 1”).

Measures of collective performance can be calculated in real time and provided as feedback to the participants in the form of a *group score*. The group score can be calculated based on various recorded data, including player actions (e.g., position, orientation, and interactions with objects), changes players make to game objects, and information available to players (e.g., game objects and other players in their field of view).

3.3 Network

The experimenter and all participants run their own instance of HuGoS and connect to a server. Computing generally happens at the local client, while synchronized environment variables and explicit messages between players are mediated by the server (see Fig. 15 in “Appendix 1”). We integrate HuGoS with Photon Unity Networking (PUN) by Exit Games,⁴ to enhance flexibility and usability for the experimenter. PUN automatically manages server hosting and facilitates easier definition of the variables that need to be synchronized by the server. With this server setup, experiments can be conducted in a laboratory setting or can be conducted online with participants in other locations. If using a laboratory setting, the setup can also be modified to run on a local server or local area network. There is no theoretical limit on the number of participants that the server can support. However, in practice, the number of participants will be limited by conditions external to HuGoS or specific to the experiment setup, such as the bandwidth used in a given game (see “Networking considerations” in “Appendix 2”).

3.4 The sequence of an experiment

Once connected to the server, the experimenter accesses a *control panel* where certain options can be modified, such as experiment conditions and the number of participants. Once the experimenter has connected and modified these options, participants connect

⁴ <https://www.photonengine.com/>.

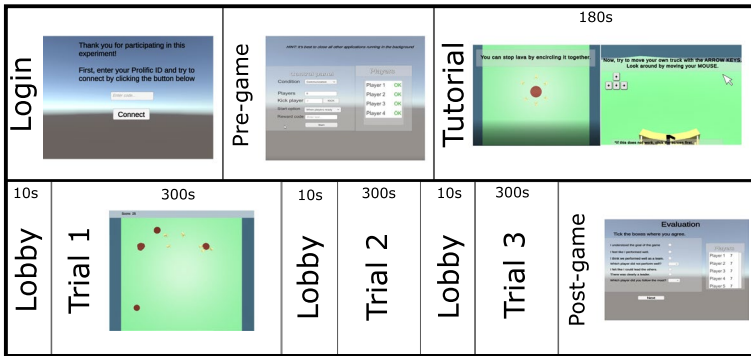


Fig. 3 Experiment sequence for case study 1

to the platform with their participant IDs (retrieved from, e.g., the recruitment platform) and enter a lobby, where they are given a pre-game questionnaire. The experimenter can start the game once the correct number of participants has joined and completed the questionnaire.

The game experience is organized into *scenes*. When the experimenter starts the game, all participants are transferred from the lobby scene to the first tutorial scene, where they receive a visual explanation of the task. In the second tutorial scene, they control an avatar while receiving instructions. After the tutorial, participants are redirected to the lobby scene to wait for the first trial. The experimenter chooses the number of trials to conduct. The length of each trial can be fixed or can depend on player behavior (e.g., the trial could end when participants complete a certain task). After each trial, participants are redirected to the lobby scene, where they can see their group score from the previous trials. When all trials are complete, participants enter the final scene, where they are given a post-game questionnaire about their experience. In this scene, participants can also be given information to receive remuneration for their time. See Fig. 3 for an illustration of the experiment sequence for case study 1.

4 Methodology of case studies to assess HuGoS

The features of HuGoS⁵ are designed to support comprehensive studies on human swarm intelligence. We conduct three case studies to demonstrate the capacity of HuGoS in the experimentation scope defined in Sect. 3.1. Each case study is a variant of a general setup, in which participants coordinate their actions to contain *lava spills* in a dynamic environment. The three case studies focus on collective decision-making, messaging and signaling, and stigmergy. Anonymous participants for these case studies are recruited via Prolific (“Appendix 2” for details). Each group of participants completes an experimental session that includes three separate trials of one case study, each lasting five minutes. The experimental sessions are run with group sizes ranging from four to nine participants.

⁵ HuGoS is available on GitHub at <https://github.com/NicolasCoucke/HuGoS-code-MAIN>.

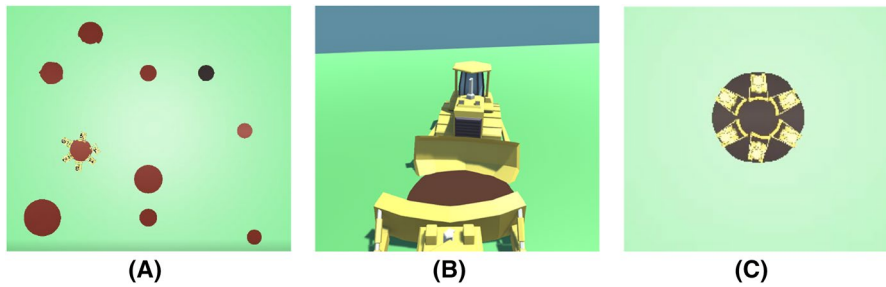


Fig. 4 Case study 1. **a** Experimenter’s view of the environment. **b** A participant’s first-person view while barricading a lava spill. **c** Experimenter’s view of a lava spill encircled by 6 players

4.1 HuGoS setup

In all case studies, each participant controls a *bulldozer* avatar in a shared environment. Using the mouse or touchpad, each participant can rotate their avatar without restriction, to change their first-person field of view and the avatar’s orientation. All avatar functions other than rotation are triggered by keystrokes. By using the arrow keys, each participant can move their avatar forward, backward, right, or left, relative to the avatar’s current orientation. In some case studies, participants can use the space bar to trigger a *crown*⁶ to appear above their avatar, seen by all participants. The crown can be used as a simple tool for boolean communication with the other participants. In some case studies, participants can alternatively communicate via chat messages. In other case studies, participants can use the space bar to pick up a block in their avatar’s vicinity, or to put down a block their avatar is already carrying.

In the shared environment, the participants use their avatars to work on the task of stopping *lava spills*, which appear spontaneously and then grow larger if they are not barricaded. The locations, sizes, and growth speeds of the lava spills are unknown to the participants prior to the experiment; they can only access this information by observing the environment. In order to stop a spill, participants must fully barricade it, using either their bulldozer avatars or blocks that they have picked up.

The task requires participants to coordinate as a group. Where bulldozers are used to barricade, the size of a spill requires six bulldozers to surround it simultaneously. Where blocks are used, all participants can place blocks at the same spill sites. In each session, there are between four and ten participants. All participants co-occupy one arena and can see the same information. The avatars each have a unique ID, displayed on the front of the bulldozer and visible to all participants and themselves.

The arena is 150 m × 150 m and is enclosed. Each bulldozer avatar has a footprint of 3.8 m × 5.2 m and has a speed of 6 m/s when in motion. The participant’s field of view is 75° horizontally and 60° vertically, centered above the avatar’s heading (see example field of view in Fig. 4b). In case studies where blocks are used, the blocks are 1.6 m cubes and can be picked up by an avatar if its heading is within 5 m of the block. At any given time, there can be up to 14 active lava spills in the arena. The spills are circular and start with

⁶ Bulldozer and crown models were retrieved from *Poly by Google* (at <https://poly.google.com/>) under a CC-BY 3.0 license.

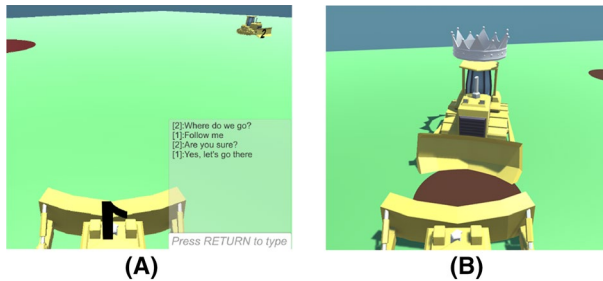


Fig. 5 Case study 2: Participants have additional communication capabilities via either text messaging or elementary signaling. **a** A participant’s first-person view of the chat window (lower right of window) that they can use for in-game text messaging. **b** A participant’s first-person view of another avatar, that is signaling a desire to lead by displaying a crown

a radius between 2 and 4 m. An invisible barrier prevents participants from entering the starting circle. While active, the lava spills increase in radius at a constant rate predefined for each spill, between 0.01 and 0.1 m/s. The lava spills are spread out across the arena, in arbitrary locations defined by the experimenter. The spill locations, times of appearance, and growth rates vary between the three trials of one session. The setup is identical in each of the sessions, which have unique groups of human participants.

In trials where spills are barricaded by bulldozers, participants can temporarily barricade the growth of the spill in a certain direction by touching it with the front edges of their bulldozers. The spill will continue to grow in directions that have not been barricaded. Participants can inactivate a spill by simultaneously barricading 90% of its starting circumference (not its current edge after growth). In this case, the spill permanently stops growing in all directions, and its color changes from red to black, indicating to all participants that it is now inactive. In order to achieve this, the size of the spill requires that multiple bulldozers touch it simultaneously. In trials where blocks are used, a spill can similarly be barricaded by avatars placing blocks around it. The spill always grows in directions where it has not yet been barricaded. To completely barricade it, the blocks need to form a closed loop around the spill. Once a block has barricaded part of a spill’s edge, that block is stationary and cannot be picked up again.

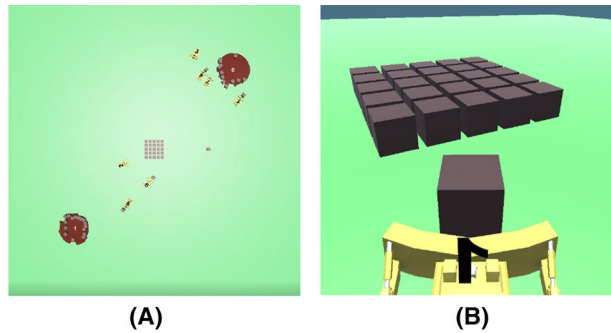
In this setup, we integrate HuGoS with several external infrastructures. Participants were recruited via Prolific, an online recruiting platform, and then were redirected to an external website where they accessed a WebGL instance of the experiment, via the Photon cloud server. Data from the experiments were saved to the experimenter’s device. A more extensive discussion, including other possible implementations, is found in “Appendix 2”.

4.2 Case study setups

4.2.1 Case study 1: basic collective decision-making

The first case study focuses on coordination and collective decision-making, with minimal communication capabilities. Participants barricade spills with their avatars. Participants have to achieve a consensus on the spill they will enclose, and have to coordinate their avatar movements to construct a circle around the spill. The size and spawn sequence of

Fig. 6 Case study 3: Participants can use blocks to contain a spill. **a** Experimenter’s view of a lava spill being partially contained with blocks. **b** A participant’s first-person view while transporting a block



the spills varies in the three trials. In the first and second trials, a larger spill size requires six bulldozers to barricade, while in the third trial, a smaller spill requires only three bulldozers. In the third trial, the best score can be achieved if participants split themselves into two groups. The variation in spills facilitates the study of speed and accuracy in collective decision-making. Additionally, because the best performance in the third trial requires two groups, this setup can be used to study changing group structure in a dynamic environment.

4.2.2 Case study 2: messaging and signaling

The second case study is identical to the first, except that participants have the added capability of explicit communication, either through text communication or elementary signaling. In text communication, the players can exchange messages using an in-game chat. Participants can interrupt their avatar control and start typing a message by pressing the return key; they can press the return key again to send the message and resume control of their avatar. Once sent, the message is visible in the message box for all the other participants (Fig. 5a). In elementary signaling, participants can choose to display a crown above their avatar (Fig. 5b), and are told that this can indicate their desire to lead others. They can activate and deactivate the crown at any time by pressing the space bar. After deactivating the crown, a participant has to wait 4 s before reactivating it, in order to prevent ‘flashing’ of the crown. The two types of communication can be used to study the effect of different communication abilities on group dynamics and performance.

4.2.3 Case study 3: stigmergy

In the third case study, participants use a form of stigmergic communication, and can barricade spills by placing blocks. The center of the environment is filled with an unlimited pile of blocks (Fig. 6a). Participants can pick these up and release them again using the space bar (Fig. 6b). This setup can be used to study indirect coordination via observation of the modifications made by others, rather than observation of avatar motion.

4.3 Assessment metrics

With HuGoS, we aim to develop a broad tool that is flexible and easy to use, and that captures data about human behavior with sufficient detail to support research on individual actions as well as group dynamics. To assess HuGoS, we provide the case study results and

demonstrate example analyses of individual and group behavior. We also evaluate the overall performance and usability of HuGoS.

4.3.1 Individual behavior

Each participant has a limited repertoire of actions they can use to interact with the environment and other participants. All these actions are captured and available for analysis (e.g., avatar position and rotation, time of signaling, or time and content of text messages). The unique IDs, positions, and properties of game objects (i.e., bulldozers, spills, and blocks) are also captured, along with participants' interactions with those objects (e.g., the IDs and properties of objects in the FoV). A more extensive description of data and analysis types can be found in sections "Data types" and "Analysis types" in "Appendix 1".

4.3.2 Group behavior

Group behavior is primarily assessed according to a performance score that tracks participant success at the given task. Group behavior is also assessed using network analysis, according to the degree of centralization occurring in the communication graph.

4.3.2.1 Performance score In each trial, all participants are scored as a single group. The score represents the group's performance at the task of stopping lava spills. The score is tracked continuously and is displayed to each participant throughout the game. The score G is defined as the total surface area that the group has prevented from being covered by lava, for all spills that have appeared. At time t , the score G_t is calculated as:

$$G_t = \sum_{i=1}^n A_{i_t}^{\max} - A_{i_t}^{\text{actual}}, \quad (1)$$

where $A_{i_t}^{\max}$ is the surface area that would have been covered by spill i at time t if no barricades had been placed, and $A_{i_t}^{\text{actual}}$ is the actual surface area covered by spill i at time t .

For example, if participants were to not interact with any spills, then $A_{i_t}^{\max}$ and $A_{i_t}^{\text{actual}}$ would be of equal value, and the score would be $G_t = 0$. If participants barricade part of the spill, G_t will increase, because $A_{i_t}^{\max}$ is increasing faster than $A_{i_t}^{\text{actual}}$ (see Fig. 11). New spills appear constantly and grow at different speeds, requiring participants to constantly evaluate the best spill to target. Participants receive rough information about score calculation. Specifically, they are informed that better scores can be achieved if they target larger spills, spills that are growing more quickly, spills that are closer. The final performance score of a given trial reflects not only the speed and effectiveness of physical maneuvering and coordination, but also the speed and accuracy of the collective evaluation of spills.

4.3.2.2 Network analysis Communication networks can represent explicit communication such as text messages, or implicit communication such as presence in the FoV. Network centralization reflects dynamics of self-organization in the group, such as the emergence of implicit ad hoc leadership. We assess player communication as directed networks and using indegree network centralization. Indegree network centralization C of a network with n nodes is calculated based on Freeman (1978), as:

$$C = \frac{\sum_{i=1}^n (n-1) - \text{deg}(p_i)}{(n-1)^2}, \quad (2)$$

where $\text{deg}(p_i)$ is the indegree (i.e., number of incoming communication connections) of player i .

4.3.3 Overall performance and usability of HuGoS

For overall performance and usability, we assess connection issues in the case studies and compare them to a controlled latency study with known clients and connection speeds. We also assess HuGoS using participant responses to questionnaires. When participants join the game, they are given a questionnaire on personality traits related to leadership. After the last trial, they receive a questionnaire on their game experience. Finally, assess the advantages and limitations of HuGoS from the point of view of a potential experimenter, in terms of performance, usability, and flexibility.

5 Results and assessment

Our dataset includes 117 participants (42 females) with a mean age of 25.3 (SD = 7.8). We first summarize the results of each case study (all data are available in the supplementary materials⁷), analyze individual and group behavior, and compare the results of the three case studies. Second, we evaluate HuGoS in terms of participant connection and latency, and participant questionnaire responses. Finally, we give an assessment of the overall performance, usability, and flexibility of HuGoS as a tool for experimenters.

5.1 Case study 1: basic collective decision-making

In case study 1, participants need to coordinate their actions to collaboratively barricade spills. This task can be completed more effectively if participants improve their speed and effectiveness in reaching consensus about the next spill to barricade. This could be challenging, as participants cannot communicate directly, and must rely on observing the actions of others. We analyze players' coordination of their positions, as well as the consensuses and group dynamics that support this coordination. We present the results of one example trial, with 8 participants.

Figure 7a gives the Euclidean distance between each player and the next barricaded spill, over time, during an entire example trial. Each red circle represents an instance in which a spill is successfully barricaded and deactivated. Figure 7a shows that the players' positions repeatedly converge toward the next spill. It takes players a much longer time to reach a consensus and barricade the first spill than later spills. After the second spill, players seem to have learned how to take cues from each other and reach consensus, as the time to converge on a new spill becomes much shorter. Figure 7b–c shows the results of one example trial in terms of a directed connectivity graph, where nodes represent players, and an edge from player a to player b corresponds to player b being present in player a 's field

⁷ The supplementary materials are available in a Zenodo data repository: <http://doi.org/10.5281/zenodo.4384805>.

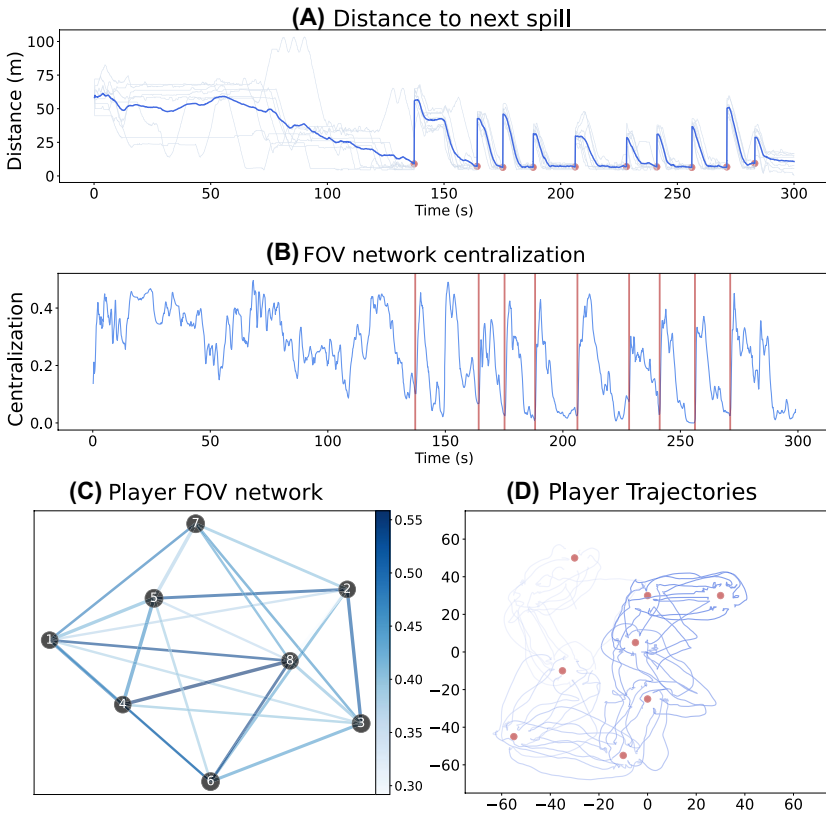


Fig. 7 A trial of case study 1 with eight participants. **a** Euclidean distance over time, from each player to the spill that will be barricaded next (red circles indicate barricade completions). The plot gives both the individual players (light blue) and the average of all players in the trial (dark blue). **b** The indegree network centralization of the player field of view (FoV) network over time, during one trial. This measure is used as an indication of player network clustering. Vertical red lines indicate barricade completions. **c** The FoV network between players. The darkness of connections represents the percentage of time those two players were in each other's FoV. **d** xy positions of all players in one trial, used to show player motion trajectories over time (darker color lines indicate later times). Spill locations are indicated by red circles (Color figure online)

of view. In Fig. 7b, indegree network centralization C during one trial is plotted over time. In Fig. 7c, the connectivity graph is given, with the weight of connections (indicated by the darkness of the lines) corresponding to the percentage of time two players were mutually included in each other's fields of view. The results in Fig. 7b–c are relevant to the analysis of collective decision-making, because, if one player is more often in others' fields of view, this player could be expected to have more influence on group behavior. Network centralization can therefore be taken as an implicit measure of the presence of leadership in group dynamics. Figure 7d shows the xy positions of players during one example trial (red circles indicate spill locations). To improve plot legibility, a moving average filter with a window of 0.6s is applied to the xy positions of each player. Players' motion trajectories over time are indicated by line color (darker lines are later in time). This plot shows the motion coordination achieved between players, as they generally move toward the same next spill, but also stay fairly distant from one another, avoiding collisions and interference.

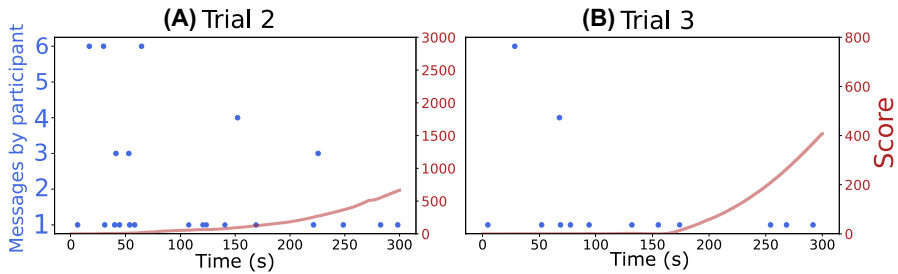


Fig. 8 Text messaging communication in case study 2 (cf. session 5 in Fig. 11), during two example trials. Messages (each blue circle is one message) sent by each participant (indicated by participant ID on the y axis), compared to the group performance score G over time (red line). Participant 1 continues to send messages throughout both trials, and seems to have taken a leadership role (Color figure online)

Overall, the results of case study 1 show that players collaborate fairly well using only observation of each other's movements and the environment. At each instance that a spill is deactivated, and consensus must be reached about which spill to barricade next, there are up to 14 active spills present in the environment. Throughout the trial, players consistently reach a consensus about the next spill and then coordinate their positions to complete a barricade.

5.2 Case study 2: messaging and signaling

Case study 2 follows the same setup as case study 1, except that participants are now able to communicate explicitly. The additional communication between players could facilitate coordination and improve the effectiveness of reaching a consensus. We present the results of two example experiment sessions: one in which participants could send text messages, and another in which participants could send a boolean signal via crown. We examine these two types of communication and their relationships to group behavior and performance.

In sessions with text communication, all messages are broadcast to all other participants. We present the results of the second and third trials of a single session; the group of human participants is the same in both trials and has already interacted as a group during one complete trial. Figure 8 shows each message that each participant (indicated by participant IDs 1–6) broadcasts over time in both trials (blue circles), compared to the group performance score G (in red). In the second trial, shown in Fig. 8a, participants seem to have initially struggled with barricading a spill. Several participants broadcast messages in the beginning of the trial, while the score does not increase. Once the score begins increasing, most participants stop sending messages, and only participant 1 continues sending messages throughout the whole trial. One interpretation of these results is that participants first went through a deliberation phase, after which they agreed on a course of action that is led by participant 1. Figure 8b shows the same results for the third trial (performed immediately after the second). Again, participants initially seem to struggle to increase their score, although there are far fewer participants sending messages. Presumably, this happens because the environment has changed to contain smaller spills, which required participants to change their strategy, which was perhaps negotiated by the previously chosen leader. Toward the second half of the trial, participants succeed in coordinating, resulting in a score increase. Participant 1 seems to have maintained a leadership role and continues

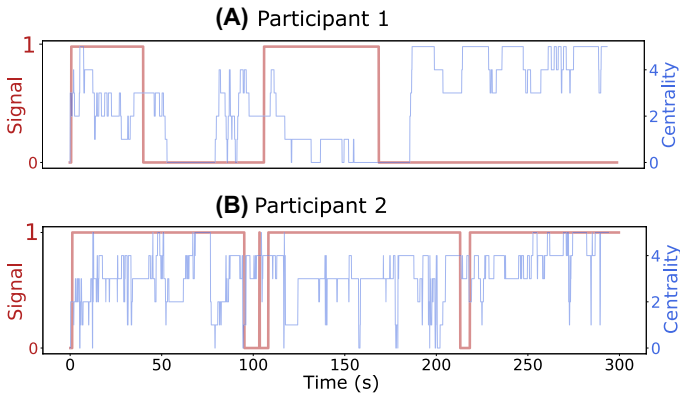


Fig. 9 Signaling in case study 2, for two participants in one example trial. Each subplot represents one participant. A participant's crown is displayed when the signal is 1, and is hidden when the signal is 0 (see red line). The blue line represents the participant's indegree (Color figure online)

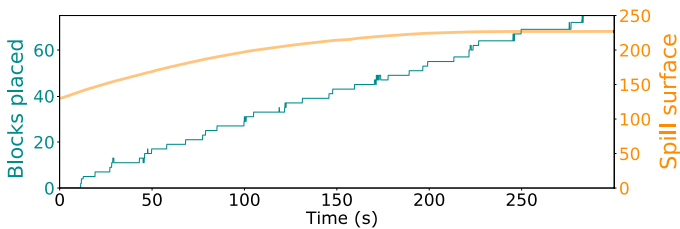


Fig. 10 An example trial of case study 3, with eight participants. The cumulative number of blocks placed at a lava spill by all participants (green), compared to the spill surface area in m^2 (yellow). Participants place blocks at a roughly consistent rate and have stopped spill growth 250 s into the trial (Color figure online)

sending messages throughout the trial. Participants' full chat logs are available in the supplementary materials.⁸

In sessions with boolean signaling (crown on, or crown off), each participant can view signals from other participants currently in their FoV. Figure 9 gives the signaling behavior of two participants during an example trial (signal activation in red), compared to that participant's indegree—i.e., number of other participants visible in the FoV—in the communication network (in blue). The first participant, shown in Fig. 9a, activates the crown signal for two periods during the trial. This participant has a large indegree during a long period following the end of the second activation. The second participant, shown in Fig. 9b, activates the crown signal during almost the whole trial, with two brief intermissions. This participant has fairly high indegree throughout the trial, but with more variation than participant 1. A causal analysis of whether signaling systematically influences a player's centrality in the communication network is beyond the scope of this paper. Participants' full

⁸ The supplementary materials are available in a Zenodo data repository: <http://doi.org/10.5281/zenodo.4384805>.

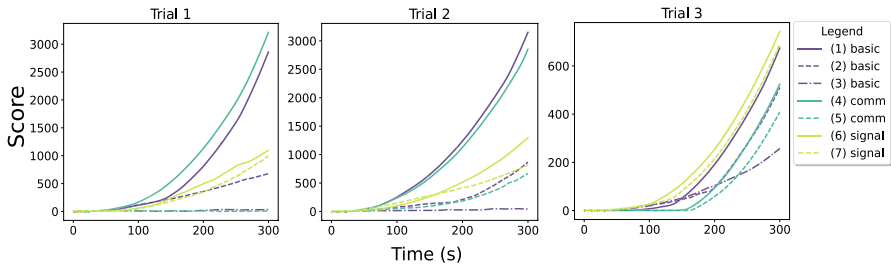


Fig. 11 Score progression in seven sessions of case studies 1 and 2, each consisting of three 5-min trials. The legend indicates a session ID number, and the experimental condition in that session: *basic* indicates case study 1; *comm* indicates case study 2 with text messaging; *signal* indicates case study 2 with crown signaling

signaling logs and the data associated with the communication network are available in the supplementary materials.⁹

Together, these results indicate that participants make use of the provided opportunity to explicitly communicate, during completion of the task. Further research should be conducted to elucidate the relationships between communication and collective behavior.

5.3 Case study 3: stigmergy

In case study 3, participants use blocks to barricade spills. Each participant selects a spill to target when transporting blocks, and can only coordinate with others by observing their motions and the blocks they have placed. By influencing the behavior of others through environment modification, the participants can engage in stigmergic coordination. We describe the results of one example trial with eight participants and two lava spills. Figure 10 shows the cumulative number of blocks placed at one spill (in green), compared to the surface area of that spill (in yellow). The spill surface area initially increases, because its growth is not restricted by any barricades. The surface area stagnates when a sufficient number of blocks are placed around the spill. The stagnation indicates that the participants coordinated their block placements well enough to form a complete barricade loop around the spill. The increase in placed blocks occurs at an approximately consistent rate over time. Near the end of the trial, when the spill surface remains approximately constant, the intervals between block placements become only slightly longer, perhaps indicating that participants are taking more time to find a good position for the next block placement.

Collectively, participants were able to completely barricade a spill by coordinating the placement of blocks. A more detailed analysis of the coordination mechanisms could be conducted, for instance, by combining data represented in Fig. 10 with positional and FoV data similar to Fig. 7.

⁹ The supplementary materials are available in a Zenodo data repository: <http://doi.org/10.5281/zenodo.4384805>.

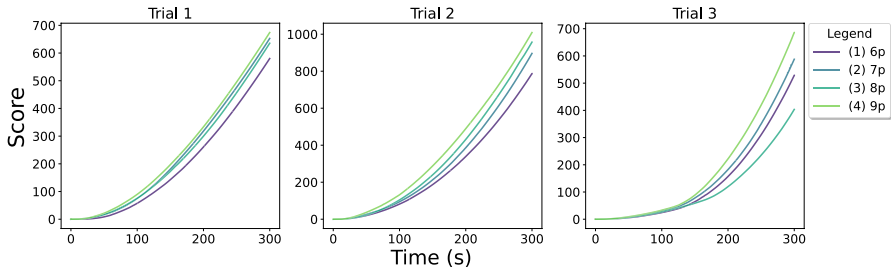


Fig. 12 Score progression across four sessions of case study 3. The legend indicates a session ID number, and the number of participants in that session

5.4 Group performance in all case studies

In total, across the three case studies, 131 participants recruited via Prolific took part in 20 sessions, each consisting of three trials. The group performance score G was recorded in all trials. Scores from case studies 1 and 2 can be compared directly, as these sessions were identical apart from the communication capabilities. All experiment data are available in the supplementary materials.¹⁰

Figure 11 compares the score progressions for seven sessions from case studies 1 and 2. The score varies greatly between different sessions. For example, participants in session 4 seem to quickly succeed in coordinating. Once they barricade a few spills, their score starts to increase rapidly. Conversely, participants in session 3 did not manage to adequately coordinate, leaving them without a superlinear increase in score. Participants from this session were able to barricade some spills in the third trial, presumably because the smaller spills did not require the full group to succeed in coordination. Participants in session 6 achieved the highest score in the third trial, which indicates that they were the most successful at splitting their group into two smaller sub-groups. In summary, the large variation in scores can reflect a difference in group cohesion, coordination, or strategy between sessions. Further analysis into the underlying variables causing these differences could be conducted with the data described in Sects. 5.1, 5.2 and 5.3.

Four sessions of case study 3 were conducted, with 6, 7, 8, and 9 participants. Figure 12 shows the score progression of all five sessions across the three trials. When considering session 2 in Fig. 12, for example, the end score of a session seems proportional to the number of players in that session. The score differences in Fig. 12 are less pronounced than in Fig. 11, presumably because performance did not depend on a group's ability to achieve a consensus and synchronously encircle a spill, but rather resulted from participants' asynchronous coordination of block placement through stigmergy. These results indicate that, for experiments with different aims, the score can reflect different aspects of collective behavior.

¹⁰ The supplementary materials are available in a Zenodo data repository: <http://doi.org/10.5281/zenodo.4384805>.

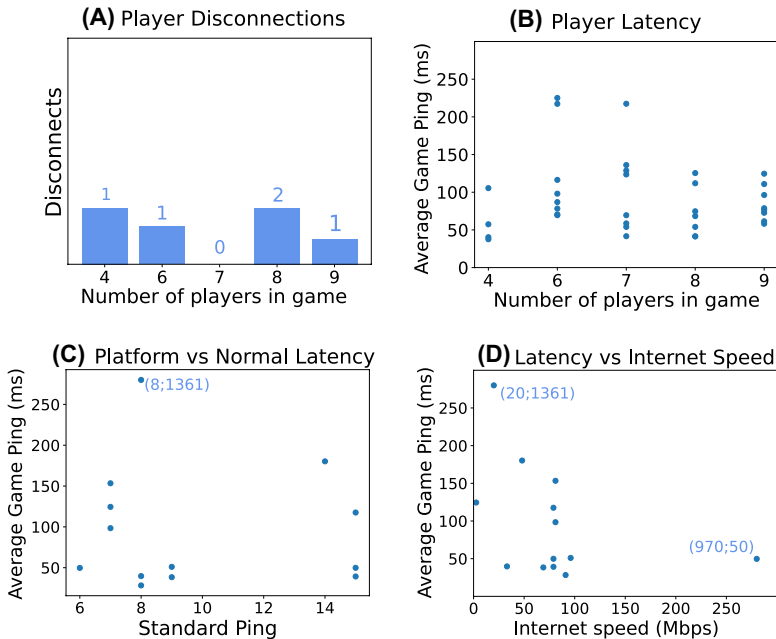


Fig. 13 Assessment of connection issues. **a** The number of player disconnections experienced by anonymous participants recruited from Prolific, in each of the five sessions of case study 3. **b** The latency (ping), defined as the time taken for a message to travel from the player to the server and back, for anonymous participants recruited from Prolific. Each value represents a player's average ping during the first trial. The values are plotted for five sessions of case study 3, with different numbers of players. **c** In-game ping in a latency test with known participants, compared with the player's internet speed. **d** In-game ping in a latency test with known participants, compared to a player's ping on a standard internet speed test. For (c) and (b), outliers are plotted at the edge of the plot with their value

5.5 Assessment of connection and latency

A performance requirement for HuGoS is that participants stay connected throughout an experiment, with an acceptable latency between the participants' game instances. In principle, disconnections could occur due to a player having a poor internet connection, issues on the server side, bugs in the back-end code, an overflow of traffic caused by the game, or the player voluntarily deciding to leave. When more players are connected, it presumably becomes more likely one of them will disconnect voluntarily or due to an internet connection problem. When more players join, the traffic through the network also increases. Figure 13a shows the number of disconnections against the number of players in five sessions of case study 3. As participants recruited over Prolific are not always available to be contacted when connection issues occur, it is not always possible to determine why a player disconnected. In most cases, players were able to reconnect to the game server after a disconnection. In addition to the case studies described above, we ran a series of quality-assurance connection tests with known participants (the connection tests replicated two sessions of case study 3, with a total of 13 participants). In those connection tests, none of the participants experienced a disconnection. Therefore, we infer that most disconnections experienced by anonymous players recruited through Prolific were attributable to a poor internet connection or a voluntary (perhaps inadvertent) disconnection.

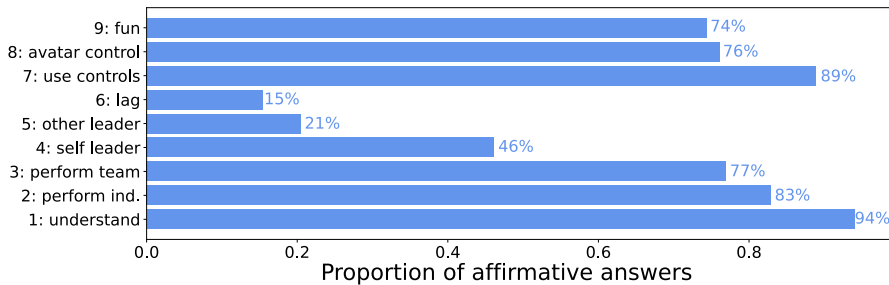


Fig. 14 Results of post-game questionnaires. (1) I understood the goal of the game. (2) I feel like I performed well. (3) I think we performed well as a team. (4) I felt like I could lead the others. (5) There was clearly a leader. (6) Did you experience lag/slow/bad connection? (7) I used my keyboard and mouse to control my avatar. (8) I could move my avatar where I wanted. (9) The game was fun

Another important factor to ensure smooth gameplay is the latency between clients and servers. In principle, the latency might be influenced by the number of players connected, the geographical location of players, and the players' internet speed. To assess this, we measured latency in terms of *ping*—the time in milliseconds it takes to send a message from the client back and forth to the server—in five sessions of case study 3. Figure 13b shows the latency experienced, according to the numbers of players in the session. Most players experienced average pings of less than 150 ms. We also ran a series of latency tests with known participants (replicating two sessions of case study 3, with a total of 13 participants). For these 13 known participants, we compared latency between client and game server to the client's standard ping and internet speed in Mbps, measured with a standard internet connection test.¹¹ The in-game latency with respect to unloaded latency and speed is shown in Fig. 13c, d. No clear impact from either the number of players or the players' internet speed is apparent from the latency results.

Average in-game latency for most players was in the range 0–200 ms. Most online games can be adequately played with delays up to 500 ms (Claypool and Finkel 2014).

5.6 Participant questionnaire responses

At the end of each trial, participants were asked to fill in a questionnaire about their experience during the experiment. In total, we received completed questionnaires from 117 players. The answers to the questions are summarized below and in Fig. 14. Participants' full responses are available in the supplementary materials.¹²

In one set of yes/no questions, participants were asked about the performance and strategies of themselves and others within the task. 84% indicated that they thought they performed well as an individual, while 77% indicated that they performed well as a group. Another set of questions asked about leadership within the task. 21% of participants indicated that there was a clear leader in the group, and 44% indicated that they could lead others in some capacity. In further research, these answers could be compared to actual

¹¹ Internet connection tests were conducted using <https://fast.com/>.

¹² The supplementary materials are available in a Zenodo data repository: <http://doi.org/10.5281/zenodo.4384805>.

performance in the trials (Figs. 11, 12), to assess participants' ability to self-assess. They could also be compared to the player networks (see Fig. 7c), to assess the relationship between players' judgments of leadership in the group and actual observed behaviors. In an open question, participants were also asked to report their strategy for achieving the best performance (available in the supplementary materials).

Participants were also asked questions about their experience with gameplay during the experiment. 16% of participants indicated that they experienced problems with connection or lag at some point during the experiment. 89% indicated that they could correctly use their mouse and keyboard to control the avatar and 76% indicated they felt they could always move their avatar where they wanted. 94% reported understanding the goal of the task, while 74% reported having enjoyed the experience. The issues with avatar control that 24% of players seemed to have, could be due to connection speed (as part of the 16%) or potentially due to a deficiency in the tutorial at the beginning of the game. The open-source version of HuGoS will be continually updated to address such issues as they become apparent. Also note that the default response to the questions was negative and that participants had to change the answer in order to give an affirmative.

5.7 Assessment of performance, usability, and flexibility

We chose to build HuGoS in Unity because Unity provides a reliable, accessible,¹³ well-documented, and well-supported platform for both experimenters and participants. Unity supports all common operating systems (including Windows, macOS, Linux, iOS, and Android) and supports WebGL¹⁴ (Web Graphics Library) in all common browsers (Chrome, Firefox, Safari). Participants in an experiment can join via a web browser, without having to download or install any specialized software. Unity's support and documentation ensure that when participants run HuGoS in a browser via WebGL and cannot be directly monitored by an experimenter, user keystrokes will reliably be recorded and sent to the game server. The system requirements for both experimenters (i.e., in Unity Editor) and participants (i.e., in Unity Players) are minimal,¹⁵ although the exact performance, speed, and rendering quality will of course depend on the user's system. Importantly, Unity also provides experimenters with an intuitive user interface and open-source repository of code examples, increasing the accessibility of HuGoS to experimenters with various levels of programming experience. Unity even provides a visual scripting interface, via its Bolt¹⁶ product. Given the interdisciplinarity of this topic, we regard this as a crucial usability feature—HuGoS needs to be as accessible as possible to researchers in many fields (e.g., psychology or anthropology), regardless of their programming background.

There is no technical limit on the size of the environment in Unity; it can be set as large as required for the experiment. If desired, the environment and task can be programmed to automatically adjust to changing game specifications. However, there is of course a practical limit on the size of the environment that can be populated by game objects, in terms of time and cost overhead involved for the experimenter. To push back against this limitation, however, experimenters could potentially populate infinitely large environments with game

¹³ Unity is open-access when used in an academic capacity.

¹⁴ https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API.

¹⁵ <https://docs.unity3d.com/Manual/system-requirements.html>.

¹⁶ <https://unity.com/products/unity-visual-scripting>.

objects, with the help of procedural content generation (cf. Shaker et al. 2016; Liu et al. 2020).

Beyond the functions supported by Unity, HuGoS provides an out-of-the-box solution for multi-player experiments in scenarios relevant for swarm robotics, including the user interface for the experimenter. The HuGoS source code and tutorial is available on Github.¹⁷ Experimenters can download and use the basic HuGoS setup with the case studies described in this paper. The only actions required to run basic HuGoS are: (1) update the local file path for data storage, (2) make a free account on the Photon game server, and (3) define a new application in Photon and link it to the local instance of the HuGoS Unity project. Some key changes can be made with minimal adjustments, including the number of trials, length of trials, and content of questionnaires. For more extensive changes, custom models and packages can be easily added, by referring to the documentation of HuGoS and Unity. Visual recordings of the bird's eye experimenter's view can be made using the on-screen capturing tool Open Broadcaster Software[®],¹⁸ or equivalent.

6 Discussion

We have introduced HuGoS, a novel multi-user virtual environment built in Unity, designed for conducting experiments with human participants interacting as avatars. We specifically designed HuGoS to facilitate a wide scope of experiments that we consider of interest to the domain of human swarm intelligence. In three case studies, we have shown how the features embedded in HuGoS enable experiments across this scope. In all case studies, participants completed a task that required both physical coordination and observation of the environment. In the first case study, we showed that participants could complete rudimentary collective decision-making in this setup, reaching consensus to complete a task that required cooperation. In the second case study, we showed that participants could use make use of two additional channels of communication during the same collective decision-making. In the third case study, we showed that participants could complete a task that required asynchronous coordination through modification of the environment. The base ingredients of these case studies can be used to design many of the scenarios within the scope of human swarm intelligence.

Section 5 demonstrates the capabilities of HuGoS in terms of data analysis. The data captured from the case studies can comprehensively represent the collective behavior and performance of players. Additionally, detailed data on each player's actions and observations can be captured and linked to dynamics at the group level. We also implemented questionnaires before and after the experiments. In future studies, participants' responses to the questionnaires can then be linked to the behavioral data observed and captured during the experiments.

By using a virtual environment, we can implement scenarios that would be either impossible or too costly to implement with human participants in real setups. Also, the dynamical interactions between participants' avatars are a better simulation of embodied social interactions than existing online multi-participant studies that focus on discrete decisions. Yet, our virtual environment, like most other existing online studies, leaves out many

¹⁷ <https://github.com/NicolasCoucke/HuGoS-code-MAIN>.

¹⁸ <https://obsproject.com/>.

types of interactions that would be present in real settings, such as eye contact, speech characteristics, and body language. Their absence might decrease the ecological validity of findings on human behavior (Hermans et al. 2019). Accordingly, we do not propose that HuGoS is a replacement for in-person studies. Rather, studies in HuGoS might be complementary to in-person studies. Our virtual environment enables us to isolate specific interaction types and study their impact on collective behavior, while removing factors that are challenging to quantitatively capture (also in in-person studies), such as body language and speech characteristics.

We have illustrated the potential of HuGoS for controlled experiments with human participants by performing online experiments that participants could run as an online browser game from their personal device. Online experiments necessarily require us to give up some degree of control over participants' equipment, internet connection, and voluntary behavior. Participants might have variable screen resolutions, computer mice, and internet connections. These factors can be better controlled when performing experiments in one shared computer room, where the experimenter has better control and overview of conditions experienced by participants, and where every player has the same workstation and connection speed (cf. Zhao et al. 2018). The choice between online and on-site experiments then becomes a trade-off between logistical ease of recruiting large numbers of participants online, and the control and uniformity of on-site experiments.

In terms of technical specifications, we will aim to improve the latency of HuGoS in future work. Our current delays do not normally exceed 200 ms, and in real-time strategy games such as World of Warcraft, delays can be larger than 500 ms without affecting player performance, as performance depends on decisions made on longer timescales (Claypool 2005). However, in fast-paced games such as first-person shooters, player performance might decrease for latencies as low as 100 ms (Claypool and Finkel 2014). Many participants in our pilot studies had latencies larger than that. However, given that the tasks in our experiments do not require such fast coordination, we do not expect this latency to affect player performance. Yet, to achieve dynamic real-time interactions between players, we would ideally like to see latency around 100 ms. Future versions of HuGoS will aim to mitigate delay by further limiting traffic that passes through the server. Latency could also be diminished by organizing experiments on a local area network. In the current version of HuGoS, the effects of latency are partly diminished by interpolating, e.g., avatar positions between network updates in order to achieve the apparently smooth interactions between players.

Beyond the workstation and connection conditions of participants, their engagement in the study and other voluntary behavior can also be a factor. To reduce the need for monitoring participant engagement during the online experiments, we incentivized participants with a possible bonus payment for better performance. Additionally, we designed the experiment to be intrinsically motivating by making sure that the tasks have a clear goal that is neither too difficult nor too easy to achieve, giving participants sufficient control over the outcome of the task, and giving them regular feedback about their actions (Nakamura and Csikszentmihalyi 2014; Jung et al. 2010). Given that the experiments are designed to be intrinsically motivating, future experiments might also be conducted as "citizen science" (cf. Cooper et al. 2010; Sørensen et al. 2016), where participants take part in scientific studies voluntarily, to advance science by contributing to novel solutions or theories. Performance trackers such as leaderboards might be an additional motivator (Wang and Sun 2012). However, a game's intrinsic impetus has often been shown to contribute more to motivation and performance than external rewards such as bonus payments and leaderboards (Nakamura and Csikszentmihalyi 2014; Jung et al. 2010).

In addition to motivation, many other factors might influence participants' behavior and performance. When a group of participants starts a game as naïve players, a learning curve is always present. We expect participants to roughly converge to similar strategies once they learn how to play the game. Therefore, while the case studies reported here last for only 20 min—consisting of a tutorial and three trials of five minutes each—future studies may benefit from longer experiment times. However, with longer game times, new issues may arise such as user fatigue, developing stronger social relationships with other players, or a higher chance of disconnections. The influence of understanding the game, or developing strategies, can be investigated by observing the difference between sessions where participants go through a tutorial, and sessions where they must figure out the task without instruction—similar to what a group of AI agents would have to do if using reinforcement learning. Interesting manipulations might also be done by introducing naïve participants during an already started experiment, adding an experienced player to a group of naïve players, or exchanging participants between game sessions that have converged on different strategies. With minor adjustments, HuGoS supports these possibilities by keeping participants in the room for some time without letting them participate in the game and by running multiple rooms with the same game simultaneously, between which participants can be exchanged.

Many social psychological factors can influence how participants interact with each other. For example, experiencing a shared identity with other participants might enhance performance. A shared identity might be established by asking participants to imagine that they had already experienced a certain event together or by having similar avatar appearances (Titlestad et al. 2019). Interesting manipulations could be done where avatar appearance is varied systematically. Creating two different groups with different avatar characteristics that compete on a certain task could also yield interesting results.

When conducting experiments in a virtual environment, participant behavior is heavily impacted by whether they believe that other avatars are controlled by humans (Blasovich et al. 2002). When performing experiments solely with human-controlled avatars, participants should be informed that other avatars are also human players. In cases where autonomous agents are used, an interesting manipulation could be to let some avatars be controlled by autonomous agents, while participants are told that they are human controlled (cf. Shirado and Christakis 2017).

In short, we hope that the presentation of the virtual environment, together with the presented case studies, illustrates how HuGoS can be used to conduct a wide range of experiments and analysis, in a way that is useful to the research community. Although the virtual environment provides some level of control over participants' range of behaviors, many factors such as personality, culture, shared identity, understanding, and motivation still have to be taken into account. We do not consider these factors merely as artifacts to neutralize; they are important modulators of human cognition that could be instrumental in understanding successful human strategies in swarm intelligence tasks.

7 Conclusion

We have designed and presented HuGoS, a multi-user virtual environment that supports the study of human interactions and group behaviors relevant to the topic of swarm intelligence. HuGoS is a versatile tool that allows implementation of a large number of possible scenarios. We have shown the functionality of HuGoS with anonymous human participants

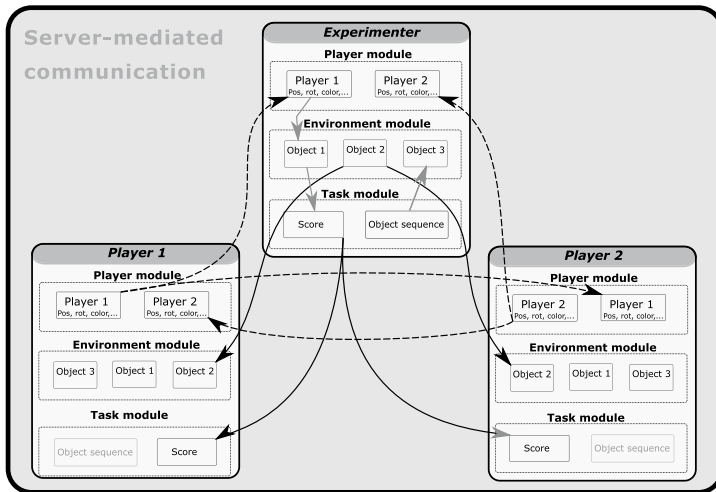


Fig. 15 Client and server communication: All clients run their own instances of the game. The experimenter is the *master client* and determines the course of the experiment. Black arrows indicate information flow mediated by the server. Full black arrows indicate predominant information flow from the experimenter to the players. Dashed black arrows indicate information originating mainly from players. Gray arrows indicate local information flow that is not (yet) mediated by the network

in a coordination task, under conditions of (1) dynamic best-of- n collective decision-making, (2) additional messaging or signaling, and (3) stigmergic interactions. The software is open-source and can be easily adapted to other experiment types. With this contribution, we hope to encourage further research into human swarm intelligence, including unique aspects of human psychology that are not usually studied under swarm intelligence.

Appendix 1: Detailed features of HuGoS

This appendix provides a more detailed description of the HuGoS architecture described in Sect. 3. The sections below describe the general architecture, before moving to the avatar capabilities, the data that can be captured, and eventually the types of analysis that can be conducted with the captured data. Sections “Avatar capabilities”–“Analysis types” in “Appendix 1” are largely reproduced from our previous work (Coucke et al. 2020).

Multi-player Unity implementation

HuGoS is built in Unity, a 3D game development platform that can support intelligent agents in a physically realistic game environment (Juliani et al. 2018). In Unity, basic building blocks of virtual environments are termed *game objects*. Each game object represents a physical 3D object within the game environment that is subject to physics engines (when desired) and can additionally be equipped with specific *behaviors*, defined through specific C# back-end scripts. These back-end scripts can be used to define fully autonomous artificial behaviors for the game objects and to define player controls (e.g., keystrokes) and their impact on game objects. Depending on the behaviors defined via these scripts, we

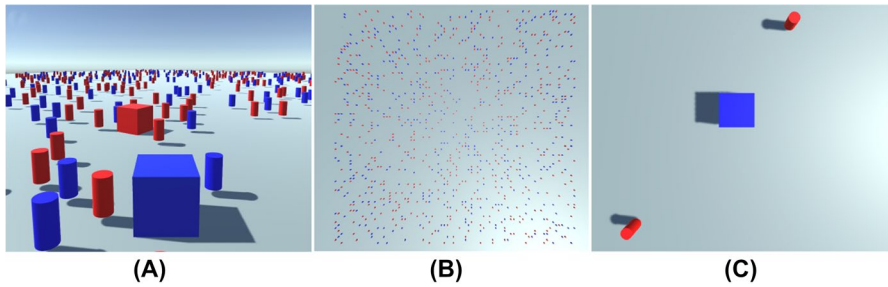


Fig. 16 Simple environment setup of a collective decision-making scenario. Each participant has a third-person view of a block avatar that they control. Participants estimate the percentage of blue/red landmarks (cylinders) in the environment and indicate their opinion by changing the color of their avatar (for a detailed explanation, see Coucke et al. 2020). **a** Each player has an oblique view of their avatar and surroundings. **b** Top view of the whole environment. **c** Limited top-down view of an avatar. This figure is reproduced from our previous work (Coucke et al. 2020)

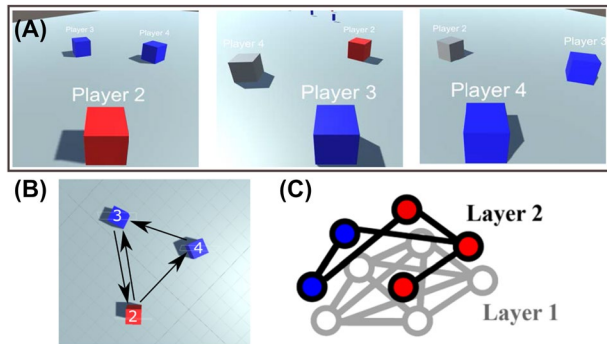
define the categories of game objects in HuGoS to be: (i) passive immobile (e.g., obstacle), (ii) passive mobile (e.g., building block), (iii) controlled by simple rule-based behaviors but immobile, (iv) mobile and equipped with a controller to act as an artificial agent, or (v) mobile and controlled by a human player. We refer to game objects in HuGoS as *avatars* if they act as artificial agents or are controlled by human players. Using Unity’s networking capabilities, we organize the multi-user architecture of HuGoS as follows (see Fig. 15).

Each game object can have multiple behaviors attached to it. For example, an avatar might have one behavior to move according to player controls and another behavior to change color according to an environmental stimulus. Game objects can possess many behaviors—they have a negligible impact on the game file size during the initial download. However, the number of simultaneously updated or activated behaviors is limited by the maximum traffic on the server (see section “Networking considerations” in “Appendix 2”).

Game objects influence each other both according to their interactions in the 3D physics engine (e.g., collisions) and according to their programmed behaviors that govern rules of interaction. For example, in order for a player to pick up and move a building block, both the player and the block object must be equipped with behaviors that allow for this interaction. Physics engine calculations and behavioral interactions are executed locally on the client who initiates the interaction. The results from that interaction (e.g., player *a* carries block *x*) are then synced to all players over the network (see Fig. 15). After the physics engine calculations and behavioral interactions are complete, the new positions of the game objects are synced over the network. To ensure smooth interactions between players during an experiment, the latency between the server and the clients should be kept as small as possible (see Sect. 5.5).

The visual appearance of game objects—both avatars and passive objects in the environment—can be easily manipulated according to the needs of an experiment. Game objects can be adjusted in several ways, while the game mechanics remain identical. Game objects are built from meshes and materials—e.g., a mesh in Wavefront ASCII object format (OBJ) and a material in Material Template Library format (MTL). Meshes and materials for game objects can be built in third-party tools and imported into Unity. For easier changes to the visual appearance of game objects, materials with plain colors can be easily defined within Unity; an experimenter can make use of the mesh modeling tool ProBuilder that Unity provides, or can source mesh models from any third-party open-access CAD

Fig. 17 Manipulating player communication networks. **a** Each player can see other player's avatar positions; **b** but the exchange of opinions/signals is governed by a super-imposed network structure. **c** There can be different layers to control different communication abilities



library (e.g., GrabCAD,¹⁹ containing over 4 million models). Visual appearance may have a substantial impact on player experience, and therefore, an experimenter must have flexibility and control over these features. Some features might also be designed to change during an experiment, such as the color of an object changing to indicate a player's current performance at a given task (e.g., see Sect. 4.2.1).

Game objects, such as those in the environment, can be programmed to automatically change during runtime, according to the task or the status of other game objects. For instance, the experimenter can program given parts of the task and environment to scale according to the number of player avatars that join the experiment. Game objects can also be instantiated at any time during the experiment. For example, in a simple estimation task, all objects might be instantiated at the beginning of the experiment (see Fig. 16b). For a dynamic task, the experimenter can program the options to be added at different times during the experiment (see Sect. 4).

An experiment session is initiated once the experimenter opens an instance of HuGoS and connects to the server. The connection established by the experimenter creates a *room* on the server that can be joined by a predefined number of players from their own HuGoS instances. Most of the calculations in the game happen locally on the client instance. Only variables that are important for events that have to be synchronized between clients are passed through the server. An illustration of this architecture is shown in Fig. 15. The course of the experiment is mostly synchronized through messages originating from the experimenter client (the *master client*) to the player clients (full arrows). Changes originating from the players (such as movement) are indicated by dashed arrows. These changes get passed to other players to achieve a synchronized game experience and also to the experimenter to keep track of task progression, calculate the score, and store data.

The activities taking place in each instance are divided into three modules: the player module, the environment module, and the task module. The player module tracks player actions and mediates interactions between players. The environment module instantiates and tracks game objects in the scene, including changes made to them by either players or controllers. The task module determines the sequence of task-related events in the environment and passes them to the environment module. By tracking variables in the player module and environment module, the task module also keeps track of the task progression (e.g., the score). The relevant variables of each module are synced between clients (see Fig. 15).

¹⁹ <https://grabcad.com/library>.

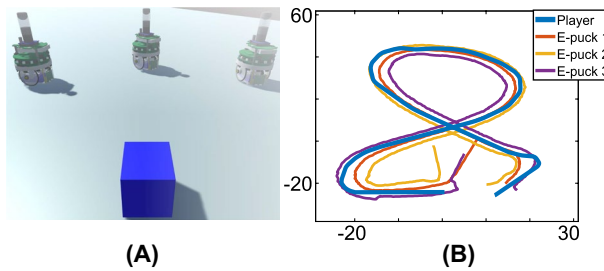


Fig. 18 Implementation of robot models as autonomous agents. **a** Three robot models are in the same environment as a player-controlled avatar (blue cube). **b** The robot models follow the player-controlled avatar (blue line) at a certain distance. This figure is reproduced from our previous work (Coucke et al. 2020) (Color figure online)

Avatar capabilities

Each player controls an avatar that is situated in the virtual environment. The capabilities of the avatars in a given experiment setup are defined in the player module. In HuGoS, players have a first-person or third-person view of their avatar through a virtual camera that follows the avatar position and rotation. Players move their avatar by pressing four user-customizable keys (e.g., WSAD) and rotate via the left/right arrow keys or cursor movement. Depending on the experiment scenario, specific additional actions can be activated for the avatars. For example, the player can be permitted to manipulate the environment by clicking on game objects to grab them, then moving, and releasing the cursor to move them. Indirect communication between players can occur via changes to the environment, for instance, by moving game objects or by changes to display features of the player’s avatar, such as color. Direct communication can also be permitted—and limited as desired—by sending written messages. In the player module of HuGoS, the players’ environment perception can be controlled firstly by changing the field of view (FoV) of the player. A player that has a limited top-down avatar FoV (Fig. 16c) can only perceive the environment in a small perimeter, while a player that has first-person avatar FoV (Fig. 16a) can see a much greater proportion of the environment (Fig. 4b) in the viewed direction. Unlimited top-down FoV—similar to the view of the experimenter—is also possible, giving a player global view (Fig. 16b). Additionally, game objects can be programmed to be invisible to players, or to be visible only for a subset of players.

Player interactions can be modulated by changing the structure of the *player networks* in the player module, which are directed graphs. If a player network is fully connected, for instance, then every player can interact with all other players in the way associated with that network. Player networks manage different types of interaction and have independently defined structures. For example, a fully connected network might be defined for viewing avatar positions, while a sparsely connected network might be defined for viewing avatar colors (Fig. 17c). Player networks also govern explicit message passing between players. As connections are directional (e.g., player 1 might be able to see player 2, while player 2 cannot see player 1), the information privileges of players can be made hierarchical (Fig. 17b). Certain players can have higher node indegrees or outdegrees. The structure of player networks can be changed during experiment runtime, and can optionally be triggered by the players. For instance, players might be permitted to ‘follow’ another player by clicking on its avatar, causing their own decisions to automatically copy those of the followed

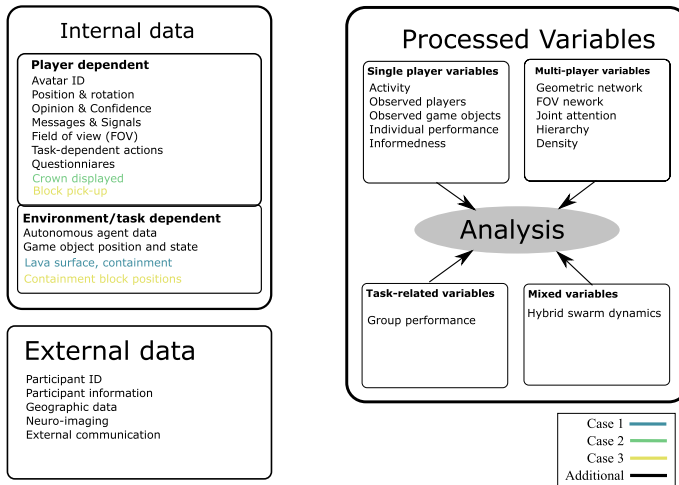


Fig. 19 Primary variables from the environment can be used for analyses specific to the experimental conditions, for calculation of secondary variables, and to conduct further analyses

player, until that player is un-followed (for an implementation, see Coucke et al. 2020). The ability to control communication links between players also allows for comparison between limited communication networks and fully connected communication networks. This can facilitate the study of information cascades, bias in the group, or dysfunctional dynamics that may lead to low performance. Avatars can also act as autonomous agents when their behavior is controlled by algorithms. In some cases, these autonomous avatars might have the same visual appearance as player avatars. In other cases, they might be accurate models of actual robots. These autonomous can interact with each other, the environment, and with player-controlled avatars. Figure 18 shows an implementation where three robots interact with a human-controlled avatar in the same virtual environment. The control of the robots is given in Algorithm 1 (for details, see Coucke et al. 2020).

Algorithm 1: Robot following player

```

if distance to player is larger than 6m then
  | Move in direction of player
for all other robots do
  | if distance to robot is smaller than 4m then
  | | Move away from robot

```

Data types

Data about the players, environment, and task are logged for analysis. Each player has a unique anonymized player ID (defined by the recruiting service in most cases), and each avatar has an avatar ID. These two IDs are important in cases where players switch avatar identities between trials, so that the behaviors of specific players can be analyzed separately from the features accumulated by a shared avatar. Additionally, the player IDs would

be important when the experiments would be conducted in a lab environment (e.g., players occupy the same physical environment and can communicate, or players' physiological data is monitored, such as EEG). Avatar capabilities, positions, orientations, fields of view, and actions are all logged, according to the avatar ID. These logs enable the calculation of other simple data about avatars, such as which other avatars are in one avatar's FoV. Messages passed by avatars are also logged, including the content, time, sender avatar ID, and receiver avatar ID. All other player interactions are also tracked and logged as events—for instance, a player choosing to follow another player—again including content, time, and sender and receiver IDs. Changes in the environment are also logged, including positions and states of all game objects. When artificial agents such as robots are included in a setup, their positional data can be logged in the same way as other avatars (Fig. 18b). Additionally, any data specific to that agent can be logged. For instance, in a setup with models of e-puck robots (Mondada et al. 2009), proximity sensing and motor control might be logged.

Analysis types

The data logged as primary variables (i.e., recorded directly) allow many secondary variables to be calculated and analyzed during runtime or post-processing (Fig. 19). Here, we use task performance as an illustrative example. Task performance can be continuously calculated by the task module, according to the specific scenario. For example, in a flocking scenario the task performance would depend on player positions; or in decision-making, on player opinions. Once task performance is calculated, additional analysis might assess, for instance, how this performance relates to the in-game behavior of players. Player behavior might be represented by distances between avatars, the network of implicit connections between individuals that occur when avatars enter each others' fields of view, or the network of direct messages between players with connection weights representing message frequency. The primary variables also allow for analysis of individual behavior, which can be used to give feedback to players during the experiment. For instance, in a collective decision-making scenario, comparing individual opinion to overall task performance yields relative player performance. If this is provided as feedback to players, players can use it to determine and display their opinion confidence. If the calculated player performance is not provided to the player when the player determines opinion confidence, then a comparison of these two variables will yield the player's *self-assessment* (i.e., the ability to evaluate their own performance). Using player IDs, out-of-game data can also be used in post-analysis. For example, each player might be asked to fill in a questionnaire about personality traits or subjective experience during the game. In an extended out-of-game setup, gameplay could even be linked to real-time physiological recordings, such as eye-tracking, ECG or EDA tracking of stress (Weibel et al. 2018), or neural recordings via EEG or fMRI. Such extensions could be used to analyze the connection between individual cognitive mechanisms and collective performance during gameplay.

Using the data gathered, several behaviors can be studied both implicitly and explicitly. For example, players who wish to lead others could display a crown to signal their desire to lead. They could also try to lead others by acting as an example to be emulated. Various leadership options could be used to study social learning patterns using HuGoS. To analyze social learning explicitly, participants could be given a control option to click on another player's avatar and activate a *follow* behavior, making the follower automatically adopt the choice of the chosen leader. Social learning could also be studied implicitly. For

example, the field of view, position, and performance data from an experiment could allow the experimenter to infer whether a player discovers a new strategy by watching other players, or by executing their own trial-and-error process.

Appendix 2: Detailed setup of additional infrastructure

To carry out these case studies experimentally, the HuGoS platform was integrated with external tools that provided the networking service, data recording, and participant recruitment.

Networking considerations

As described in Sect. 3.3, the networking part of HuGoS is built on PUN.²⁰ Every game takes place in one *room*. Multiple rooms can be opened simultaneously. This means that multiple experiment sessions can be run simultaneously, in different rooms. The limit on the number of players is not likely to arise from explicit restrictions by Photon. Rather, the number of players that a room can support is the result of the ‘server side limit for client buffers’ which is 500 KB. Since every player receives information from all the other players in the experiment, this puts the actual limit on the number of players and attributes that can be synchronized. The maximum number of simultaneous participants in a shared environment thus depends on the requirements of the specific experiment.

Data extraction

Data from the experiments can be recorded either locally or online. In the local version, all the data about the task and the participants are stored in a csv file on the experimenter’s workstation. In the online version, the data of all participants are sent to an online file (e.g., Google Docs) via a webhook; in this case, the experiment can take place without the presence of the experimenter, but it generates extra traffic. In the experiments presented in this paper, data were recorded locally.

Participant recruitment

In the current version, HuGoS is fully accommodated to run online experiments. Participants are recruited via an online platform and can access the experiment in a browser window. This implementation has the advantage that many participants can be recruited with minimal time and cost. Alternatively, HuGoS could be easily adapted to conduct experiments in laboratory settings which would allow the experimenter to have more control over the participant behavior and internet connection. Since these approaches are well established (e.g., Zhao et al. 2018; Boos et al. 2019), we do not discuss them here. Instead, we investigate HuGoS’s potential to facilitate experiments in a completely online way—including the difficulties and limitations.

²⁰ We used the free open-access version that allows up to 20 concurrent users; using the paid license version, Photon can in principle support up to 50,000 concurrent users.

An important challenge when recruiting participants online is to make a certain number of participants simultaneously log in and stay connected to the platform. Addressing this challenge requires making choices as to the following: (1) the way through which to recruit participants, (2) how to maintain the right number of participants in the trial, and (3) how to reimburse the participants.

Recruitment process

We discuss three approaches for planning experiments and recruiting participants: instantaneous recruitment, planned recruitment, and ad hoc voluntary participation. The first two approaches can be done with online recruiting platforms. In these case studies, we used the online recruiting platform *Prolific*²¹ to recruit participants. We opted for Prolific instead of other platforms such as Amazon Mechanical Turk because Prolific enables the selections of participants based on past performance and participation (Palan and Schitter 2018).

In the first approach, participants can be allowed to join a game lobby in real time until enough players joined to start a session. This approach only works if a large number of participants are instantly available. If not, participants joining earlier will have to wait too long in the lobby and might leave again. In our experiments, participants usually joined fast enough to make all players stay. The experimenter can define the number of participants required for a given study. Once they registered for a study, participants are given a link to the website where they can access their HuGoS instance by entering their Prolific ID. At the end of the experiment, participants receive a code that they can enter on Prolific to receive reimbursement.

The second approach for recruiting participants is to design a schedule with different time slots that participants can join. This approach works both with online recruiting platforms and other avenues such as student recruitment at universities. This approach provides the possibility to compose groups of participants based on questionnaires administered at some time before the experiment. A disadvantage of this approach is that the right number of participants rarely shows up at the scheduled time, especially in the case of online platforms.

A third approach makes use of online *citizen science* (e.g., Cooper et al. 2010; Heck et al. 2018). In this approach, participants can voluntarily participate in the study at any time. This would require volunteers to be able to start studies at any time on a website that is constantly active and captures data of any game played. The ad hoc volunteering might make it difficult to have the desired number of players to start a session at any time.

The right number of participants

Since the participants' behavior and internet connection cannot be controlled, the planned number of participants will not always show up. When working with instantaneous recruitment, the number of participants that can connect to the game server within several minutes is usually lower than the number recruited—even when all participants join the study at the same time. This might be due to some participants' slow internet connection, or by the participant's own decision to delay connecting to the server. Additionally, participants might disconnect during the experiment (see Sect. 5.5). Therefore, we always opted for

²¹ <https://www.prolific.co/>.

redundancy in players; for example, a session of 7 or 8 spots was created for a session with 6 participants.

Payment and ethics

In these case studies, participants were rewarded £2.5 for their participation, with a possible bonus of maximum £2 for good performance and extra waiting time. All participants indicated their consent to participate in the study by accepting a written statement. These pilot studies were approved by the ethical committee of the Université libre de Bruxelles (permission 126/2020).

Acknowledgements The authors would like to thank the many volunteers who tested early versions of the case studies and provided invaluable feedback.

Funding This work was partially supported by the program of Concerted Research Actions (ARC) of the Université libre de Bruxelles. M.K. Heinrich, A. Cleeremans and M. Dorigo acknowledge support from the F.R.S.-FNRS, of which they are, respectively, postdoctoral researcher and research directors.

Availability of data and materials The data are available at <https://zenodo.org/record/4384805>.

Code availability The code used in this paper is available at <https://github.com/NicolasCoucke/HuGoS-code-MAIN>.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

Consent to participate All participants indicated their consent to participate in the study by accepting a written statement.

Consent for publication All participants indicated their consent for their anonymized data to be made available to the scientific community, by accepting a written statement.

Ethics approval These pilot studies were approved by the ethical committee of the Université libre de Bruxelles (permission 126/2020).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Almaatouq, A., Becker, J., Houghton, J.P., Paton, N., Watts, D.J., & Whiting, M. E. (2020). Empirica: a virtual lab for high-throughput macro-level experiments. arXiv preprint [arXiv:2006.11398v1](https://arxiv.org/abs/2006.11398v1).
- Bailenson, J. N., Beall, A. C., Loomis, J., Blascovich, J., & Turk, M. (2004). Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence Teleoperators and Virtual Environments*, 13(4), 428–441. <https://doi.org/10.1162/1054746041944803>.

- Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J. S., Gill, S. P. D., Lin, A., & Huyck, C. K. (2011). Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics*. <https://doi.org/10.4401/ag-5324>.
- Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*. https://doi.org/10.1207/S15327965PLI1302_01.
- Boos, M., Pritz, J., Lange, S., & Belz, M. (2014). Leadership in moving human groups. *PLoS Computational Biology*, 10(4), e1003541. <https://doi.org/10.1371/journal.pcbi.1003541>.
- Boos, M., Pritz, J., & Belz, M. (2019). The HoneyComb paradigm for research on collective human behavior. *Journal of Visualized Experiments*, 143, e58719. <https://doi.org/10.3791/58719>.
- Claypool, M. (2005). The effect of latency on user performance in real-time strategy games. *Computer Networks*, 49(1), 52–70. <https://doi.org/10.1016/j.comnet.2005.04.008>.
- Claypool, M., & Finkel, D. (2014). The effects of latency on player performance in cloud-based games. In *13th annual workshop on network and systems support for games*. IEEE. <https://doi.org/10.1109/netgames.2014.7008964>.
- Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., et al. (2010). Predicting protein structures with a multiplayer online game. *Nature*, 466(7307), 756–760. <https://doi.org/10.1038/nature09304>.
- Coucke, N., Heinrich, M.K., Cleeremans, A., & Dorigo, M. (2020). Hugos: A multi-user virtual environment for studying human–human swarm intelligence. In *International conference on swarm intelligence* (pp. 161–175). Springer. https://doi.org/10.1007/978-3-030-60376-2_13.
- Derex, M., & Boyd, R. (2015). The foundations of the human cultural niche. *Nature Communications*, 6(1), 8398. <https://doi.org/10.1038/ncomms9398>.
- Dorigo, M., Theraulaz, G., & Trianni, V. (2020). Reflections on the future of swarm robotics. *Science Robotics*, 5(49):eabe4385. <https://doi.org/10.1126/scirobotics.abe4385>.
- Drury, J. (2018). The role of social identity processes in mass emergency behaviour: An integrative review. *European Review of Social Psychology*, 29(1), 38–81. <https://doi.org/10.1080/10463283.2018.1471948>.
- Ducatelle, F., Di Caro, G. A., Förster, A., Bonani, M., Dorigo, M., Magnenat, S., et al. (2013). Cooperative navigation in robotic swarms. *Swarm Intelligence*, 8(1), 1–33. <https://doi.org/10.1007/s11721-013-0089-4>.
- Eberhart, R., Palmer, D., & Kirschenbaum, M. (2015). Beyond computational intelligence: Blended intelligence. In *Swarm/Human blended intelligence workshop (SHBI)*. IEEE. <https://doi.org/10.1109/shbi.2015.7321679>.
- Ellwart, T. (2011). *Assessing Coordination in human groups: Concepts and methods* (pp. 119–135). Springer. https://doi.org/10.1007/978-3-642-15355-6_7.
- Fowler, J. H., & Christakis, N. A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences*, 107(12), 5334–5338. <https://doi.org/10.1073/pnas.0913149107>.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7).
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63(1), 287–313. <https://doi.org/10.1146/annurev-psych-120710-100449>.
- Heck, R., Vuculescu, O., Sørensen, J. J., Zoller, J., Andreasen, M. G., Bason, M. G., et al. (2018). Remote optimization of an ultracold atoms experiment by experts and citizen scientists. *Proceedings of the National Academy of Sciences*, 115(48), E11231–E11237.
- Hermans, K., Achterhof, R., Myin-Germeyns, I., Kasanova, Z., Kirtley, O., & Schneider, M. (2019). Improving ecological validity in research on social cognition. In *Social cognition in psychosis* (pp. 249–268). Elsevier. <https://doi.org/10.1016/b978-0-12-815315-4.00010-0>.
- Hunt, E. R., Jones, S., & Hauert, S. (2019). Testing the limits of pheromone stigmergy in high-density robot swarms. *Royal Society Open Science*, 6(11), 190225. <https://doi.org/10.1098/rsos.190225>.
- Jaderberg, M., Czarnecki, W. M., Dunning, I., Marris, L., Lever, G., Castañeda, A. G., et al. (2019). Human-level performance in 3d multiplayer games with population-based reinforcement learning. *Science*, 364(6443), 859–865. <https://doi.org/10.1126/science.aau6249>.
- Jensen, J. H. M., Gajdacz, M., Ahmed, S. Z., Czarkowski, J. H., Weidner, C., Rafner, J., Sørensen, J. J., Mølmer, K., & Sherson, J. F. (2020). Crowdsourcing human common sense for quantum control. arXiv preprint [arXiv:200403296v2](https://arxiv.org/abs/200403296v2).
- Juliani, A., Berges, V. P., Vckay, E., Gao, Y., Henry, H., Mattar, M., & Lange, D. (2018). Unity: A general platform for intelligent agents. arXiv preprint [arXiv:180902627](https://arxiv.org/abs/180902627).

- Jung, J. H., Schneider, C., & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science*, 56(4), 724–742. <https://doi.org/10.1287/mnsc.1090.1129>.
- Kirschenbaum, M., & Palmer, D. W. (2015). Perceptualization of particle swarm optimization. In *Swarm/human blended intelligence workshop (SHBI)*. IEEE. <https://doi.org/10.1109/shbi.2015.7321681>.
- Krafft, P. M., Zheng, J., Pan, W., Della Penna, N., Altshuler, Y., Shmueli, E., Tenenbaum, J. B., & Pentland, A. (2016). Human collective intelligence as distributed bayesian inference. arXiv preprint [arXiv:160801987](https://arxiv.org/abs/160801987).
- Krause, J., Ruxton, G. D., & Krause, S. (2010). Swarm intelligence in animals and humans. *Trends in Ecology and Evolution*, 25(1), 28–34. <https://doi.org/10.1016/j.tree.2009.06.016>.
- Labella, T. H., Dorigo, M., & Deneubourg, J. L. (2004). Self-organised task allocation in a group of robots. In *Distributed autonomous robotic systems* (Vol. 6, pp. 389–398). Springer, Japan. https://doi.org/10.1007/978-4-431-35873-2_38.
- Lepri, B., Staiano, J., Shmueli, E., Pianesi, F., & Pentland, A. (2016). The role of personality in shaping social networks and mediating behavioral change. *User Modeling and User-Adapted Interaction*, 26(2–3), 143–175.
- Lin, A. Y. M., Huynh, A., Lanckriet, G., & Barrington, L. (2014). Crowdsourcing the unknown: The satellite search for Genghis Khan. *PLoS ONE*, 9(12), e114046. <https://doi.org/10.1371/journal.pone.0114046>.
- Liu, J., Snodgrass, S., Khalifa, A., Risi, S., Yannakakis, G. N., & Togelius, J. (2020). Deep learning for procedural content generation. *Neural Computing and Applications*, 33(1), 19–37. <https://doi.org/10.1007/s00521-020-05383-8>.
- Lofgren, E. T., & Fefferman, N. H. (2007). The untapped potential of virtual game worlds to shed light on real world epidemics. *The Lancet Infectious Diseases*, 7(9), 625–629. [https://doi.org/10.1016/s1473-3099\(07\)70212-8](https://doi.org/10.1016/s1473-3099(07)70212-8).
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025. <https://doi.org/10.1073/pnas.1008636108>.
- Mathews, N., Christensen, A. L., O’Grady, R., Mondada, F., & Dorigo, M. (2017). Mergeable nervous systems for robots. *Nature Communications*, 8, 439. <https://doi.org/10.1038/s41467-017-00109-2>.
- Michel, O. (2004). Cyberbotics Ltd. Webots™: Professional mobile robot simulation. *International Journal of Advanced Robotic Systems*, 1(1), 40–43.
- Mondada, F., Bonani, M., Raemy, X., Pugh, J., Cianci, C., Klaptocz, A., Magnenat, S., Zufferey, J. C., Floreano, D., & Martinoli, A. (2009). The e-puck, a robot designed for education in engineering. In *Proceedings of the 9th conference on autonomous robot systems and competitions* (Vol. 1, pp. 59–65). IPCB: Instituto Politécnico de Castelo Branco.
- Moussaïd, M., Helbing, D., Garnier, S., Johansson, A., Combe, M., & Theraulaz, G. (2009). Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B: Biological Sciences*, 276(1668), 2755–2762. <https://doi.org/10.1098/rspb.2009.0405>.
- Moussaïd, M., Kapadia, M., Thrash, T., Sumner, R. W., Gross, M., Helbing, D., & Hölscher, C. (2016). Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of the Royal Society Interface*, 13(122), 20160414. <https://doi.org/10.1098/rsif.2016.0414>.
- Moussaïd, M., Schinazi, V. R., Kapadia, M., & Thrash, T. (2018). Virtual sensing and virtual reality: How new technologies can boost research on crowd dynamics. *Frontiers in Robotics and AI*, 5, 82. <https://doi.org/10.3389/frobt.2018.00082>.
- Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In *Flow and the foundations of positive psychology* (pp. 239–263). Springer. https://doi.org/10.1007/978-94-017-9088-8_16.
- Navajas, J., Niella, T., Garbulsky, G., Bahrami, B., & Sigman, M. (2018). Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour*, 2(2), 126–132. <https://doi.org/10.1038/s41562-017-0273-4>.
- Nebel, S., Schneider, S., & Rey, G. D. (2016). Mining learning and crafting scientific experiments: A literature review on the use of minecraft in education and research. *Educational Technology and Society*, 19, 355–366.
- Nebel, S., Schneider, S., Beege, M., Kolda, F., Mackiewicz, V., & Rey, G. D. (2017). You cannot do this alone! Increasing task interdependence in cooperative educational videogames to encourage collaboration. *Educational Technology Research and Development*, 65(4), 993–1014. <https://doi.org/10.1007/s11423-017-9511-8>.

- Nölle, J., Fusaroli, R., Mills, G. J., & Tylén, K. (2020). Language as shaped by the environment: Linguistic construal in a collaborative spatial task. *Palgrave Communications*, 6(1), 27. <https://doi.org/10.1057/s41599-020-0404-9>.
- O'Bryan, L., Beier, M., & Salas, E. (2020). How approaches to animal swarm intelligence can improve the study of collective intelligence in human teams. *Journal of Intelligence*, 8(1), 9. <https://doi.org/10.3390/jintelligence8010009>.
- OpenAI, Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., de Oliveira Pinto, H. P., Raiman, J., Salimans, T., Schlatter, J., Schneider, J., Sidor, S., Sutskever, I., Tang, J., Wolski, F., & Zhang, S. (2019). Dota 2 with large scale deep reinforcement learning. arXiv preprint [arXiv:1912.06680v1](https://arxiv.org/abs/1912.06680).
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>.
- Pinciroli, C., Trianni, V., O'Grady, R., Pini, G., Brutschy, A., Brambilla, M., et al. (2012). ARGoS: a modular, parallel, multi-engine simulator for multi-robot systems. *Swarm intelligence*, 6(4), 271–295.
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., & Ng, A. Y. (2009). ROS: An open-source robot operating system. In *ICRA workshop on open source software* (vol. 3, p. 5). <https://doi.org/10.5772/5618>.
- Quinn, A. J., & Bederson, B. B. (2011). Human computation: A survey and taxonomy of a growing field. In *Proceedings of the international conference on human factors in computing systems, association for computing machinery* (pp. 1403–1412). <https://doi.org/10.1145/1978942.1979148>.
- Rand, D. G., Arbesman, S., & Christakis, N. A. (2011). Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences*, 108(48), 19193–19198. <https://doi.org/10.1073/pnas.1108243108>.
- Rosenberg, L., Baltaxe, D., & Pescetelli, N. (2016). Crowds vs swarms, a comparison of intelligence. In *2016 swarm/human blended intelligence workshop (SHBI)* (pp. 1–5). IEEE. <https://doi.org/10.1109/shbi.2016.7780278>.
- Rosenberg, L. B. (2015). Human swarms, a real-time method for collective intelligence. In *Proceedings of the European conference on artificial life 2015* (pp. 658–659). The MIT Press. <https://doi.org/10.7551/978-0-262-33027-5-ch117>.
- Rubenstein, M., Cornejo, A., & Nagpal, R. (2014). Programmable self-assembly in a thousand-robot swarm. *Science*, 345(6198), 795–799. <https://doi.org/10.1126/science.1254295>.
- Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(2), 226–233. <https://doi.org/10.1016/j.cognition.2009.08.009>.
- Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences*, 104(18), 7361–7366. <https://doi.org/10.1073/pnas.0702077104>.
- Shaker, N., Togelius, J., & Nelson, M. J. (2016). *Procedural content generation in games*. Springer. <https://doi.org/10.1007/978-3-319-42716-4>.
- Shirado, H., & Christakis, N. A. (2017). Locally noisy autonomous agents improve global human coordination in network experiments. *Nature*, 545(7654), 370–374. <https://doi.org/10.1038/nature22332>.
- Sørensen, J. J. W. H., Pedersen, M. K., Munch, M., Haikka, P., Jensen, J. H., Planke, T., et al. (2016). Exploring the quantum speed limit with computer games. *Nature*, 532(7598), 210–213. <https://doi.org/10.1038/nature17620>.
- Stolk, A., Verhagen, L., Schoffelen, J. M., Oostenveld, R., Blokpoel, M., Hagoort, P., et al. (2013). Neural mechanisms of communicative innovation. *Proceedings of the National Academy of Sciences*, 110(36), 14574–14579. <https://doi.org/10.1073/pnas.1303170110>.
- Tavakoli, A., Nalbandian, H., & Ayanian, N. (2016). Crowdsourced coordination through online games. In *2016 11th ACM/IEEE international conference on human–robot interaction (HRI)* (pp. 527–528). <https://doi.org/10.1109/HRI.2016.7451839>.
- Thrash, T., Kapadia, M., Moussaid, M., Wilhelm, C., Helbing, D., Sumner, R. W., & Hölscher, C. (2015). Evaluation of control interfaces for desktop virtual environments. *Presence Teleoperators and Virtual Environments*, 24(4), 322–334. https://doi.org/10.1162/pres_a_00237.
- Titlestad, K., Snijders, T., Durrheim, K., Quayle, M., & Postmes, T. (2019). The dynamic emergence of cooperative norms in a social dilemma. *Journal of Experimental Social Psychology*, 84, 103799. <https://doi.org/10.1016/j.jesp.2019.03.010>.
- Valentini, G., Ferrante, E., & Dorigo, M. (2017). The best-of-n problem in robot swarms: Formalization, state of the art, and novel perspectives. *Frontiers in Robotics and AI*, 4. <https://doi.org/10.3389/frobt.2017.00009>.

- Vasile, C., Pavel, A., & Buiu, C. (2011). Integrating human swarm interaction in a distributed robotic control system. In *2011 IEEE international conference on automation science and engineering* (pp. 743–748). IEEE. <https://doi.org/10.1109/CASE.2011.6042493>.
- Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., et al. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology*, *07*, 2039. <https://doi.org/10.3389/fpsyg.2016.02039>.
- Walker, P., Amraii, S. A., Chakraborty, N., Lewis, M., & Sycara, K. (2014). Human control of robot swarms with dynamic leaders. In *2014 IEEE/RSJ international conference on intelligent robots and systems* (pp. 1108–1113). IEEE. <https://doi.org/10.1109/IROS.2014.6942696>.
- Wang, H., & Sun, C. T. (2012). Game reward systems: Gaming experiences and social meanings. In *Proceedings of DiGRA 2011 conference: Think design play, DiGRA* (vol. 6).
- Weibel, R. P., Grübel, J., Zhao, H., Thrash, T., Meloni, D., Hölscher, C., & Schinazi, V. R. (2018). Virtual reality experiments with physiological measures. *Journal of Visualized Experiments*, *138*, e58318. <https://doi.org/10.3791/58318>.
- Werfel, J., Petersen, K., & Nagpal, R. (2014). Designing collective behavior in a termite-inspired robot construction team. *Science*, *343*(6172), 754–758. <https://doi.org/10.1126/science.1245842>.
- Williams, J. P., & Kirschner, D. (2012). Coordinated action in the massively multiplayer online game world of warcraft. *Symbolic Interaction*, *25*(3), 340–367. <https://doi.org/10.1002/j.1533-8665.2012.00022.x>.
- Yannakakis, G. N., & Togelius, J. (2018). Modeling players. In *Artificial Intelligence and Games* (pp. 203–255). Springer. https://doi.org/10.1007/978-3-319-63519-4_5.
- Zhang, Y., Wang, S., Oğuz, S., Heinrich, M., Wang, X., Garone, E., & Dorigo, M. (2021). *Self-reconfigurable hierarchical frameworks for bearing-based formation control of robot swarms*. Technical Report TR/IRIDIA/2021-004, IRIDIA, Université Libre de Bruxelles, Brussels, Belgium.
- Zhao, H., Thrash, T., Wehrli, S., Hölscher, C., Kapadia, M., Grübel, J., Weibel, R. P., & Schinazi, V. R. (2018). A networked desktop virtual reality setup for decision science and navigation experiments with multiple participants. *Journal of Visualized Experiments*. <https://doi.org/10.3791/58155>.
- Zhao, H., Thrash, T., Grossrieder, A., Kapadia, M., Moussaïd, M., Hölscher, C., & Schinazi, V. R. (2020). The interaction between map complexity and crowd movement on navigation decisions in virtual reality. *Royal Society Open Science*, *7*(3), 191523. <https://doi.org/10.1098/rsos.191523>.
- Zhu, W., Allwright, M., Heinrich, M. K., Oğuz, S., Christensen, A. L., & Dorigo, M. (2020). Formation control of UAVs and mobile robots using self-organized communication topologies. In *Swarm Intelligence—Proceedings of ANTS 2020—Twelfth International Conference* (vol. 12421, pp. 306–314). Springer, Berlin, Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-030-60376-2_25.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.