UNIVERSITÉ LIBRE DE BRUXELLES
École Polytechnique de Bruxelles
IRIDIA - Institut de Recherches Interdisciplinaires
et de Développements en Intelligence Artificielle

Effects of the Interaction with Robot Swarms on the Human Psychological State

Gaëtan Podevijn

Promoteur de Thèse:
Prof. **Marco Dorigo**

Thèse présentée en vue de l'obtention du titre de
Docteur en Sciences de l'Ingénieur

Année académique 2016-2017

Effects of the Interaction with Robot Swarms on the Human Psychological State

Abstract

Human-swarm interaction studies how human beings can interact with a robot swarm—a large number of robots cooperating with each other without any form of centralized control. In today's human-swarm interaction literature, the large majority of the works investigate how human beings can issue commands to and receive feedback from a robot swarm. However, only a few of these works study the effect of the interaction with a robot swarm on human psychology (e.g., on the human stress or on the human workload). Understanding human psychology in human-swarm interaction is important because the human psychological state can have significant impact on the way humans interact with robot swarms (e.g., a high level of stress can cause a human operator to freeze in the middle of a critical task, such as a search-and-rescue task).

Most existing works that study human psychology in human-swarm interaction conduct their experiments using robot swarms simulated on a computer screen. The use of simulation is convenient because experimental conditions can be repeated perfectly in different experimental runs and because experimentation using real robots is expensive both in money and time. However, simulation suffers from the so-called *reality gap*: the inherent discrepancy between simulation and reality. It is therefore important to study whether this inherent discrepancy can affect human psychology—human operators interacting with a simulated robot swarm can react differently than when interacting with a real robot swarm.

A large literature in human-robot interaction has studied the psychological impact of the interaction between human beings and single robots. This literature could in principle be highly relevant to human-swarm interaction. However, an inherent difference between human-robot interaction and human-swarm interaction is that in the latter, human operators interact with a large number of robots. This large number of robots can affect human psychology—human operators interacting with a large number of robots can react differently than when interacting with a single robot or with a small number of robots. It is therefore important to understand whether the large number of robots that composes a robot swarm affects human psychology. In fact, if this is the case, it would not be possible to directly apply the results of human-robot interaction research to human-swarm interaction.

We conducted several experiments in order to understand the effect of

the reality gap and the effect of the group size (i.e., the number of robots that composes a robot swarm) on the human psychological state. In these experiments our participants are exposed to swarms of robots and are purely passive—they do not issue commands nor receive feedback from the robots. Making the interaction passive allowed us to study the effects of the reality gap and of the group size on the human psychological state without the risk that an interaction interface (such as a joystick) influences the psychological responses of the participants (and thus limiting the visibility of both the reality gap and group size effects). In the reality gap experiments, participants are exposed to simulated robot swarms displayed either on a computer screen or in a virtual reality environment, and to real robot swarms. In the group size experiments, participants are exposed to an increasing number of real robots.

In this thesis, we show that the reality gap and the group size affect the human psychological state by collecting psychophysiological measures (heart rate and skin conductance), self-reported (via questionnaires) affective state measures (arousal and valence), self-reported workload (the amount of mental resource needed to carry out a task) and reaction time (the time needed to respond to a stimulus). Firstly, we show with our results that our participants' psychophysiological measures, affective state measures, workload and reaction time are significantly higher when they interact with a real robot swarm compared to when they interact with a robot swarm simulated on a computer screen, confirming that the reality gap significantly affects the human psychological state. Moreover, we show that it is possible to mitigate the effect of the reality gap using virtual reality—our participants' arousal, workload and reaction time are significantly higher when they interact with a simulated robot swarm displayed in a virtual reality environment as opposed to when it is displayed on a computer screen. Secondly, we show that our participants' psychophysiological measures and affective state measures increase when the number of robots they are exposed to increases.

Our results have important implications for research in human-swarm interaction. Firstly, for the first time, we show that experiments in simulation change the human psychological state compared to experiments with real robots. Secondly, we show that a characteristic that is inherent to the definition of swarm robotics—the large number of robots that composes a robot swarm—significantly affects the human psychological state. Finally, our results show that psychophysiological measures, such as heart rate and skin conductance, provide researchers with more information on human psychology than the information provided by using traditional self-reported measures (collected via psychological questionnaires).

À Caroline, Nolan, Johanne, Maman et Papa.

Acknowledgments

FIRST OF ALL, I thank my thesis supervisor, Marco Dorigo. Thank you Marco for trusting in me when I proposed a thesis subject that mainly focused on psychology. Not many supervisors would have taken this risk. I am glad and proud that this risk paid off.

For the same reason, and for many others, I would like to express my sincerest thanks to Rehan O'Grady. You have been a fantastic co-supervisor during these last six years (we started working together in 2010, when I was a first year Master student!). During all these years, you have always been extremely patient with me. Your advices and your help allowed me to grow as a scientist, but more importantly as a person.

Thank you Mauro Birattari for always being avalaible when I needed external advices and external help. Thank you also for allowing me to co-supervsise Master students with you and Gio. And thank you very much for giving me the taste of Brazilian jiu-jitsu. Though we will not see each other at IRIDIA anymore, we will still see each other at Kaizen! Oss.

Thank you Carole Fantini-Hauwel for getting involved in my Ph.D. Without you, I would never have been able to produce this work. Thanks for the meetings, the coffees, your advices and your help for the psychological side of my research.

Thank you Nithin Mathews and Audrey Gilles. Nithin, it was great working with you. Your calm and your perpetual positive spirit are inspiring. Audrey, thanks a lot for the experiments we conducted together. It was a lot of fun!

Thank you Roman Miletitch for being an awesome office mate. Your "presence" in the office made my days full of fun. I will miss our endless philosophical discussions about life and work. We also should continue making side projects we will never finish. Also: café?

Thank you to the generation of IRIDIAns who were there when I arrived and who left during the course of my Ph.D.—it was inspiring seeing you succeed in your scientific career: Carlo, Eliseo, Giovanni, Arne, Vito, Manuele, Jérémie. Thanks to the IRIDIAns of "my" generation—I will keep great memories of all of you: Dj, Gabri, Leslie, Anthony, Lorenzo, Alberto, Federico, Joe. And thanks to the new generation: Marcolino, Ken, Volker, Anthoine; IRIDIA is in good hands with you guys.

Merci à mes parents et à ma soeur pour m'avoir soutenu durant toutes mes études et pour avoir toujours cru en moi. Et merci de m'avoir toujours permis de faire ce que je voulais faire.

Merci à Steph, Nico, Amé, Arnaud et François (dit Bus) pour tous les moments "off" qui m'ont permis de déconnecter.

Enfin, merci à Caroline pour avoir toujours été présente à mes côtés. Merci d'avoir été compréhensive pendant les périodes les plus difficiles.

Composition of the jury

Hugues Bersini
Université Libre de Bruxelles, Brussels, Belgium

Mauro Birattari
Université Libre de Bruxelles, Brussels, Belgium

Marco Dorigo (*thesis supervisor*)
Université Libre de Bruxelles, Brussels, Belgium

Carole Fantini-Hauwel
Université Libre de Bruxelles, Brussels, Belgium

Rehan O'Grady
RISE Financial Technologies
Université Libre de Bruxelles, Brussels, Belgium

Carlo Pinciroli
Worcester Polytechnic Institute, Massachusetts, USA

Bram Vanderborght
Vrije Universiteit Brussel, Brussels, Belgium

Contents

1
Introduction

A swarm robotics system comprises a large number of relatively simple robots that rely on decentralized control to carry out complex tasks by interacting and cooperating with each other. Swarm robotics systems have three main desirable properties: i) *robustness* to individual failures, ii) *scalability* in the number of robots, and iii) *flexibility* to the environment in which they operate.

To date, swarm robotics has been an active cutting edge research field on which millions of euros are spent annually. For example, the European Union has funded numerous projects such as Swarm-Bots (Dorigo, 2001–2005), I-SWARM (Seyfried, 2004–2008), Swarmanoid (Dorigo, 2006–2010), GUARDIANS (Penders, 2006–2010), SYMBRION (Winfield, 2008–2013), sFly (Scaramuzza, 2009–2012), ASCENS (Wirsing, 2010–2015), E-Swarm (Dorigo, 2010–2015), CoCoRo (Schmickl, 2011–2014), DiODE (Marshall, 2015–2020), Flora Robotica (Hamann, 2015–2019) and DEMIURGE (Birattari, 2016–2021).

Swarm robotics systems are well suited for real world applications such as search-and-rescue, demining, harvesting, grass mowing or even construction. However, swarm robotics is still confined to research laboratories. A critical challenge to tackle in order to make swarm robotics systems ready for real world applications is enabling humans to interact with robot swarms. Human-swarm interaction (i.e., the interaction between humans and robot swarms) is vital for real world applications because swarms of robots will never be fully autonomous. Humans usually have a high-level understanding of a task, while a robot swarm only has a low-level understanding of the task it carries out. This high-level understanding can be leveraged by human operators to take real-time and ethical decisions that could not be taken by a robot swarm. In a search-and-rescue task for instance, a human operator can give priority to sites in which there is higher chance of finding victims. This decision could not be taken by a robot swarm because it does not have a high-level understanding of the task. And even if it had this high-level understanding, this decision could not be taken by a robot swarm for ethical reasons (this decision could lead to the situation where victims would not be found, or found too late, resulting in loss of human lives).

An important step towards an effective human-swarm interaction system is to understand to what extent the interaction with a robot swarm impacts human psychology. Understanding the impacts of the interaction with an interactive system—and in the case of this thesis, with a swarm robotics system—on human psychology is essential for two reasons (Carroll, 1997). Firstly, it allows researchers to have a global comprehension of humans interacting with a robot swarm. This global comprehension can be used by researchers to frame the development of future human-swarm interaction systems. Secondly, it allows researchers to evaluate the usability of human-swarm interaction systems. In the current human-swarm interaction literature, the study of the impacts of the interaction with a robot swarm on human psychology is largely ignore.

The goal of this thesis is to have a first understanding of human psychology in human-swarm interaction. As we see in Chapter 2, only a few works study human psychology in human-swarm interaction research. These works, moreover, are conducted in simulation—a group of participants interact with a simulated robot swarm, in a simulated environment, displayed on a computer screen. In the human-robot interaction literature (presented in Chapter 2), however, when human operators interact with a robotics system in simulation, their psychological reactions are significantly different than when they interact with a real robotics system. Before conducting any human-swarm interaction experiments, therefore, it is important to understand whether simulation and reality also provoke different psychological reactions in the context of human-swarm interaction. In this thesis, we study the effect of interacting with a simulated robot swarm and with a real robot swarm on human psychology (henceforth referred to as the *reality gap*). We hypothesize that when humans interact with a robot swarm in simulation, their psychological responses are lower than when they interact with a real robot swarm. By confirming this hypothesis, we would show that the current methodology to conduct human-swarm interaction experiments is flawed and that human-swarm interaction experiments should be conducted with real robots.

In human-robot interaction, researchers already studied the psychological effects of the interaction between a human operator and a single robot (see a selection of studies in Chapter 2). The results of the experiments conducted in human-robot interaction could be significantly pertinent to human-swarm interaction. However, there is a fundamental difference between human-robot interaction and human-swarm interaction: in human-swarm interaction, human operators interact with a large number of robots. As of today, we do not know whether this large number of robots—a characteristic inherent to the definition of swarm robotics—could affect human psychology. In this thesis, we study the effect of the increasing number of robots on human psychology. We hypothesize that when humans are exposed to a large number of robots, their psychological reactions are higher than when they are exposed to a single robot, or to a smaller group of robots. If this is the case, there would be significant implications for research in human-swarm interaction. Since the large number of robots is inherent to the definition of swarm robotics, it would not make any sense to reduce the number of robots in order to decrease the human operators' psychological responses. Researchers who propose novel interaction interfaces, therefore, should focus their efforts on developing interaction interfaces that do not increase even more the human operators' psychological responses.

In this thesis, we study the human psychological state by analysing a combination of objective and subjective measures. Objective measures are measures that can not be consciously modified by humans. In this thesis, objective measures are psychophysiological measures (heart rate and skin conductance) and reaction time measures. Subjective measures are measures that a participant can consciously modify because they are reported via a questionnaire (the participants may decide not to report the truth). In this thesis, subjective measures are affective state measures (arousal and valence) and workload measure (see Chapter 3 for a definition of each of these measures).

In order to confirm the hypothesis that the reality gap affects the human psychological state, we present two experiments in which we compare the psychological state of a group of participants interacting with real robot swarms and with simulated ones. We show that the reality gap effect significantly affects our participants psychological state—our participants' psychophysiological measures, affective state measures, workload and reaction time measures are significantly higher when they interact with a real robot swarm compared to with a simulated robot swarm. Moreover, we propose virtual reality as an alternative to computer-based simulation to mitigate the reality gap effect—we show that some of the psychological reactions (arousal, reaction time and workload) of our participants are significantly higher when they interact with simulated robot swarms in a virtual reality environment compared to with simulated robot swarms displayed on a computer screen.

In order to confirm the hypothesis that the number of robots (the group size) affects the human psychological state, we present an experiment in which a group of participants interact with an increasing number of robots. In this experiment, we avoid the reality gap effect by conducting the experiment with swarms of real robots. We show that the greater the number of robots, the more the human psychological state is affected—our participants' psychophysiological and affective state measures are higher when they interact with the highest number of robots. These results have important implications for research in human-swarm interaction, as the large number of robots that compose a robot swarm is an inherent characteristic of swarm robotics.

In this thesis, we study the reality gap effect and the group size effect in a specific form of interaction—in a passive interaction. In a passive interaction, human operators are only exposed to a swarm of robots and do not issue any commands to nor receive any feedback from a swarm of robots. We decided to use passive interaction because it allows us to isolate the reality gap effect and the group size effect from potential confounding factors that would be due to an additional interaction interface (e.g., a joystick).

1.1 Thesis Structure and Related Scientific Contributions

In this section, we describe the structure of this manuscript, and present the related scientific articles published in international conferences and journals.

In Chapter 2, we first review the existing literature on human-swarm interaction. Then, we review the related literature to our research on the effect of the reality gap. We also review the literature in human-robot interaction that contains experiments in which psychophysiological measures are used. Finally, we review the literature related to the effect of the robot group size on the human psychological state.

In Chapter 3, we describe the materials and methods used to develop and analyse our experiments. Firstly, we present the robotic platform used in our experiments, the experimental environments and the software used to simulate robot swarms. Secondly, we present the measures and the tools that we used to analyse our experiments.

In Chapter 4, we present two experiments that we conducted in order to understand what the effect of the reality gap is on human psychology. This chapter is based on two scientific articles published in an international journal and in an international conference:

- **G. Podevijn**, R. O'Grady, C. Fantini-Hauwel, M. Dorigo. Investigating the Effect of the Reality Gap on the Human Psychophysiological State in the Context of Human-swarm Interaction. *PeerJ Computer Science*, 2:e82, 2016.

- **G. Podevijn**, R. O'Grady, C. Fantini-Hauwel, M. Dorigo. Human Responses to Stimuli Produced by Robot Swarms - The Effect of the Reality Gap on Psychological State. In *Proceedings of the 13th International Symposium on Distributed Autonomous Robot Systems, DARS 2016*, Springer, in press.

In Chapter 5, we present an experiment conducted with a robot swarm composed of up to 24 real robots. In this study, we show that the number of robots in a robot swarm has significant impacts on human psychology. The results presented in this chapter have been published in an international journal:

- **G. Podevijn**, R. O'Grady, N. Mathews, A. Gilles, C. Fantini-Hauwel, M. Dorigo. Investigating the Effect of Increasing Robot Group Sizes on the Human Psychophysiological State in the Context of Human-swarm Interaction. *Swarm Intelligence*, 10(3), pp. 193–210, Springer, 2016.

In Chapter 6, we discuss the limitations of our research. We propose several future work directions in order to overcome these limitations and present our conclusions.

1.2 Other Scientific Contributions

Together with Prof. Marco Dorigo and Dr. Rehan O'Grady, we started to study human-swarm interaction during my Master's thesis at IRIDIA. During the first year of my doctoral studies, we published two papers based on the materials of my Master's thesis:

- **G. Podevijn**, R. O'Grady, and M. Dorigo. Self-organised Feedback in Human-swarm Interaction. In *Workshop on Robot Feedback in Human-*

Robot Interaction: How to Make a Robot Readable for a Human Interaction Partner, Ro-Man 2012, Paris, France, 2012.

- **G. Podevijn**, R. O'Grady, Y. Nashed, and M. Dorigo. Gesturing at Subswarms: Towards Direct Human Control of Robot Swarms. In *Proceedings of the Fourteenth Annual Conference on Towards Autonomous Robotic Systems, TAROS 2013*, pp. 390–403, volume 8069 of LNCS, Springer, 2013.

In the first paper, we proposed to leverage the same self-organised mechanisms used in swarm robotics in order for a swarm of robots to provide a human operator with meaningful feedback. We named this type of feedback *self-organised feedback*. In the second paper, we presented a framework that enables interaction between human beings and robot swarms. In this paper, the human operator occupies the same space as the robots, and can engage in bi-directional communication with the robots (i.e., the human operator can issue commands to the robots and can receive feedback from the robots) without the need for any intermediate interface such as a graphical user interface. The framework presented in this paper is based on the idea that the human operator should be able to interact with a swarm of robots as if it were a single entity. We developed a gesture-based interaction interface for human operators to issue commands to robot swarms. These commands are converted into meaningful individual commands by distributed algorithms running on each robot. In this paper, we validated the proposed framework by conducting an experiment in which 18 participants had to guide simulated robots to complete a resource allocation scenario. We also validated the framework by conducting an experiment with real robots.

During my four years at IRIDIA, I also had the opportunity to collaborate with Dr. Mauro Birattari and Mr. Gianpiero Francesca. We collaborated on the analysis of `AutoMoDe`. `AutoMoDe` is an approach proposed by G. Francesca and M. Birattari to automatically design control software for robot swarms. This collaboration resulted in the publication of a conference paper and of a journal paper:

- G. Francesca, M. Brambilla, A. Brutschy, L. Garattoni, R. Miletitch, **G. Podevijn**, A. Reina, T. Soleymani, M. Salvaro, C. Pinciroli, V. Trianni, and M. Birattari. An Experiment in Automatic Design of Robot Swarms: AutoMoDe-Vanilla, EvoStick, and Human Experts. In *Proceedings of the Ninth International Conference on Swarm Intelligence, ANTS 2014*, pp. 25–37, volume 8667 of LNCS, Springer, 2014.

- G. Francesca, M. Brambilla, A. Brutschy, L. Garattoni, R. Miletitch, **G. Podevijn**, A. Reina, T. Soleymani, M. Salvaro, C. Pinciroli, F. Mascia, V. Trianni, and M. Birattari. AutoMoDe-Chocolate: A Method for the Automatic Design of Robot Swarms that Outperforms Humans. *Swarm Intelligence*, 9(2–3):125–152, Springer, 2015.

During this collaboration, I acted, together with other Ph.D. students of IRIDIA, as a human expert in the programming of control software for robot swarms. In these papers, the control software produced by two specializations of `AutoMoDe` (i.e., `AutoMoDe-Vanilla` and `AutoMoDe-Chocolate`) were compared to the control software produced by human experts and by other automatic design approaches (e.g., `EvoStick`). Results have shown that while `AutoMoDe-Vanilla` is not able to outperform human experts (but is able to outperform other automatic design approaches), `AutoMoDe-Chocolate` (an improved version of `AutoMoDe-Vanilla`) performs better than human experts in the design of control software for robot swarms.

I also had the opportunity to co-supervise the Master's thesis of a student from the Faculté des Sciences Appliquées. Together with Dr. Andreagiovanni Reina (Ph.D. student at that time) and Dr. Mauro Birattari, we defined an original research idea in human-swarm interaction and we co-supervised the student's research during his final academic year.

- Anthony Debruyn. *Human-Swarm Interaction: An Escorting Robot Swarm that Diverts a Human Away from Dangers one Cannot Perceive.* M.Sc. Thesis in Computer Science Engineering, Université Libre de Bruxelles, Belgium, 2015.

2

Context and Related Literature

In this chapter, we describe the context of the research presented in Chapter 4 and Chapter 5. In Section 2.1, we review the current literature in human-swarm interaction. As we will see, few of the works presented in this section conducted experiments with participants and the majority of the works that conducted experiments with participants used simulated robots in a simulated environment. Moreover, few of the works that conducted experiments with participants have investigated the psychology of human operators interacting with a robot swarm.

Due to the reality gap effect, the psychological state of the participants conducting an experiment in simulation can be biased. To date, there is no work in the human-swarm interaction literature (other than our work presented in Chapter 4) that investigates the effect of the reality gap on the human psychological state. The effect of the reality gap, however, is studied when humans interact with a single robot—with a service robot or with a companion robot for instance. In Section 2.2, we review these works.

In this thesis, we study the human psychological state by using psychophysiological measures (see Chapter 3). Psychophysiological measures have been used in a single human-swarm interaction study other than our studies, but have been used more frequently in human-robot interaction to investigate the human psychophysiological state when humans interact with a single robot. In Section 2.3, we review these studies.

We also present the first work in human-swarm interaction that studies the effect of the number of robots on the human psychophysiological state. Following the conclusions of our research on the reality gap, we conducted an experiment with swarms of real robots (see Chapter 5). The effect of the number of robots was already studied in human-swarm interaction and in human multi-robot interaction. However, these works did not use psychophysiological measures and did not use real robots. In Section 2.4, we review these works.

2.1 Human-swarm Interaction

The decentralized nature of a swarm robotics system along with the large number of robots that compose a swarm robotics system make the interaction with these systems challenging. In the last five years, the challenges encountered during the interaction between humans and robot swarms have motivated researchers to propose novel approaches to interacting with robot swarms. Today, human-swarm interaction has become an active and independent field of research.

As the remaining of this section shows, today's human-swarm interaction literature is very scattered. There are several research directions and within each direction, the literature shows a disparity of both the results and the methodology used to obtain these results. A reason for this disparity is that, as of today, there is no concrete real world applications of swarm robotics. It is not clear how swarms of robots are going to be used to carry out real world tasks. And without a clear understanding of the use of swarm robotics systems, it is extremely difficult to understand how humans are going to interact with these systems.

In this section, we review the different research directions in human-swarm interaction. In contrast to the survey of Kolling et al. (2016), we devote particular attention to researches that i) validate their approaches with experiments in simulation or with experiments with real robots and ii) study the impacts of their approaches on the human psychological state. At the end of this section, we summarize in Table 2.1 the researches that validated their approach by conducting experiments with a group of participants (with simulated robots or with real robots) and those that studied the psychological impact of their approach.

To the best of our knowledge, there is only one research in human-swarm interaction that studied the human psychological state with psychophysiological measures (Karavas and Artemiadis, 2015). In this research, the authors investigated the effect of the swarm's cohesion on the human electroencephalographic (EEG) activity. The authors conducted an experiment with two participants. In this experiment, the participants had to watch a simulated robot swarm moving from one edge of a computer screen to another. The participants conducted the experiment multiple times with different swarm's cohesions. In each experiment, after a certain amount of time, the robot swarm changed its direction of motion. The participants were asked to press a button when the robot swarm's direction of motion changed. The authors measured the participants' reaction time to press the button and the participants' EEG activity. The results show that the swarm's cohesion affected both the participants' reaction time and the participants' EEG activity. Unfortunately, the experiment was conducted with a very small number of participants, did not include any subjective measures (i.e., psychological questionnaires) and was conducted in simulation only.

We organised the literature in human-swarm interaction into five categories: robot swarm control, interaction interfaces, bandwidth limitation and neglect benevolence, level of automation, and formal verification. In the remaining of this section, we review the literature related to each of these five categories.

2.1.1 Robot Swarm Control

A great deal of attention has been given to controlling (e.g., guiding) a robot swarm. Two well studied methods for controlling a robot swarm are based on the control of a leader robot or multiple leader robots. In these methods, a human operator controls a single robot or a subset of robots in order to influence the other robots of the robot swarm (i.e., robots that are not controlled by the human operator).

To the best of our knowledge, there is no study that compares these two control methods. The following works show that there is no consensus on which control method (a single robot or a subset of robots) is more appropriate to control a robot swarm.

Bashyal and Venayagamoorthy (2008) designed an experiment based on a radiation source localization problem (i.e., a swarm of robots must detect the localization of a radiation source). The authors conducted an experiment with 5 participants. In this experiment, a human operator is asked to select and guide a leader robot in order for the robot swarm to explore seven key-locations defined in an environment. These key-locations represent seven areas in which radiation sources could be present. The goal is for the human operator to reach these seven key-locations as fast as possible. The results of the experiment showed that when a human operator can guide a leader robot from a robot swarm, the seven key-locations were reached 50% faster than when the robot swarm was fully autonomous. The experiment was conducted in simulation only.

Walker et al. (2013a) studied the effect of two propagation methods when a human operator guides a leader robot (i.e., methods to propagate a command issued by a human operator to all the robots of a swarm). In their work, a human operator guides a leader robot by changing the leader robot's velocity and heading through a graphical interface. They compare the so-called *flooding propagation method* to the so-called *consensus propagation method*. In the flooding propagation method, the robots of the swarm controlled by a human operator all set their velocity and heading to the leader robot's velocity and heading. In the consensus propagation method, the robots of the swarm all set their velocity and heading to the average velocity and heading of their neighbours. The authors conducted an experiment with 18 participants. Each participant had to guide a robot swarm towards specific locations of an environment with the flooding propagation method and with the consensus propagation method. The authors showed that the participants were able to reach more locations with the flooding propagation method than with the consensus propagation method. An analysis of the leader's position in the swarm also showed that the participants had better results (i.e, they reached more locations) when the leader robot was positioned closer to the front of the swarm. The authors also evaluated the effect of the propagation method on the participants' psychological state by measuring the participants' workload (see Section 3.2.3). The participants' workload was lower in the flooding propagation method than in the consensus propagation method. The experiment was conducted in simulation only.

Walker et al. (2014) suggested to use multiple leader robots that are dynamically selected (i.e., the leader robots can automatically change over time). The authors conducted an experiment with 17 participants. As in Walker et al. (2013a), the participants of the experiment had to guide a robot swarm in order to reach different locations of an environment. In contrast to Walker et al. (2013a), only the consensus propagation method was used. In this experiment, the authors showed that when each robot of the swarm was 1-hop away from a leader robot (i.e., each non-leader robot is the neighbour of a leader robot), the performance of the human operator (i.e., reaching specific locations in an environment) was higher than when each robot was 2-hop or 3-hop away from a leader robot (i.e., there are at most 2 or 3 other robots between a non-leader robot and a leader robot). The experiment was conducted in simulation only.

Kapellmann-Zafra et al. (2016b) designed an experiment in which a human operator helps a robot swarm to aggregate in a location of an obstructed environment. The human operator can control a single leader robot at a time. In this experiment, the operator controls the leader robot via a graphical interface. The operator can control the leader robot's motion or can change its behaviour (e.g., aggregation, follower, gossip). When the leader robot's behaviour is changed, the behaviour is broadcasted to the other robots of the robot swarm. The graphical interface provides the operator with very limited feedback about the robot under control (i.e., either camera feedback or proximity sensor values). The authors conducted an experiment with 42 participants. The authors used the number of robots in the largest aggregated group of robots as a performance metric. The results showed that, untrained participants (i.e., participants with no prior training with the experiment) without access to the environment's global state (i.e., bird's eye view of the environment) did not have a significant influence on the robot swarm. When the untrained participants did have access to the global state of the environment, however, they had a significant influence on the robot swarm. The results also showed that trained participants (i.e., participants with prior training with the experiment) and expert participants (i.e., developers of the experimental software) without access to the global state of the environment had a significant influence on the robot swarm. The experiment was conducted in simulation only.

De la Croix and Egerstedt (2012) studied the effect of different initial robot swarm's network topologies when a human operator has to change the initial robot swarm's network topology to either an ellipse or to a wegde by controlling a single leader robot at a time. The authors conducted an experiment with 18 participants. The authors used the mean least square (mean LSQ) metric to measures the final formation of the robot swarm. They also measured the participants' workload and collected the participants' self-assessed difficulty rating (i.e., a score between 0 and 20 representing the difficulty of the task). The results showed that the initial network topology of a robot swarm had a significant impact on the human operator's performance (assessed by the mean LSQ metric), workload and self-assessed difficulty rating. The experiment was conducted in simulation only.

Salomons et al. (2016) presented a research in which a human operator can manage a robot swarm via a leader robot (i.e., assign a task to the robot swarm, request the robot swarm to count how many robots are in the robot swarm, create subgroups of robots). In this research, the leader robot is either selected randomly or is selected by a human operator (by providing a robot's identification number). Once a leader is selected, the other robots of the robot swarm (i.e., the non-leader robots) stop and wait for incoming commands (e.g., a new task). The authors did not conduct any experiment with participants. However, they validated their approach with real robots.

Other works by Goodrich et al. provided a more theoretical approach to controlling a swarm with leader robots. Goodrich et al. (2012a,b) identified different types of robots—*human aware* and *human blind*. Human aware robots are either robots that are controlled by a human operator only, or robots that are controlled by a human operator and influenced by nearby robots. Human blind robots are influenced only by nearby robots. The authors also studied the effect of controlling a robot swarm with two types of leaderships—*lead by attraction* (also called *leader-style*) and *lead by repulsion* (also called *predator-style*). None of these works conducted an experiment with participants and none of these works conducted an experiment with real robots.

Controlling a robot swarm with a leader robot or with multiple leader robots is not the only approach that has been investigated in human-swarm interaction. To the best of our knowledge, among the works that use another approach to controlling a robot swarm, only one conducted an experiment with participants.

Kolling et al. (2013) proposed and compared two control methods—*selection* and *beacon* control. With the selection control method, a human operator selects a robot swarm (e.g., by drawing a rectangular zone around a robot swarm in a graphical interface) and then controls the selected robot swarm. With the beacon control method, a human operator exerts an indirect influence on the robot swarm by adding virtual beacons in the environment (via a graphical interface). The authors conducted an experiment with 32 participants. The participants had to use each control method (i.e., selection and beacon) in order to collect information placed at different locations of an environment. Each information was assigned to a specific score. The sum of the scores assigned to the information collected by the participants was used as a performance metric. The participants performed the experiment in different types of environments (with and without obstacles) and with a different number of robots (from 50 to 200). The results showed that the participants performed the experiment better with the selection control method than with the beacon control method. Participants also seemed to have better performance with the beacon control method when the number of robots in the robot swarm increased. The experiment was conducted in simulation only.

The following works did not conduct any experiment with participants. Hexmoor et al. (2005) proposed a control method for Unmanned Aerial Vehicles (UAV) based on the setting of different UAV's parameters. The parameters are conformity (how quickly a UAV reacts to a human operator's command), sociability (how is a UAV influenced by nearby UAVs), commitment (how committed to a task a UAV is) and disposition (to what extent a UAV abandons a task). Their approach was tested in simulation only. Walter et al. (2006) also proposed a control method based on parameter settings. In this work, the authors allow human operators to change parameters associated to "virtual pheromones" (i.e., a virtual pheromone encodes data exchanged between robots) in order to influence a robot swarm (e.g., robots can be attracted or repelled by virtual pheromones). A graphical interface allows a human operator to change the virtual pheromones' parameters such as evaporation rate and propagation rate. Their approach was tested in simulation only. Goodrich et al. (2011) and Kira and Potter (2009) investigated the possibility of modifying parameters such as repulsion and attraction (to nearby robots) and the radius of a robot swarm. Kira and Potter (2009) validated their approach with six real robots. Ayanian et al. (2014) presented a control method that allows a human operator to move and scale a virtual prism that surrounds a robot swarm. The human operator moves and scales the virtual prism using a tablet interface. The results of the prism transformations are transferred to the robot swarm in order for the robots to move accordingly. The authors validated their approach with three real robots.

Controlling a robot swarm is important because swarms of robots are not fully autonomous—for instance, humans must have the possibility to guide a robot swarm to its task locations. Unfortunately, the literature on control methods in human-swarm interaction currently fails at identifying which control methods are the most appropriate. Future work should focus on comparing existing control methods and identifying whether they are appropriate in certain conditions more than in others.

2.1.2 Interaction Interfaces

Another research direction that received significant attention is the interfaces that allow human operators to interact with a robot swarm. These interfaces render it possible for a human operator to issue commands to and receive feedback from a robot swarm or a subgroup of a robot swarm. Of the 15 works presented below, only 4 conducted experiments with participants.

The majority of the works presented in this section address technical issues related to the interaction interfaces (e.g., how can a robot swarm decode gestures, how can a human select a group or a subgroup of robots). However, with the exception of Podevijn et al. (2013), none of the following works study the interaction interface's usability. Studying the interaction interface's usability is vital for three reasons. Firstly, it enables the implementation of interaction interfaces that can be used by human operators with effectiveness, efficiency and satisfaction (ISO, 2000). Secondly, it can help identify for what types of interaction an interface is better at. For instance, a gesture-based interface can be less efficient for *action commands* (e.g., selecting a particular group of robot in a robot swarm), while a gesture-based interface can be more efficient for *analog commands* (e.g., "go faster") (Podevijn et al., 2013). Finally, it can help decrease human psychological reactions such as the workload (Gerhardt-Powals, 1996).

Different types of interaction interfaces are studied in the literature and some works propose a combination of multiple interaction interfaces (i.e., multimodal human-swarm interaction interfaces). The interaction interfaces can be divided into five categories: gesture interfaces, face engagement interfaces, voice interfaces, haptic interfaces and graphical interfaces (both displayed on a computer screen or in a dedicated headset).

Giusti et al. (2012) presented a hand gesture-based interaction system. The goal of their work is to allow robots to decode hand gestures. Each robot decodes a human's hand gesture. Because of sensing noises and because the robots are placed in different positions, different robots do not necessarily decode the same hand gesture. After each robot individually decodes a hand gesture, a distributed consensus algorithm is used in order for all the robots to agree on the hand gesture. The authors tested their hand gesture-based interface on real robots but no experiment with participants was conducted. Nagi et al. (2015) developed a novel distributed consensus algorithm for all the robots of a swarm to agree on a hand gesture.

Nagi et al. (2012) extended Giusti et al. (2012) by focusing on a machine learning algorithm that allows robots to learn hand gestures directly from a human operator (instead of having learned hand gestures based on an offline large data set of hand gesture pictures).

In these last studies, the authors have validated their algorithms with real robots, but no experiment with participants was conducted. The following study conducted an experiment with participants.

Podevijn et al. (2013) presented a framework that enables direct communication between human operators and robot swarms. The authors use a gesture-based interaction interface for human operators to issue commands to a robot swarm. They evaluated their framework with an experiment conducted with 18 participants. The experiment was based on a resource allocation and guidance scenario. In this scenario, the participants had to move selected robot groups of different sizes to specific locations in an environment. The authors evaluated the time taken by the participants to complete the experiment and asked the participants to evaluate the gesture-based interaction with a usability questionnaire. The experiment was conducted in simulation but the authors validated their framework with real robots.

Researchers found that some interaction interfaces are more appropriate to a task than to another. For instance, face engagement (i.e., looking at a robot) is more appropriate to select a robot or a group of robots, while voice recognition is more appropriate to provide commands to the robots. In the following, we present the works that propose multimodal human-swarm interaction interfaces (i.e., a combination of different interaction interfaces).

Pourmehr et al. (2013) presented a multimodal human-swarm interaction interface to select flying robots with face engagement and by speaking the number of robots the operator wants to select. In order to increase the size of a selected group, the authors also proposed a method for adding robots to a selected group with face engagement. The study was validated with real robots but no experiment with participants was conducted.

In the same vein, Couture-Beil et al. (2010) used face engagement for selecting an individual robot from a group of robots and used gesture recognition for giving commands to the selected robot. Monajjemi et al. (2013) extended Couture-Beil et al. (2010) by allowing a human operator to select multiple robots by face engagement and gesture recognition. Nagi et al. (2014) also proposed a combination of face detection and hand gestures to select a single UAV or a group of UAVs. The feasibility of the four aforementioned studies have been tested on real robots but no experiment with participants was conducted.

Haas et al. (2011) proposed a multimodal human-swarm interaction interface in which a human operator can issue commands to a robot swarm via a speech-based interface (i.e., voice recognition) and via a touched-based interface (i.e., graphical interface). The commands issued by a human operator modify the robot swarm's environment (i.e., they place different types of objects in the robot swarm's environment). The authors conducted an experiment with 12 participants. In this experiment, each command had to be issued by speaking out loud the type of object (e.g., waypoints, targets, hot spots) to be placed in the robot swarm's environment and by touching on a screen the location in the environment where the object had to be placed. The participants could issue a command by first speaking out loud the type of object and then by touching the location on the screen, or by first touching the location on the screen and then speaking out loud the type of object. The results showed that the majority of the participants started by touching the location on the screen. The results also showed that the temporal binding (i.e., time difference between speaking out loud the object and touching the location) was shorter when the participants started by touching the location first. The experiment was conducted in simulation only.

Kapellmann-Zafra et al. (2016a) proposed a multimodal human-swarm interaction with a Google Glass device. In this work, a human operator wears a Google Glass device in order to control a single robot (i.e., a leader). The goal is for the human operator to help a robot swarm move an object in an environment. The Google Glass device allows a human operator to select a leader robot and to move the leader robot in the environment via touch and voice commands. The authors did not conduct any experiment with participants. They, however, used real robots in order to validate their approach.

Haptic devices are useful for human operators to receive feedback from a robot swarm while controlling a robot swarm. There are two researches that investigate the use of haptic devices to control a robot swarm and receive feedback from a robot swarm.

Nunnally et al. (2013) conducted an experiment with 32 participants. In this experiment, the participants had to guide a robot swarm to different locations in an environment. The participants were divided into two groups. In one group, haptic feedback and visual feedback (in a graphical interface) were provided to the participants during the experiment. In the other group, only visual feedback was provided to the participant. The feedback represented the aggregate repulsive force of each individual robot (i.e., repulsive force from obstacles or nearby robots). The authors used the environment coverage as a metric of performance (the more the swarm covered the environment, the better the performance). The authors also administered a questionnaire to the participants in order for the participants to subjectively report the utility of the haptic feedback. The results showed that participants with both types of feedback (i.e., haptic and visual) performed better. The questionnaire's results also showed that the participants evaluated the haptic feedback as being useful. The experiment was conducted in simulation only.

Setter et al. (2015) focused on the mapping between the manipulability (a number that provides information on whether it is easy to control a robot swarm or not) and the haptic force feedback (i.e., a function of manipulability providing force feedback). The authors investigated the difference between a linear mapping and four exponential mappings (with different values of a parameter that controls the force rate of change). The authors conducted an experiment with 10 participants. In this experiment, the participants had to use a haptic device to move a robot swarm to different locations in an environment. The participants moved the robot swarm by controlling a single robot (i.e., a leader). They performed the experiment with each of the five mapping functions. The authors calculated the distance between the leader and the locations and collected the time taken by the participants to conduct the experiment as performance metrics. The authors also evaluated the participants' psychological state by measuring the participants' workload. The results showed that the participants performed better and had a lower workload with the exponential mapping than with the linear mapping. The experiment was conducted with real robots.

The following works rely on a secondary display that provides a human operator with a real-time representation of both the environment and the robot swarms. In such approaches, therefore, the human operator does not interact with the real robots in their real environment, but with a modelled representation of both the robots and the environment. In order to create a modelling layer, it is necessary to collect telemetry data about the robots (i.e., their position and orientation) and data about the environment (i.e., size and obstacles). And to be useful for human-swarm interaction purposes, such data must be collected and modelled in real-time. Simulated human-swarm interaction approaches have used the omniscience afforded by robotic simulators to collect all of the relevant data. However, in the real-world, external tracking infrastructure would be required (e.g., GPS or external cameras). Such tracking infrastructure is often infeasible in the dynamic, a priori unknown environments for which swarm robotics systems are best suited.

Daily et al. (2003) proposed a human-swarm interaction interface in which human operators wear an optical see-through head-worn device which receives simple robots' messages. When the device receives these messages, it analyzes them and augments the environment with a visual representation of these messages. No experiment with participants was conducted. A similar system is used by Naghsh et al. (2008) who proposed an augmented reality interface for firefighters helped in their mission by a robot swarm. The firefighters' helmets are augmented by a visual device, giving them direction information to potential hazards. The authors did not conduct any experiment with participants. McLurkin et al. (2006) proposed a graphical user interface based on real-time strategy video games where the user controls an army of hundreds of individuals. Their graphical user interface displays virtual robots in a virtual environment and provides the user with extra debugging information for each individual robot (e.g., waypoints for individual robots, global positioning). The authors note that it can be difficult to display such a large amount of data while ensuring that the user still has a clear understanding of what is going on in the swarm. The authors did not conduct any experiment with participants but validated their approach with real robots.

2.1.3 Bandwidth Limitation and Neglect Benevolence

The effects of bandwidth limitation and neglect benevolence (i.e., when human operators must wait a robot swarm to stabilize in a specific state before issuing their commands, see Nagavalli et al. (2014) for a formal definition) have also been studied in human-swarm interaction. Brown and Goodrich (2014) proposed a classifier to distinguish between two types of robot swarm's motions—a torus motion (robots in a robot swarm are moving in a clockwise or anticlockwise rotation but the centroid of the robot swarm is not moving) and a flock motion (robots in a robot swarm are moving in the same direction). The classifier is able to detect these two types of motion under bandwidth limitation, that is, by considering only a subset of the robots in a robot swarm. These results would allow a human operator to have a clear understanding of a robot swarm motion, but no experiment with participants was conducted. Nunnally et al. (2012), however, conducted an experiment with 25 participants to investigate the effect of bandwidth limitation in the context of a foraging task (i.e., a task that requires a human operator to collect several targets in different locations of an environment). In this study, the positions of the robots a human operator was controlling (via a graphical interface) were not updated in real time. The authors considered three types bandwidth limitations—low bandwidth (only one robot's position was updated on the graphical interface at a time), medium bandwidth (only an estimate of the centroid and orientation of the robot swarm were updated on the graphical interface), and high bandwidth (all robots' positions were updated at the same time on the graphical interface). The number of targets collected was used as a performance metric. The results showed that, under low bandwidth, the participants detected less targets than under medium bandwidth. However, the authors did not find any statistical difference when their participants conducted the experiment under medium bandwidth and high bandwidth. These results suggest that it is not necessary for a human operator to visualize each and every robot's position. The experiment was conducted in simulation only.

Neglect benevolence is a concept that received attention in the literature. Walker et al. (2012) conducted an experiment with 21 participants. In this experiment, the participants had to control a robot swarm in the context of a foraging task. The results showed that when an operator provided the robot swarm with commands frequently (i.e., without waiting for the robot swarm to stabilize), the area covered by the robots was significantly lower than when the operator waited for the swarm to stabilize. The experiment was conducted in simulation only. Nagavalli et al. (2015) conducted an experiment in order to investigate if human operators can detect the optimal time to issue commands to a robot swarm. In this experiment, 44 participants had to detect the optimal time to issue a command in order for a robot swarm to move from a spatial configuration to another (e.g., to move from a circle configuration to a cross configuration). The results show that when participants received visual help (a line drawn between the current robot's position and the final robot's position), they were able to issue the commands at a better time than without visual help. The experiment was conducted in simulation only.

Walker et al. (2016a) built up their research on Nagavalli et al. (2015)'s results. Walker et al. (2016a) argue that it is vital for a human operator to firstly understand the dynamic of a robot swarm before being able to understand when is the best moment to issue a command. The authors conducted an experiment with 32 participants. In this experiment, the participants had to watch several videos. Each video showed a robot swarm executing a robot swarm behaviour (rendezvous, flocking or dispersion). Different level of noise was added to each robot swarm behaviour (i.e., some of the robots of the robot swarm were moving randomly). The goal of the participants was to recognize, under different level of noise, which robot swarm behaviour (rendezvous, flocking or dispersion) they were watching. The authors found that the rendezvous behaviour was the easiest robot swarm behaviour to recognize. The experiment was conducted in simulation only. Walker et al. (2016b) conducted a similar study to that of Walker et al. (2016a). In this study, the goal was to determine whether participants were able to predict the future robot swarm's state based on different types of visualizations—full (each robot's position and heading are displayed on a graphical interface), centroid/ellipse (an ellipse showing the boundary of the robot swarm and a cross located at the centroid of the robot swarm are displayed on a graphical interface) and two additional visualizations based on two different algorithms showing only a subset of the robot swarm. The robot swarm was performing either a rendezvous behaviour, a flocking behaviour or a dispersion behaviour. The goal of the participants was to draw the final shape of the robot swarm after 30 seconds. The results show that the dispersion behaviour was the behaviour that was predicted by the participants with higher accuracy. The authors also showed that the full and centroid/ellipse visualizations did not significantly change the performance of their participants. The experiment was conducted in simulation only. Manning et al. (2015) provided a discussion on different types of robot swarm visualizations.

2.1.4 Level of Automation

Researchers have also investigated to what extent a robot swarm needs the influence of a human operator, i.e., what is the robot swarm's level of automation. These research are based on previous works in levels of automation in human-computer interaction (Sheridan and Verplank, 1978). Sheridan and

Verplank (1978) identified 10 levels of automation, where the lowest level suggested that a human operator had to make all decisions, and the highest level suggested that the human operator had no decision to make (in between, the human operator has different levels of supervision).

In the context of human-swarm interaction, Cummings (2004) distinguished *levels of automation* and *levels of autonomy*. The levels of automation is the level of human decision making required by an autonomous system (e.g., computer or robot). The levels of autonomy are specific to multi-robot systems and robot swarm systems. These levels determine the intra-robot autonomy, i.e., the level of collaboration between robots in a robot group. The authors argue that the levels of automation and levels of autonomy are not necessarily inter-dependent. When the level of autonomy is at its minimum, the levels of automation is not impacted and can vary along the 10-level scale of automation. When the level of autonomy is at its maximum, however, the level of automation should be high—the human operator should be considered only as a supervisor and should not make low level decisions (e.g., decisions on a specific individual robot). Because in a robot swarm the inter-robot collaboration is relatively high (high level of autonomy), the authors state that swarm robotics systems tend to have a high level of automation. The author did not conduct any experiment with participants.

Coppin and Legras (2012) proposed the notion of autonomy spectrum. The autonomy spectrum enables different control mechanisms and different levels of automation for a single task (a task that has to be carried out by a robot swarm). It is, therefore, possible for a human operator to interact with a robot swarm at different levels of automation (e.g., a human operator can interact at Sheridan's level 1 (with full control) of automation or at level 8 (with supervisory control) of automation). The authors conducted an experiment with 23 participants in order to validate their approach. In this experiment, the participants had to help a robot swarm to patrol and to pursuit intruders. The participants were allowed to change the level of pheromone the robot swarm was influenced by. The authors concluded that the participants' performance was positive but that further research was required. The experiment was conducted in simulation.

Walker et al. (2013b) conducted an experiment to compare high level of automation and a low level of automation. The authors conducted an experiment in which 20 participants had to move a robot swarm in order to find different targets. In the high level of automation, the participants were only able to issue a command to disperse the robot swarm in the environment. In the low level of automation, the participants could select a subset of the robot swarm and assign this subset of robots to a goal direction. The authors used the number of targets found by the participants as a performance metric. The authors concluded that a balance between high level of automation and low level of automation provided the best results on their participants. The experiment was conducted in simulation only.

In the aforementioned works, the authors only studied to what extent different levels of automation affect the human operators' performances. We believe that an important aspect to consider is the psychological effect of different levels of automation. Future work in this research direction should determine whether, for instance, low-level of automation significantly increases the human operators' psychological responses by using a methodology similar to the one used in this thesis (i.e., a combination of objective psychophysiological measures and subjective psychological measures). These results could show that a human-swarm interaction system that allows a human operator to interact with a robot swarm with a high-level of automation is beneficial for the human operator's psychological state (e.g., decrease the level of stress).

2.1.5 Formal Verification

Currently there is no work that formally verifies the interaction between a human and a robot swarm. Formal verification is important, for instance, to guarantee performance Tabibian et al. (2014) and Sycara et al. (2015) conducted an experiment in which their participants had to issue one command from two possible alternatives ("deploy" or "rendezvous") to a robot swarm in order for the robot swarm to cover an environment. They use the results of this experiment in order to develop analytical models of the human operators interacting with a robot swarm. Analytical models of the human operators is a first step towards formal verification. The experiment was conducted in simulation only.

As in the majority of the studies we presented in this literature review, studies that investigate formal verification do not take into account human psychology. Guaranteeing performance in human-swarm interaction, for instance, should help decrease frustration as well as psychological responses. In future work in formal verification, human psychology could also be studied to guarantee a stable human operator's psychological state. The psychological measures used in this thesis (i.e., combination of objective psychophysiological and subjective self-reported measures) would be appropriate.

As shown in Table 2.1, only one of the nineteen experiments was conducted with real robots. Moreover, the majority of these experiments did not include any psychological measures. To the best of our knowledge, we are the first who conducted experiments in human-swarm interaction that focused on the psychological aspect of the human operator interacting with robot swarms composed of up to 24 real robots.

2.2 Reality Gap in Human-robot Interaction

In the current body of research in human-swarm interaction, more and more researchers validate their work by performing experiments with participants. However, as we have seen in Section 2.1, a large majority of the existing experiments are performed exclusively in simulation, with human operators interacting with simulated robots on a computer screen, see Table 2.1.

To the best of our knowledge, our research on the effect of the reality gap on the human psychological state presented in Chapter 4 is the first of its kind in the context of human-swarm interaction. However, the question of the psychological reaction differences when humans interact with a real robot or with a simulated robot has been already addressed in multiple studies in human-robot interaction, and more specifically in social robotics. In this section, we review these studies.

Our work is different from the works presented in this section in that the

Study	Experiment	Dependent Variables
Karavas and Artemiadis (2015)	simulation	reaction time, EEG activity
Bashyal and Venayagamoorthy (2008)	simulation	time-on-task (time to reach locations)
Walker et al. (2013a)	simulation	number of locations reached, workload
Walker et al. (2014)	simulation	number of locations reached
Kapellmann-Zafra et al. (2016b)	simulation	number of robots in an aggregated group
De la Croix and Egerstedt (2012)	simulation	mean LSQ, workload, self-assessment difficulty rating
Kolling et al. (2013)	simulation	number of information collected
Haas et al. (2011)	simulation	order of speech-based and touch-based commands, temporal binding
Podevijn et al. (2013)	simulation	time-on-task, usability questionnaire
Nunnally et al. (2013)	simulation	environment coverage, utility questionnaire
Setter et al. (2015)	real robots	distance to locations, time-on-task, workload
Nunnally et al. (2012)	simulation	number of targets collected
Nagavalli et al. (2015)	simulation	participant's time compared to the optimal time computer by an offline algorithm
Coppin and Legras (2012)	simulation	intruders progress, idleness of the robots
Walker et al. (2013b)	simulation	number of targets found
Walker et al. (2016a)	simulation	recognition of a robot swarm behaviour
Walker et al. (2016b)	simulation	accuracy of predicting a robot swarm state
Tabibian et al. (2014)	simulation	environment coverage
Sycara et al. (2015)	simulation	environment coverage

Table 2.1: Summary of research that contain an experiment conducted with participants.

interaction with a robot swarm is inherently different from the interaction with a social robot for two reasons. Firstly, because there is no social interaction between human beings and robot swarms. Secondly, because a swarm of robots is composed of a large number of relatively simple robots.

In social robotics, the goal of the robot designers is for the robot to socially interact with humans (Hegel et al., 2009). Most of the works that address the question of the human psychological reaction differences between the interaction with real robots and simulated robots in social robotics tend to show that humans prefer to interact with a real robot than with a simulated robot.

In the following research, all authors used a measure of “enjoyment”. The enjoyment is assessed either by a self-developed questionnaire, or by following the game flow model (a model developed to evaluate player enjoyment in games (Sweetser and Wyeth, 2005)). When a robot provides humans with help and instructions on a given task, Kidd and Breazeal (2004), Wainer et al. (2007) and Fasola and Matarić (2013) all reported that humans had a more enjoyable experience (assessed by a self-developed questionnaire) with a real robot compared to a simulated robot. Pereira et al. (2008) and Leite et al. (2008) also show that humans had a more enjoyable experience with a real robot than with a simulated robot when their participants were playing chess against the robot (both assessed by the game flow model). In Powers et al. (2007), the participants of the authors’ study conversed with a real robot and with a simulated robot about health habits. The participants reported to have a more enjoyable conversation with the real robot than with the simulated robot (assessed by a self-developed questionnaire). Wrobel et al. (2013) performed an experiment in which elder participants play a card game against a computer, a real robot and a simulated robot. In their results, their participants reported more joy playing against the computer than against the real robot or the simulated robot. However, their participants had a more enjoyable experience playing against the real robot than against the simulated robot (assessed by the game flow model).

Other research compared real robots and simulated robots based on different aspects of the interaction. For instance, in the context of an interaction with a kiosk providing information, participants seemed to pay more attention to a kiosk providing information with a physical hand than with a kiosk providing information on a screen (Ju and Sirkin, 2010). Kennedy et al.

(2015) showed that children payed more attention to a real robot teaching them new information than to a simulated robot teaching them new information. In the context of a game scenario between children and a robot, children payed more attention to the real robot than to the simulated one (Looije et al., 2012). In the same context, Jost et al. (2012) showed that children's response time was higher when playing with a real robot than with a simulated robot. Fujimura et al. (2010) also showed that participants had a higher response time when participants had to select an object a real robot was pointing at. Komatsu and Abe (2008) showed that a real robot had more persuasion than a simulated robot when the robot was trying to engage the participant in a second task while the participant was already performing a primary task.

A more comprehensive survey about the psychological differences when humans interact with real robots and simulated robots is provided in Li (2015). In addition to presenting studies that compare the interaction with real robots and simulated robots, this survey also presents studies that compare the interaction with real robots physically present in an environment and real robot displayed on a video screen.

The majority of the aforementioned studies show that, in the context of human-robot interaction, humans react significantly differently when they interact with a real robot than with a simulated one. Human operators have a higher level of arousal, perform better at a task, and are more engaged when they interact with a real robot, rather than with a simulated robot. Though there are inherent differences between human-robot interaction and human-swarm interaction (i.e., the number of robots human operators interact with is different, and there is no social interaction with robot swarms), in Chapter 4, we also show that human operators react differently when they interact with a real robot swarm than with a simulated one (they have higher levels of arousal, workload and skin conductance when interacting with a real robot swarm compared to with a simulated one). These reaction differences can have profound consequences on the use of these robotic systems (be it swarm or not). In the context of a high-risk task which involves human lives for example, it is vital to keep the human operator's engagement at the maximum level for saving a maximum of lives. The aforementioned results in the context of social robotics, and the results we present in Chapter 4 in the context of swarm robotics, suggest that interacting with simulated robots (or with a simulation-based representation of robots) could diminish the human operator's engagement.

2.3 Psychophysiological Measures in Human-swarm Interaction and Human-robot Interaction

In human-robot interaction, psychophysiological measures are already used to evaluate the psychological state of human beings interacting with a robot. As shown in the remaining of this section, different types of psychophysiological measures are used in human-robot interaction (e.g., heart rate, skin conductance level, deltoid muscle activity, ocular behaviour, skin temperature, respiratory activity, body movements, ...). In human-swarm interaction, however, psychophysiological measures were used in a single study before us (Karavas and Artemiadis, 2015). A reason why psychophysiological measures are not commonly used in human-swarm interaction is because using psychophysiological measures is resource and time consuming (it requires specific hardware, it takes time to place the sensors on the participants and to analyse the data).

In the researches we present in Chapter 4 and Chapter 5, we use two psychophysiological measures—heart rate and skin conductance level (formally defined in Chapter 3). We decided to use these two psychophysiological measures because they are simple to monitor and because they are the two more commonly psychophysiological measures used in the literature. The results we obtained with these two psychophysiological measures show that psychophysiology is a methodology that is appropriate to study human-swarm interaction. Future human-swarm interaction studies should, therefore, consider using psychophysiology to study the human psychological state. In the remaining of this section, we present the works in human-robot interaction that use psychophysiology to study the human psychological state during the interaction with a robot.

Psychophysiological measures have been used in studies on the interaction between humans and a robot manipulator arm. In Kulić and Croft (2003), the authors monitored the blood volume pressure, the skin conductance, the chest cavity expansion and contraction and the corrugator muscle activity (i.e., used to control the eye brows) of four participants in order to study the participants' attention and expectation during the interaction. In Kulić and Croft (2005, 2007b), the authors monitored the participants' skin conductance, heart rate and corrugator muscle activity during the interaction with a robot manipulator arm. The authors used a fuzzy inference engine in order to estimate the participants' emotional state to different motion types of the robot. In Kulić and Croft (2007a), the authors improved their emotional state estimator by using a hidden Markov model instead of a fuzzy inference engine. In all these studies, the participants were passive during the interaction with the robot manipulator arm. In Dehais et al. (2011), the authors evaluated the interaction between a human and a robotic arm in an active interaction scenario with an object hand-over task (i.e., a robotic arm gives an object to a human). The authors monitored the participants' skin conductance, deltoid muscle activity and ocular behaviour during the interaction with the robot arm moving at different speeds.

Researchers have also used psychophysiological measures in order to evaluate the use of companion robots. Aminuddin et al. (2016) have proposed an experiment in which they monitor skin conductance in order to examine whether the interaction with a cuddly companion robot is able to reduce stress in humans. In this experiment, the authors are interested in determining what aspects of the interaction with the cuddly companion robot reduce the stress (e.g., stroking the robot, talking to the robot). At the time of writing this manuscript, the results of the study are not yet available. Robinson et al. (2015) have studied the effect of the interaction with a cuddly companion robot on the blood pressure and heart rate of elderly persons. The authors conducted their experiments in a retirement home with 17 elderly persons (over 71 year old). Over a period of ten minutes, the participants had a cuddly companion robot on their lap. The authors reported that the blood pressure and the heart rate of their participants were significantly lower after having interacted with the cuddly companion robot. Lu and Hsu (2015) used psychophysiological measures to evaluate the use of a companion robot in the context of elderly care. They used heart rate, skin conductance and skin temperature to evaluate different types of companion robots (non-humanoid and humanoid robots). The psychophysiological responses of the participants seemed lower when they were interacting with a non-humanoid companion robot.

Research in robotic system for rehabilitation services also use psychophysiological measures. Tiberio et al. (2012) studied the affective response of older people suffering from Mild Cognitive Impairment (MCI) to a telepresence robot. In their experiments, two groups of elder participants were compared—one group being composed of healthy older adults and one group being composed of MCI older adults. Each participant had to interact with the experimenter directly, and then with the experimenter through a telepresence robot. The authors studied the participants' stress during the interaction by monitoring the participants' heart rate and heart rate variability. The results showed that the group composed of MCI older adults were more stressed during the interaction with the experimenter through the telepresence robot than during the interaction with the experimenter. Novak et al. (2010) have conducted an experiment based on a virtual rehabilitation task with a haptic robotic arm. The goal of their experiment was to determine which psychophysiological measures would be the more reliable to study the psychophysiological state of humans suffering from stroke. They monitored the participants' heart rate, skin conductance, skin temperature and respiration. Their results suggest that skin conductance is the psychophysiological measure that provides the more reliable information on the psychophysiological state of human suffering from stroke. Goljar et al. (2011) have performed a similar study to that of Novak et al. (2010). In their study, the authors compared a control group (healthy participants) to two groups of stroke participants (subacute stroke and chronic stroke). The participants had to reach and grasp a virtual moving object displayed on a screen. A virtual hand that followed the participants' hand motion was also displayed on the screen. A haptic robotic interface allowed the participants to feel the virtual objects in their hands. The authors monitored the participants' heart rate, heart rate variability, skin conductance, skin temperature and respiration. The authors showed that stroke groups had weaker psychophysiological responses than the control group. Shirzad and Van der Loos (2016) proposed a method for a robot to adapt the difficulty of a rehabilitation task to the engagement of a human being performing the physical rehabilitation task. Their method enables human beings to remain engaged in the rehabilitation task. The robot evaluated the human's level of engagement in order to adapt the difficulty of the rehabilitation task. The authors conducted an experiment in which they monitored the participants' skin temperature, skin conductance and respiration rate. The authors used these physiological responses in order to infer the participants' level of engagement during the rehabilitation task.

Studies in service robotics for healthcare (e.g., a robot that assists a nurse for medicine delivery) have also used psychophysiological measures. Zhang et al. (2010) have collected heart rate and skin conductance in an experiment in which they evaluated the effect of different service robot interfaces (i.e., different types of anthropomorphic features such as camera for eyes, human-life face, synthesized or digitized human voice). The authors have shown that the more anthropomorphic a service robot is, the greater the psychophysiological responses are. Swangnetr et al. (2010) have also studied the effect of different service robot interfaces and found similar results to those of Zhang et al. (2010). Kraft and Smart (2016) designed an experiment with a teleoperated robot used in health care. The authors studied whether their participants trusted more a teleoperated robot when the teleoperator (i.e., a human that controls the robots from another room) was visible to the participants. The authors collected the participants' skin conductance during the experiment. The participants had higher skin conductance values when the teleoperator was not visible to the participants. Swangnetr and Kaber (2013) have developed novel algorithms that allow a service robot to infer the human emotional state based on psychophysiological measures. In their study, the authors used neural networks in order to classify human emotional states based on heart rate and skin conductance measures. The authors concluded that their algorithms can be used by service robots to evaluate in real time the patients' emotional state. Evaluating the patients' emotional state in real time could allow service robots to adapt their behaviours to the patients in order to improve their healthcare.

Adapting the robots' behaviour to the human's psychophysiological state is well studied in the literature. In the early 2000s, the first frameworks for robots to monitor and adapt to the human psychophysiological state are proposed (Rani et al., 2002, 2004, Sarkar, 2002). Rani and Sarkar (2005) have conducted an experiment in which a robot adapts to the human anxiety. The human anxiety was evaluated based on psychophysiological measures (heart activity, skin conductance, skin temperature, corrugator muscle activity). The authors designed an exploration task in which a human operator and a robot had to work in close collaboration. In the case the robot detected anxiety in the human operator, the robot either raised an alarm, moved towards the human operator or engaged into a discussion with the human operator. In this study, the author pre-recorded the participants physiological activity during a cognitive task because it was not possible to always infer anxiety from the participants during the experiments. The pre-recording of the physiological activity was then provided to the robots during the experiments. In Rani et al. (2006), the authors presented a robot-based basketball experiment. In this experiment, a basketball hoop was attached to a robotic arm. During the experiment, the basketball hoop was constantly moving. The participants were asked to shoot multiple baskets into the hoop. The level of anxiety was monitored in order to adapt the robotic arm movements (the more the participants was anxious, the less the robot arm was moving). Munekata et al. (2015) have used skin conductance in order for a social robot to detect boredom during an interaction (the interaction consists in a simple discussion). The authors made the assumption that low skin conductance values is related to boredom. When the social robot detected that the participant it was interacting with was getting annoyed by the interaction (witnessed by a drop in the participant's skin conductance), the social robot changed its discussion topic in order to get the participant's interest back. Özcan et al. (2016) have proposed a novel type of companion robot for children suffering of autism spectrum. This companion robot is augmented with physiological sensors (heart rate, heart rate variability, skin conductance, skin temperature and body movements), allowing the companion robot to adapt its behaviour to the children psychophysiological state, for example, by providing the children with affective stimuli.

Psychophysiological measures have been used in other types of research in human-robot interaction. For instance, Harriott et al. (2015), Novak et al. (2015) and Teo et al. (2016) all used psychophysiological measures to determine the human operator's level of workload. Yang and Dorneich (2015) studied the effect of time delay (i.e., the time between a human operator issues a command to a robot and the time the robot responds to the command) on the human psychophysiological state in the context of a guidance scenario (a human operator had to guide a wheeled robot in a maze). Broadbent et al. (2011) compared the psychophysiological measures of human operators who had a human-like self-representation of a healthcare robot to those who did not have human-like self-representation of a healthcare robot. They showed that human beings who had a human-like representation of healthcare robot had higher psychophysiological reactions when they saw a real non human-like healthcare robot. Lazzeri et al. (2015) studied the psychophysiological reactions of human beings faced to different emotional stimuli of a robot.

2.4 Group Size Effect in Multi-robot Systems

In this section we review the related work to our study on the effect of the increasing robot group size (i.e., the number of robots in a swarm) on the human psychophysiological state (presented in Chapter 5). To the best of our knowledge, there is only a single study in human-swarm interaction that is related to our own research. In this section, we also review related studies but in the field of human multi-robot interaction.

In the literature, the distinction between multi-robot systems and swarm robotics systems is not well defined. In this thesis, our goal is not to formally define the distinction between these robotic systems. However, we make the assumption that, in contrast to a swarm robotics system, robots in a multi-robot system act as individual entities, i.e., they do not necessarily have to cooperate with one another in order to complete tasks successfully. Moreover, we also make the assumption that, in human multi-robot interaction, a human operator controls each robot individually. These assumptions are made based on the existing literature on multi-robot systems.

In the context of human-swarm interaction, Pendleton and Goodrich (2013) studied the human psychological state by studying the workload level of human operators controlling a varying number of robots. The authors studied the effect of the robot swarm size on the human workload level by collecting data using dedicated psychological questionnaires only (they did not collect any psychophysiological data). They performed an experiment in which participants guided 20, 50 and 100 simulated robots in a simulated environment to gather information about a series of locations in the environment. In their study, the authors used dedicated psychological questionnaires to study the human workload when the size of the robot swarm controlled by a human was increased. The results of their study showed that human workload does not depend on the number of robots when interacting with a swarm robotics system. In order to validate their methodology, the authors also performed preliminary tests on two real robots. However, an experiment based on real robots was not conducted. We believe that these results may only be conveying a limited message about the human psychological state as the experiment was limited to simulated robots, a simulated environment, and a questionnaire-based data collection methodology because:

- as shown in Chapter 4, experiments performed with simulated robots and environment suffer of the reality gap effect;
- participants might not always answer psychological questionnaires objectively, i.e., participants might not answer what they felt during the experiment, but they might answer what they believe the experimenters would like them to answer (Bethel et al., 2007).

The interaction with an increasing number of robots is also studied in the field of human multi-robot interaction. Humphrey et al. (2007) developed an experiment in which their participants had to search for and detect bombs with a group of six robots and a group of nine robots. The authors show that the participants' workload was significantly higher when they had to search for the bombs with nine robots compared to with six robots. Velagapudi et al. (2008) conducted search and rescue experiments with four, eight and twelve simulated robots. The results of their experiments revealed that the participants' workload increased when the number of robots increased. In an experiment that simulated a transportation task (i.e., a task in which robots must transport objects from one location to another), Adams (2009) also shows an increase of workload with an increase of the number of robots—the participants' workload was higher when they were doing the experiment with four robots compared to two robots and one robot.

2.5 Discussion and Conclusions

In this chapter, we provided an overview of the current literature in human-swarm interaction. As depicted by this overview, the field of human-swarm interaction is scattered—there are many different research directions and within each research direction, there are no consensus on the methods to use to provide human operators with an effective human-swarm interaction system (e.g., when is it better to control a single leader robot versus multiple leader robots). A reason why the field is scattered is because researchers in swarm robotics do not know today how swarm robotics systems are going to be used for real world applications. Swarm robotics is still confined in research laboratories in which only abstract tasks are taken into consideration. The absence of real world applications makes it difficult for human-swarm interaction researchers to even hypothesize how a human operator will interact with a swarm of robots.

The work we present in this thesis is not impacted by the absence of real world applications. In this thesis, we are not interested in studying how human operators can issue commands to or receive feedback from a robot swarm. We are interested to understand what are the effects of the interaction with a robot swarm on the human psychological state. Specifically, we study the impacts of the reality gap (i.e., the inherent discrepancy between reality and simulation) and of the group size (i.e., the number of robots that constitute a robot swarm) on the human psychological state. As we showed in Section 2.2, a large majority of human-swarm interaction experiments is conducted in simulation. We show in Chapter 4 that human beings react differently when they interact with real robot swarms than with simulated ones. We show in Chapter 5 that the number of robots significantly affects different psychological measures. These results highlight the importance of human psychology in the design of new interaction systems.

In order to study the effects of the interaction with a robot swarm on human psychology, we use psychophysiological measures. In this chapter, we saw that psychophysiological measures are already used in human-robot interaction. In human-robot interaction, many different psychophysiological measures are used, such as skin conductance level, heart rate, deltoid muscle activity, respiratory activity or skin temperature. As it was uncertain that psychophysiological measures would be appropriate to study human-swarm interaction, we decided to use the two more commonly used measures in the literature—skin conductance level and heart rate. The results we present in Chapter 4 and in Chapter 5 show that, in the context of human-swarm interaction, psychophysiological measures provide more information on the human psychological state than the information provided by traditional psychological measures collected via questionnaires.

In the following chapter, we present the materials and methods we used to conduct our research. We present the robotic hardware and the environments in which the experiments were conducted. We also present in more details the psychological measures used in our experiments (psychophysiological measures, affective state measures, workload and reaction time).

3
Materials and Methods

In this chapter, we present the hardware, the software and the data collection methods used to design, implement and conduct the experiments of Chapter 4 and Chapter 5. In Section 3.1, we introduce the e-puck robotic platform and the experimental environments that we utilized to conduct our experiments. We also describe the software used to design our experiments in simulation—on a computer screen and in a virtual reality environment. In Section 3.2, we present the measures collected on our participants during the experiments. We also present the tools used to collect these measures.

3.1 Robot and Environments

In this section, we first describe in Section 3.1.1 the e-puck robot platform—the robotic platform used in our experiments. In these experiments, the e-puck robots were placed in two types of environments that we present in

Section 3.1.2. As explained in more details in Chapter 4, the experiments conducted to study the reality gap effect require the participants to interact with simulated robots in a simulated environment. We describe in Section 3.1.3 the tools and software used to simulate both the e-puck robot and the environments.

3.1.1 Robotic Platform

We use the open-hardware e-puck robot platform designed at the École Polytechnique Fédérale de Lausanne, Switzerland. The e-puck robot was designed for educational and research purposes. It is particularly well suited for research in swarm robotics because it is small, extensible and relatively cheap. The e-puck robot is cylindric and has a diameter of 7 cm. It comes with a dsPIC microcontroller embedding a 16-bit processor running at 64 Mhz. Several sensors and actuators are available on the e-puck. In our experiments, we used only a subset of the available sensors and actuators. We refer the reader to Mondada et al. (2009) and Gutiérrez et al. (2008) for the complete list of sensors and actuators. The sensors and actuators used in our experiments are the following.

Proximity sensors The proximity sensors are used to detect and avoid obstacles such as walls and nearby robots. There are 8 infrared proximity sensors evenly placed all around the e-puck, allowing it to detect obstacles all around its chassis.

Ground sensors The ground sensors are used to detect the color of the ground (shades of gray) under the e-puck robot. The ground sensors are part of an extension board. This extension board is equipped with 3 infrared sensors (i.e., the ground sensors) that face the ground under the e-puck.

Wheel actuators The wheel actuators are used to control the e-puck motion. The wheels are controlled by two stepper motors. They enable the e-puck to move at a maximal linear speed of 12 cm/s.

RGB LEDs The RGB LEDs are used to communicate with other robots or to report simple information to a human operator. There are 3 RGB LEDs mounted on an extension board placed on the top of the e-puck.

Embedded Linux A Linux extension board extends the e-puck microcontroller. It augments the e-puck with an ARM Cortex-A8 processor running at 600 MHz and with 256 Mb of RAM. The Linux extension board also enables code controllers developed in simulation to be executed on the e-puck (Garattoni et al., 2015), see Section 3.1.3.

We show in Figure 3.1 an e-puck used in our research. The e-puck used in our research also has an additional camera extension board and a range and bearing board that we did not use in our experiments.

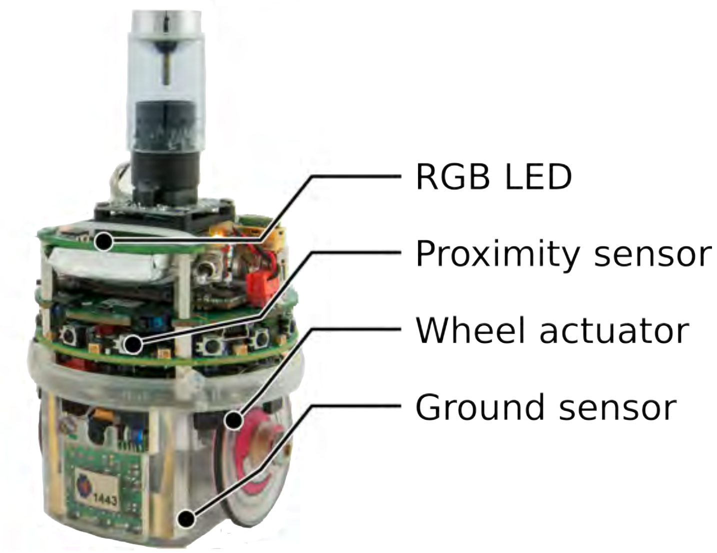


Figure 3.1: An e-puck robot. The RGB LEDs is used in our experiments to provide visual stimuli to the participants. The proximity sensors are used to detect and avoid walls and nearby robots. The wheel actuators are used to control the robot's motion. The ground sensors are used to detect the ground color under the e-puck (shades of gray).

3.1.2 Environments

All our experiments require a swarm of e-puck robots to be placed in a closed environment. We use two different environments. In the experiments presented in Chapter 4, we use a square environment of dimension 2 m × 2 m. This environment is shown in Figure 3.2 *(left)*. In the experiment presented in Chapter 5, we also use a square environment of dimension 2 m × 2 m. The environment is shown in Figure 3.2 *(right)*. This environment has four covered *hidden zones* of 25 cm width each. These four covered hidden zones are placed adjacent to the inner environment walls. A curtain is installed at the entrance of each hidden zone. These hidden zones and the curtains are used to hide the e-puck robots from the participants prior to the beginning of the experiments. In front of each hidden zone entrance, we also added a dark area of 175 cm × 20 cm. These dark areas are used to prevent the robots from accidentally returning back into a hidden zone during the experiment. We refer the reader to Chapter 5 for more information.

Figure 3.2: Environments used in our experiments with real robot swarms. Left: The environment used in the experiments presented in Chapter 4. Right: The environment used in the experiments presented in Chapter 5.

3.1.3 Simulation

In this section, we describe the software and the tools used to develop the simulation-based experiments presented in Chapter 4. We first present ARGoS, a swarm robotics simulator that allowed us to display our simulation-based experiments on a computer screen. Then, we present the virtual reality framework that allowed us to implement and display our simulation-based experiments in virtual reality.

ARGoS

The "Autonomous Robots Go Swarming" (ARGoS) simulator is a swarm robotics simulator developed within the EU Swarmanoid project (Dorigo et al., 2013, Pinciroli et al., 2012). As of writing the lines of this manuscript, the current version of the simulator is ARGoS 3. Its installation instructions are available on its website[1].

ARGoS is an open-source software developed in C++ running on Linux and Mac OSX. It was developed to meet 3 requirements of a swarm robotics simulator: *accuracy*, *efficiency* and *flexibility*. Accuracy measures the similarity of the simulation compared to the reality. Efficiency measures the run-time performance of the simulation. Flexibility refers to the ability of the simulator to be extended.

The flexibility of ARGoS makes it a highly modular software. Thanks to its modularity, ARGoS provides the user with the ability to customize the simulator. For instance, it is possible for a user to add new robots, to modify existing robots (e.g., modify their sensors and actuators), to add new visualization interfaces (ARGoS already supports an OpenGL interface, a PoV-Ray interface and a text-based interface), new physics engines (four physics engine are currently supported) and new media (i.e., inter-robot communication capabilities). The architecture of ARGoS also allows a user to run the same controller code in simulation and on real robots. The controller code calls the robots' sensors and actuators via the control interface which is the same for simulated robots or real robots. In Figure 3.3, we show the simulator's architecture diagram.

In our experiments, we used ARGoS 3, the supported version of the e-puck robot, a 2D-kinematics physics engine and the OpenGL visualization interface.

Virtual Reality

The ARGoS simulator does not allow simulation-based experiments to be executed on a virtual reality platform. Therefore, we developed the virtual

[1] http://www.argos-sim.info (last access: June 2016).

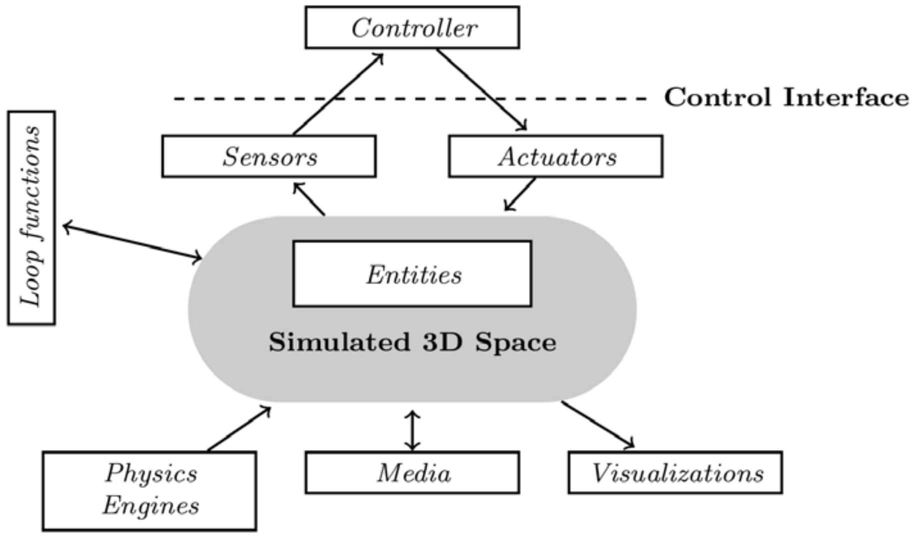


Figure 3.3: ARGoS Architecture. *Source:* from Pinciroli (2014); permission received from the author to use the figure.

reality-based experiments with additional software. We used Unity3D (Unity Technologies), SketchUp (Trimble) and the Google Cardboard VR plugin for Unity3D (Google Inc.). Unity3D is a game engine offering a programming framework in C#. We used Unity3D to develop the logic of the experiments (i.e., the robots' behaviour, see Chapter 4 for more details) and to render a 3D view of the environment. However, Unity3D does not provide any 3D model of the e-puck robot and does not support virtual reality by default (the cameras do not update their view based on the user head movements). In order to have a 3D model of the e-puck robot in our virtual reality experiments, we created a minimalistic model of the e-puck robot in SketchUp—a 3D modelling software. We then imported this 3D model into Unity3D. We show in Figure 3.4 the minimalistic 3D model of the e-puck robot. We enabled the virtual reality visualization by using the Google Cardboard VR plugin for Unity3D. This plugin makes it possible to track the user head movements to update the camera views accordingly.

We run our virtual reality-based experiments on a Google Nexus 5 device running Android 6.0. The Google Nexus 5 was inserted in a Google Cardboard (see Figure 3.5). The Google Cardboard is an inexpensive head set

Figure 3.4: A minimalistic simulated e-puck robot modeled in SketchUp.

created to turn a smartphone into a virtual reality device.

Figure 3.5: A Google Cardboard. *Source:* Evan Amos / Wikimedia Commons / Public Domain.

3.2 Data Collection

In this section, we present the measures collected on our participants during the experiments, and the tools that we used to collect these measures. In Section 3.2.1, we first introduce psychophysiology. Then we present the psychophysiological measures collected on our participants and the hardware used to collect these psychophysiological measures. We are also interested in investigating our participants self-reported affective state (i.e., in order to determine whether our participants are conscious of their psychophysiological state). In Section 3.2.2, we describe the affective state measure and we present the psychological questionnaire we used to collect this measure. Subsequently, we present two additional measures used in one of the two experiments of Chapter 4—the workload measure in Section 3.2.3 and the reaction-time measure in Section 3.2.4.

3.2.1 Psychophysiology

Psychophysiology aims at understanding and explaining human social and psychological behaviours through the study of physiological responses of the human body (Cacioppo et al., 2007). Physiological responses are activated by the autonomic nervous system. The autonomic nervous system is divided into the sympathetic nervous system and the parasympathetic nervous system. The sympathetic nervous system is considered to be responsible for the activation of the fight-or-flight physiological responses (i.e., physiological responses in case of stress). The parasympathetic nervous system, on the other hand, is considered to be responsible for maintaining physiological responses to a normal activity (i.e., the physiological responses at rest).

Psychophysiology is a field of research that started in the mid-1960s with the creation of the first scientific medium of communication on psychophysiology: the journal of *Psychophysiology*. In the first issue of the journal, psychophysiology was defined as *a method for bringing both physiological and psychological aspect of behavior into a single field of discourse* (Ax, 1964) in which *the dependent variable* (i.e., what is measured by the experimenter) *is a physiological measure and the independent variable* (i.e., what is manipulated by the experimenter) *a "behavioural"* [variable] (Stern, 1964). In today's experiments, however, physiological variables can also be considered as the independent variable in order to study behavioural modifications (Stern et al., 2001).

Psychophysiology has already been used in human-computer interaction to evaluate the interaction with technology (Dirican and Göktürk, 2011, Mandryk et al., 2006, Scheirer et al., 2002, Sykes and Brown, 2003, Ward and Marsden, 2003, Wastell and Newman, 1996), and in human-robot interaction to evaluate the interaction with robotic systems, see Chapter 2.

In human-robot interaction, psychophysiology is a method that can be used in three types of research (Bethel et al., 2007). In the first type of research, psychophysiology is used to detect and identify the human operator's emotion. In the second type of research, psychophysiology is used to modify, in real-time, a robot's behaviour with respect to psychophysiological responses of a human operator. Finally, psychophysiology is used in the third type of research in order to evaluate human operators' psychophysiological reactions to a robotic system. The research presented in this thesis lies in the third type of research.

Using psychophysiological measures to study the interaction with a robotic system has advantages and disadvantages (Bethel et al., 2007). Firstly, it offers a non-invasive method to the experimenter to study the participants' psychological states. Secondly, since it is very difficult to intentionally manipulate our own psychophysiological responses (e.g., intentionally decrease our own heart rate), psychophysiological measures are considered objective. However, psychophysiological measures can be difficult to acquire and interpret. The sensitivity of the sensors or a misplacement of the sensors can have important impacts on the physiological responses that are recorded. Moreover, the individual psychophysiological state of the participants at the moment of the experiment can be different from a participant to another (Bethel et al., 2007, Kidd and Breazeal, 2005).

Psychophysiological Measures

In the research presented in Chapter 4 and Chapter 5, we measure the responses of two different physiological activities: the *electrodermal activity* and the *cardiovascular activity*. The electrodermal activity and the cardiovascular activity are two common physiological activities used in the literature to study the human psychophysiological state.

The electrodermal activity is the electrical activity of the skin. The electrical activity of the skin can be used as a measure of the sympathetic nervous system. The electrical activity of the skin is related to the sweat glands of the skin. Eccrine is a special type of sweat glands mostly found in the soles of the feet and in the palm of the hands. Eccrine's primary role is to thermoregulate the body, but it has been found that eccrine is responsive to the sympathetic nervous system activity (while other types of sweat glands are more responsive to temperature variations) (Stern et al., 2001). When the eccrine sweat glands produce sweat in response to the sympathetic nervous system activity, the electrical resistance of the skin decreases and its reciprocal, the electrical conductance, increases. In our research, we study the electrodermal activity by monitoring the skin conductance level (SCL). The SCL is the tonic level of the skin conductivity (i.e., it is the level of skin conductance in absence of stimuli) and is measured in microsiemens (µS). In this thesis, we measure the change of SCL over time. An increase of the SCL is only due to an increase of the sympathetic nervous system activity. It is, therefore, a measure of choice to study the human fight-or-flight response. SCL has also been correlated to the affective state arousal (Boucsein, 2012) and to the cognitive workload (i.e., an increase of the SCL suggests an increase of the cognitive workload) (Kramer, 1991).

The cardiovascular activity is the activity of the cardiovascular system. The cardiovascular system is composed of the heart and of a blood distribution system. There are different measures of the cardiovascular activity. Blood pressure, heart rate and heart rate variability are the most common measures used in the literature (Cacioppo et al., 2007). The blood pressure is the pressure applied by the blood on the walls of the blood vessels. The blood pressure is at its maximal value when the heart contracts (i.e., systolic blood pressure) and at its minimum value between two heart beats (i.e., diastolic blood pressure). The heart rate is the number of heart beats per unit of time. It is usually measured in beats per minute (BPM). Heart rate has been found to be correlated to the cognitive workload (i.e., an increase of heart rate suggests an increase of workload) (Kramer, 1991). Blood pressure, and heart rate can vary due to either a variation of the sympathetic nervous system, a variation of the parasympathetic nervous system, or a combination of both (Cacioppo et al., 2007). They are, therefore, more difficult to analyse and interpret than the SCL. The heart rate variability is the time variability between heart beats. Heart rate variability can be analysed both with time domain methods (e.g., standard deviation and square root of the mean-squared difference of successive heart beat intervals) and frequency domain methods (e.g., frequency division of the heart rate variability into low frequencies and high frequencies). Frequency domain methods enable the analysis of the interaction between the sympathetic nervous system and the parasympathetic nervous system. It provides, therefore, more information than the blood pressure and heart rate. However, in order to accurately analyse the heart rate variability, it is recommended to obtain 5 minutes of recordings (Task Force of the European Society of Cardiology et al., 1996). In the research presented in Chapter 4 and Chapter 5, we use the heart rate to study the participants' cardiovascular activity.

Acquisition

We monitored our participants' physiological responses with a PowerLab 26T (ADInstruments Ltd.) data acquisition system augmented with a GSR Amp device (see Figure 3.6). In the experiments presented in Chapter 4 and Chapter 5, we connected via USB the PowerLab 26T to a laptop computer running Mac OSX Yosemite. We used the software LabChart 8 to record the physiological responses acquired by the PowerLab 26T data acquisition system.

We collected our participants' heart rate by monitoring their blood vol-

Figure 3.6: Left: the ADInstruments PowerLab 26T. Right: the ADInstruments GSR Amp device. Images used with permission of ADInstruments Ltd.

ume pulse (BVP). The blood volume pulse is the change in the pulsatile blood flow. We used an infrared photoelectric sensor (i.e., a photopletismograph) to measure the blood volume pulse of our participants. The blood volume pulse can be retrieved by the photopletismograph from the peripheral parts of the human body such as on the fingers. Photopletismographs are sensible to body motions and can, therefore, cause motion artifacts (Elgendi, 2012). The heart rate is computed based on the blood volume pulse. Firstly, we calculate the inter-beat interval. The inter-beat interval is the time in seconds between two peaks in the blood volume pulse. Then, we calculate the heart rate by dividing 60 by the inter-beat interval. For instance, if the inter-beat interval of an individual is 1 s, this individual's heart rate is 60 BPM. Figure 3.7 shows the blood volume pulse of a participant during a time window of 10 s.

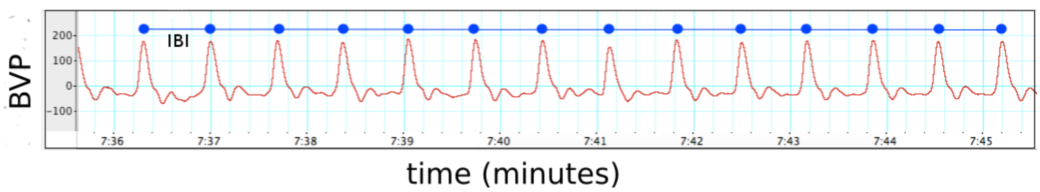


Figure 3.7: The graph of a participant's blood volume pulse during 10 seconds. The BVP does not have a standard unit. The x-axis is the time in minutes since the beginning of the recording. The time between two peaks (depicted with two dots connected with a line on the picture) is called the inter-beat interval (IBI). The participant's heart rate (the number of beats per minute) is computed by dividing 60 by the inter-beat interval. In this example, the mean heart rate of the participant during these 10 s is of 87 BPM.

To monitor the electrodermal activity of our participants, we used brightly polished stainless steel bipolar electrodes connected to the GSR Amp device. In order to monitor the skin conductance, the GSR Amp device applies a direct constant voltage between the bipolar electrodes. The constant voltage is small enough (i.e., 22 mV) to prevent the participants from feeling it. As the voltage is known and constant (22 mV), the GSR Amp device can measure the current between the bipolar electrodes. When the current is known, the GSR Amp device can calculate the conductance of the skin by applying Ohm's law (conductance is the current measured between the electrodes divided by the constant voltage applied by the GSR Amp device between the electrodes). Figure 3.8 shows the skin conductance of a participant during a time window of 10 s.

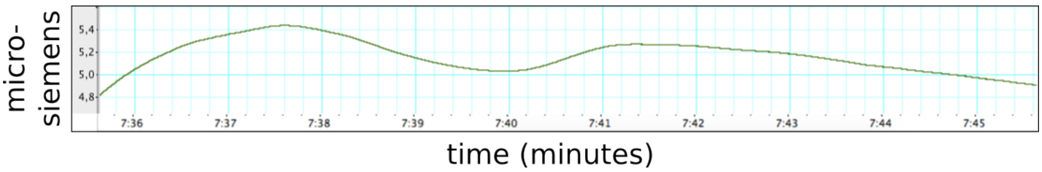


Figure 3.8: The graph of a participant's skin conductance during 10 s. The skin conductance's unit is the microsiemens (y-axis). The x-axis is the time in minutes since the beginning of the recording. The skin conductance is computed by measuring the current flowing between two electrodes and by dividing this current by a constant voltage applied between the electrodes. The average skin conductance level of this participant during these 10 s is of 5.17 µS.

3.2.2 Self-reported Affective State

In the research presented in Chapter 4 and Chapter 5, we also use self-reported measures to study our participants' affective state. We gather data from our participants using a dedicated psychological questionnaire to determine whether our participants are subjectively conscious of their psychophysiological reaction changes and whether these reaction changes are positive (i.e., our participants report to have a positive experience) or negative (i.e., our participants report to have a negative experience).

Affective State Measures

We measure our participants' affective state with two scales: *valence* and *arousal*. Valence is the cognitive judgement (i.e., pleasure or displeasure) of an evaluation such as the interaction with robots considered in our research. Higher valence values correspond to greater pleasure, while lower valence values correspond to a less pleasurable experience. The arousal scale assesses the mental alertness and the level of physical activity or level of excitation (Mehrabian, 1996) felt during an evaluation. Higher arousal values correspond to a higher excitation state, while lower arousal values correspond to a lower excitation state.

Acquisition

We used the Self-Assessment Manikin (SAM) questionnaire to collect our participants' self-reported affective state (i.e., valence and arousal) (Lang, 1980). In the version of the SAM questionnaire used in this study (see Figure 3.9), each scale is composed of 9 pictures. Each picture in the valence and arousal scale represents a value of valence or arousal, respectively. The left-most picture represents the lowest level, and the right-most picture represents the highest level of valence or arousal that can be chosen by the participant.

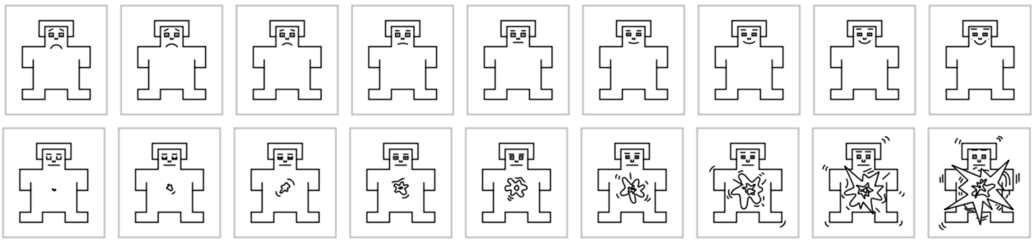


Figure 3.9: Self-Assessment Manikin scales. Top: The valence scale. The left-most picture corresponds to the lowest level of valence. The right-most picture corresponds to the highest level of valence. Bottom: The arousal scale. The left-most picture corresponds to the lowest level of arousal. The right-most picture corresponds to the highest level of arousal. These pictures are taken from and available at http://www.pxlab.de (last access: April 2016).

Each picture of each scale (i.e., valence and arousal) corresponds to a numerical score. Numerical scores vary from 1 to 9. In the valence scale, 1 corresponds to the lowest level of valence (i.e., displeasure is maximal) and 9 corresponds to the highest level of valence (i.e., pleasure is maximal). In the arousal scale, 1 corresponds to the lowest level of arousal (i.e., excitement is minimal) and 9 corresponds to the highest level of arousal (i.e., excitement is maximal).

3.2.3 Workload

Each task a human operator carries requires a specific amount of mental resources. However, the amount of mental resources of a human being is limited. The amount of mental resources required by a human operator is referred to as his workload. The more a task requires mental resources, the more the workload increases.

Acquisition

In this thesis, we acquire our participants' workload level via a subjective psychological questionnaire—the NASA Task Load Index Scale (NASA-TLX) questionnaire (Hart and Staveland, 1988). The NASA-TLX questionnaire is a multidimensional questionnaire developed to assess the workload experienced by participants during a task. The questionnaire is based on six bipolar scales. The six scales of the NASA-TLX questionnaire are:

- Mental demand: to what extent the task requires a mental activity.
- Physical demand: to what extent the task requires a physical activity.
- Temporal demand: to what extent the task is rapid.
- Performance: to what extent the participant successfully performs the task.
- Effort: to what extent the task requires mental and physical effort to perform the task.

- Frustration level: to what extent the participant feels frustrated (i.e., insecure, discouraged, stressed) during the task.

Each of these six scales is graded from “low” to “high”, except the performance scale which is graded from “good” to “poor”. These scales are associated to a numerical value between 0 and 100 for analysis purposes (a value of 0 corresponding to “low” or “good” and a value of 100 corresponding to “high” or “poor”).

The analysis of the NASA-TLX questionnaire provides a global workload number varying between 0 and 100. In the original version of the NASA-TLX questionnaire, each of the six scales had to be assigned with a weight. This weight was determined by the participants (i.e., participants had to compare each possible pair of dimensions and report, for each pair, the most relevant dimension for his or her understanding of workload). The workload number was then computed by multiplying each scale by its corresponding weight, by summing up the weighted scales and by dividing the summation by 15 (i.e., the number of pairs of dimensions).

More recently, various studies have eliminated the weighting process from the questionnaire—participants do not have to compare each pair of dimensions. In the version of the NASA-TLX questionnaire without the weighting process, the workload number is computed by dividing the summation of the raw scores (i.e., the values chosen by the participants) by 6. This modification results in a simpler and faster application of the questionnaire. When the NASA-TLX questionnaire is used in this form, it is usually referred to as the NASA Raw Task Load Index Scale (NASA-RTLX) (Hart, 2006). In the research presented in Chapter 4, we use the NASA-RTLX version of the questionnaire.

As stated in Section 3.2.1, workload has been found to be correlated to both heart rate and skin conductance (Kramer, 1991). However, recent research in human-robot interaction found inconsistencies in different studies were physiological measures were used to analyse participants' workload (Harriott et al., 2015).

3.2.4 Reaction Time

Reaction time is the time between the occurrence of a stimulus and a response to that stimulus (Gawron, 2008, MacKenzie, 2012). Reaction time differs depending on the type of stimulus (e.g., visual, auditory). For instance, for auditory stimuli, mean reaction time is of 160 ms and of 190 ms for visual stimuli (Welford, 1980).

We can differentiate three types of reaction time: i) *simple reaction time*, where there is only one stimulus to which a participant has to respond, ii) *choice reaction time*, where there are multiple stimuli to which a participant has to respond with different types of response possibilities and iii) *recognition reaction time*, where there are multiple stimuli but the participant has to respond only to a subset of them. In the research presented in Chapter 4, we use the first type of reaction time, where participants must respond to a visual stimulus.

Acquisition

In Chapter 4, we collect our participants' reaction time by measuring the time taken by our participants to press a button after a visual stimulus occurs. The button the participants have to press is connected to a computer via Bluetooth. A software computes the time, in milliseconds, between the visual stimulus and the moment a participant pressed the button. Figure 3.10 shows the Bluetooth device and the button the participants had to press.

Figure 3.10: Bluetooth game-pad used to collect the reaction time. The button the participants had to press after each stimulus is encircled in red in the picture (it is the right-most button on the device).

3.3 Conclusions

In this chapter, we presented the hardware, software and the psychological measures used to conduct our experiments. We presented the e-puck robotic platform—a robotic platform commonly used in swarm robotics. We presented the environments used to conduct the reality gap experiments (i.e., the study of the effect of the reality gap on the human psychology) and to conduct the group size experiments (i.e., to study the effect of the number of robots on human psychology).

We presented in details the psychological measures we used in our experiments and we explained how we collected these psychological measures. We presented the two psychophysiological measures (heart rate and skin conductance level), the affective state measures (valence and arousal), the workload measure (the amount of mental resource need to carry out a task) and the reaction time (the time needed to respond to a stimulus).

In the next two chapters, we use these measures to study the reality gap effect (Chapter 4) and the group size effect (Chapter 5). We show that the reality gap significantly affects the participants' skin conductance level, arousal, valence, workload and reaction time and that the number of robots affects the participants' skin conductance, heart rate and arousal.

4

Effect of the Reality Gap

As we have seen in Chapter 2, the majority of the experiments in human-swarm interaction are conducted exclusively in simulation, with human operators interacting with simulated robots on a computer screen. See Table 2.1 for a summary of the experiments with participants in human-swarm interaction research.

Simulation is a convenient choice for swarm roboticists for two reasons. Firstly, simulation allows experimental conditions to be replicated perfectly in different experimental runs. Secondly, gathering enough real robots to make a meaningful robot swarm is often prohibitively expensive in terms of both money and time. However, conducting experiments in simulation suffers from a potentially fundamental problem—the inherent discrepancy between simulation and the reality (the reality gap).

In this chapter, we study the effect of the reality gap on human psychology. Even though the effect of the reality gap is well studied in human-robot interaction (with companion robots or healthcare robots for instance, see Chapter 2), it is not clear what the effect of the reality gap is in human-swarm interaction.

More specifically, we expect the participants' psychological responses to be significantly higher when they interact with a real robot swarm compared to when they interact with a simulated robot swarm. Our goal is to confirm this expectation and, if confirmed, to study how can the effect be mitigated (i.e., how can we diminish the difference of psychological reactions between reality and simulation).

We present two experiments in which humans interact with a real robot swarm, with a simulated robot swarm displayed on a computer screen, and with a simulated robot swarm displayed in virtual reality (within a virtual reality headset). We show in Figure 4.1 an example of our experiments. In the first experiment, our participants are purely passive—they are asked to watch attentively a swarm of robots moving in an environment. In the second experiment, our participants are not purely passive anymore—in addition to watching attentively a swarm of robots moving in an environment, they are asked to push a button each time a robot from the robot swarm illuminates its LEDs.

(a) **(b)** **(c)**

Figure 4.1: Example of an experiment. **(a)** A participant interacts with a swarm consisting of 20 real robots. **(b)** A participant interacts with a simulated swarm of 20 robots displayed on a computer screen. **(c)** A participant is attached to a virtual reality head set and interacts with a simulated swarm of 20 robots. The participant shown in this figure is the author of the manuscript and did not take part in the experiment. The pictures shown in this figure were taken for illustration purposes.

The first experiment, in which the participants are purely passive, allows us to study the effect of the reality gap on human psychology without the risk that an interaction interface (e.g., joystick, keyboard, voice or gestures) is the strongest measurable reaction, drowning out the difference in reaction to the reality gap. In this experiment, we study the psychological impact of the reality gap using psychophysiological measures and self-reported measures. The second experiment allows us to study the effect of the reality gap on human psychology in a context where our participants are more involved in the interaction than in the first experiment. In this second experiment, the task assigned to the participants (pushing a button when a robot illuminates its LEDs) allows us to study the impact of the reality gap on the participants' psychophysiological state, workload level and reaction time.

In addition to studying the effect of the reality gap on the human psychology, we also investigate how to mitigate this effect in simulation. Indeed, it is not always possible for researchers to conduct real robot experiments, essentially because real robots are expensive and because real robots experiments are time consuming. This is the reason why we also asked our participants to perform our experiments in a virtual reality environment. If the psychological reactions of our participants interacting with a robot swarm in a virtual reality environment are significantly stronger than when they are interacting with a robot swarm displayed on a computer screen (i.e., stronger psychophysiological reactions and higher workload and reaction time), then we can reasonably argue that virtual reality mitigates the effect of the reality gap. In other words, we expect our participants' psychological reactions to be i) significantly higher when they interact with a real robot swarm compared to a simulated one (be it displayed on a computer screen or in a virtual reality environment), and ii) significantly higher when they interact with a simulated robot swarm displayed in a virtual reality environment compared to with a simulated robot swarm displayed on a computer screen.

This chapter is structured as follows. In the first section, we present the experiment in which our participants are purely passive. In Section 4.1.1 we present the hypotheses used to design the experimental scenario presented in Section 4.1.2. In Section 4.1.3 and in Section 4.1.4 we describe the robots' behaviour and the measures collected on our participants, respectively. We report some statistics about our participants in Section 4.1.5. We explain the experimental procedure in Section 4.1.6 and we report the results of the experiment in Section 4.1.7. In the second section of this chapter, we present the experiment in which our participants are not purely passive anymore. This second section follows the same structure as the first section. We discuss the results of these two experiments and conclude this chapter in Section 4.3.

4.1 Experiment 1: Psychological Responses in a Passive Supervision Task

In this experiment, our goal is to show that the human psychophysiological state is affected by the reality gap when the interaction with a robot swarm is purely passive.

4.1.1 Hypotheses

We based this experiment on the following hypotheses:

- The psychophysiological reactions of humans are stronger when they interact with a real robot swarm than when they interact with a simulated robot swarm.

- The psychophysiological reactions of humans are stronger when they interact with a simulated robot swarm displayed in virtual reality than when they interact with a simulated robot swarm displayed on a computer screen.

Confirming the first hypothesis would imply that human-swarm interaction experiments should be conducted with real robots instead of with simulated robots. Confirming the second hypothesis would imply that in order to mitigate the effect of the reality gap in simulation, it is better for a researcher to simulate a robot swarm in virtual reality because it provokes more realistic psychophysiological reactions compared to simulated robot swarms displayed on a computer screen.

4.1.2 Experimental Scenario

We designed an experimental scenario that allowed us to study the two hypotheses presented in the previous section. This experimental scenario is divided into three sessions. In each session, a participant must monitor (i.e., watch attentively) a swarm consisting of 20 robots. In the so-called *Real Robots* session, the robot swarm is composed of real robots (see Figure 4.2a). In the so-called *Screen Simulation* session, the robot swarm is simulated in 2D and displayed on a computer screen (the participants see the robots swarm from the top view, see Figure 4.2b). In the so-called *Virtual Reality* session, the robot swarm is simulated in 3D and displayed in a virtual reality environment (the participants wear the virtual reality headset presented in Section 3.1.3 and see the robot swarm as they would see it in reality, see Figure 4.2c). We decided to compare a 2D (top-view) simulation with the reality because this is how the majority of the experiments in human-swarm interaction display the robot swarm[1]. We decided to compare a 3D simulation (virtual reality) to the 2D simulation and to reality in order to test our second hypothesis. During the three sessions (i.e., *Real Robots*, *Screen Simulation*, *Virtual Reality*), the participants must monitor the robots for a period of 60 s.

The order a participant encounters the sessions is random. Prior to beginning an experiment, the experimenter assigns the sessions to the participant by randomly selecting an order (i.e., a random function was programmed to randomly permute the three sessions)[2].

4.1.3 Robot Behaviour

At the beginning of each session (i.e., *Real Robots*, *Screen Simulation*, *Virtual Reality*), 20 e-puck robots are randomly placed in the environment. When an

[1]Please note that this particular choice introduces the question of whether any differences observed between virtual reality and simulation are due to the different perspectives. As discussed in Chapter 6, this aspect will be considered in future work.

[2]A problem with this technique is that the order of the sessions is not guaranteed to be balanced. In the future, the experimenter should use a proper method to randomly assign the sessions to the participants, such as using a Latin Square.

(a) (b) (c)

Figure 4.2: Robots and environments for each of the three sessions. **(a)** View of the real robots and environment. The view is displayed from the participant's perspective. **(b)** Top view of the robots and of the environment simulated on a computer and displayed on a screen. **(c)** View of the robots and of the environment simulated in virtual reality. The view is displayed from the participant's perspective.

experiment starts, the 20 e-puck robots perform a *random walk with obstacle avoidance* behaviour for a period of 60 seconds. Each robot executes the two following steps: i) it drives straight with a constant velocity of $10\,\mathrm{cm\,s^{-1}}$, and ii) it changes its direction when it encounters either a robot or an obstacle in the direction of movement (i.e., it turns in place until the obstacle is no longer detected in the front part of its chassis). The robots perform random walk with obstacle avoidance in the environment described in Section 3.1.2 and depicted in Figure 4.2.

4.1.4 Measures

In this experiment, we are interested in understanding the effect of the reality gap on our participants' psychophysiological state. In order to study this effect, we measure our participants' heart rate and skin conductance level during the interaction with a robot swarm (see Section 3.2.1). We also use the SAM questionnaire to study the participants' subjective affective state. This questionnaire also allows us to determine whether our participants are conscious of their psychophysiological state changes (see Section 3.2.2).

4.1.5 Participants

We recruited 28 participants from the campus population of the Université Libre de Bruxelles. All participants were between 18 and 29 years old with an average age of 22.75 years old ($SD = 3.28$) and came from different faculties of the university (e.g., law, science, psychology, economics). None of the participants had a background in robotics. We excluded potential participants with cardiovascular problems. Our participants received an informed consent form explaining that they would have been filmed[3] during the experiment and that their physiological responses would be collected for research purposes only (see Annex B). At the end of the experiment, we offered a 7 € financial incentive for participation.

4.1.6 Experimental Procedure

We conducted our experiments in the robotics experiment room of the artificial intelligence laboratory at the Université Libre de Bruxelles (IRIDIA). Upon arrival, we explained to the participant that she or he was going to monitor, i.e., watch attentively, a swarm of robots with three different types of visualization interfaces (i.e., in reality with real robots, in simulation on a computer screen and in a virtual reality headset). We then showed to the participant the swarm of robots displayed in the three visualization interfaces. The participant was allowed to look at the real robots, at a computer screen displaying a top view of a swarm of robots, and was allowed to wear the virtual reality headset. Once the participant was familiar with the three visualization interfaces, we presented and explained how to answer the SAM questionnaire. In order for the participant to clearly understand the notions of valence and arousal, we orally associated to the valence scale and to the arousal scale two bipolar adjectives: unhappy-happy and unsatisfied-satisfied for valence and relaxed-stimulated and calm-excited for arousal. These bipolar adjectives are highly correlated with the notions of valence and arousal (Bradley and Lang, 1994). Then, we invited the participant to

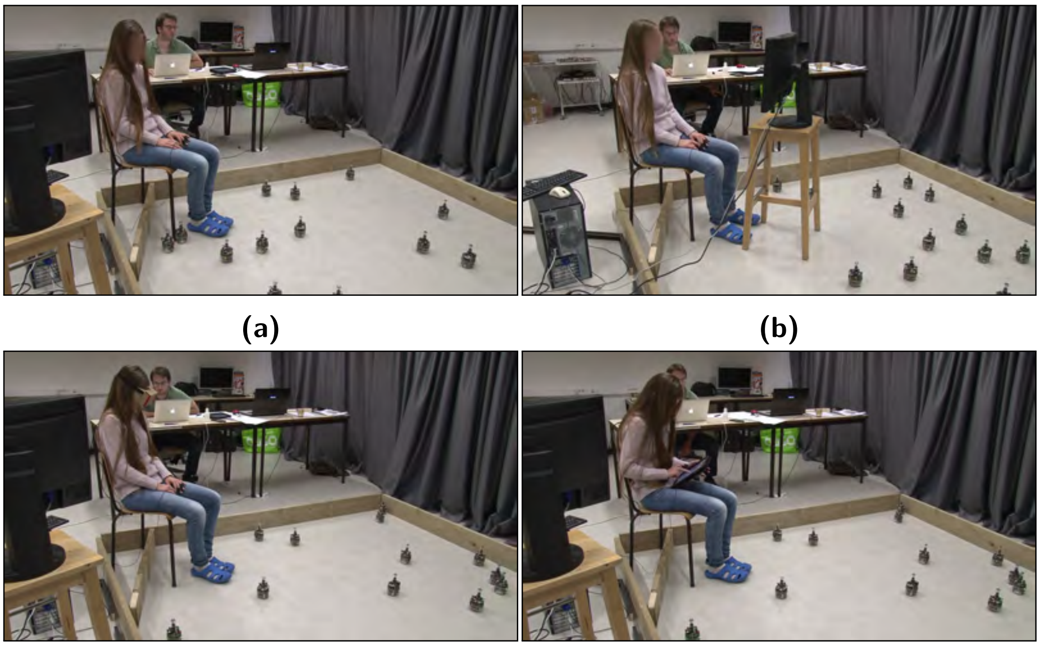
[3] We did not use the video recordings in our analysis. We recorded our experiments to have a visual history in case an experiment failed (e.g., robot crashes).

read and sign the consent form. In order for the physiological sensors to work properly, we asked the participant to wash their hands in clear water (i.e., with no soap) and to remain seated on a chair placed in a corner of the environment used for the *Real Robots* session (see Figure 3.2). We then attached the participant to the physiological sensors. We proceeded with a 5 minute rest period in order to collect the participant's physiological baseline (i.e., physiological responses at rest). After the 5 minute rest period, we started the first session. After each session, we asked the participant to answer the SAM questionnaire. Before starting the next session of the experiment, we collected the participant's baseline during an additional 3 minute rest period. This 3 minute rest period allowed the participant to get back to a normal physiological activity. During the whole duration of the experiment, the participant remained seated on the same chair. During the *Real Robots* session, the participant was immersed in the environment in which the robots were randomly moving. Prior to the *Screen Simulation* session, we placed a computer screen in front of the participant. Prior to the *Virtual Reality* session, we attached the virtual reality headset to the participant. See Figure 4.3 for an example of a participant during the experiment.

After the experiment ended, we detached the sensors from the participant and conducted a brief interview. During the interview, we explained to the participant the goal of the study and we answered questions the participant asked. We finished the experiment by thanking the participant and by giving the participant the 7 € incentive. The entire experiment's duration was approximately 30 minutes per participant.

4.1.7 Data Analysis and Results

Out of the 28 participants who took part to the experiment, we had to remove the physiological data (heart rate and skin conductance) of 5 participants due to sensor misplacement. We, however, kept the self-reported data (valence and arousal values reported by the SAM questionnaire) of these 5 participants. In the following of this section, therefore, we analyse the psychophysiological data of 23 participants (15 female and 8 male) and the



(a) (b) (c) (d)

Figure 4.3: A participant during the experiment. **(a)** A participant during the *Real Robots* session. **(b)** A participant during the *Screen Simulation* session. **(c)** A participant during the *Virtual Reality* session. **(d)** A participant completing the SAM questionnaire. These pictures are snapshots taken from a video recording of an experiment. The face of the participant is blurred for ethical reasons.

self-reported data of 28 participants (17 female and 11 male).

Physiological responses can vary between individuals. Therefore, it is difficult to compare the physiological responses of an individual with those of another. In order to compare the physiological responses between our participants, we conduct all our analyses on the difference between our participants' physiological responses at rest (i.e., the participants' baseline) and during the experiment.

In the following of this section, we first analyse the data and present the results of the reality gap effect. Then, we analyse our data in order to investigate whether or not some of the dependent variables (i.e., heart rate, skin conductance, arousal and valence) are pairwise correlated. Finally, we analyse our data in order to study potential gender effect (i.e., whether females and males differ in their results) and any potential session order effect (i.e., whether the participants become habituated to the experiment).

Reality Gap Effect

We analysed our data with the R software (R Core Team, 2015) by performing a repeated measures design analysis. Because the data was not normally distributed, we did not use the repeated measure ANOVA test (as the test assumes a normal distribution). Rather, we used a non-parametric Friedman test to analyse both the psychophysiological data and the self-reported data (i.e., the SAM questionnaire). The Friedman test is a rank-based test that does not make any assumption on the distribution of the data. In our case, the Friedman test's null hypothesis states that there are no differences between the three sessions *Real Robots*, *Screen Simulation* and *Virtual Reality*. The alternative hypothesis states that there are at least two sessions that are different. When the Friedman test is significant, we can reject the null hypothesis in favour of the alternative hypothesis. The alternative hypothesis, however, does not allow us to determine which sessions differ.

In order to determine which sessions significantly differ, we proceeded with a post-hoc Nemenyi test. The Nemenyi test compares the Friedman mean rank differences of two groups to a critical difference (CD) value. The critical difference value depends on the sample size (i.e., the number of participants), the number of data sets (i.e., the number of sessions) and on a value derived from the Studentized range statistic. Equation 4.1 shows the Nemenyi test for two groups i and j:

$$|R_i - R_j| > \underbrace{\frac{q_{\infty,k,\alpha}}{\sqrt{2}}\sqrt{\frac{k(k+1)}{6N}}}_{Critical\ Difference\ (CD)} \ , \tag{4.1}$$

where R_i and R_j are the Friedman mean ranks of group i and j and $q_{\infty,k,\alpha}$ is the Studentized upper quantile for an infinite level of freedom, for k number of groups and for the level of significance α (in this research, we set α to 0.05). N is the sample size. The Nemenyi post-hoc test already takes into account the Type-I error, that is, to declare the test significant while it is not. The Nemenyi post-hoc test can be easily interpreted graphically with a so-called *critical difference diagram* (Demšar, 2006). A critical difference diagram is composed of an horizontal axis on which the groups' mean ranks are plotted. A bar whose length corresponds to the critical difference value is also plotted on the diagram. Two groups are significantly different when the distance between the two groups on the horizontal axis is equal or greater to the length of the bar representing the critical difference value. When two groups are not significantly different, then these two groups are connected with a thick line. Table 4.1 summarises the results by showing the median and the Friedman mean ranks and the inference statistics of the Friedman tests (i.e., p-value and χ^2). We also show the boxplots of each measure in Figure 4.4.

Dependent Variable	n	RR	SS	VR	χ^2	p
Heart Rate	23	0.39 (1.87)	1.44 (2.13)	1.01 (2)	$\chi^2(2) = 0.78$	.67
SCL	23	4.91 (2.65)	0.77 (1.57)	1.54 (1.78)	$\chi^2(2) = 15.2$	< **.001**
Arousal	28	6 (2.45)	3 (1.39)	5 (2.16)	$\chi^2(2) = 19$	< **.001**
Valence	28	7 (2.73)	5 (1.55)	6 (1.71)	$\chi^2(2) = 24.87$	< **.001**

Table 4.1: Descriptive statistics of the psychophysiological data and of the self-reported data. RR stands for *Real Robots*, SS stands for *Screen Simulation* and VR stands for *Virtual Reality*. We report the median and the Friedman mean rank (in parentheses) of the three sessions (*Real Robots*, *Screen Simulation*, *Virtual Reality*). We also report the inference statistics of the Friedman test (i.e., χ^2 and p).

Heart rate and skin conductance level – The results of the Friedman test on the psychophysiological data do not show any main effect of the reality gap on our participants' heart rate ($\chi^2(2) = 0.78$, $p = .67$), see Figure 4.5a. The results show, however, a main effect of the reality gap on our participants' skin conductance level ($\chi^2(2) = 15.2$, $p < .001$). A Nemenyi post-hoc test confirms that our participants' skin conductance level was statistically significantly different between the *Virtual Reality* session and the *Real Robots* session ($CD = 0.69$, $p < .05$) and between the *Screen*

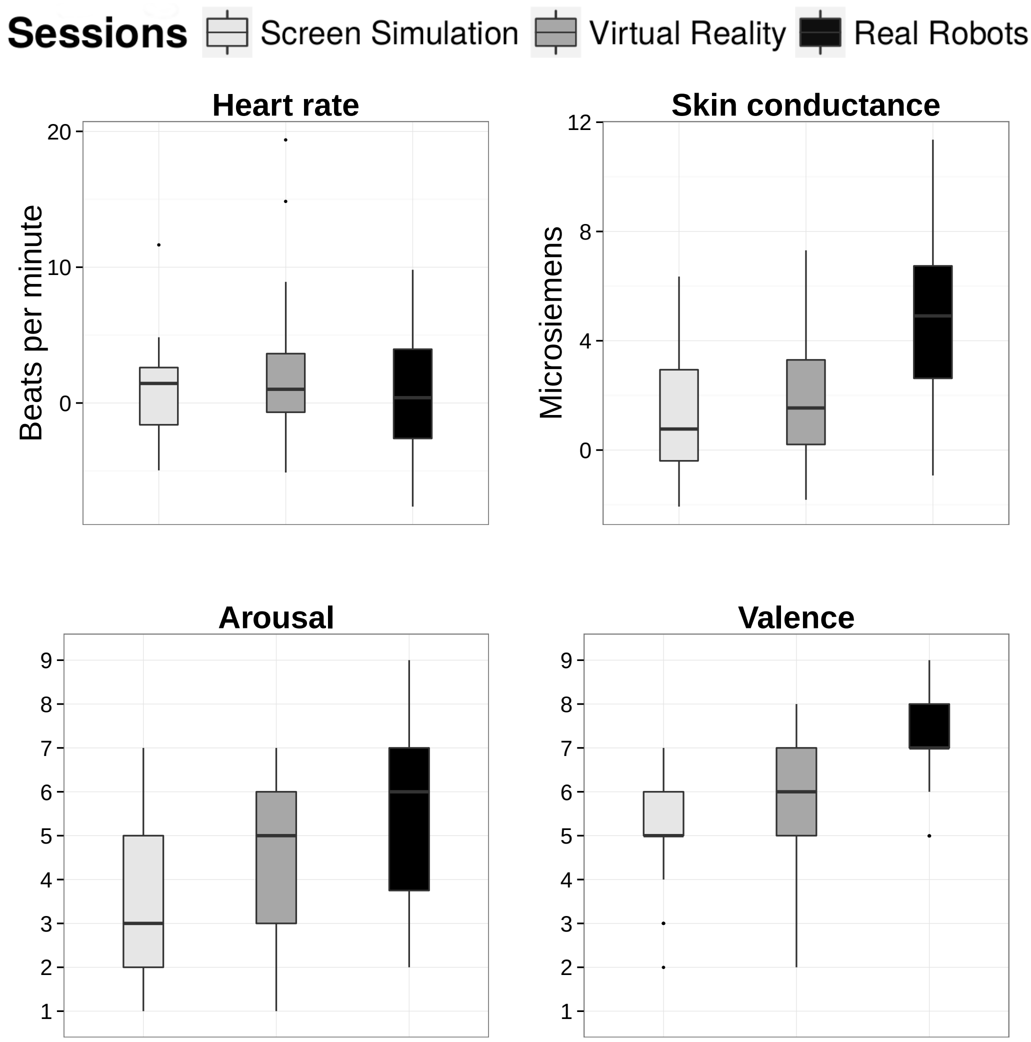
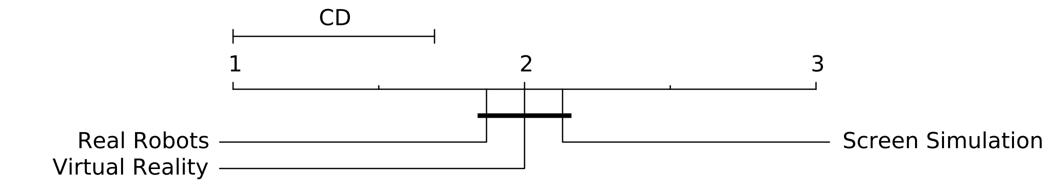


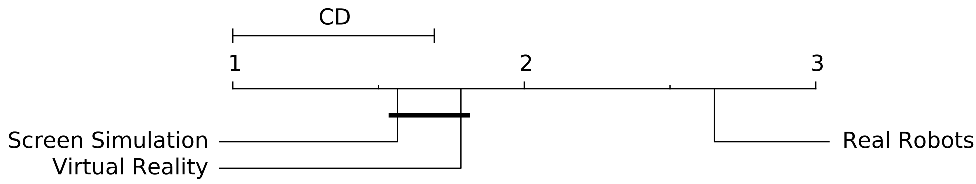
Figure 4.4: Boxplots showing the heart rate values (top left), skin conductance level values (top right), arousal values (bottom left) and valence values (bottom right) of all three sessions. The median value of each session is shown using the bold horizontal line in the box. Outliers are represented using dots.

Simulation session and the *Real Robots* session ($CD = 0.69$, $p < .001$). The Nemenyi post-hoc test does not show any statistically significant difference between the *Screen Simulation* session and the *Virtual Reality* session ($CD = 0.69$, $p = .74$), see Figure 4.5b.

SAM questionnaire – The results of the Friedman test on the self-



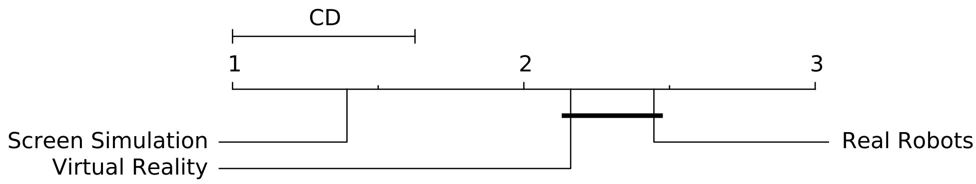
(a) Critical difference diagram showing pairwise statistical differences of the heart rate data. There are no statistical differences between any of the three sessions.



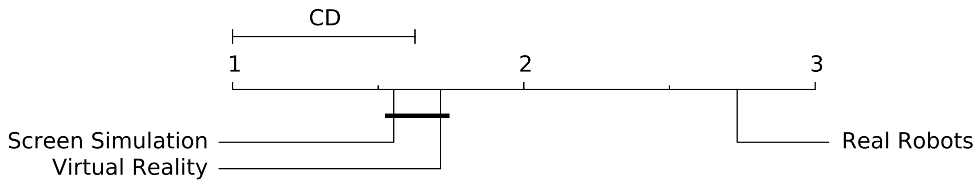
(b) Critical difference diagram showing pairwise statistical differences of the skin conductance level data—*Screen Simulation* and *Real Robots*, and *Virtual Reality* and *Real Robots* are statistically significantly different.

Figure 4.5: Critical difference diagrams showing pairwise statistical differences for **(a)** heart rate and **(b)** skin conductance level. When two sessions are connected with a thick line, the two sessions are not statistically different.

reported data also show a main effect of the reality gap on our participants' arousal ($\chi^2(2) = 19$, $p < .001$) and on our participants' valence ($\chi^2(2) = 24.87$, $p < .001$). A Nemenyi post-hoc test on the arousal data highlights a statistically significant difference between the *Screen Simulation* session and the *Real Robots* session ($CD = 0.63$, $p < .001$) and between the *Screen Simulation* session and the *Virtual Reality* session ($CD = 0.63$, $p < .05$). The Nemenyi post-hoc test does not show any statistically significant difference between the *Virtual Reality* session and the *Real Robots* session ($CD = 0.63$, $p = 0.53$), see Figure 4.6a. The Nemenyi post-hoc test shows that our participants' self-reported valence was statistically significantly different between the *Virtual Reality* session and the *Real Robots* session ($CD = 0.63$, $p < .001$) and between the *Screen Simulation* session and the *Real Robots* session ($CD = 0.63$, $p < .001$). The Nemenyi post-hoc test does not show any statistically significant difference between the *Screen Simulation* session and the *Virtual Reality* session ($CD = 0.63$, $p = .8$), see Figure 4.6b.



(a) Critical difference diagram showing pairwise statistical differences of the arousal data—*Screen Simulation* and *Virtual Reality*, as well as *Screen Simulation* and *Real Robots* are statistically significantly different.



(b) Critical difference diagram showing pairwise statistical differences of the valence data—*Screen Simulation* and *Real Robots*, as well as *Virtual Reality* and *Real Robots* are statistically significantly different.

Figure 4.6: Critical difference diagrams showing pairwise statistical differences for **(a)** arousal and **(b)** valence.

Correlations, Gender Effect and Order Effect

In order to calculate a correlation between the psychophysiological data and the self-reported data (e.g., correlation between skin conductance and arousal) we only took into account the self-reported data of the participants whose psychophysiological data had not been rejected (due to sensor misplacement). For the correlation test between arousal and valence we took the 28 participant data points. We did not find any correlation within each of the three sessions (i.e., there was no correlation for any pairwise dependent variable within the *Real Robots* session nor the *Screen Simulation* session nor the *Virtual Reality* session). We, therefore, investigated whether there was some correlation when the data of each condition was pooled together (e.g., we aggregated skin conductance values from the three sessions). Regarding correlation between psychophysiological data and self-reported data, we found a correlation between skin conductance and valence ($\rho = .42$, $p < .001$) and a weak correlation between skin conductance and arousal ($\rho = .253$, $p = .03$). There was no correlation between heart rate and valence and between heart rate and arousal. Concerning the self-reported data, we found a correlation between arousal and valence ($\rho = .32$, $p = .002$). We did not find any correlation between heart rate and skin conductance.

We analysed the gender effect by splitting into two groups the males' and females' results of each dependent variable (i.e., heart rate, skin conductance, arousal and valence) for each condition (i.e., *Real Robots*, *Screen Simulation*, *Virtual Reality*). We compared these two groups with a Wilcoxon rank-sum test. We did not find any statistically significant difference between males and females in any condition, for any dependent variable.

We studied the session order effect as follows. For each condition and for each dependent variable, we separated into three groups the results of the participants who encountered the session first, second or third, respectively. We compared the three groups with a Kruskall-Wallis test—a non-parametric test similar to a Friedman test but for independent groups. We did not find any statistically significant difference among the three groups in any session, for any dependent variable, suggesting that the session order had no significant effect on our participants.

4.2 Experiment 2: Psychological Responses in a Supervision Task with Visual Stimuli

In the previous experiment, we have seen that the reality gap affects the human psychophysiological state when the interaction with a robot swarm is purely passive. In this section, we present an experiment that allows us to show that the human psychophysiological state, as well as the human workload and reaction time are also affected when the interaction is not purely passive.

4.2.1 Hypotheses

We based this experiment on 2 hypotheses:

1. The psychophysiological reactions, workload and reaction time of humans are higher when they interact with a real robot swarm than with

a simulated one.

2. The psychophysiological reactions, workload and reaction time of humans are higher when they interact with a simulated robot swarm displayed in a virtual reality environment than with a simulated robot swarm displayed on a computer screen.

Confirming the first hypothesis would allow us to show that, not only the reality gap has an effect on the human psychophysiological state, but equally importantly, that it has also an effect on the human workload and reaction time. Confirming the second hypothesis would complete the results of experiment 1. In experiment 1, we showed that our participants' arousal was significantly higher in virtual reality than with the computer screen. By showing that the psychophysiological reactions, workload and reaction time are higher when humans interact with a robot swarm in a virtual reality environment rather than with a robot swarm displayed on a computer screen, we would strengthen our confidence that virtual reality can mitigate the effect of the reality gap in simulation.

4.2.2 Experimental Scenario

In order to test our 2 hypotheses, we designed an experimental scenario similar to that of experiment 1. In the experimental scenario used in experiment 1, our participants' task was to monitor a swarm consisting of 20 robots during a period of 60 s. The difference with the experimental scenario used in experiment 1 is that, in this experimental scenario, our participants have to press a button each time a robot illuminates its LEDs in red.

As in experiment 1, this experimental scenario is divided into the three sessions *Real Robots*, *Screen Simulation* and *Virtual Reality*, that our participants encounter randomly.

4.2.3 Measures

In this experiment, in addition to studying the effect of the reality gap on the psychophysiological state (by monitoring our participants' heart rate and skin conductance level and by asking them to answer the SAM questionnaire, see Section 3.2.1 and Section 3.2.2) we also study the effect of the reality gap on the workload and reaction time. We study the effect of the reality gap on the workload using the NASA-RTLX questionnaire (see Section 3.2.3). We study the effect of the reality gap on the reaction time by measuring the time our participants take to press a button after the appearance of a visual stimulus coming from a robot, that is, when a robot of the swarm illuminates its LEDs in red (see Section 3.2.4).

4.2.4 Robot Behaviour

As in experiment 1, the 20 robots perform a random walk with obstacle avoidance behaviour over 60 s. In this experiment, the robots also provide the participants with visual stimuli: randomly, one robot at a time illuminates its LEDs in red. The probability for a robot to illuminate its LEDs in red is computed by a software running on an external computer (there is a TCP communication link between the software and each robot). The software computes this probability as follows. Every 100 milliseconds, with a probability of 0.02, the software randomly chooses a robot's identification number (each robot has a unique identification number). With a probability of 0.98, the software does not choose any robot's identification number. When the software selects an identification number, it sends a signal (i.e., a message via the TCP communication link) to the robot associated to that identification number. When a robot receives a signal, it illuminates its LEDs in red for 2 seconds. When the software chooses an identification number, it also makes sure to wait 2 additional seconds in order to prevent two robots from being illuminated at the same time.

4.2.5 Participants

In this experiment, we recruited 37 participants. These participants came from the campus population of the Université Libre de Bruxelles. No participant had a robotic background and none of them had participated to the first experiment. They came from various faculties of the university. They were between 17 and 30 years old with an average age of 23.2 years old ($SD = 3.54$). People with current or anterior cardiovascular problems could not participate to the experiment. Our participants had to read and sign an informed consent form explaining that we monitored their physiological activity during the experiment (see Annex B). We offered a 7 € financial incentive for their participation.

4.2.6 Experimental Procedure

The experimental procedure is similar to that of experiment 1, described in Section 4.1.6. The experiments took place in IRIDIA, the artificial intelligence laboratory of the Université Libre de Bruxelles. We started the experiment by explaining to the participant the supervision task (i.e., watch a swarm of robots attentively and press a button each time a robot in the swarm illuminates its LEDs in red). Then, we showed to the participant the three visualization interfaces used in each session (the real robots in the real environment, the simulated robots displayed on a computer screen and the simulated robots displayed in a virtual reality headset). We allowed the participant to carefully look at the robots in each visualization interface in order to get familiarised with each of them. Once familiarised with the three visualization interfaces, we explained how to answer the SAM and the NASA-RTLX questionnaires. After the participant signed the consent form, we asked the participant to wash their hands in clear water (i.e., without soap) and to take a seat on a chair placed in a corner of the environment used in the *Real Robots* session. The participant remained seated on the chair during the whole duration of the three sessions. Once seated, we attached the physiological sensors to the participant's non-dominant hand. Prior to the first session, we collected the participant's baseline (i.e., physiological responses at rest) during 5 minutes. After these 5 minutes, we proceeded with the first session. After the first session, we administrated the SAM and the NASA-RTLX questionnaires to the participant. We pursued by collecting the participant's baseline during 3 minutes in order for the participant to get back to their baseline physiological activity. We followed the same procedure for the second and third session. After the experiment, we explained the goal of the experiment to the participant and we answered potential questions. The whole experiment's duration was approximately 30 minutes per participant.

4.2.7 Data Analysis and Results

Due to the fact that our participants had to press a button during the experiment, the heart rate data became too noisy to be usefully analysed (the pulse transducer sensor is extremely sensible to small movements). We, therefore, decided not to analyse our participants' heart rate data. Out of the 37 participants, we had to remove the skin conductance data of 6 participants due to sensor misplacement. We also removed the SAM questionnaire data and the NASA-RTLX questionnaire data of 3 participants due to an error of the experimenter in the administration of the questionnaires. We finally removed the reaction time data of 4 participants due to a hardware problem with the button. We performed, therefore, our statistical analyses on 31 skin conductance data (17 female and 14 male), 34 SAM and NASA-RTLX questionnaire data (19 female and 15 male) and on 33 reaction time data (19 female and 14 male). The analysis is conducted on the difference between the participants' physiological responses at rest and during the experiment.

In the following of the section, we first present our analysis and results on the reality gap effect. Then, we present the results of a correlation analysis (in order to determine whether any dependent variables were pairwise correlated) and on potential gender and session order effects.

Reality Gap Effect

We analysed our data with the R software by performing a repeated measure design analysis. We used the non-parametric Friedman test to determine whether the reality gap has a significant effect on our participants' measures (i.e., skin conductance, arousal, valence, NASA-TLX and reaction time). In case of statistical significance of the Friedman test, we performed a Nemenyi post-hoc test to evaluate the significance of the differences between sessions.

In Table 4.2, we summarise the results by giving the median and the Friedman's mean ranks and the inference statistics of the Friedman tests (i.e., p-values and χ^2). We also report the boxplots of each measure in Figure 4.7.

Dependent Variable	n	RR	SS	VR	χ^2	p
SCL	31	4.54 (2.54)	1.47 (1.71)	1.93 (1.74)	$\chi^2(2) = 14$	$<$**.001**
Arousal	34	5 (2.31)	3 (1.33)	5 (2.35)	$\chi^2(2) = 25.35$	$<$ **.001**
Valence	34	7 (2.14)	7 (1.9)	7 (1.9)	$\chi^2(2) = 1.38$	1
NASA-RTLX	34	27.5 (2.29)	18.33 (1.23)	35 (2.47)	$\chi^2(2) = 32.25$	$<$ **.001**
Reaction Time	33	0.87 (2)	0.72 (1.39)	1.02 (2.6)	$\chi^2(2) = 24.24$	$<$ **.001**

Table 4.2: Descriptive statistics of the psychophysiological data, of the self-reported data and of the reaction time data. We report the median and the Friedman's mean rank (in parentheses) of the three sessions (*Real Robots*, *Screen Simulation*, *Virtual Reality*). We also report the inference statistics of the Friedman test (i.e., χ^2 and p value).

Skin conductance level – The results of the Friedman test on the skin conductance level show a main effect of the reality gap on our participants ($\chi^2(2) = 14$, $p < .001$). A Nemenyi post-hoc test on the skin conductance level data highlights a statistically significant difference between the *Virtual Reality* session and the *Real Robots* session ($CD = 0.59$, $p < .05$) and between the *Screen Simulation* session and the *Real Robots* session ($CD = 0.59$, $p < .05$). The Nemenyi post-hoc test does not show any statistically significant difference between the *Screen Simulation* session and the *Virtual Reality* session ($CD = 0.59$, $p = .9$), see Figure 4.8.

SAM questionnaire – The Friedman test on the SAM questionnaire data reports a main effect of the reality gap on our participants' arousal ($\chi^2(2) = 25.35$, $p < .001$). The Nemenyi post-hoc test on the arousal data shows that there is a statistically significant difference between the *Screen Simulation* session and the *Real Robots* session ($CD = 0.57$, $p < .001$), between the *Screen Simulation* session and the *Virtual Reality* session ($CD =$

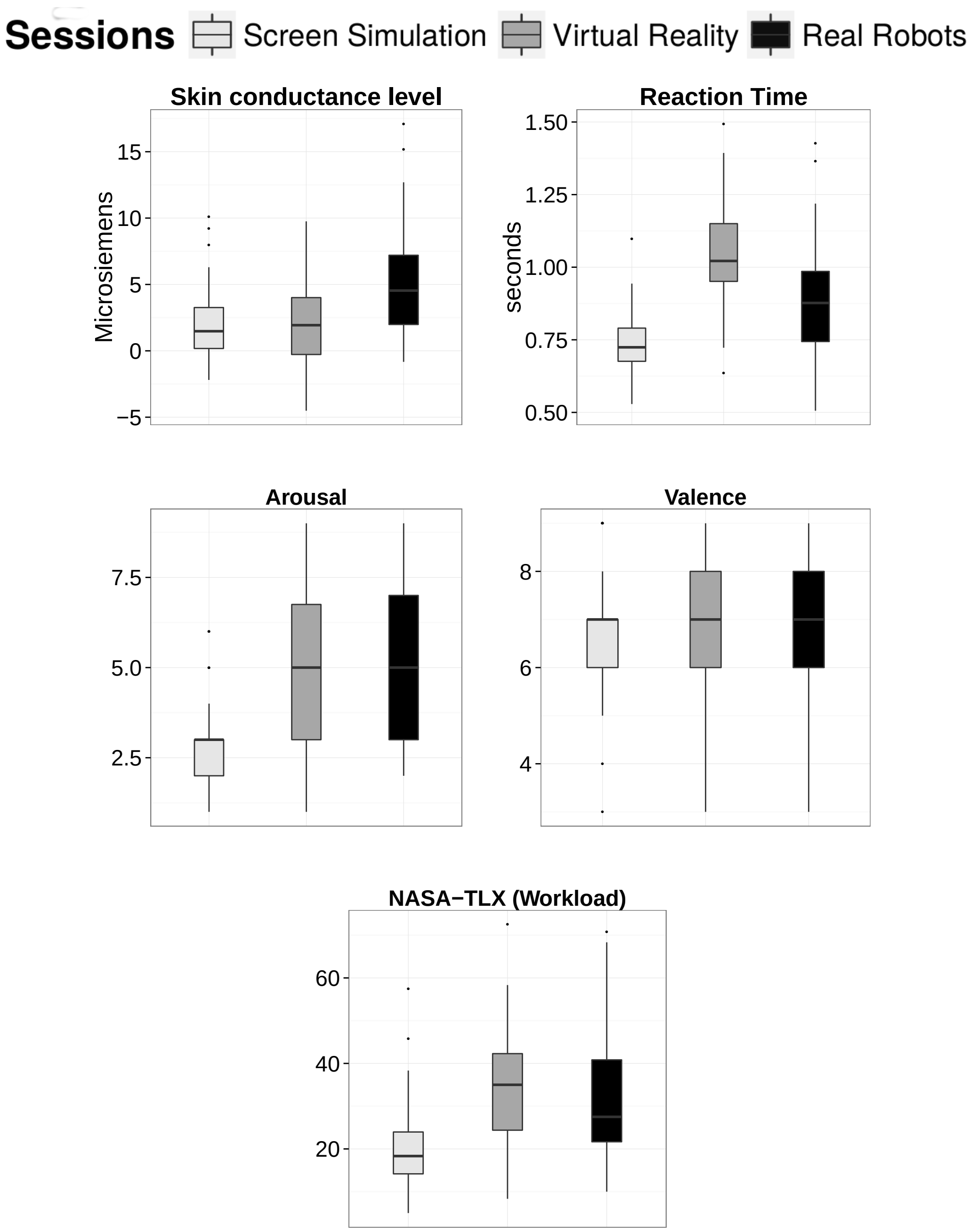


Figure 4.7: Boxplots showing the skin conductance level (top left), reaction time (top right), arousal values (middle left), valence values (middle right) and workload (bottom) of all three sessions. The median value of each session is shown using the bold horizontal line in the box. Outliers are represented using dots.

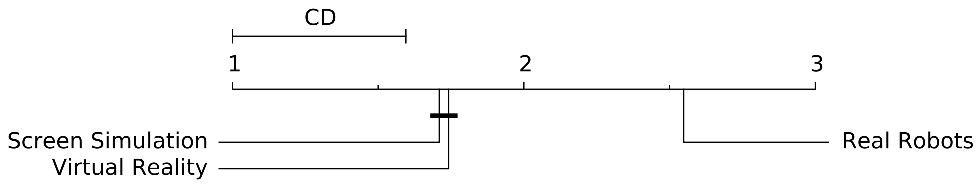
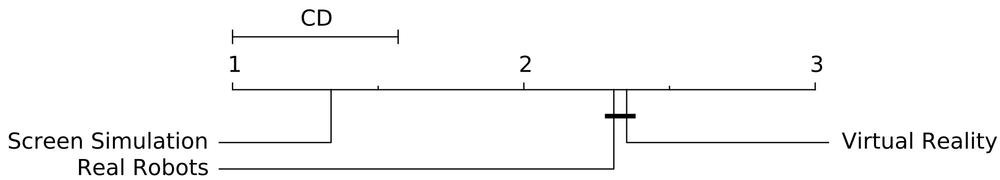
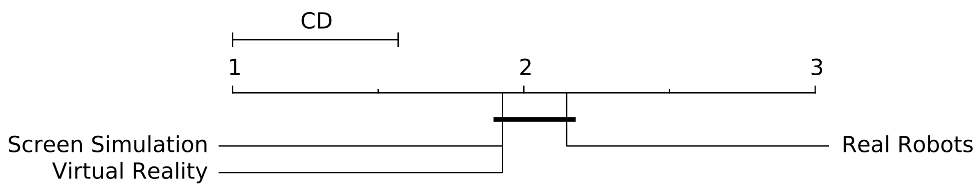


Figure 4.8: Critical difference diagram showing pairwise statistical differences of the skin conductance level data—*Screen Simulation* and *Virtual Reality* are statistically significantly different, as well as *Virtual Reality* and *Real Robots*.

0.57, $p < .001$). The Nemenyi post-hoc test does not show any statistically significant difference between the *Virtual Reality* session and the *Real Robots* session ($CD = 0.57$, $p = .9$), see Figure 4.9a. The Friedman test does not show any main effect of the reality gap on our participants' valence ($\chi^2(2) = 1.38$, $p = 1$), see Figure 4.9b.



(a) Critical difference diagram showing pairwise statistical differences of the SAM questionnaire's arousal data—*Screen Simulation* and *Real Robots* are statistically significantly different, as well as *Screen Simulation* and *Virtual Reality*.



(b) Critical difference diagram showing pairwise statistical differences of the SAM questionnaire valence data. There are no statistical differences between any of the three sessions.

Figure 4.9: Critical difference diagrams showing pairwise statistical differences for **(a)** arousal and **(b)** valence.

NASA-RTLX questionnaire – The results of the Friedman test on our participants' workload show a main effect of the reality gap ($\chi^2(2) = 32.25$, $p < .001$). The Nemenyi post-hoc test shows a statistically significant difference in the workload level of our participants between the *Screen Simulation* session and the *Real Robots* session ($CD = 0.57$, $p < .001$) and between the *Screen Simulation* session and the *Virtual Reality* session ($CD = 0.57$, $p < .001$). The Nemenyi post-hoc test does not show any statistically significant difference between the *Virtual Reality* session and the *Real Robots* session ($CD = 0.57$, $p = .7$), see Figure 4.10. We report in Figure 4.11 the score of each dimension of the NASA-RTLX questionnaire for each session.

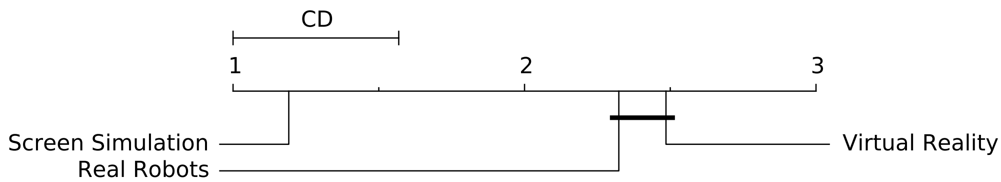


Figure 4.10: Critical difference diagram showing pairwise statistical differences of the NASA-RTLX questionnaire data—*Screen Simulation* and *Real Robots* are statistically significantly different, as well as *Screen Simulation* and *Virtual Reality*.

Reaction time – Finally, the results of the Friedman test on our participants' reaction time report a main effect of the reality gap ($\chi^2(2) = 24.24$, $p < .001$). The Nemenyi post-hoc test shows a statistically significant difference between the *Real Robots* session and the *Virtual Reality* session ($CD = 0.57, p < .05$), between the *Screen Simulation* session and the *Real Robots* session ($CD = 0.57$, $p < .05$) and between the *Screen Simulation* session and the *Virtual Reality* session ($CD = 0.57$, $p < .001$), see Figure 4.12.

Correlations, Gender Effect and Order Effect

As for the first experiment, we analysed our data in order to determine whether any dependent variables (i.e., skin conductance, arousal, valence, workload and reaction time) were pairwise correlated. We performed this analysis by aggregating the data of the three sessions. We found a correlation between reaction time and arousal ($\rho = 0.26$, $p < .05$), between workload and reaction time ($\rho = 0.43$, $p < .001$), between workload and arousal ($\rho = 0.37$, $p < .001$) and between workload and skin conductance ($\rho = 0.26$, $p < .05$).

We were also interested in determining whether the gender of our par-

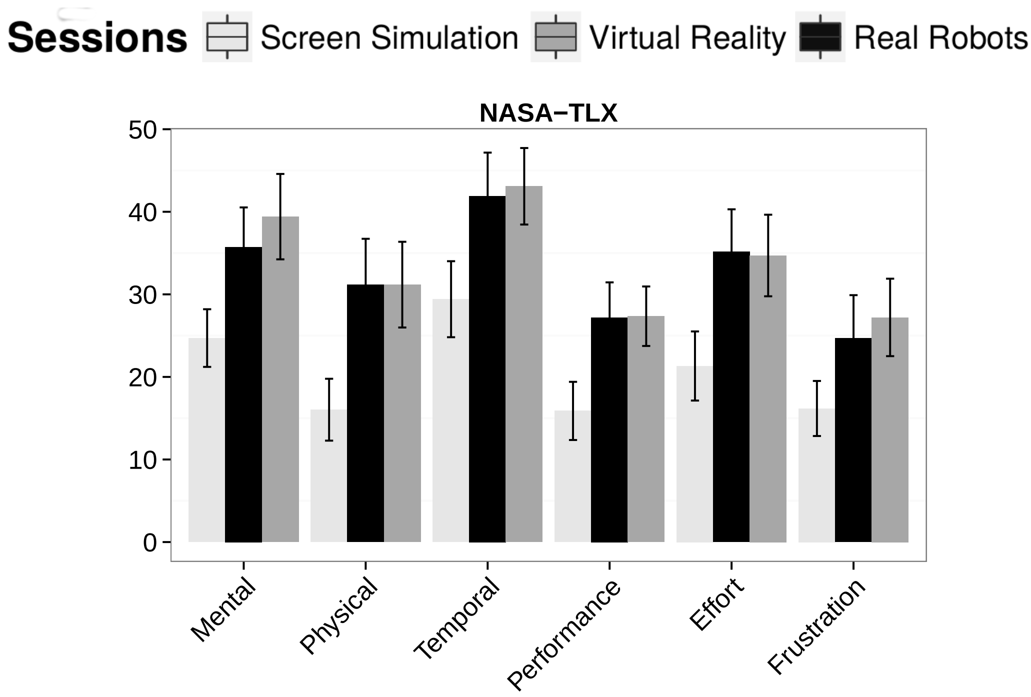


Figure 4.11: NASA-RTLX individual dimension's results. The error bar is the standard error for each session (standard deviation divided by the squared root of the sample size).

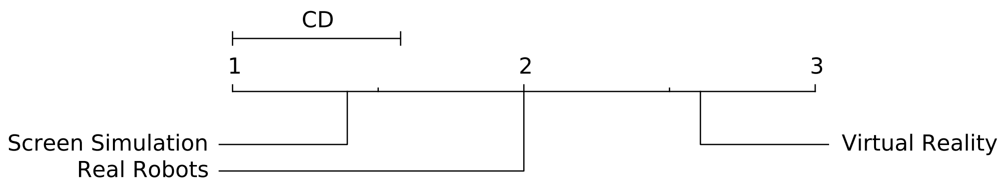


Figure 4.12: Critical difference diagram showing pairwise statistical differences of the reaction time data. Each session is statistically significantly different from the other.

ticipants and the order our participants encountered the sessions had any effects on their results. We proceeded with the same analysis as in the first experiment. As in the results of the first experiment, we did not find any effect of the gender nor of the session order on our participants' results.

4.3 Discussion and Conclusions

In this chapter, we presented two experiments on the effect of the reality gap on the psychological state of humans interacting with a robot swarm. The first experiment allowed us to show that our participants' psychophysiological reactions were stronger when they were passively interacting with a real robot swarm rather than when they were interacting with a simulated one displayed on a computer screen. Thanks to the additional task asked to our participants in the second experiment (i.e., pushing a button when a robot illuminates its LEDs), we showed that our participants' workload and reaction time were also higher when our participants interacted with a real robot swarm rather than with a simulated robot swarm displayed on a computer screen. Moreover, in the second experiment, we found similar psychophysiological reactions to those of experiment 1—our participants psychophysiological reactions were also higher when they were interacting with a real robot swarm rather than with a simulated one displayed on a computer screen. These results show that it is vital to take into account the reality gap when researchers design human-swarm interaction experiments.

A solution to avoid the reality gap would be to conduct all human-swarm interaction experiments with real robots. However, real robots experiments are expensive and time consuming. It is, therefore, not realistic to expect researchers to conduct human-swarm interaction experiments with dozens or hundreds of real robots. For this reason, we decided to investigate the possibility of using virtual reality in order to mitigate the effect of the reality gap. To the best of our knowledge, virtual reality has never been used in the research field of human-swarm interaction and is also little studied in social robotics (Li, 2015). In our experiments, our results suggest that virtual reality could be considered to mitigate the effect of the reality gap. In the first experiment, our participants reported higher arousal values when they were interacting with a simulated robot swarm in a virtual reality environment than with a simulated robot swarm displayed on a computer screen. In the second experiment, our participants also reported higher arousal values when they were interacting with the robots swarm in a virtual reality environment.

In addition to higher arousal values, results of the second experiment also show higher workload and reaction time values when our participants were interacting with a robot swarm in a virtual reality environment than when they were interacting with a robot swarm displayed on a computer screen.

Though the interaction with a robot swarm is fundamentally different than the interaction with a single robot (because of the number of robots human operators interact with, and because there is no social interaction with a robot swarm), our results show similar findings than studies in human-robot interaction—human beings react differently when they interact with real robots than with simulated robots. In addition to having a significant impact on the research methodology in human-swarm interaction (i.e., researchers should use real robots to conduct their experiments), these differences in reactions also have consequences on the development of interaction interfaces. For instance, in a search-and-rescue task, providing human operators with a simulation-like interface (i.e., an interface that shows a simulated representation of real robots), could make human operators less engaged in the task because all they see would be simulated robots only.

In the next chapter, we present a human-swarm interaction experiment in which humans interact with robot swarms of increasing sizes. In this experiment, we avoid the reality gap by conducting our experiments with real robots only.

5

Effect of the Increasing Group Size

In this chapter, we present the results of our research on the effect of increasing group size on the human psychological state. The goal of this research is to address this basic question—is the psychology of human beings affected by the number of robots to which they are exposed? Surprisingly, this fundamental question has been largely ignored in human-swarm interaction research. Some research has considered the role of human psychology in human-swarm interaction (see Section 2.1). However, the main focus of existing studies was on human workload, i.e., the mental effort required to deal with robot swarms under various different circumstances (De la Croix and Egerstedt, 2012, Pendleton and Goodrich, 2013, Setter et al., 2015). Only Karavas and Artemiadis (2015) focused on the human electroencephalographic activity when the robot swarm's cohesion varied. Compared to these previous studies, our contribution is twofold. Firstly, we answer a more basic question—what is the psychological effect on a human being of being confronted with increasing number of robots. Secondly, we adopt a rigorous, objective methodology. As discussed in Chapter 4, the reality gap has an effect on the human psychological state—humans react differently when they interact with a real robot swarm than with a simulated one. Therefore, in contrast to the majority of existing studies in human-swarm interaction, we use exclusively real robots. Equally importantly, while the majority of the previous studies have relied only on subjective questionnaires to determine psychological effects, we use a combination of objective psychophysiological measures and subjective psychological measures.

In this research, we measure the psychological state of twenty-four participants. We measure our participants' psychological state during a purely passive interaction with an increasing number of robots. For the same reasons as in Chapter 4, restricting our participants to passively interact with a swarm allows us to study the effect of the robot group size in the simplest form of interaction, that is, without the risk of increasing any psychological reactions with an extra interaction interface (e.g., joystick, keyboard, voice commands).

This chapter is organised as follows. In Section 5.1, we present the hypothesis on which we based the design of our experimental scenario presented in Section 5.2. In Section 5.3 we describe the measures collected on our participants. We give the descriptive statistics of our participants in Section 5.4. In Section 5.5, we explain the individual robot behaviour. We describe the experimental procedure of the experiment in Section 5.6. In Section 5.7, we explain the analysis of our data and we present the results of our research. Finally, in Section 5.8, we discuss our contributions and conclude the chapter.

5.1 Hypothesis

One of the main characteristics of a swarm robotics system is the relatively large number of robots that constitutes the system. We believe that the number of robots in a robot swarm can have important psychological implications during the interaction between a human operator and the robot swarm. We, therefore, designed our experiment based on the following hypothesis:

- The psychophysiological reactions of humans become stronger when the number of robots in a robot swarm increases.

Confirming this hypothesis would suggest that the psychological responses of humans is affected by the number of robots to which they are exposed. This result should have profound implications in the future of human-swarm interaction since the relatively large number of robots in a robot swarm is an inherent characteristic of swarm robotics and can not, therefore, be avoided.

5.2 Experimental Scenario

We designed an experimental scenario based on the aforementioned hypothesis. In this scenario, a participant is seated in the same environment as the robots. We divide the experiment into three sessions. In each session, we increase the number of robots. As in Velagapudi et al. (2008), we did not randomize the sessions across the participants because our principal focus is on the effect of increasing robot group sizes. While the first session includes only a single robot (hereafter referred to as the *1-robot* session) the second and third sessions (hereafter referred to as the *3-robot* session and the *24-robot* session, respectively) include a total of three and twenty-four robots respectively. Each participant is exposed to each of the three groups of robots (i.e., *1-robot*, *3-robot*, *24-robot*) for a period of 45 s. In Figure 5.1, we show an example of a participant at the beginning of the experiment (Figure 5.1 (a)), interacting with one robot (Figure 5.1 (b)), with three robots (Figure 5.1 (c)), and finally with twenty-four robots (Figure 5.1 (d)).

5.3 Measures

In this research, we want to study the effect of increasing robot group sizes on our participants' psychophysiological state. As in Chapter 4, we study our participants' psychophysiological state by monitoring their heart rate and skin conductance level. We also investigate whether our participants are conscious of their psychophysiological state by administrating the SAM questionnaire.

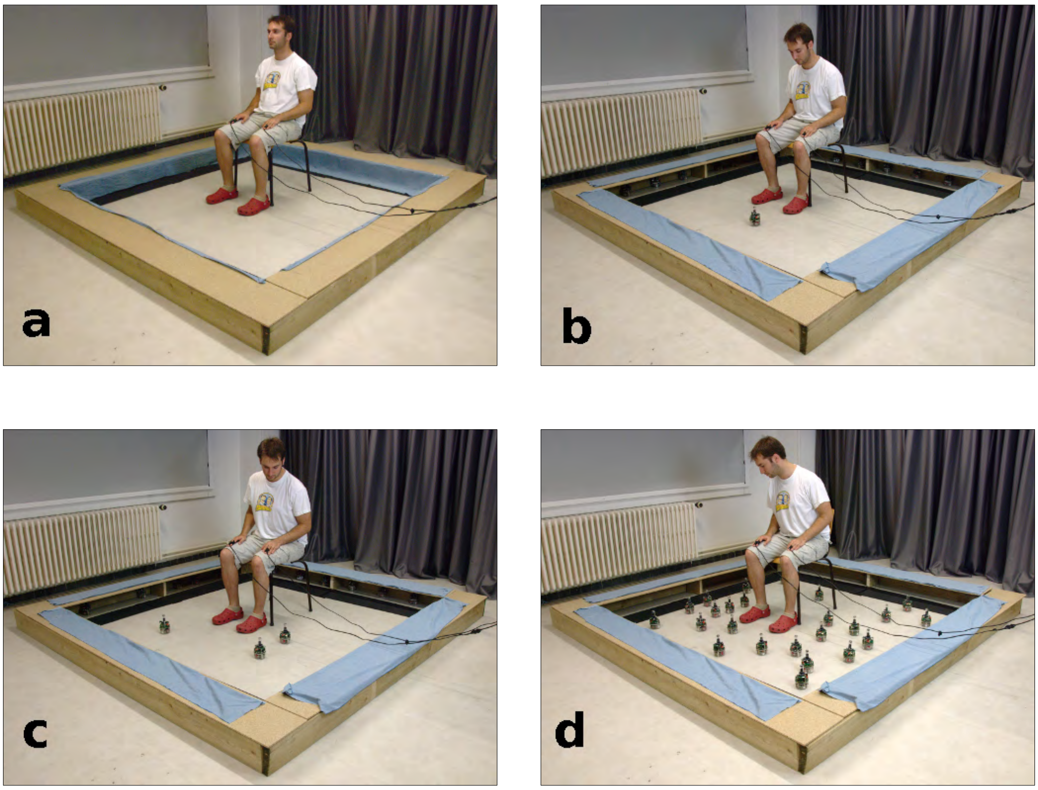


Figure 5.1: An example of an experiment in progress. **(a)** At the beginning of the experiment, the participant is attached to two physiological sensors. **(b)** The experiment begins with one robot moving around the participant. **(c)** Subsequently, two more robots appear and the participant is exposed to a group of three robots. **(d)** Finally, twenty-one robots appear and the participant is exposed to a group of twenty-four robots. The participant shown in this figure is the author of this thesis and did not take part in the experiment. The pictures shown in this figure were taken for illustration purposes.

5.4 Participants

We recruited 25 participants from the overall population of the Université Libre de Bruxelles. None of our participants had a background in robotics and none of them had participated to the experiments of Chapter 4. Participants were between 18 and 45 years old with an average age of 25.04 years (SD = 5.16). We considered current or anterior cardiovascular problems that could act on the central nervous system as exclusion criteria. Our participants received an informed consent form explaining that they were filmed[1] during the experiment and that their physiological responses were being collected for research purposes only (see Annex C). We offered a 7 € financial incentive for participation.

5.5 Robot Behaviour

During an experiment, robots gradually drive out of the hidden zones (see Section 3.1.2) in order for a participant to be exposed to three different robot group sizes, i.e., one robot, three robots and twenty-four robots. At the beginning of an experiment, one robot drives out of the hidden zone and becomes visible to the participant. Then, two robots drive out of the hidden zones at the same time so that the participant is exposed to a group of three robots (i.e., the first robot that started the experiment, joined by two additional robots). In order for the participant to be exposed to a group of twenty-four robots, twenty-one robots drive out of their hidden zone at the same time.

Once the robots become visible to a participant, they execute the same *random walk with obstacle avoidance* behaviour as in Chapter 4. Additionally, the obstacle avoidance behaviour is also triggered when the robots enter the black area (see Section 3.1.2). When a black area is detected by a robot, the robot performs a U-turn in order to avoid to enter the hidden zone. Doing so, all robots always remain visible to the participant.

In order for all of our participants to be subjected to similar experimental conditions, the robots that appear in the *1-robot* session and in the *3-robot* session were always coming from identical locations and hidden zones. As shown in Figure 5.2, the single robot of the *1-robot* session was always coming from the front of the participant and the two additional robots of the *3-robot* session were always coming from the front left and the front right of the participant.

[1] As in Chapter 4, the video recordings were not used during the analysis.

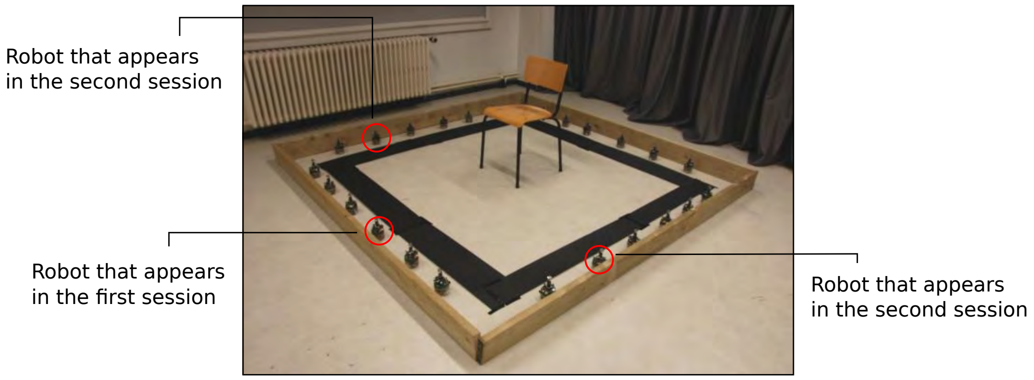


Figure 5.2: Initial position of the robots used in the experiments. The robot in front of the chair (encircled in the picture) is the robot that comes out in the 1-robot session. The two robots on the left and on the right of the chair (encircled in the picture) are the robots that come out in the 3-robot session. The other robots (i.e., those not encircled) are those that come out in the 24-robot session. The four boards of wood that cover the robots (i.e., that render the robots invisible) have been removed when taking this picture.

5.6 Experimental Procedure

As for the experiments described in Chapter 4, all our experiments were conducted at IRIDIA, the artificial intelligence laboratory of the Université Libre de Bruxelles. Upon arrival, a brief explanation of the procedure of the experiment was given to the participant. We explained to the participant that the experiment was divided into three sessions and that in each session, a certain number of robots would move around them. We asked the participant to read and sign the consent form and to wash their hands in clear water. Then, we placed the participant on the chair, we attached the physiological sensors to the participant and we explained the SAM questionnaire. Before starting the experiment, we recorded the participant's baseline during a period of 5 minutes. We then proceeded with the *1-robot* session. After this session, we asked the participant to choose an image in the valence scale and an image in the arousal scale that correspond to their subjective psychological state. The SAM questionnaire was attached to the wall in front of the participant. A number from 1 to 9 was written at the bottom of each image of the SAM questionnaire. In order for the experimenter to record the valence and arousal data, the participant had to speak out loud the number of the image he or she chose. We followed the same procedure for the *3-robot* session and for the *24-robot* session.

Figure 5.3 shows a participant during an experiment while the participant was confronted with a group of twenty-four robots. The entire experiment's duration was approximately 30 minutes per participant.

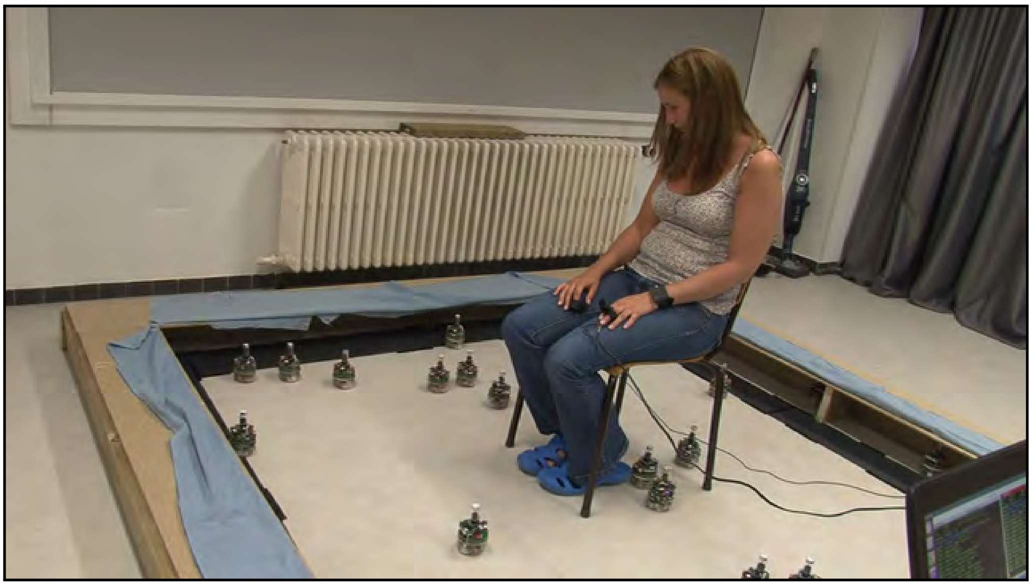


Figure 5.3: A participant during the 24-robot session. The picture is a snapshot taken from the video recording of an experiment. The participant gave her written consent to the use of the picture.

After the experiment ended, we detached the sensors from the participant and conducted a brief interview with the participant. During the interview, we explained to the participant the goal of the study. We also asked the participant to describe his or her experience with the experiment and we answered the participant's questions.

5.7 Data Analysis and Results

For one of the participants, the robots were misplaced and were visible prior to the beginning of the experiment. Therefore, we rejected this participant's data (both psychophysiological data and self-reported data) from our analysis in order to avoid possibly biased data. For another participant, the psychophysiological data was very noisy, probably due to a misplacement of the sensors. We did not take into account this participant's psychophysiological data (i.e., heart rate, skin conductance level), but we kept the participant's self-reported data (i.e., valence and arousal of the SAM questionnaire). We, therefore, used the psychophysiological data of 23 participants (11 male and 12 female) and the self-reported data of 24 participants (11 male and 13 females). As in Chapter 4, the analysis is conducted on the difference between the participants' physiological responses at rest and during the experiment.

In the following of this section, we first analyse the effect of the increasing group size on the human psychophysiological state. Then, we analyse our data in order to detect any potential habituation effects in our participants. Finally, we present the results of a correlation analysis and of a gender effect analysis.

5.7.1 Group Size Effect

We analysed our data with the R software (R Core Team, 2015). We used the non-parametric Friedman test to analyse both the psychophysiological data and the self-reported data (i.e., the SAM questionnaire). In our case, the Friedman test's null hypothesis states that the three sessions *1-robot*, *3-robot* and *24-robot* are not different. The alternative hypothesis states that at least two sessions are different. In the case of the Friedman test is significant, we proceeded with a pairwise comparison of the three sessions with a Nemenyi post-hoc test.

In Table 5.1, we summarise the results of the psychophysiological and self-reported data (i.e., median and Friedman's mean rank of heart rate, skin conductance level, arousal and valence) in each session (i.e., *1-robot*, *3-robot*, *24-robot*) as well as the inference statistics of the Friedman tests (i.e., p-values and χ^2). In Figure 5.4, we show the boxplots of the three sessions.

Heart rate – The analysis of the heart rate data shows a main effect of

Dependent Variable	n	1-robot	3-robots	24-robots	χ^2	p
Heart Rate	23	-1.88 (1.87)	-2.48 (1.69)	0.17 (2.43)	$\chi^2(2) = 6.87$	$<$ **.05**
SCL	23	4.45 (1.61)	4.76 (1.61)	6.73 (2.78)	$\chi^2(2) = 21.13$	$<$ **.001**
Arousal	24	3.5 (1.56)	4 (1.62)	6 (2.8)	$\chi^2(2) = 27.88$	$<$ **.001**
Valence	24	6.5 (1.97)	6.5 (1.89)	7 (2.12)	$\chi^2(2) = 0.8$	.6

Table 5.1: Descriptive statistics of the psychophysiological data and of the self-reported data. We report the median and the Friedman's mean rank (in parentheses) of the three sessions (1-robot, 3-robot, 24-robot). We also report the inference statistics of the Friedman test (i.e., χ^2 and p).

the number of robots on our participants ($\chi^2(2) = 27.88, p < .001$).[2] The Nemenyi post-hoc test on the heart rate data revealed that our participants' heart rate was statistically significantly different between the *3-robot* session and the *24-robot* session ($CD = 0.69, p < .05$). Our participants' heart rate was not statistically significantly different between the *1-robot* session and *3-robot* session ($CD = 0.69, p = .8$). Finally, our participants' heart rate was not statistically significantly different between the *1-robot* session and the *24-robot* session ($CD = 0.69, p = 0.13$), see Figure 5.5.

Skin conductance level – The analysis of the skin conductance level also confirmed a main effect of the number of robots on our participants ($\chi^2(2) = 21.13, p < .001$). The Nemenyi post-hoc test on the skin conductance level data revealed a statistically significant difference between the *3-robot* and *24-robot* sessions ($CD = 0.69, p < .001$), and between the *1-robot* and *24-robot* sessions ($CD = 0.69, p < .001$). The skin conductance level was not statistically significantly different between the *1-robot* and *3-*

[2] A reason for the heart rate values (the difference between the heart rate values during the baseline and the heart rate values during the three sessions) to be negative is that, in situations that generates affective responses, the heart rate first decreases before increasing (Bradley and Lang, 2000). In our case, the heart rate decrease was more prominent than the following heart rate increase.

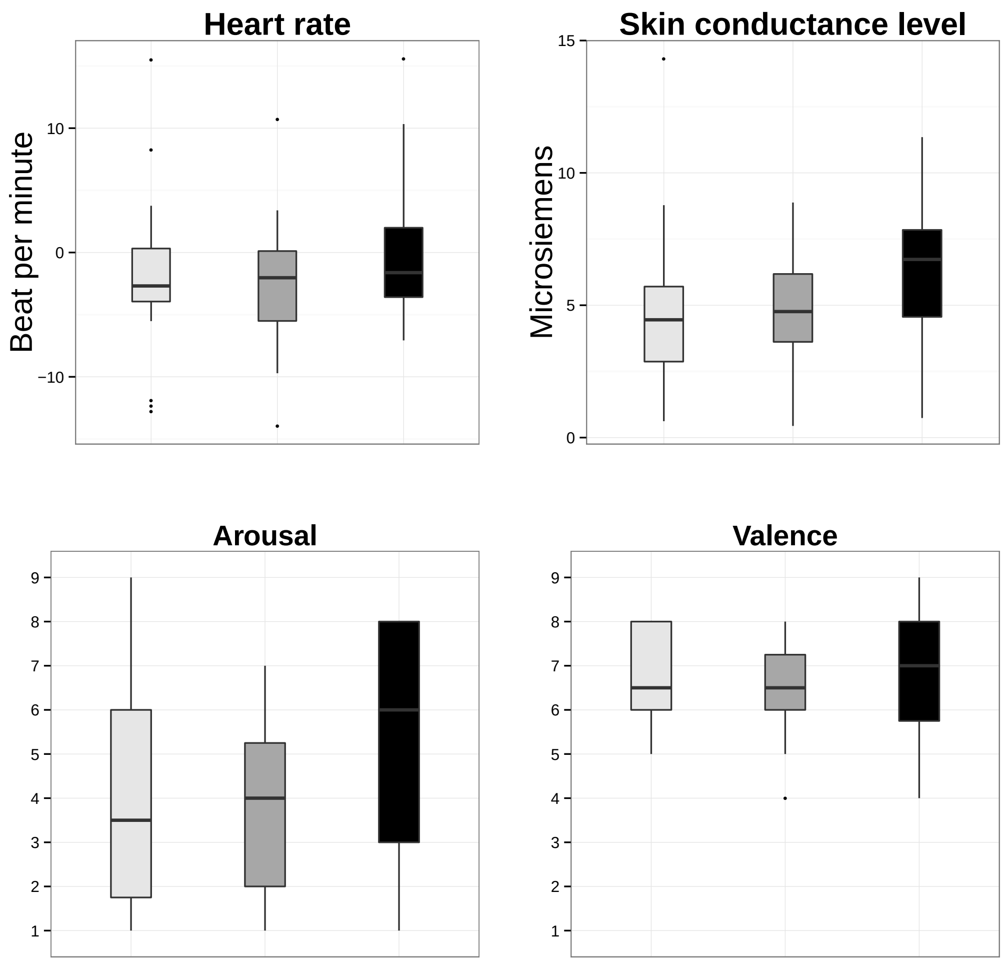


Figure 5.4: Boxplots showing the heart rate values (top left), skin conductance level values (top right), arousal values (bottom left) and valence values (bottom right) of all three sessions (1-robot, 3-robot, 24-robot). The median value of each session is shown using the bold horizontal line in the box. Outliers are represented using dots.

robot sessions ($CD = 0.69, p = 1$), see Figure 5.6.

SAM questionnaire – The results of the Friedman test on the self-reported arousal data confirm a main effect of the number of robots on our participants ($\chi^2(2) = 27.88, p < .001$). The Nemenyi post-hoc test showed that our participants' self-reported arousal was statistically significantly different between the *1-robot* and *24-robot* sessions ($CD = 0.67, p < .001$) and between the *3-robot* and *24-robot* sessions ($CD = 0.67, p < .001$). The

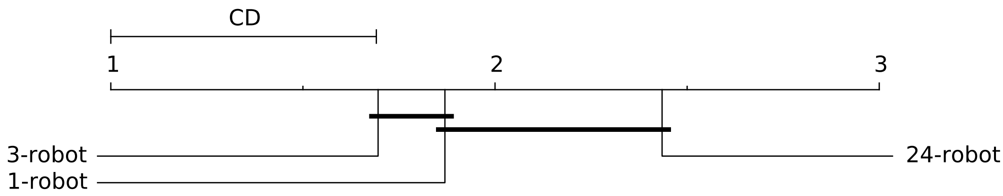


Figure 5.5: Critical difference diagram showing pairwise statistical differences of the heart rate data—3-robot and 24-robot are statistically significantly different.

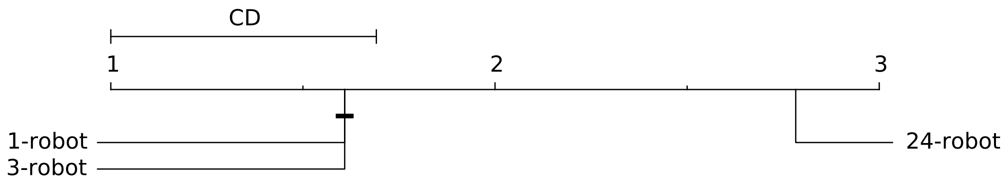


Figure 5.6: Critical difference diagram showing pairwise statistical differences of the skin conductance level data—1-robot and 24-robot, and 3-robot and 24-robot are statistically significantly different.

self-reported arousal was not statistically significantly different between the *1-robot* and *3-robot* sessions ($CD = 0.67, p = .9$), see Figure 5.7. The analysis of the self-reported valence does not show any main effect of the number of robots on our participants ($\chi^2(2) = 0.8, p = .6$), see also Figure 5.8.

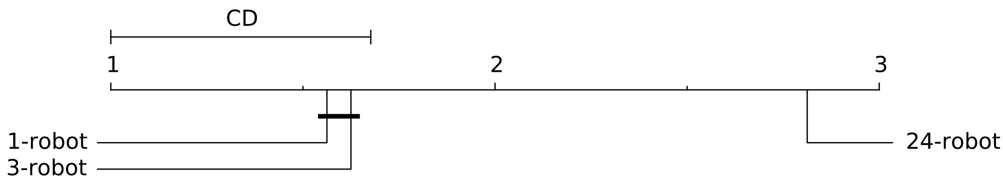


Figure 5.7: Critical difference diagram showing pairwise statistical differences of the SAM questionnaire's arousal data—1-robot and 24-robot, and 3-robot and 24-robot are statistically significantly different.

5.7.2 Habituation Effect

We should account for the possibility that the psychophysiological reactions we observed were attributable to an initial surprise effect that the participants felt on being exposed to robots, which could then wear off as the

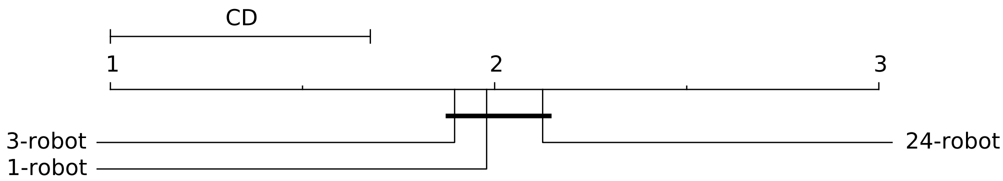


Figure 5.8: Critical difference diagram showing pairwise statistical differences of the SAM questionnaire's valence data—there are no statistical differences between any of the three sessions.

participants became accustomed to the robots—the so called "habituation effect". If our participants were indeed "surprised" by the robots, we would expect their skin conductance level to rise quickly during the first seconds of the experiment. In order to detect if our participants were surprised by the robots, we computed the mean of all of our participants' skin conductance values during the whole duration of each session. We show the results of all three sessions in Figure 5.9. As depicted in Figure 5.9, in each of the three sessions, the graph peaks before the 10 first seconds of each session. Then, the skin conductance values decrease and remain stable until the end of the session.

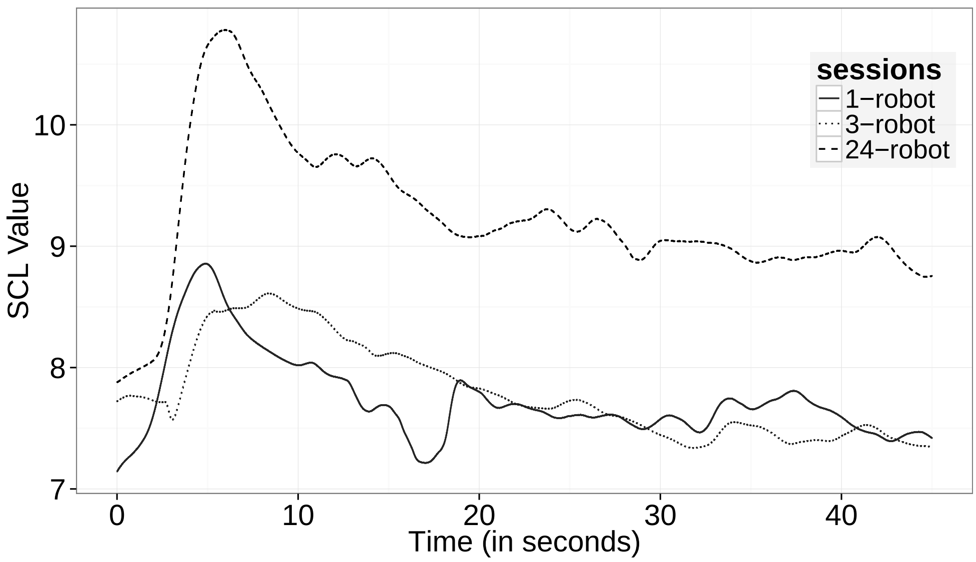


Figure 5.9: Mean skin conductance level values (all participants) over time.

The peak within the 10 first seconds suggests that our participants were surprised by the robots. After this peak, though, the stabilization of our participants' skin conductance suggests a "habituation effect" within each session—during each session, our participants get habituated to the robots moving around them. However, the data does not suggest any habituation effect between sessions with few robots (1 and 3) and the session with many robots (24). Therefore, the data does not contradict our hypothesis that an increasing number of robots affects the human psychophysiological state—as depicted in Figure 5.9, the *24-robot* session's curve clearly remains above the two other curves (i.e., *1-robot* and *3-robot*) during the entire duration of the experiment.

5.7.3 Correlations and Gender Effect

We did not find any correlation between valence and skin conductance, valence and heart rate, arousal and heart rate, nor arousal and valence. However, there was a significant correlation between skin conductance and heart rate ($\rho = .29$, $p = .01$) and there was a marginally significant correlation between the skin conductance and arousal ($\rho = .229$, $p = .059$)[3].

As in Chapter 4, we also investigated whether the gender of our participants had any effect on their results. As in Chapter 4, we studied the gender effect by splitting our participants into two groups—males and females. The Wilcoxon rank-sum test did not report any significant difference between groups, suggesting that the gender of our participants did not have any effect on their results.

5.8 Discussion and Conclusions

The work we presented in this chapter contributes to the human-swarm interaction literature in that it is the first time the effect of the number robots

[3]In order to be scientifically correct, we should not consider a p-value significant nor "marginally" significant if its value is greater than the predetermined significance level (in our case $\alpha = .05$). However, we wanted to report this result because we believe it might be nonetheless interesting for any researcher having done correlation analyses between arousal and skin conductance.

on human psychology was studied by using psychophysiological measures. Moreover, to the best of our knowledge, it is the first human-swarm interaction experiment performed with up to twenty-four real robots.

The hypothesis we aimed to test in this study was the following: the psychological response of humans is affected by the number of robots to which they are exposed. Twenty-four participants were exposed to three robot groups of varying sizes: 1 robot, 3 robots and finally 24 robots. The results of our experiment confirm this hypothesis. Our results show that our participants' heart rate and skin conductance level were significantly higher when the number of robots increased to twenty-four robots. Moreover, our results also show that our participants were conscious of their psychophysiological state change, as they reported significantly higher arousal values when exposed to the twenty-four robots.

Since the large number of robots is inherent to the definition of swarm robotics, it will be challenging to mitigate the effect of the number of robots on human psychology. Moreover, in our experiments, our participants were completely passive—there was no bidirectional communication between the participants and the robots (i.e., the participants did not issue any commands and the robots did not provide any feedback). We expect that in an active interaction scenario (i.e., with a bidirectional communication), human psychology will be even more affected. For instance, feedback provided by each individual robot could be overwhelming for human operators, increasing their psychological responses. Therefore, researchers should focus on diminishing as much as possible the effects of non-inherent characteristics of swarm robotics (e.g., issue commands or receive feedback). In the case of feedback for instance, a solution could be to use *self-organised feedback* (Podevijn et al., 2012), where swarms of robots leverage self-organised techniques to provide human operators with a swarm-level feedback.

6
Conclusions

In this chapter, we first present the limitations and future work of our research. Then, we present our conclusions.

6.1 Limitations and Future Work

In Chapter 4, we have shown that the reality gap—the inherent discrepancy between simulation and reality—affects the psychological state of humans who perform a supervision task with a robot swarm. More specifically, we have shown, with two distinct experiments, that the human psychophysiological state, workload and reaction time measured for the case of interaction with a swarm of real robots and for the case of interaction with a simulated robot swarm displayed on a computer screen were significantly different. These results show that the reality gap effect must be carefully considered in human-swarm interaction.

As discussed in Section 4.3, it is often not practically feasible to conduct all human-swarm experiments using real robots. We, therefore, propose virtual reality as an alternative to simulation displayed on a computer screen. We showed that virtual reality could mitigate the reality gap effect—our participants' arousal, workload and reaction time were significantly higher when they were interacting with the simulated robot swarm in the virtual reality environment than when they were interacting with the simulated robot swarm displayed on the computer screen. However, we should qualify these results because our participants' reaction time was also significantly higher when they were interacting with the robot swarm in the virtual reality environment compared to when they were interacting with the real robot swarm. Though these results do not contradict our hypothesis, we believe more research is necessary to better understand the use of virtual reality in human-swarm interaction studies.

In addition, we should account for the possibility that the difference of perspectives that we used in our experiments (top-view in the 2D *Screen Simulation* session and similar to the reality in the 3D *Virtual Reality* session) has also an effect on the participants' psychophysiological state, workload and reaction time. Future work should investigate whether the difference of perspectives has a significant impact on the human psychological state. For instance, we could replicate the experiments presented in Chapter 4 by replacing the 2D top-view perspective of the *Screen Simulation* session with a 3D perspective of the robots and of the environment.

In Chapter 5, we have shown that the psychological response of humans is affected by the number of robots to which they are exposed. Our results confirm this hypothesis, and furthermore show that greater numbers of robots provoke a stronger response.

Though our results have shown an effect of the group size on the human psychological state, we did not consider all possible variables that could also influence the psychological state. For instance, the size of the arena was kept constant during the experiments. It would be interesting in the future to study whether increasing the size of the arena while keeping the group size constant would decrease the effect on the psychological state. Another variable that could influence the psychological state is the robots' behaviour.

In the experiment presented in Chapter 5, our robots were executing a basic swarm behaviour—a random walk. Future work should focus on other swarm behaviours such as flocking, foraging, search and rescue, and so on. A third variable to consider is the location of the participant during the experiment. We wanted our participants to be immersed in the environment. However, allowing participants to move in the arena could decrease the effect of group size—for instance if they feel more comfortable moving to a corner of the arena or even outside the arena. Other variables could influence the human psychological state (e.g., the size of the robots, the noise produced by the robots, the participants' prior experience with robotic systems). These variables, or combinations of these variables (with the group size for instance) should be considered in future work.

One interesting aspect of the results obtained in Chapter 4 and Chapter 5 that also deserves attention in future work is the nature of the psychological state. Psychophysiological measures are considered objective in that they are by large not under the conscious control of the participant being measured. However, the exact relationship between the psychophysiological measures and the psychological state that provoked the physiological response is not always clear. For instance, in Chapter 5, we were expecting to see a primarily stress-based response to the robots. In fact, however, the valence values reported by the SAM questionnaire we asked our participants to fill out suggest that our participants had a positive experience—on a scale of 1 to 9 (1 being the less happy and 9 being the more happy), they reported valence values of 6.66, 6.62 and 6.87 for the *1-robot*, *3-robot* and *24-robot* sessions respectively. We believe that our participants reported these valence values because they were not actively interacting with the robots, i.e., they did not have to control the robots. In the experiments presented in Chapter 4 and Chapter 5, the participants passively interacted with the robots (i.e., the participants did not send commands to the robots), instead of actively interacting with the robots (i.e., the participants would send commands to the robots). We chose passive over active interaction to reduce the number of variables that could impact the experiment, so as to increase our confidence that it was the reality gap or the number of robots that was affecting our participants' psychological state. For example, in an interactive scenario, frustration due to interaction difficulties or failures might also have introduced psychological effects. For each study (i.e., reality gap and group size effects), the effect of active interaction should be investigated in future work. For instance, we should study whether actively interacting with a swarm of robots (e.g., by guiding a swarm) negatively changes the psychological state of our participants. We would expect, for example, frustration or anger (i.e., low valence values) when a participant has to guide a swarm of robots.

6.2 Conclusions

Human-swarm interaction is a field of research that became active during the past five years. Researchers have realized that if we want robot swarms to become useful for real world applications, studying the interaction with these robot swarms becomes vital. Unfortunately, today swarm robotics is limited to research laboratories and has no concrete real world applications. It is, therefore, difficult to understand how swarms of robots are going to be used and how humans will interact with these swarms of robots. The absence of real world applications and the absence of understanding how robot swarms are going to be used have rendered the field of human-swarm interaction scattered. As we have seen in Chapter 2, there are many research directions that are investigated. In these research directions, there are no consensus on the methods to use in order to provide human operators with effective human-swarm interaction systems. For instance, different methods are studied to control a robot swarm (e.g., by controlling either a single robot, or by controlling multiple leader robots), but none of these methods are compared, making it difficult to understand what is the best method to create an effective human-swarm interaction system.

It is our contention that before creating any human-swarm interaction systems, it is vital to understand how human beings react to swarms of robots. The results presented in this thesis confirm that it is important to study the impact of these swarms of robots on human psychology. Even more so when the impact is due to a characteristic that is inherent to the definition of swarm robotics, such as the number of robots that composes a robot swarm. The psychological impact of a characteristic inherent to the definition of swarm robotics is hard to mitigate. For instance, it would not make any sense to reduce the number of robots a human operator is interacting with in order to decrease the human operator's psychological responses. Therefore, it will be important to take into account the psychological impact when designing a human-swarm interaction system in order to avoid increasing even more the human operator's psychological reactions.

In this thesis, the main contribution has been to make a step forward in the understanding of the psychological impact of the interaction between humans and robot swarms. We conducted three experiments. For the first time in human-swarm interaction, the experiments were all conducted with robot swarms consisting of twenty real robots for two of the experiments and of up to twenty-four real robots for the third one. Experiments with real robots are incredibly time consuming, all the more so when we have to set up monitoring equipment for psychophysiological measurement. For this reason, the majority of the studies published in the literature do neither—standard practice is to use simulated robots with psychological questionnaires (see Chapter 2).

In our experiments, we used a combination of objective psychophysiological measures (physiological responses such as skin conductance and heart rate) and subjective self-reported measures (with two psychological questionnaires). With the exception of human-computer interaction, psychophysiological measures are still not commonly used in the literature. In human-computer interaction, psychophysiology is a methodology that is starting to move out of the research laboratories—as suggested by more and more human-computer interaction textbooks that propose psychophysiology for designing and studying new systems humans can interact with (Bainbridge, 2004, Dix et al., 2003, Salvendy, 2012, Tullis and Albert, 2008). In the same way psychophysiology is used to create today's real-world human-computer interaction systems, we strongly believe that psychophysiology (in combination with psychological questionnaires) should be used to create tomorrow's real-world human-swarm interactions systems.

A
Companion Website

In research, it is vital for researchers to be able to replicate their experiments or other scientists' experiments. Successfully replicating an experiment significantly increases the confidence we can place in the results.

Recently, the Open Science Collaboration group, headed by Brian Nosek, has replicated 100 experiments in psychology. Their results, published in Science, are astonishing: only a third of these replications showed statistical significant results. Hence, two third of their replications failed to find similar results as the original researches (Open Science Collaboration, 2015).

In this thesis, we dedicated significant time and effort to making our experiments reproducible. We described the experimental scenarios and experimental procedures of our three experiments as clearly and with as many details as possible. However, describing the experimental scenarios and the experimental procedures is not enough to make an experiment reproducible. Describing and releasing the tools and the software source code used to conduct an experiment is equally important. Therefore, we have created a companion website to this thesis. In this annex, we provide an overview of the companion website. The companion website is available at: http://iridia.ulb.ac.be/~gpodevijn/phd/supp.

A.1 Goals

The companion website is an integral part of this thesis and is intended to be used by researchers that are interested in replicating our experiments, or by researchers interested in extending our research.

We created this companion website with three goals in mind. Firstly, to make the software used to conduct the experiments presented in Chapter 4 and Chapter 5 accessible (e.g., the robot controllers, the virtual reality Android application). Secondly, to explain, step-by-step, how to install and use the software. Thirdly, to make the data we collected during our experiments and the tools used to analyse this data (e.g., the R scripts) accessible.

A.2 Structure

The companion website is divided into five parts. Each part is always accessible from anywhere in the website via the menu placed on the top of the screen.

Home

This part contains a short summary of each experiment, a link to each experiment instructions, a link to the materials web page in which all the software can be downloaded and a link to the instructions to install the ARGoS simulator (see Figure A.1).

Reality gap

This part is subdivided into two parts. Each part contains the instructions and the data of the experiments presented in Chapter 4 (see Figure A.2).

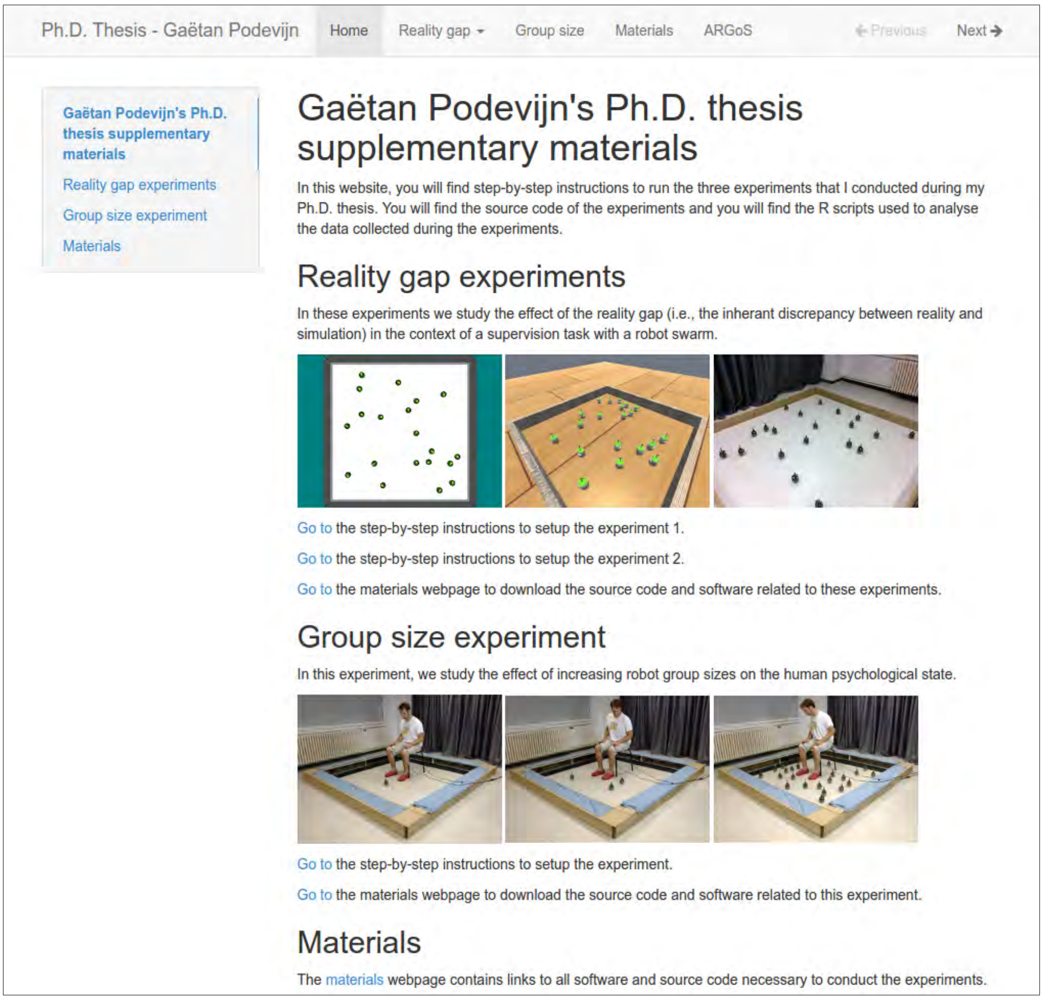


Figure A.1: Screenshot of the home page. There is a brief summary of each experiment and a link to the experiments' instructions web page. There is also a link to the materials web page on which all the software can be downloaded and a link to the instructions to install the ARGoS simulator.

Group Size

This part contains the instructions and the data of the experiment presented in Chapter 5 (see Figure A.3).

Materials

This part aggregates all the software (and software source code) used to setup the experiments (see Figure A.4).

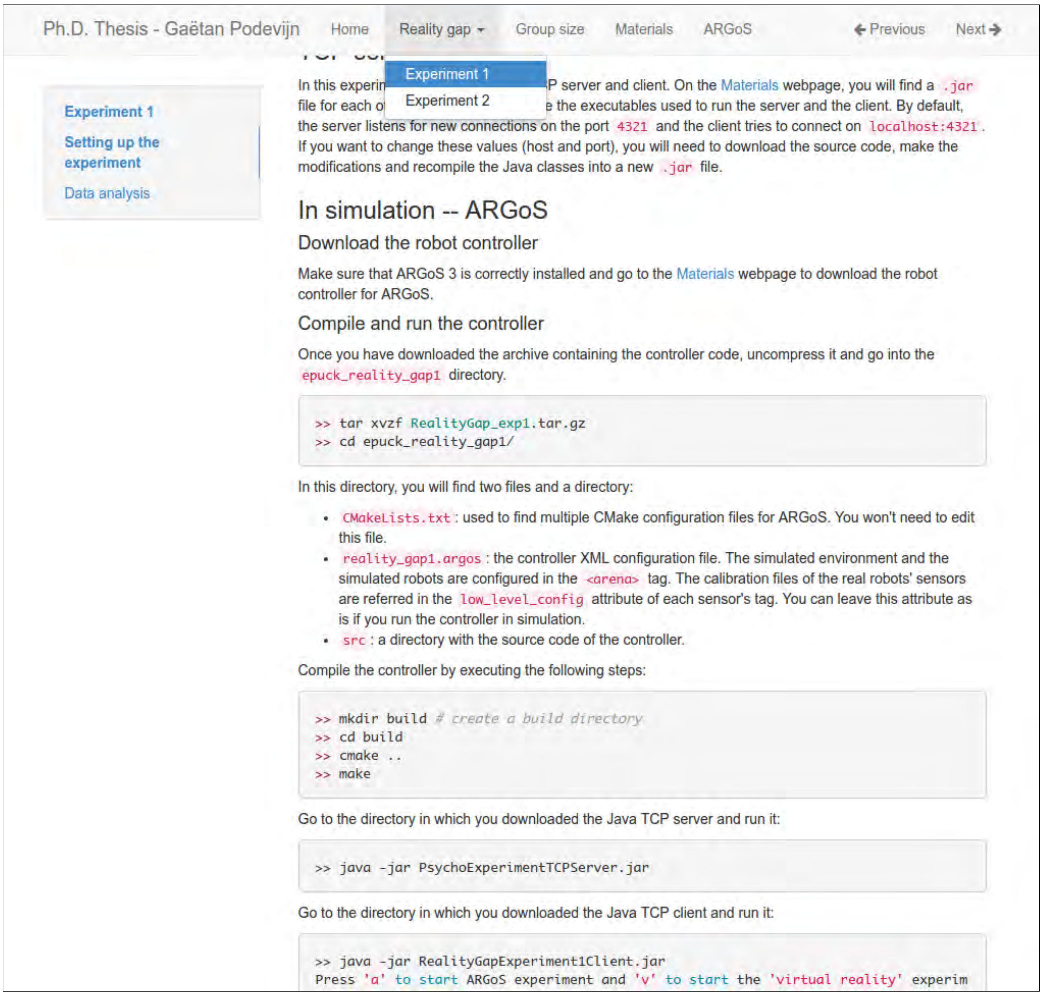


Figure A.2: Screenshot of the reality gap's first experiment instruction page. This page contains step-by-step instructions to run the experiment in simulation and with the real robots. The menu placed on the top of the screen allows the user to navigate to the second experiment web page. The data of each experiment and the scripts to analyse the data are available on each experiment web page.

ARGoS

This part contains the instructions to install the ARGoS simulator and the ARGoS-Epuck plugin (a plugin used to use the real e-puck robot platform and the e-puck robot platform in simulation) (see Figure A.5).

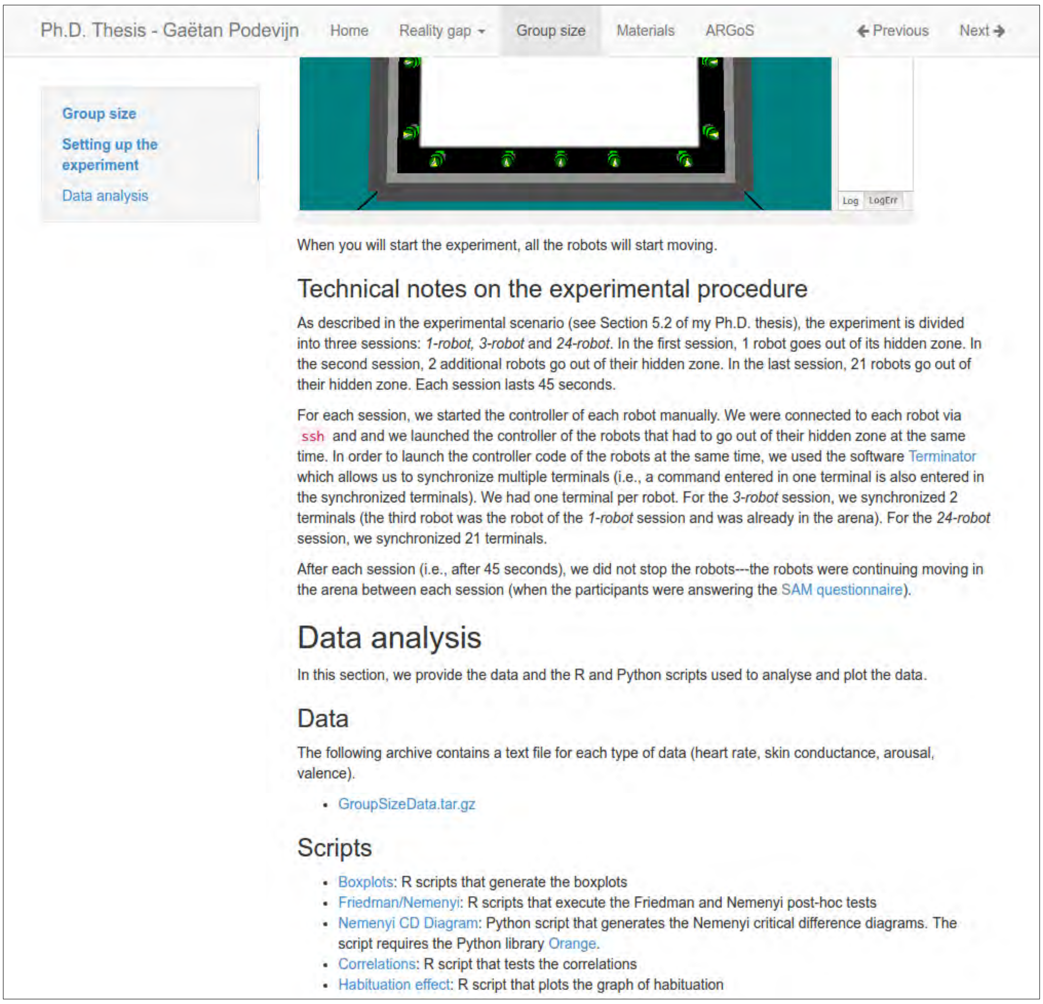


Figure A.3: Screenshot of the group size experiment web page. This web page contains the instructions to run the experiment both in simulation and with the real robots. The data of the group size experiment and the scripts to analyse the data are available from this web page.

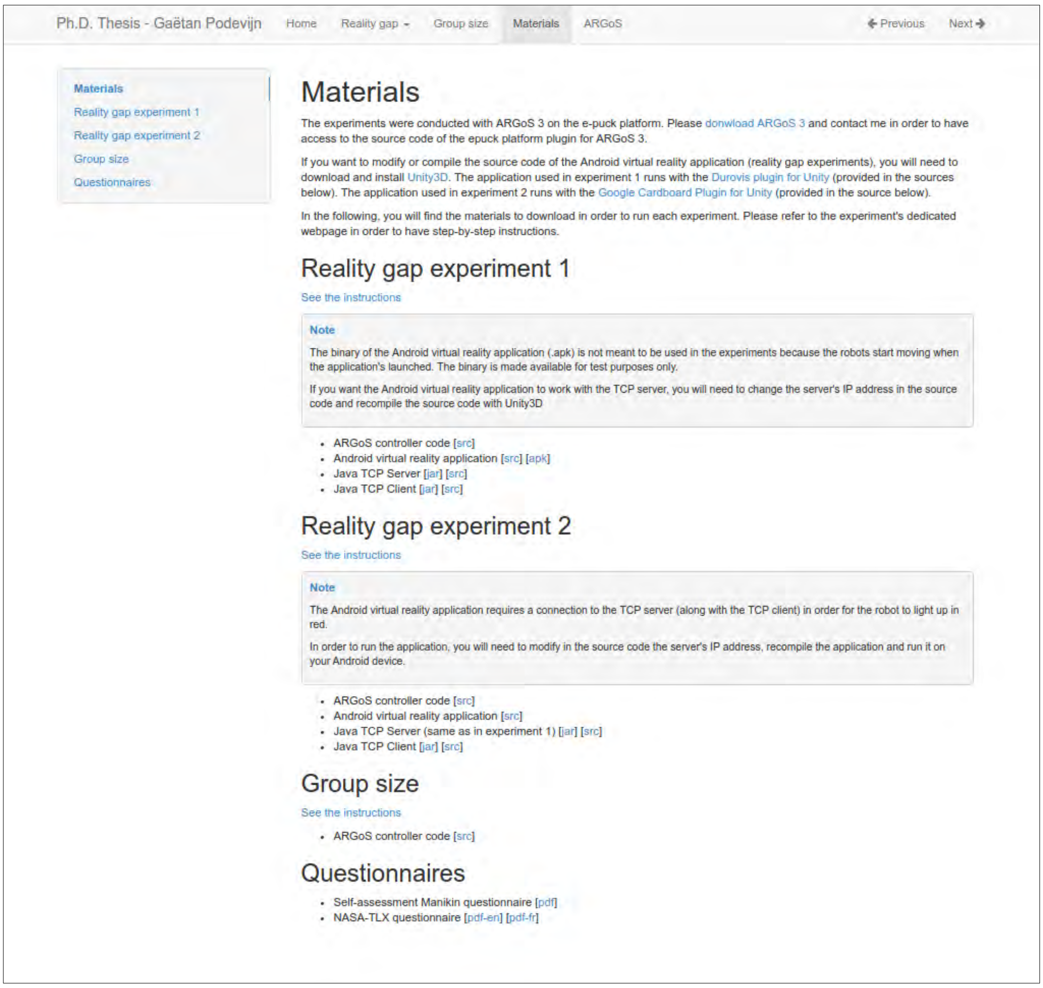


Figure A.4: Screenshot of the materials web page. This web page contains all the software we used to conduct the experiments. It also contains a link to the questionnaires (SAM and NASA-RTLX questionnaires).

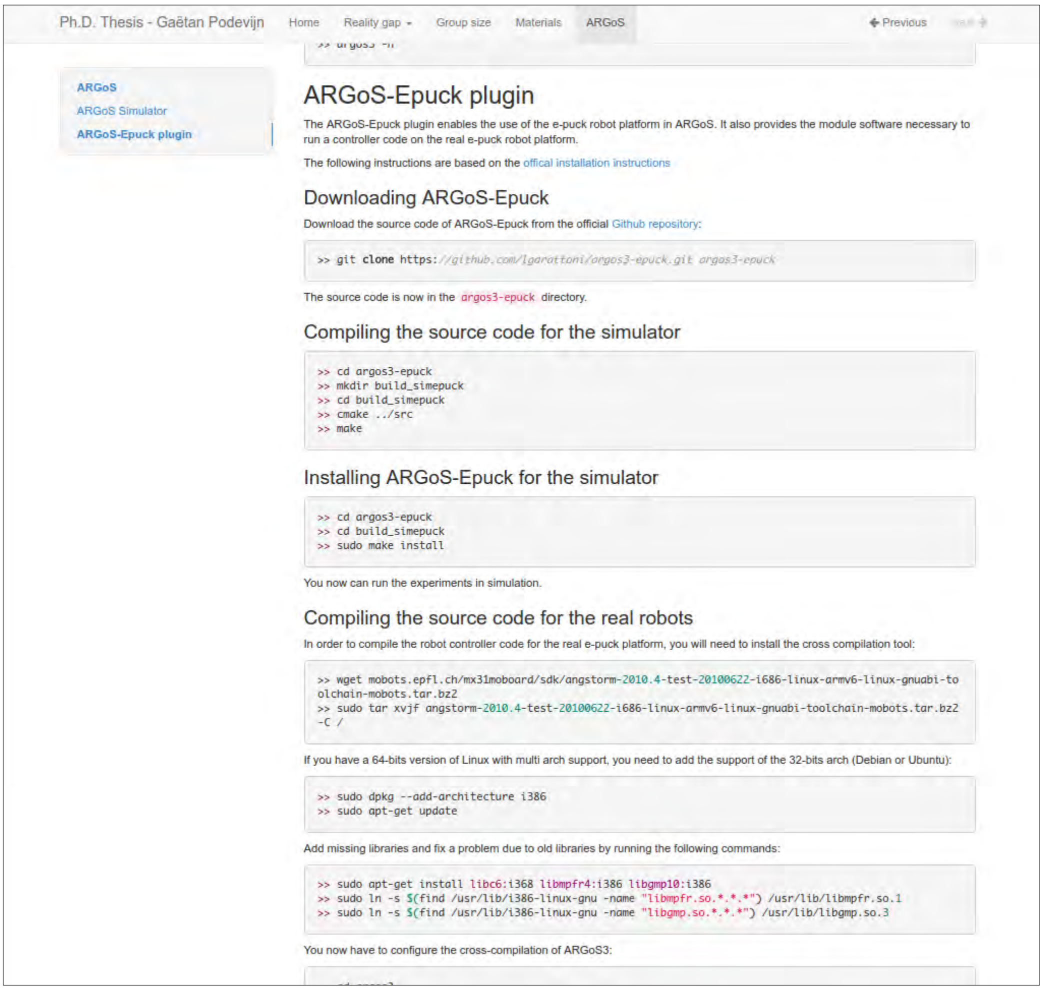


Figure A.5: Screenshot of the ARGoS installation instructions web page. This web page contains the instructions to install both the simulator and the ARGoS-Epuck plugin.

B
Reality Gap

B.1 Ethics Committee Approval

The experiments presented in Chapter 4 were approved by the Ethics Committee of the Faculty of Psychology, Université Libre de Bruxelles. The approval number is: 061/2015. The ethics committee's approval letter is shown in Figure B.1. This letter is in French.

B.2 Consent Form and Information Letter

Before the experiments, we asked the participants to read and sign a consent form (shown in Figure B.2) and an information letter (shown in Figure B.3). These documents are in French.

ULB

COMITE D'ETHIQUE DE LA FACULTE DES SCIENCES PSYCHOLOGIQUES ET DE L'EDUCATION
Président : P. Peigneux
Secrétaire : O. Klein
Membres effectifs : A Bazan, B Dan, V Carette, C Hellemans, C Leys,
Membres suppléants : C Colin, S Kahn, L Licata, I Merckaert, C Mottrie, S Pohl
Adresse de contact :
Prof P. Peigneux, CP191, Avenue F. D. Roosevelt 50, B-1050 Bruxelles (Belgique)
Tel +32 (2) 650 26 39 (secrétariat 4581)
Fax +32 (2) 650 22 09
Email : Philippe.Peigneux@ulb.ac.be

UNIVERSITÉ LIBRE DE BRUXELLES, UNIVERSITÉ D'EUROPE

Bruxelles, le 6 janvier 2016

Demande d'Avis Ethique 061/2015 : «Reality-gap in human-swarm interaction user studies»

Demandeur : PODEVIJN, Gaëtan

Promoteur : DORIGO, Marco

Monsieur,

Le Comité d'Ethique Facultaire a examiné votre demande sous référence dans le cadre de votre projet de recherche académique et émet un avis favorable à votre projet de recherche.

Au nom du Comité d'Ethique Facultaire, je vous souhaite tout le succès possible dans votre entreprise et vous prie d'agréer, Monsieur, l'expression de mes meilleurs sentiments.

Pour le Comité d'Ethique Facultaire,
Prof. Philippe Peigneux, Président

MEMBRE DE L'ACADÉMIE UNIVERSITAIRE WALLONIE-BRUXELLES ET DU PÔLE UNIVERSITAIRE EUROPÉEN BRUXELLES WALLONIE

Figure B.1: Ethics committee's approval letter.

UNIVERSITÉ LIBRE DE BRUXELLES

IRIDIA Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle

Formulaire de consentement

Reality-gap in human-swarm interaction user studies

Gaëtan Podevijn, Rehan O'Grady, Carole Fantini-Hauwel, Marco Dorigo

Objectifs

Le but de cette recherche est d'étudier les différences de réactions lorsqu'un être humain est amené à intéragir avec un essaim de robots (i.e., un groupe de robots autonome) via différent types d'interfaces utilisateurs.

---

1. Je soussigné, ……………………………………………………………………………………………...déclare avoir lu le document d'information et accepte de participer à l'étude "Reality-gap in human-swarm interaction user studies" de G. Podevijn, R. O'Grady, C. Fantini-Hauwel et M. Dorigo.
2. J'ai reçu une explication concernant la nature, le but, la durée de l'étude et j'ai été informé de ce que l'on attend de ma part.
3. Les catégories de données qui seront utilisées dans le cadre de cette étude sont :
   a. Données physiologiques: fréquence cardiaque et activité électrodermale
   b. Réponses données au questionnaire sur la charge cognitive
   c. Enregistrement vidéo de l'expérience
4. J'accepte que ces données fassent l'objet de traitements ultérieurs à des fins scientifiques, en relation directe avec les objectifs de la recherche ci-dessus mentionnés, dans le respect de la loi belge du 8 septembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel. Mon nom, mes réponses aux questionnaires, mes résultats obtenus, mes données physiologiques et mes informations personnelles seront gardés confidentiels. Les responsables scientifiques de cette étude et les personnes qui traiteront les données s'engagent à respecter cette confidentialité de données.
5. J'accepte que les résultats de cette étude, qui seront toujours anonymisés, soient diffusés à des fins scientifiques et en respectant les règles déontologiques de la communauté scientifique.
6. Je peux à tout moment demander la consultation des données à caractère personnel collectées ou leur rectification sans frais. Ces données seront conservées durant le temps nécessaire à leur analyse et ce, jusqu'à un maximum de dix années. Le responsable du traitement de ces données (Gaëtan Podevijn) peut être contacté à l'adresse suivante : gpodevij@ulb.ac.be
7. L'expérience dure approximativement 30 minutes. Je peux décider d'arrêter l'expérience à tout moment sans avoir à justifier ma décision. Dans ce cas, je peux demander de détruire les données récoltées.
8. Je consens de mon plein gré à participer à cette étude.

Veuillez précéder votre signature de la mention "Lu et approuvé."	Date
______________________________	______________________________

Responsable de l'expérience: Gaëtan Podevijn, service IRIDIA-CoDE, ULB.

Figure B.2: Consent form used in the experiments presented in Chapter 4.

Lettre d’information au participant

Reality-gap in human-swarm interaction studies

Gaëtan Podevijn, Rehan O’Grady, Carole Fantini-Hauwel, Marco Dorigo

Madame, Monsieur,

Vous allez participer à une expérience menée en collaboration par le laboratoire IRIDIA (institut de recherches interdisciplinaires et de développements en intelligence artificielle) de la faculté Polytechnique de l’ULB et par le Centre de Recherche en Psychologie clinique, Psychopathologie et Psychosomatique, de la Faculté de Psychologie et des Sciences de l’Éducation de l’ULB. La présente lettre a pour but d’apporter l’ensemble des informations relatives à l’expérience, à son déroulement et à votre participation.

Présentation

L’interaction homme-robot swarm (human-swarm interaction) est un sujet de recherche visant à étudier la manière dont les êtres humains peuvent intéragir avec des systèmes de robots en essaim (swarm robotics). Ces systèmes sont constitués d’un nombre de petits robots autonomes (pouvant varier d’une dizaine à plusieurs centaines).

Actuellement, très peu d’études se concentrent sur la manière dont un être human réagit psychologiquement lorsqu’il intéragit avec ces systèmes de robots. Le but de cette expérience est d’étudier la manière dont un être humain réagit face à ces robots lorsqu’il intéragit avec eux au moyen de différents types d’interfaces utilisateurs.

Déroulement

Cette étude est divisée en trois parties. Dans chacune de ces parties, vous serez amené à réaliser une tâche de supervision (qui vous sera expliquée oralement) d’un essaim de robots. Après chacune de ces parties, nous vous inviterons à répondre à un questionnaire (qui vous sera expliqué oralement). Durant chaque partie de l’expérience, nous collecterons vos données physiologiques (activité électrodermale et activité cardiaque). Vous serez également filmé. Toutes ces données resterons anonymes (la vidéo ne sera pas rendue publique).

Tout au long de l’expérience votre collaboration et votre implication maximale sont requises et doivent être des plus sincères et des plus pertinentes.

Je vous remercie d’avance de votre participation et de votre attention. Je me tiens à votre disposition à tout moment pour de plus amples informations.

Gaëtan Podevijn, IRIDIA, ULB
Responsable de l’étude
gpodevij@ulb.ac.be

Figure B.3: Information letter used in the experiments presented in Chapter 4.

C
Group Size

C.1 Consent Form and Information Letter

Before the experiments, we asked the participants to read and sign a consent form (shown in Figure C.1) and an information letter (shown in Figure C.2). These documents are in French.

Formulaire de consentement éclairé au participant

Je soussigné(e) (en lettres d'imprimerie S.V.P), résidant à :

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

Déclare avoir lu l'information qui précède et accepter de participer à l'étude (veuillez cocher la case correspondante à votre décision)

- [ ] OUI
- [ ] NON

Déclare avoir connaissance que l'expérience est filmée et accepter que ces images soient traitées à des fins scientifiques (veuillez cocher la case correspondante à votre décision)

- [ ] OUI
- [ ] NON

Date et signature :

Figure C.1: Consent form used in the experiment presented in Chapter 5.

Formulaire d'information au participant

Cette étude est menée en collaboration par l'Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle et le Centre de Recherche en Psychologie clinique, Psychopathologie et Psychosomatique, de la faculté de psychologie et des sciences de l'éducation de l'ULB. Elle a pour but d'investiguer les relations entre l'être humain et les essaims de robots appelés couramment « swarmrobots ».

Cette expérience ne présente aucun danger pour l'être humain et est non-invasive. Vous serez mis en présence de robots et vous aurez à répondre à quelques questionnaires. Aussi, des mesures physiologiques telles que le rythme cardiaque, la variabilité cardiaque, l'activité électrodermale et la température corporelle seront prises à l'aide d'un bracelet électronique qui se porte comme une montre ainsi qu'à l'aide de capteurs qui seront placés au niveau des doigts. Enfin, l'entièreté de l'expérience sera filmée. Les données et images ainsi récoltées sont anonymes et confidentielles. Seuls les chercheurs travaillant sur cette étude analyseront celles-ci à des fins scientifiques.

L'expérience dure approximativement 20 minutes. Vous avez le droit de quitter celle-ci à tout moment sans avoir à justifier votre décision auprès des expérimentateurs.

Figure C.2: Information letter used in the experiment presented in Chapter 5.

References

J.A. Adams. Multiple Robot / Single Human Interaction: Effects on Perceived Workload. *Behaviour & Information Technology*, 28(2):183–198, 2009.

R. Aminuddin, A. Sharkey, and L. Levita. Interaction With the Paro Robot May Reduce Psychophysiological Stress Responses. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interation*, pages 593–594, Piscataway, NJ, 2016. IEEE Press.

A.F. Ax. Goals and Methods of Psychophysiology. *Psychophysiology*, 1(1): 8–25, 1964.

N. Ayanian, A. Spielberg, M. Arbesfeld, J. Strauss, and D. Rus. Controlling a Team of Robots with a Single Input. In *Proceedings of IEEE/RSJ International Conference on Robotics and Automation (ICRA)*, pages 1755–1762, Piscataway, NJ, 2014. IEEE Press.

W. S. Bainbridge, editor. *Berkshire Encyclopedia of Human-Computer Interaction.* Berkshire Publishing Company, Limited, Great Barrington, MA, USA, 2004.

S. Bashyal and G.K. Venayagamoorthy. Human Swarm Interaction for Radiation Source Search and Localization. In *Swarm Intelligence Symposium*, pages 1–8, St. Louis, MO, 2008. IEEE.

C.L. Bethel, K. Salomon, R.R. Murphy, and J.L. Burke. Survey of Psychophysiology Measurements Applied to Human-robot Interaction. In *Proceedings of the 16th IEEE International Symposium on Robot and Human interactive Communication (RO-MAN 2007)*, pages 732–737, Piscataway, NJ, 2007. IEEE Press.

M. Birattari. DEMIURGE: automatic design of robot swarms., 2016–2021. ERC Consolidator. `http://demiurge.be`.

W. Boucsein. *Electrodermal activity.* Springer Science & Business Media, 2012.

M. M. Bradley and P. J. Lang. Measuring Emotion: Behavior, Feeling, and Physiology. In Richard D. R. Lane, L. Nadel, G. L. Ahern, J. Allen, and Alfred W. Kaszniak, editors, *Cognitive Neuroscience of Emotion*, pages 25–49. Oxford University Press, 2000.

M. M. Bradley and P.J. Lang. Measuring Emotion: the Self-assessment Manikin and the Semantic Differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

E. Broadbent, Y.I. Lee, R.Q. Stafford, I.H. Kuo, and B.A. MacDonald. Mental Schemas of Robots as More Human-like are Associated with Higher Blood Pressure and Negative Emotions in a Human-robot Interaction. *International Journal of Social Robotics*, 3(3):291–297, 2011.

D.S. Brown and M.A. Goodrich. Limited Bandwidth Recognition of Collective Behaviors in Bio-inspired Swarms. In *Proceedings of the International Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*, pages 405–412. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

J.T. Cacioppo, L.G. Tassinary, and G. Berntson. *Handbook of Psychophysiology*, volume 2. Cambridge University Press, New York, 2007.

J.M. Carroll. Human-computer interaction: Psychology as a science of design. *Annual Review of Psychology*, 48(1):61–83, 1997.

G. Coppin and F. Legras. Autonomy Spectrum and Performance Perception Issues in Swarm Supervisory Control. *Proceedings of the IEEE*, 100(3): 590–603, 2012.

A. Couture-Beil, R.T. Vaughan, and G. Mori. Selecting and Commanding Individual Robots in a Multi-Robot System. In *Proceedings of the Canadian Conference on Computer and Robot Vision (CRV)*, pages 159–166, Los Alamitos, CA, 2010. IEEE Computer Society Press.

M.L. Cummings. Human supervisory control of swarming networks. In *Proceedgins of the 2nd Annual Conference on Swarming: Autonomous Intelligent Networked Systems*, pages 1–9, 2004.

M. Daily, Y. Cho, K. Martin, and D. Payton. World embedded interfaces for human-robot interaction. In *Proceedings of the 36th Annual Hawaii*

International Conference on System Sciences, pages 6–11, Los Alamitos, CA, 2003. IEEE Computer Society Press.

J.-P. De la Croix and M. Egerstedt. Controllability Characterizations of Leader-Based Swarm Interactions. In *AAAI Fall Symposium Series Technical Reports*. AAAI, 2012.

F. Dehais, E.A. Sisbot, R. Alami, and M. Causse. Physiological and Subjective Evaluation of a Human-robot Object Hand-over Task. *Applied Ergonomics*, 42(6):785–791, 2011.

Janez Demšar. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine learning research*, 7:1–30, 2006.

A.C. Dirican and M. Göktürk. Psychophysiological Measures of Human Cognitive States Applied in Human Computer Interaction. *Procedia Computer Science*, 3(0):1361 – 1367, 2011.

A. Dix, J.E. Finlay, G.D. Abowd, and R. Beale. *Human-Computer Interaction (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.

M. Dorigo. The swarm-bot project, 2001–2005. Future and Emerging Technologies (FET-OPEN), http://www.swarmanoid.org.

M. Dorigo. The swarmanoid project., 2006–2010. Future and Emerging Technologies (FET-OPEN), http://www.swarmanoid.org.

M. Dorigo. E-SWARM: Engineering Swarm Intelligence Systems, 2010–2015. ERC Advanced. http://www.e-swarm.org.

M. Dorigo, D. Floreano, L. M. Gambardella, F. Mondada, S. Nolfi, T. Baaboura, M. Birattari, M. Bonani, M. Brambilla, A. Brutschy, D. Burnier, A. Campo, A. L. Christensen, A. Decugnière, G. A. Di Caro, F. Ducatelle, E. Ferrante, A. Förster, J. Guzzi, V. Longchamp, S. Magnenat, J. Martinez Gonzales, N. Mathews, M. Montes de Oca, R. O'Grady, C. Pinciroli, G. Pini, P. Rétornaz, J. Roberts, V. Sperati, T. Stirling, A. Stranieri, T. Stützle, V. Trianni, E. Tuci, A. E. Turgut, and F. Vaussard. Swarmanoid: A Novel Concept for the Study of Heterogeneous Robotic Swarms. *IEEE Robotics & Automation Magazine*, 20(4):60–71, 2013.

M. Elgendi. On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews*, 8(1):14–25, 2012.

J. Fasola and M.J. Matarić. A Socially Assistive Robot Exercise Coach for the Elderly. *Journal of Human-Robot Interaction*, 2(2):3–32, 2013.

R. Fujimura, K. Nakadai, M. Imai, and R. Ohmura. PROT – An Embodied Agent for Intelligible and User-friendly Human-robot Interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3860–3867, Los Alamitos, CA, 2010. IEEE Computer Society Press.

L. Garattoni, G. Francesca, A. Brutschy, C. Pinciroli, and M. Birattari. Software Infrastructure for E-puck (and TAM). Technical Report TR/IRIDIA/2015-004, IRIDIA, Université Libre de Bruxelles, Brussels, Belgium, 2015.

V.J. Gawron. *Human Performance, Workload, and Situational Awareness Measures Handbook*. CRC Press, 2008.

Jill Gerhardt-Powals. Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human-Computer Interaction*, 8(2):189–211, 1996.

A. Giusti, J. Nagi, L. Gambardella, and G.A. Di Caro. Cooperative sensing and recognition by a swarm of mobile robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 551–558, Los Alamitos, CA, 2012. IEEE Computer Society Press.

N. Goljar, M. Javh, J. Poje, J. Ocepek, D. Novak, J. Ziherl, A. Olensek, M. Mihelj, and M. Munih. Psychophysiological Responses to Robot Training in Different Recovery Phases After Stroke. In *Proceedings of the IEEE International Conference on Rehabilitation Robotics (ICORR'11)*, pages 1–6, Piscataway, NJ, 2011. IEEE Press.

M. A. Goodrich, B. Pendleton, S. Kerman, and P.B. Sujit. What Types of Interactions do Bio-inspired Robot Swarms and Flocks Afford a Human? In *Proceedings of the Robotics Science and Systems VIII*, pages 105–112, Cambridge, MA, 2012a. MIT Press.

M.A. Goodrich, B. Pendleton, P.B. Sujit, J. Pinto, and J.W. Crandall. Toward Human Interaction with Bio-inspired Teams. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1265–1266. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

M.A. Goodrich, S. Kerman, and S.-Y. Jun. On Leadership and Influence in Human-Swarm Interaction. In *AAAI Fall Symposium: Human Control of Bioinspired Swarms*. AAAI, 2012b.

Á. Gutiérrez, A. Campo, M. Dorigo, D. Amor, L. Magdalena, and F. Monasterio-Huelin. An Open Localization and Local Communication Embodied Sensor. *Sensors*, 8(11):7545–7563, 2008.

E. C. Haas, K. Pillalamarri, C.C. Stachowiak, and M. Fields. Multimodal Controls for Soldier/Swarm Interaction. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, pages 223–228, Piscataway, NJ, 2011. IEEE Press.

Heiko Hamann. Flora Robotica - Societies of Symbiotic Robot-Plant Bio-Hybrids as Social Architectural Artifacts, 2015–2019. Future and Emerging Technologies (FET-PROACTIVE), http://www.florarobotica.eu.

C. E. Harriott, T. Zhang, G. L. Buford, and J. A. Adams. Measuring Human Workload in a Collaborative Human-Robot Team. *Journal of Human-Robot Interaction*, 4(2):61–96, 2015.

S.G. Hart. NASA-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage Publications, 2006.

S.G. Hart and L.E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52:139–183, 1988.

F. Hegel, C. Muhl, B. Wrede, M. Hielscher-Fastabend, and G. Sagerer. Understanding Social Robots. In *Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions (ACHI 2009)*, pages 169–174. IEEE, 2009.

H. Hexmoor, B. McLaughlan, and M Baker. Swarm Control in Unmanned Aerial Vehicles. In *Proceedings of the 2005 International Conference on Artificial Intelligence (ICAI)*, volume 2, pages 911–917, Georgia, USA, 2005. CSREA Press.

C.M. Humphrey, C. Henk, G. Sewell, B.W. Williams, and J. Adams. Assessing the Scalability of a Multiple Robot Interface. In *Proceedings of the 2nd ACM/IEEE International Conference on Human-robot Interaction (HRI 2007)*, pages 239–246, Piscataway, NJ, 2007. IEEE Press.

ISO, editor. *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices*. 2000.

C. Jost, V. André, B. Le Pévédic, A. Lemasson, M. Hausberger, and D. Duhaut. Ethological Evaluation of Human-robot Interaction: Are Children More Efficient And Motivated With Computer, Virtual Agent or Robots? In *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1368–1373. IEEE, 2012.

W. Ju and D. Sirkin. Animate Objects: How Physical Motion Encourages Public Interaction. In *International Conference on Persuasive Technology*, pages 40–51, Berlin, Germany, 2010. Springer.

G. Kapellmann-Zafra, J. Chen, and R. Groß. Using Google Glass in Human–Robot Swarm Interaction. In *Proceedings of the 17th Annual Conference on Towards Autonomous Robotic Systems (TAROS)*, volume 9716 of *Lecture Notes in Computer Science*, pages 196–201, Berlin, Germany, 2016a. Springer.

G. Kapellmann-Zafra, N. Salomons, A. Kolling, and R. Groß. Human-robot swarm interaction with limited situational awareness. In *Proceedings of the 10th International Conference on Swarm Intelligence (ANTS)*, volume 9882 of *Lecture Notes in Computer Science*, pages 125–136. Springer, Berlin, Germany, 2016b.

G.K. Karavas and P. Artemiadis. On the Effect of Swarm Collective Behavior on Human Perception: Towards Brain-swarm Interfaces. In *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 172–177, Piscataway, NJ, 2015. IEEE Press.

J. Kennedy, P. Baxter, and T. Belpaeme. Comparing Robot Embodiments in a Guided Discovery Learning Interaction with Children. *International Journal of Social Robotics*, 7(2):293–308, 2015.

C.D. Kidd and C. Breazeal. Effect of a Robot on User Perceptions. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and System (IROS)*, volume 4, pages 3559–3564, Los Alamitos, CA, 2004. IEEE Computer Society Press.

C.D. Kidd and C. Breazeal. Human-robot Interaction Experiments: Lessons Learned. In Kerstin Dautenhahn and Rene te Boekhorst, editors, *Proceedings of Robot Companions: Hard Problems and Open Challenges in*

Robot-Human Interaction, Symposium at Social Intelligence and Interaction in Animals, Robots and Agents (AISB), pages 141–142, 2005.

Z. Kira and M. A. Potter. Exerting human control over decentralized robot swarms. In *Proceedings of the 4th International Conference on Autonomous Robots and Agents (ICARA)*, pages 566–571. IEEE, 2009.

A. Kolling, K. Sycara, S. Nunnally, and M. Lewis. Human Swarm Interaction: An Experimental Study of two Types of Interaction with Foraging Swarms. *Journal of Human-Robot Interaction*, 2(2):103–128, 2013.

A. Kolling, P. Walker, N. Chakraborty, K. Sycara, and M. Lewis. Human Interaction With Robot Swarms: A Survey. *IEEE Transaction on Human-Machine Systems*, 46(1):9–26, 2016.

T. Komatsu and Y. Abe. Comparing an On-screen Agent with a Robotic Agent in Non-face-to-face Interactions. In *Proceedings of the International Workshop on Intelligent Virtual Agents*, pages 498–504, Berlin, Germany, 2008. Springer.

K. Kraft and W.D. Smart. Seeing is Comforting: Effects of Teleoperator Visibility in Robot-mediated Health Care. In *Proceedings of the 11th ACM/IEEE International Conference on Human Robot Interaction*, pages 11–18, Pascataway, NJ, 2016. IEEE Press.

A. F. Kramer. Physiological Metrics of Mental Workload: A Review of Recent Progress. In Diane Damos, editor, *Multiple-Task Performance*, chapter 11, pages 279–328. Taylor and Francis, London, 1991.

D. Kulić and E.A. Croft. Estimating Intent for Human-robot Interaction. In *Proceedings of the IEEE International Conference on Advanced Robotics*, pages 810–815. IEEE, 2003.

D. Kulić and E.A. Croft. Anxiety Detection During Human-robot Interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, pages 616–621, Los Alamitos, CA, 2005. IEEE Computer Society Press.

D. Kulić and E.A. Croft. Affective State Estimation for Human-Robot Interaction. *IEEE Transactions on Robotics*, 23(5):991–1000, 2007a.

D. Kulić and E.A. Croft. Physiological and Subjective Responses to Articulated Robot Motion. *Robotica*, 25(01):13–27, 2007b.

P. J. Lang. Behavioral Treatment and Bio-behavioral Assessment: Computer Applications. In J. B. Sidowski, J. H. Johnson, and T. H. Williams, editors, *Technology in Mental Health Care Delivery Systems*, pages 119–137. Ablex, 1980.

N. Lazzeri, D. Mazzei, A. Greco, A. Rotesi, A. Lanatà, and D.E. De Rossi. Can a Humanoid Face Be Expressive? A Psychophysiological Investigation. *Frontiers in Bioengineering and Biotechnology*, 3, 2015.

I. Leite, A. Pereira, C. Martinho, and A. Paiva. Are Emotional Robots More Fun to Play With? In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, pages 77–82, Piscataway, NJ, 2008. IEEE Press.

J. Li. The Benefit of Being Physically Present: A Survey of Experimental Works Comparing Copresent Robots, Telepresent Robots and Virtual Agents. *International Journal of Human-Computer Studies*, 77:23–37, 2015.

R. Looije, A. Van der Zalm, M.A. Neerincx, and R.-J. Beun. Help, I Need Some Body the Effect of Embodiment on Playful Learning. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, pages 718–724, Piscataway, NJ, 2012. IEEE Press.

J.-M. Lu and Y.-L. Hsu. The Use of Psychophysiological Methods in the Evaluation of Mental Commitment Robots for Elderly Care. *Journal of Industrial and Production Engineering*, 32(7):449–456, 2015.

I Scott MacKenzie. *Human-Computer Interaction: An Empirical Research Perspective*. Newnes, 2012.

R.L. Mandryk, K.M. Inkpen, and T.W. Calvert. Using Psychophysiological Techniques to Measure User Experience with entertainment Technologies. *Behaviour & Information Technology*, 25(2):141–158, 2006.

M.D. Manning, C.E. Harriott, S.T. Hayes, J.A. Adams, and A.E. Seiffert. Heuristic Evaluation of Swarm Metrics' Effectiveness. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 17–18, New York, NY, 2015. ACM Press.

James Marshall. DiODe - Distributed Algorithms for Optimal Decision-Making, 2015–2020. ERC Consolidator. http://diode.group.shef.ac.uk.

J. McLurkin, J. Smith, J. Frankel, D. Sotkowitz, D. Blau, and B. Schmidt. Speaking Swarmish: Human-robot Interface Design for Large Swarms of Autonomous Mobile Robots. In *AAAI Spring Symposium*, pages 72–75, Palo Alto, CA, 2006. AAAI Press.

Albert Mehrabian. Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. *Current Psychology*, 14(4):261–292, 1996.

V.M. Monajjemi, J. Wawerla, R. Vaughan, and G. Mori. Hri in the Sky: Creating and Commanding Teams of UAVs with a Vision-mediated Gestural Interface. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 617–623, Los Alamitos, CA, 2013. IEEE Computer Society Press.

F. Mondada, M. Bonani, X. Raemy, J. Pugh, C. Cianci, A. Klaptocz, S. Magnenat, J. Zufferey, D. Floreano, and A. Martinoli. The E-Puck, a Robot Designed for Education in Engineering. In *Proceedings of the 9th Conference on Autonomous Robot Systems and Competitions*, pages 59–65. Instituto Politècnico de Castelo Branco, Portugal, 2009.

N. Munekata, M. Tsujimoto, and T. Ono. Robot Behavior Control System Based on User's Attitude Using SCL. In *Proceedings of the 24th IEEE Internation Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 197–201, Piscataway, NJ, 2015. IEEE Press.

S. Nagavalli, L. Luo, N. Chakraborty, and K. Sycara. Neglect benevolence in human control of robotic swarms. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 6047–6053, Piscataway, NJ, 2014. IEEE Press.

S. Nagavalli, S. Chien, M. Lewis, N. Chakraborty, and K. Sycara. Bounds of Neglect Benevolence in Input Timing for Human Interaction with Robotic Swarms. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 197–204, New York, 2015. ACM Press.

A.M. Naghsh, J. Gancet, A. Tanoto, and C. Roast. Analysis and Design of Human-robot Swarm Interaction in Firefighting. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, pages 255–260, Piscataway, NJ, 2008. IEEE Press.

J. Nagi, H. Ngo, A. Giusti, L.M. Gambardella, J. Schmidhuber, and G.A. Di Caro. Incremental learning using partial feedback for gesture-based human-swarm interaction. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, pages 898–905, Piscataway, NJ, 2012. IEEE Press.

J. Nagi, A. Giusti, L. Gambardella, and G. A. Di Caro. Human-Swarm Interaction Using Spatial Gestures. In *Proceedings of the 27th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3834–3841, Los Alamitos, CA, 2014. IEEE Computer Society Press.

J. Nagi, H. Ngo, L. Gambardella, and G. A. Di Caro. Wisdom of the Swarm for Cooperative-Decision Making in Human-Swarm Interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Piscataway, NJ, 2015. IEEE Press.

D Novak, J. Ziherl, A. Olenšek, M. Milavec, J. Podobnik, M. Mihelj, and M. Munih. Psychophysiological Responses to Robotic Rehabilitation Tasks in Stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(4):351–361, 2010.

D. Novak, B. Beyeler, X. Omlin, and R. Riener. Workload Estimation in Physical Human-robot Interaction Using Physiological Measurements. *Interacting with Computers*, 27(6):616–629, 2015.

S. Nunnally, P. Walker, A. Kolling, N. Chakraborty, M. Lewis, K. Sycara, and M. Goodrich. Human Influence of Robotic Swarms with Bandwidth and Localization Issues. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 333–338, Piscataway, NJ, 2012. IEEE Press.

S. Nunnally, P. Walker, N. Chakraborty, M. Lewis, and K. Sycara. Using Coverage for Measuring the Effect of Haptic Feedback in Human Robotic Swarm Interaction. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 516–521, Piscataway, NJ, 2013. IEEE Press.

Open Science Collaboration. Estimating the Reproducibility of Psychological Science. *Science*, 349(6251):aac4716, 2015.

B. Özcan, D. Caligiore, V.i Sperati, T. Moretta, and G. Baldassarre. Transitional Wearable Companions: A Novel Concept of Soft Interactive Social Robots to Improve Social Skills in Children with Autism Spectrum Disorder. *International Journal of Social Robotics*, pages 1–11, 2016.

J. Penders. GUARDIANS: Group of Unmanned Assistant Robots Deployed in Aggregative Navigation supported by Scent detection, 2006–2010. Specific Targeted Research Project, http://cordis.europa.eu/project/rcn/80210_en.html.

B. Pendleton and M.A. Goodrich. Scalable Human Interaction with Robotic Swarms. In *Proceedings of the AIAA Infotech@Aerospace Conference*, pages 633–645, Va, USA, 2013. American Institute of Aeronautics and Astronautics.

A. Pereira, C. Martinho, I. Leite, and A. Paiva. iCat, The Chess Player: The Influence of Embodiment in the Enjoyment of a Game. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1253–1256. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

C. Pinciroli. *On the Design and Implementation of an Accurate, Efficient, and Flexible Simulator for Heterogeneous Swarm Robotics Systems.* PhD thesis, IRIDIA, Université libre de Bruxelles, 2014.

C. Pinciroli, V. Trianni, R. O'Grady, G. Pini, A. Brutschy, M. Brambilla, N. Mathews, E. Ferrante, G. Di Caro, F. Ducatelle, M. Birattari, L. M. Gambardella, and M. Dorigo. ARGoS: a modular, parallel, multi-engine simulator for multi-robot systems. *Swarm Intelligence*, 6(4):271–295, 2012.

G. Podevijn, R. O'Grady, and M. Dorigo. Self-organised Feedback in Human Swarm Interaction. In *Workshop on Robot Feedback in Human-Robot Interaction: How to Make a Robot Readable for a Human Interaction Partner (Ro-Man 2012)*, Paris, France, 2012.

G. Podevijn, R. O'Grady, Y.S.G. Nashed, and M. Dorigo. Gesturing at Subswarms: Towards Direct Human Control of Robot Swarms. In A. Natraj, S. Cameron, C. Melhuish, and M. Witkowski, editors, *14th Annual Conference on Towards Autonomous Robotic Systems (TAROS 2013)*, volume 8069 of *Lecture Notes in Computer Science*, pages 390–403, Berlin, Germany, 2013. Springer.

S. Pourmehr, V.M. Monajjemi, R.T. Vaughan, and G. Mori. "You two! Take off!": Creating, Modifying and Commanding Groups of Robots Using Face Engagement and Indirect Speech in Voice Commands. In *Proceedings of the 26th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 137–142, Los Alamitos, CA, 2013. IEEE Computer Society Press.

A. Powers, S. Kiesler, S. Fussell, and C. Torrey. Comparing a Computer Agent with a Humanoid Robot. In *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 145–152, New York, 2007. ACM.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL https://www.R-project.org.

P. Rani and N. Sarkar. Making Robots Emotion-sensitive – Preliminary Experiments and Results. In *IEEE International Workshop on Robot and Human Interactive Communication*, pages 1–6. IEEE, 2005.

P. Rani, J. Sims, R. Brackin, and N. Sarkar. Online Stress Detection Using Psychophysiological Signals for Implicit Human-robot Cooperation. *Robotica*, 20(06):673–685, 2002.

P. Rani, N. Sarkar, C. A. Smith, and L.D. Kirby. Anxiety Detecting Robotic System–Towards Implicit Human-robot Collaboration. *Robotica*, 22(01): 85–95, 2004.

P. Rani, C. Liu, and N. Sarkar. Affective feedback in closed loop human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, pages 335–336. ACM, 2006.

H. Robinson, B. MacDonald, and E. Broadbent. Physiological Effects of a Companion Robot on Blood Pressure of Older People in Residential Care Facility: A Pilot Study. *Australasian journal on Ageing*, 34(1):27–32, 2015.

N. Salomons, G. Kapellmann-Zafra, and R. Groß. Human Management of a Robotic Swarm. In *Proceedings of the 17th Annual Conference on Towards Autonomous Robotic Systems (TAROS)*, volume 9716 of *Lecture Notes in Computer Science*, pages 282–287, Berlin, Germany, 2016. Springer.

G. Salvendy. *Handbook of Human Factors and Ergonomics*. John Wiley & Sons, Hoboken, NJ, USA, 2012.

N. Sarkar. Psychophysiological Control Architecture for Human-robot Coordination – Concepts and Initial Experiments. In *Proceedings of the IEEE International Conference on Robotics and Automation, (ICRA)*, volume 4, pages 3719–3724. IEEE, 2002.

Davide Scaramuzza. sFly: Swarm of Micro Flying Robots, 2009–2012. Collaborative Project, http://www.sfly.org.

J. Scheirer, R. Fernandez, J. Klein, and R. Picard. Frustrating the User on Purpose: A Step Toward Building an Affective Computer. *Interacting with Computers*, 14(2):93–118, 2002.

Thomas Schmickl. CoCoRo - Collective Cognitive Robots, 2011–2014. Collaborative Project, http://cocoro.uni-graz.at/drupal.

T. Setter, A. Fouraker, H. Kawashima, and M. Egerstedt. Haptic Interactions With Multi-Robot Swarms Using Manipulability. *Journal of Human-Robot Interaction*, 4(1):60–74, 2015.

J. Seyfried. I-SWARM - Intelligent Small World Autonomous Robots for Micro-manipulation, 2004–2008. Future and Emerging Technologies (FET-PROACTIVE), http://cordis.europa.eu/project/rcn/71243_en.html.

T.B. Sheridan and W.L. Verplank. Human and Computer Control of Undersea Teleoperators. Technical report, DTIC Document, 1978.

N. Shirzad and H.F.M. Van der Loos. Evaluating the User Experience of Exercising Reaching Motions with a Robot that Predicts Desired Movement Difficulty. *Journal of Motor Behavior*, 48(1):31–46, 2016.

J.A. Stern. Toward a Definition of Psychophysiology. *Psychophysiology*, 1 (1):90–91, 1964.

R. M. Stern, W.J. Ray, and K.S. Quigley. *Psychophysiological Recording*. Oxford University Press, New York, 2nd edition, 2001.

M. Swangnetr and D.B. Kaber. Emotional State Classification in Patient-robot Interaction Using Wavelet Analysis and Statistics-based Feature Selection. *IEEE Transactions on Human-Machine Systems*, 43(1):63–75, 2013.

M. Swangnetr, B. Zhu, K.B. Taylor, and D.B. Kaber. Assessing the Effects of Humanoid Robot Features on Patient Emotion During a Medicine Delivery Task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 349–353. SAGE Publications, 2010.

P. Sweetser and P. Wyeth. GameFlow: a Model for Evaluating Player Enjoyment in Games. *Computers in Entertainment (CIE)*, 3(3):3–3, 2005.

K. Sycara, C. Lebiere, Y. Pei, D. Morrison, Y. Tang, and M. Lewis. Abstraction of Analytical Models from Cognitive Models of Human Control of

Robotic Swarms. In N.A. Taatgen, M.K. M.K. Van Vugt, J. P. Borst, and J. P. Mehlhorn, editors, *Proceedings of the 13th International Conference on Cognitive Modeling*. Penn State, 2015.

J. Sykes and S. Brown. Affective Gaming: Measuring Emotion through the Gamepad. In *Extended Abstracts on Human Factors in Computing Systems, CHI 2003*, pages 732–733. ACM, 2003.

S.B. Tabibian, M. Lewis, C. Lebiere, N. Chakraborty, K. Sycara, S. Bennati, and M. Oishi. Towards a Cognitively-based Analytic Model of Human Control of Swarms. In *AAAI Spring Symposium Series*, pages 68–73, Palo Alto, CA, 2014. AAAI Press.

Task Force of the European Society of Cardiology et al. The North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation*, 93(5):1043–1065, 1996.

G. Teo, L. Reinerman-Jones, G. Matthews, D. Barber, J. Harris, and I. Hudson. Augmenting Robot Behaviors Using Physiological Measures of Workload State. In *International Conference on Augmented Cognition*, pages 404–415. Springer, 2016.

L. Tiberio, A. Cesta, G. Cortellessa, L. Padua, and A.R. Pellegrino. Assessing Affective Response of Older Users to a Telepresence Robot Using a Combination of Psychophysiological Measures. In *Proceedings of the 21st IEEE Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 833–838, Piscataway, NJ, 2012. IEEE Press.

T. Tullis and W. Albert. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2008.

P. Velagapudi, P. Scerri, K. Sycara, H. Wang, M. Lewis, and J. Wang. Scaling Effects in Multi-robot Control. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008)*, pages 2121–2126, Los Alamitos, CA, 2008. IEEE Computer Society Press.

J. Wainer, D.J. Feil-Seifer, D. Shell, and M.J. Matarić. Embodiment and Human-robot Interaction: A Task-based Perspective. In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN)*, pages 872–877, Piscataway, NJ, 2007. IEEE Press.

P. Walker, S. Nunnally, M. Lewis, A. Kolling, N. Chakraborty, and K. Sycara. Neglect Benevolence in Human Control of Swarms in the Presence of Latency. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3009–3014, Piscataway, NJ, 2012. IEEE Press.

P. Walker, S.A. Amraii, M. Lewis, N. Chakraborty, and K. Sycara. Human Control of Leader-Based Swarms. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2712–2717, Piscataway, NJ, 2013a. IEEE Press.

P. Walker, S. Nunnally, M. Lewis, N. Chakraborty, and K. Sycara. Levels of Automation for Human Influence of Robot Swarms. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 429–433. SAGE Publications, 2013b.

P. Walker, S.A. Amraii, N. Chakraborty, M. Lewis, and K. Sycara. Human Control of Robot Swarms with Dynamic Leaders. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1108–1113, Los Alamitos, CA, 2014. IEEE Computer Society Press.

P. Walker, M. Lewis, and K. Sycara. Characterizing Human Perception of Emergent Swarm Behaviors. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Piscataway, NJ, 2016a. IEEE Press. In press.

P. Walker, M. Lewis, and K. Sycara. The Effect of Display Type on Operator Prediction of Future Swarm States. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Piscataway, NJ, 2016b. IEEE Press. In press.

B. Walter, A. Sannier, D. Reiners, and J. Oliver. UAV swarm control: Calculating Digital Pheromone Fields with the GPU. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 3(3): 167–176, 2006.

R.D. Ward and P.H. Marsden. Physiological responses to different WEB page designs. *International Journal of Human Computer Studies*, 59(1-2): 199–212, 2003.

D.G. Wastell and M. Newman. Stress, Control and Computer System Design: a Psychophysiological Field Study. *Behaviour & Information Technology*, 15(3):183–192, 1996.

AT Welford. Choice reaction time: Basic concepts. In Welford AT, editor, *Reaction times*, pages 73–128. Academic Press, New York, 1980.

A. Winfield. SYMBRION - Symbiotic Evolutionary Robot Organisms, 2008–2013. Future and Emerging Technologies (FET-PROACTIVE), `http://cordis.europa.eu/project/rcn/85478_en.html`.

M. Wirsing. ASCENS - Autonomic Service-Component Ensembles, 2010–2015. Collaborative Project, `http://www.ascens-ist.eu`.

J. Wrobel, Y-H. Wu, H. Kerhervé, L. Kamali, A-S. Rigaud, C. Jost, B. Le Pévédic, and D. Duhaut. Effect of Agent Embodiment on the Elder User Enjoyment of a Game. In *Proceedings of the 6th International Conference on Advances in Computer-Human Interactions (ACHI)*, pages 162–167. IARIA XPS Press, 2013.

E. Yang and M.C. Dorneich. The Effect of Time Delay on Emotion, Arousal, and Satisfaction in Human-Robot Interaction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 443–447. SAGE Publications, 2015.

T. Zhang, D.B. Kaber, B. Zhu, M. Swangnetr, P. Mosaly, and L. Hodge. Service Robot Feature Design Effects on User Perceptions and Emotional Responses. *Intelligent Service Robotics*, 3(2):73–88, 2010.