Swarm Intell (2016) 10:193–210
DOI 10.1007/s11721-016-0124-3

Investigating the effect of increasing robot group sizes on the human psychophysiological state in the context of human–swarm interaction

Gaëtan Podevijn[1] · Rehan O'Grady[1] · Nithin Mathews[1] · Audrey Gilles[2] · Carole Fantini-Hauwel[2] · Marco Dorigo[1]

Received: 25 January 2016 / Accepted: 4 June 2016 / Published online: 22 June 2016

Abstract We study the psychophysiological state of humans when exposed to robot groups of varying sizes. In our experiments, 24 participants are exposed sequentially to groups of robots made up of 1, 3 and 24 robots. We measure both objective physiological metrics (skin conductance level and heart rate), and subjective self-reported metrics (from a psychological questionnaire). These measures allow us to analyse the psychophysiological state (stress, anxiety, happiness) of our participants. Our results show that the number of robots to which a human is exposed has a significant impact on the psychophysiological state of the human and that higher numbers of robots provoke a stronger response.

Keywords Swarm robotics · Human–swarm interaction · Psychophysiology

1 Introduction

In the near future, swarms of autonomous robots are likely to be used in an array of professional and domestic tasks (Dorigo et al. 2013, 2014). An increasing body of research is devoted to studying what form interaction between humans and robot might take (e.g., Kolling et al. 2013; Nagavalli et al. 2015; Nagi et al. 2014, 2015; Nunnally et al. 2012; Podevijn et al. 2013). This nascent field is known as human–swarm interaction (HSI)—see Kolling et al. (2016) for a comprehensive review of the human–swarm interaction literature. However, no study to date has rigorously addressed the fundamental problem of how exposure to a swarm of robots may affect the psychological state of a human being. Intuitively, one might expect that exposure to large numbers of robots could provoke responses such

Guest editor: Roderich Groß.

✉ Gaëtan Podevijn
gpodevij@ulb.ac.be

1 IRIDIA, Université libre de Bruxelles, Brussels, Belgium

2 Research Centre of Clinical Psychology, Psychopathology and Psychosomatic, Université libre de Bruxelles, Brussels, Belgium

as stress or anxiety. Such responses are common when exposed to swarms of insects (e.g., apiphobia is the phobia of bees, myrmecophobia is the phobia of ants).

The goal of this study 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 to date. Some research has considered the role of human psychology in human–swarm interaction. However, the 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). 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 varying amounts of robots. Secondly, we adopt a rigorous, objective methodology. The majority of previous studies have used simulated robots, while our experiments use exclusively real robots (De la Croix and Egerstedt 2012; Kolling et al. 2013; Nagavalli et al. 2015; Nunnally et al. 2012; Pendleton and Goodrich 2013). More importantly, while previous studies have relied on subjective questionnaires to determine psychological effects, for the first time in the field of human–swarm interaction we use objective psychophysiological measures.

In our study, we measure the psychophysiological state of 24 participants. The psychophysiological state reflects the psychological state of a human (e.g., stress, anxiety, excitement) based on his or her physiological responses (Blascovich et al. 2011). These measures are considered objective, unlike the questionnaires used in all studies to date, which are prone to subjective interpretation by the participant (Bethel et al. 2007). We measure our participants' psychophysiological state during a passive interaction (i.e., our participants do not send commands to nor receive specific feedback from the robots) with an increasing number of robots (see Fig. 1). 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 psychophysiological reactions with an extra interaction interface (e.g., joystick, keyboard, voice commands).

The results of our user study show that humans have a stronger physiological responses when confronted with larger numbers of robots.

2 Related literature

To the best of our knowledge, there is no work that considers the human psychophysiological state in the context of human–swarm interaction. However, the use of dedicated psychological questionnaires were used in human–swarm interaction to study the human psychological state, and more specifically the human workload level, i.e., the mental effort (De la Croix and Egerstedt 2012; Pendleton and Goodrich 2013; Setter et al. 2015).

In human–swarm interaction, only Pendleton and Goodrich (2013) consider the human psychological state by studying the workload level of human operators controlling a varying number of robots. The authors show that the human workload is not dependent on the number of robots when interacting with a swarm robotics system. However, we believe that their study may only be conveying a limited message about the human psychological state as the user study was limited to simulated robots and a questionnaire-based data collection methodology. User studies performed with simulated robots do not necessarily reflect the same conditions as real robots. Furthermore, participants might not always answer psychological questionnaires objectively, that is, they might not answer what they felt during the experiment; rather, they

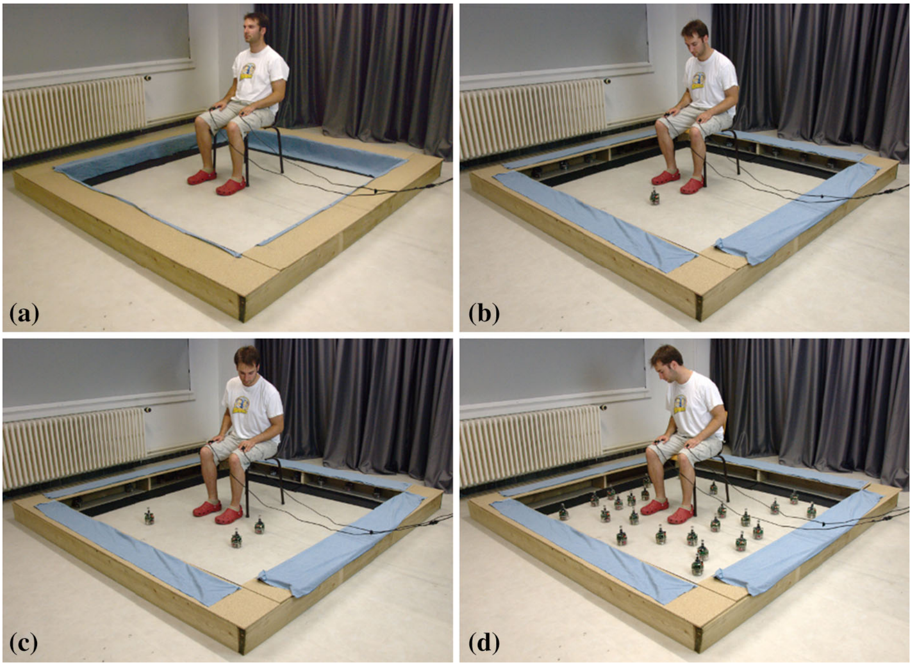

Fig. 1 An example of an experiment in progress. **a** At the beginning of the experiment, the participant is attached to two physiological sensors. **b** The experiment begins with one robot moving around the participant. **c** Subsequently, two more robots appear and the participant is exposed to a group of three robots. **d** Finally, 21 robots appear and the participant is exposed to a group of 24 robots. The participant shown in this figure is the first author of this paper and did not take part in the experiment. The pictures shown in this figure were taken for illustration purpose (Color figure online)

might answer what they believe the experimenter would like them to answer (Bethel et al. 2007).

While in swarm robotics the human operator's workload does not seem to be affected by the number of robots in a robot swarm, in multi-robot systems, the human operator's workload seems to be affected by the number of robots. 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 their 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 4, 8 and 12 simulated robots. The results of their experiments revealed that their 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—her participants' workload was higher when they were doing the experiment with four robots compared to two robots and one robot. These studies required active involvement of the participants. Active involvement clearly gives a richer interaction experience to the participant. However, it also renders the results harder to objectively interpret—it is not clear to what extent the participant is responding to intrinsic difficulties of multi-robot interaction, versus reacting to idiosyncrasies of the particular user interface used in each particular study.

In human–robot interaction in general, very few existing studies have adopted psychophysiological measures, despite the fact that they are considered much more objective than measures based on psychological questionnaires. Some researchers used psychophysiological measures to study anxiety levels of humans operating a robot manipulator arm (Dehais et al. 2011; Kulic and Croft 2005, 2007). Other researchers studied to what extent different physical features (e.g., appearance and vocal properties) of a service robot (e.g., a robot that assists a nurse for medicine delivery) can affect the patients' psychophysiological state under treatment (Swangnetr and Kaber 2013; Swangnetr et al. 2010; Zhang et al. 2010). Tiberio et al. (2012) use psychophysiological measures to study the response to a telepresence robot of elderly people suffering from mild cognitive impairment—a mobile robot mounted with a screen allowing a human to interact with the user via videoconference. Psychophysiological measures have also been used in the context of virtual rehabilitation where stroke patients used a haptic robotic interface that allowed them to feel virtual objects (Goljar et al. 2011; Novak et al. 2010). For completeness, we also mention other studies we are aware of where psychophysiology has been used in human–robot interaction, even though these studies have very different objectives from our own. Researchers have used psychophysiology to modify a robot's behaviour with respect to psychophysiological responses of a participant (Bekele and Sarkar 2014; Itoh et al. 2006; Rani and Sarkar 2005; Rani et al. 2004). Finally, psychophysiology has also been used in the field of adaptive automation (Byrne and Parasuraman 1996; Prinzel et al. 2000, 2003; Ting et al. 2010)—the psychophysiological state of a human is monitored in order to change the level of automation of a system.

3 Methodology

3.1 Experimental scenario

We designed an experimental scenario that allowed us to study the effect of an increasing number of robots on the human psychophysiological state. 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 3 and 24 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.

3.2 Measures

In our experiment, we used psychophysiological measures to determine objectively the psychological state of our participants based on physiological responses. These psychophysiological measures are considered objective because it is difficult for humans to intentionally manipulate their physiological responses (for instance to intentionally decrease heart rate). In addition to using psychophysiological measures, we also used self-reported measures (i.e., data gathered from our participants using a dedicated psychological questionnaire) to determine whether our participants are subjectively conscious of their psychological state change and whether this state change is positive (i.e., our participants report a positive experience) or negative (i.e., our participants report a negative experience). In the following two sections,

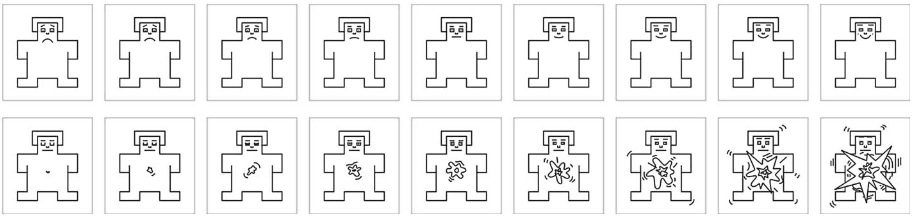

Fig. 2 Self-Assessment Manikin scales. *Top* the valence scale. The *left-most picture* corresponds to the lowest level of valence. The *right-most picture* corresponds to the highest level of valence. *Bottom* the arousal scale. The *left-most picture* corresponds to the lowest level of arousal. The *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)

we present the self-reported measures and the psychophysiological measures used in this study.

3.2.1 Self-reported measure

We used the Self-Assessment Manikin (SAM) questionnaire to collect our participants' self-reported affective state (Lang 1980). The affective state is measured with two scales: *valence* and *arousal*. In the version of the SAM questionnaire used in this study (see Fig. 2), 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.

Valence is the cognitive judgement (i.e., pleasure or displeasure) of an evaluation such as the interaction with robots considered in this study. Higher values of valence 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. Each picture of each scale 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).

3.2.2 Psychophysiological measure

The peripheral nervous system, part of the nervous system, is responsible for transmitting information between the organs and the central nervous system. The peripheral nervous system can be subdivided into the somatic nervous system and the autonomic nervous system. The somatic nervous system is associated with the voluntary control of the body movements (e.g., the muscles). The autonomic nervous system is associated with the involuntary functions of the body (e.g., cardiovascular activity). These involuntary functions of the body can, themselves, be associated with the change of several psychological states (e.g., emotion, attention, motivation, arousal) (Blascovich et al. 2011). The autonomic nervous system is made up of the sympathetic nervous system and of the parasympathetic nervous system. The role of the sympathetic nervous system is to activate physiological responses (e.g., to a stimulus). The role of the parasympathetic nervous system is opposed to the role of the sympathetic nervous system—it is responsible for maintaining physiological responses to a normal activity (i.e., the physiological responses in absence of stimulus).

We study our participants' physiological activity when they are exposed to increasing robot group sizes. In particular, we study two physiological activities of the autonomic nervous system—the cardiovascular activity and the electrodermal activity (i.e., the skin's electrical activity). We study the cardiovascular activity by measuring our participants' heart rate. We study the electrodermal activity by measuring our participants' skin conductance.

The heart rate is the number of heart beats per unit of time. It is usually measured in beats per minute (BPM). An increase of the heart rate can be due to either an increase in the sympathetic nervous system activity, an increase of the parasympathetic nervous system activity, or a combination of both. We analyse the skin conductance by measuring the skin conductance level (SCL). The skin conductance level, measured in microsiemens (μS), is a slow variation of the skin conductance over time without the occurrence of any stimuli. An increase of the skin conductance level is only due to an increase of the sympathetic nervous system activity. The skin conductance level is commonly used as a measure of the physiological arousal (Blascovich et al. 2011).

Physiological responses can significantly vary from individual to individual (e.g., heart rate and skin conductance can show different values for different individuals for similar experimental conditions). Because the physiological responses can vary from an individual to an other, it can be difficult to compare the physiological responses of an individual with those of another. In order to be able to compare the physiological responses of an individual with those of another individual, we first recorded our participants' physiological baseline, i.e., the heart rate and SCL values at rest during 5 min. Then, we ran our experiment and recorded the participants physiological values. Finally, we subtracted the mean baseline values from those recorded during the experiment. This allows us to compare the difference between our participants' physiological responses during the baseline period (which can be highly different from one participant to another) and during the experiment.

3.3 Equipment and experimental setup

3.3.1 Physiological response acquisition

The physiological responses were recorded using a PowerLab 26T (ADInstruments) data acquisition system augmented with a GSR Amp device. The PowerLab 26T was connected via USB 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 used an infrared photoelectric sensor (i.e., a photoplethysmograph) to measure the blood volume pulse (BVP) of our participants (i.e., changes in the pulsatile blood flow). The blood volume pulse can be retrieved from the photoplethysmograph from the peripheral parts of the human body such as on the fingers. We can compute the heart rate from the blood volume pulse. Firstly, we calculate the inter-beat interval (i.e., 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 shows the blood volume pulse of a participant during a time window of 10 s. The photoplethysmograph was attached to the index finger of a participant's dominant hand. The photoplethysmograph was directly connected to the PowerLab 26T.

To monitor the electrodermal activity of our participants, we used brightly polished stainless steel bipolar electrodes connected to the GSR Amp device. These bipolar electrodes

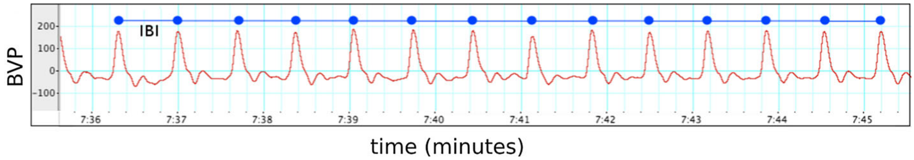

Fig. 3 The graph of a participant's blood volume pulse (BVP) during 10 s. The BVP does not have a standard unit. The x-axis is the time in minutes since the beginning of the recording. The time between two peaks (depicted with two *dots* connected with a *line* on the picture) is called the inter-beat 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 (Color figure online)

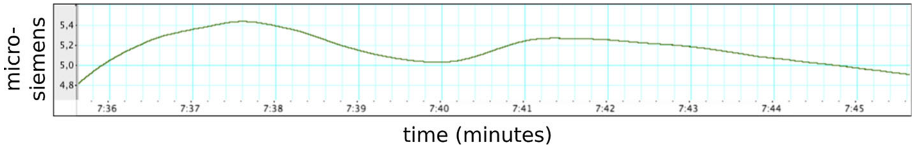

Fig. 4 The graph of a participant's skin conductance during 10 s. The skin conductance's unit is the microsiemens (y-axis). The x-axis is the time in minutes since the beginning of the recording. The skin conductance is computed by measuring the current flowing between two electrodes and by dividing this current by a constant voltage applied between the electrodes. The skin conductance level of this participant during these 10 s is of 5.17 (μS) (Color figure online)

were attached to the medial phalanxes of the index and middle fingers of a participant's non-dominant hand. 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 the 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 4 shows the skin conductance of a participant during a time window of 10 s.

3.3.2 Environment

In our experiment, each participant is immersed in an environment in which he or she is exposed to different numbers of robots. We used a square environment of dimension 200 cm $\times$ 200 cm (see Fig. 5). Adjacent to the inner environment walls, we added four covered hidden zones of 25 cm width each. These hidden zones prevent all pre-experimental visual contact between a participant—who is asked to remain seated on a chair placed in the centre of the environment—and the robots. A curtain is installed at the entrance of each hidden zone such that the robots cannot be seen by a participant while entering the experiment room. Initially, all robots are placed inside the hidden zones. When the experiment starts, the robots gradually drive out. We also added a dark area of 175 cm $\times$ 20 cm in front of each hidden zone entrance. This dark area was used to prevent the robots from accidentally returning back into a hidden zone and hence disappearing from the experiment once they were already in an area visible to the participant (see Sect. 3.3.3 for a detailed explanation of the robot behaviour implementation).

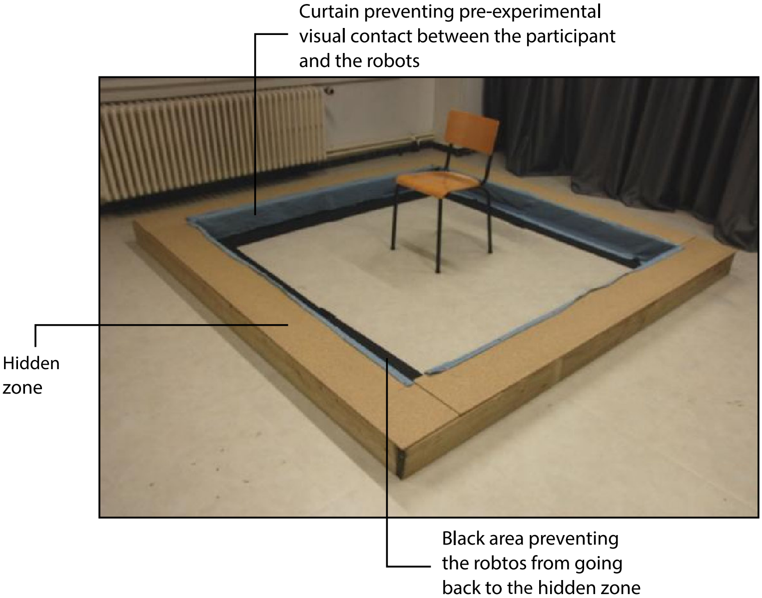

Fig. 5 The environment considered in this study with its four hidden zones and a chair for the participant. Six robots are placed under each hidden zone. The *blue* curtains prevent all visual contact between our participants and the robots prior to the beginning of an experiment. Once the participant is seated, the experimenter raises the curtains. The hidden zones are designed such that the robots remain invisible to the participant sitting on the chair until they drive out of the hidden zones (Color figure online)

3.3.3 Robot behaviour

During an experiment, robots gradually drive out of the hidden zones in order for a participant to be exposed to three different robot group sizes, i.e., 1 robot, 3 robots and 24 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 24 robots, 21 robots drive out of their hidden zone at the same time.

Once the robots become visible to a participant, they execute a *random walk with obstacle avoidance* behaviour. Each robot executes the two following steps: (i) it drives straight with a constant velocity of 10 cm/s, 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). Additionally, the obstacle avoidance behaviour is also triggered when the robots enter the black area. We used the on-board proximity sensors to detect obstacles in the direction of movement and ground sensors to detect changes in light intensity on the floor allowing the robots to register when they are about to enter the black area. A finite state machine (FSM) illustrating the robot behaviour is given in Fig. 6.

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* sessions were always coming from identical locations and hidden zones. As shown in Fig. 7, 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.

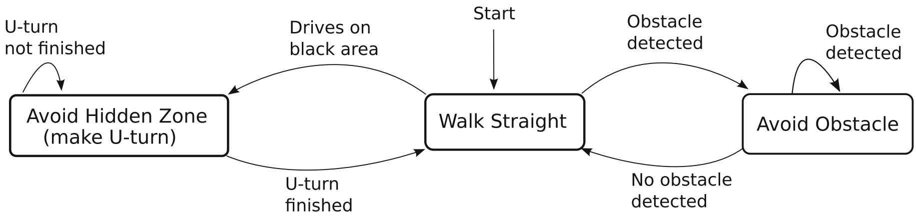

Fig. 6 An FSM-based representation of the behaviour executed by all robots

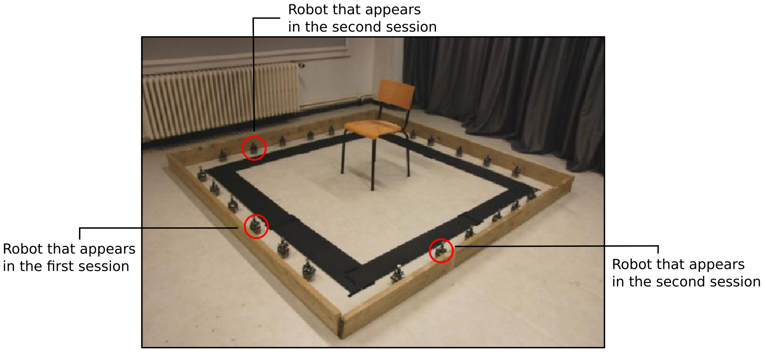

Fig. 7 Initial position of the robots used in the experiments. The robot in front of the chair (*encircled* in the picture) is the robot that comes out in the *1-robot* session. The two robots on the *left* and on the *right* of the chair (*encircled* in the picture) are the robots that come out in the *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 in the hidden zones (i.e., that render the robots invisible to the participants) have been removed for this picture (Color figure online)

3.3.4 Robot platform

The platform used in this study is the wheeled e-puck robot (see Fig. 8) equipped with an extension board. The e-puck robot is designed for educational and research purposes (Mondada et al. 2009). The extended version of the e-puck robot is 13 cm high and has a diameter of 7 cm. In this study, we used only a subset of the sensors and actuators available to the e-puck robot: the proximity sensors, the ground sensors (that comes with the e-puck extension board) and the wheel actuators. See Mondada et al. (2009) and Gutiérrez et al. (2008) for further details and for a complete list of the sensors and actuators available on the e-puck platform. We programmed the e-puck robots using the software infrastructure described in Garattoni et al. (2015).

3.4 Participants

We recruited 25 participants from the overall population of the Université libre de Bruxelles. Our participants came from different faculties of the university (e.g., law, sciences, medicine, philosophy). None of our participants had a background in robotics. 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 (i.e., we excluded potential participants with cardiovascular problems). Our participants received an informed consent form explaining that they were filmed during the

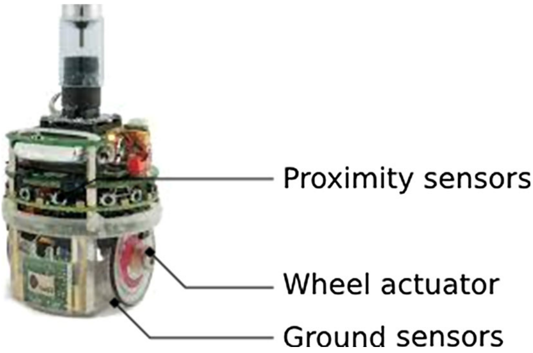

Fig. 8 An e-puck robot used in our experiments. The proximity sensors are used to detect and avoid nearby robots. The ground sensors are used to detect black areas (in order to avoid the robot to go back in a hidden zone). The wheel actuators are set to a speed of 10 cm/s (Color figure online)

experiment[1] and that their physiological responses were being collected for research purpose only. All of our participants signed the informed consent form. We offered a 7 € financial incentive for participation.

3.5 Experimental procedure

All our experiments were conducted in the robot arena of 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 then asked the participant to read and sign the consent form. We asked the participant to wash their hands with clear water (i.e., no soap) to allow the physiological sensors to function accurately. Once the participant washed their hands, we asked the participant to remain seated and we attached the participant to two physiological sensors. Since some robots may come close to the participant's feet, we told the participant to not move their legs and feet during the whole duration of the experiment. Once we attached the physiological sensors to the participant, we explained to the participant how to answer the SAM questionnaire. In order for the participant to clearly understand the notions of valence and arousal, we orally associated with 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). We then asked the participant to stay calm during a 5-min rest period. During this rest period, we recorded the participant's physiological responses that we used as the participant's baseline. After the rest period, we started the experiment with the first of the three sessions. After each session, we asked the participant to choose an image in the valence and arousal scales that correspond to their subjective psychological state. A number from 1 to 9 was written on 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. All participants completed the questionnaire within 1 min. After each session (i.e., when the participant was answering the SAM questionnaire), we kept the robots in motion in the environment. Figure 9 shows a participant during an experiment, while the participant was confronted with a group of 24 robots. The entire experiment's duration was 30 min per participant.

[1] 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).

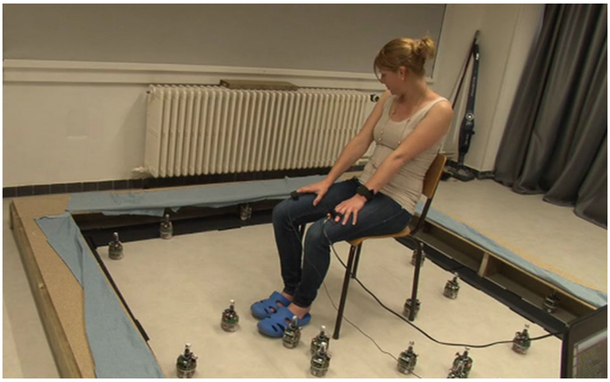

Fig. 9 A participant during the *24-robot* session. The picture is a snapshot taken from the video recording of an experiment. The participant gave her written consent to the use of the picture (Color figure online)

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 potential questions.

4 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 were 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, SCL), 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 males and 12 females) and the self-reported data of 24 participants (11 male and 13 females). We analysed our data with the R software (R Core Team 2015).

4.1 Group size effect

As each of our participants performed the experiments with three different robot group sizes, we conducted a within-subject analysis (a repeated-measure ANOVA) on both the psychophysiological measures (heart rate and skin conductance) and the self-reported measures (arousal and valence). We applied a Greenhouse-Geisser correction when the assumption of sphericity was violated (assessed with a Mauchly's test of sphericity). We selected $\alpha = .05$ to determine statistical significance. In the context of this study, the repeated-measure ANOVA's null hypothesis states that the three sessions *1-robot*, *3-robot* and *24-robot* have the same mean. The alternative hypothesis states that at least one session mean is different. When the repeated-measure ANOVA 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 their means.

In order to determine which sessions differ in their means, we proceed with a pairwise comparison of the three sessions with multiple paired t tests. The t test's null hypothesis

Table 1 Descriptive statistics of the psychophysiological data and of the self-reported data

Dependent Variable	n	1 robot	3 robots	24 robots
Heart rate (BPM)	23	−1.88 (6.2)	−2.48 (4.9)	0.16 (5.4)
SCL (μS)	23	4.77 (2.92)	4.82 (2.16)	6.23 (2.39)
Arousal	24	3.87 (2.43)	3.75 (2.11)	5.45 (2.41)
Valence	24	6.66 (1.04)	6.62 (1.05)	6.87 (1.77)

We report the mean and the standard deviation (in parentheses) of the three sessions (i.e., *1-robot*, *3-robot*, *24-robot*)

states that the mean difference between the paired values from two sessions (i.e., a value from one session paired to a value from another session) is equal to zero. The alternative hypothesis states that the mean difference of the paired values is not equal to zero. When the t test is significant, we can reject the null hypothesis in favour of the alternative hypothesis and conclude that there is a significant difference between the two sessions. Performing multiple pairwise comparisons (there are three pairwise comparisons between our three sessions) introduces the risk of increasing the Type I error, i.e., to declare the test significant, while it is not. In order to control the Type I error, we apply a Bonferroni correction to the p values obtained by the t test. In Table 1, we summarize the results by giving for each dependent variable the mean and the standard deviation (in parentheses) for each of the three sessions.

The analysis of the heart rate data shows a main effect of the number of robots on our participants $[F(2, 44) = 5.13, p \leq 0.01, \eta^2 = 0.04]$.[2] The post hoc test on the heart rate data revealed that our participants' heart rate was statistically significantly increased between the *3-robot* session and the *24-robot* session $[t(22) = -3.3, p = 0.01]$. Our participants' heart rate was not statistically significantly decreased between the *1-robot* session and *3-robot* sessions $[t(22) = 0.3, p = 1]$. Finally, our participants' heart rate was not statistically significantly increased between the *1-robot* session and the *24-robot* session $[t(22) = -1.2, p = 0.14]$, see Fig. 10 (*top-left*). The analysis of the skin conductance level also confirmed a main effect of the number of robots on our participants $[F(1.41, 31.08) = 14.4, p < 0.001, \eta^2 = 0.07]$. The post hoc test on the SCL data revealed that our participants' SCL was statistically significantly increased between the *3-robot* and *24-robot* sessions $[t(22) = -7.67, p < 0.001]$, and between the *1-robot* and *24-robot* sessions $[t(22) = -4.11, p = 0.0014]$. The SCL was not statistically significantly increased between the *1-robot* and *3-robot* sessions $[t(22) = -.14, p = 1]$, see Fig. 10 (*top-right*).

The results of the repeated-measure ANOVA on the self-reported arousal data confirm a main effect of the number of robots on our participants $[F(2, 46) = 26.02, p < 0.001, \eta^2 = 0.1]$. The post hoc test shows that our participants' self-reported arousal was statistically significantly increased between the *1-robot* and *24-robot* sessions $[t(23) = -6.21, p < 0.001]$ and between the *3-robot* and *24-robot* sessions $[t(23) = -5.74, p < 0.001]$. The self-reported arousal was not statistically significantly different between the *1-robot* and *3-robot* sessions $[t(23) = .53, p = 1]$, see Fig. 10 (*bottom left*). The analysis of the valence

[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.

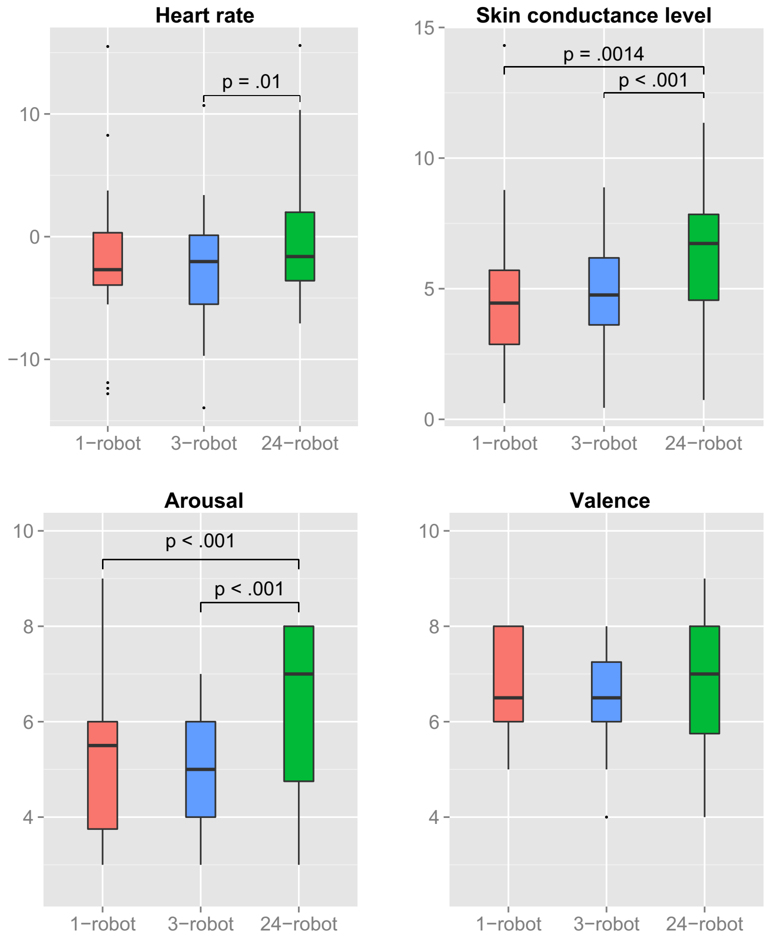

Fig. 10 *Boxplots* showing the heart rate values (*top left*), skin conductance level values (*top right*), arousal values (*bottom left*) and valence values (*bottom right*) of all three sessions (*1-robot*, *3-robot*, *24-robot*). The median value of each session is shown using the *bold horizontal line in the box*. Outliers are represented using *dots*. We also report the results of the pairwise t tests by connecting the *boxplots* of the sessions showing pairwise statistical significance (Color figure online)

does not show any main effect of the number of robots on our participants [$F(2, 46) = .32$, $p = .7$, $\eta^2 = 0.006$].

4.2 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 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

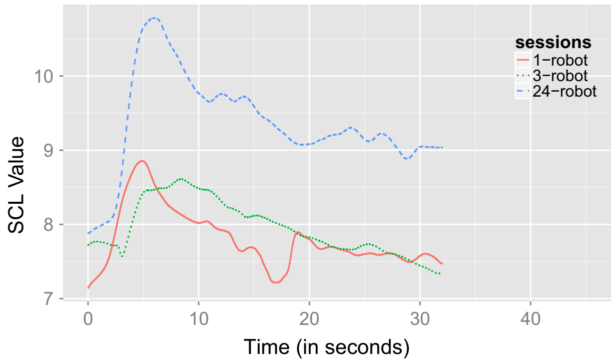

Fig. 11 Mean SCL values (all participants) over time (Color figure online)

each session. We show the results of all three sessions in Fig. 11. As depicted in Fig. 11, 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.

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 do not suggest any habituation effect between sessions with few robots (1 and 3) and the session with many robots (24). Therefore, the data do not contradict our hypothesis that an increasing number of robots affects the human psychophysiological state—as depicted in Fig. 11, 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 Discussion and conclusion

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. Our results confirm this hypothesis and furthermore show that greater numbers of robots provoke a stronger response. Firstly, there was a significant difference in our participants psychophysiological state when the number of robots increased to 24 robots. Secondly, our participants were conscious of their psychophysiological state change, as they reported significantly higher arousal values when exposed to the 24 robots.

Though we show with our results an effect of the group size on the human psychophysiological state, we did not consider all possible variables that could also influence the psychophysiological 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 psychophysiological state. Another variable that could influence the psychophysiological state is the robots' behaviour. In this experiment, 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 of the arena. Other variables could influence the human psychophysiological state (e.g., the size of the robots, the noise produced by the robots, the participants' prior experience with robotics systems). These variables, or a combination of these variables (with the group size for instance), should be considered in future work.

One interesting aspect of our results 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. We were expecting to see a primarily stress-based response to the robots. In fact, however, the valence values reported by the additional questionnaire we asked our participants to fill out (see Sect. 3.2.1), suggest that our participants had a positive experience—on a scale of 1–9 (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 our experimental scenario, the participants passively interacted with the robots (i.e., the participants do not send commands to the robots), instead of actively interacting with the robots (i.e., the participants 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 number of robots that was affecting our participants' psychophysiological state. For example, in an interactive scenario, frustration due to interaction difficulties or failures might also have introduced psychophysiological effects. The effect of active interaction will be investigated in future work. For instance, we will study whether actively interacting with a swarm of robots (e.g., by guiding a swarm) negatively changes the psychophysiological 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.

The most important contribution of this study is that we have established that robot group sizes in particular, and robot swarms in general, affects human psychology. These facts should have profound short-term and long-term implications for human–swarm interaction research. In the short-term, we advise human–swarm interaction researchers to use psychophysiological measures in all of their user studies. So far, only purely psychological questionnaires have been used to study the human workload. Using psychophysiological measures would allow human–swarm interaction researchers to have a broader and deeper understanding of the human psychology. In the long-term, it is our contention that psychophysiology should be used as an engineering methodology to develop concrete human–swarm interaction systems for real-world applications. In human–computer interaction, for instance, 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 should be used to create tomorrow's real-world human–swarm interactions systems.

Acknowledgments The authors thank Mauro Birattari for his help in the construction of the experimental environment. This work was partially supported by the European Research Council through the ERC Advanced Grant "E-SWARM: Engineering Swarm Intelligence Systems" Contract 246939) awarded to Marco Dorigo.

Rehan O'Grady and Marco Dorigo acknowledge support from the Belgian F.R.S.-FNRS. The authors also thank the 24 participants who participated in the experiment.

References

Adams, J. (2009). Multiple robot/single human interaction: Effects on perceived workload. *Behaviour & Information Technology*, *28*(2), 183–198.

Bainbridge, W. S. (Ed.). (2004). *Berkshire encyclopedia of human–computer interaction*. Great Barrington, MA: Berkshire Publishing.

Bekele, E., & Sarkar, N. (2014). Psychophysiological feedback for adaptive human–robot interaction (HRI). In S. H. Fairclough & K. Gilleade (Eds.), *Advances in physiological computing* (pp. 141–167). Springer.

Bethel, C., Salomon, K., Murphy, R., & Burke, J. (2007). 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)* (pp. 732–737). New York: IEEE Press.

Blascovich, J., Vanman, E., Mendes, W., & Dickerson, S. (2011). *Social psychophysiology for social and personality psychology*. Beverley Hills, CA: Sage.

Bradley, M. M., & Lang, P. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. In R. D. R. Lane, L. Nadel, G. L. Ahern, J. Allen, & A. W. Kaszniak (Eds.), *Cognitive neuroscience of emotion* (pp. 25–49). Oxford: Oxford University Press.

Byrne, E., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, *42*(3), 249–268.

De la Croix, J. P., & Egerstedt, M. (2012). Controllability characterizations of leader-based swarm interactions. In *AAAI fall symposium series technical reports*. AAAI.

Dehais, F., Sisbot, E., Alami, R., & Causse, M. (2011). Physiological and subjective evaluation of a human–robot object hand-over task. *Applied Ergonomics*, *42*(6), 785–791.

Dix, A., Finlay, J., Abowd, G., & Beale, R. (2003). *Human–computer interaction* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Dorigo, M., Birattari, M., & Brambilla, M. (2014). Swarm robotics. *Scholarpedia*, *9*(1), 1463.

Dorigo, M., Floreano, D., Gambardella, L. M., Mondada, F., Nolfi, S., Baaboura, T., et al. (2013). Swarmanoid: A novel concept for the study of heterogeneous robotic swarms. *IEEE Robotics & Automation Magazine*, *20*(4), 60–71.

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

Goljar, N., Javh, M., Poje, J., Ocepek, J., Novak, D., Ziherl, J., et al. (2011). Psychophysiological responses to robot training in different recovery phases after stroke. In *Proceedings of the IEEE international conference on rehabilitation robotics (ICORR'11)* (pp. 1–6). New York: IEEE Press.

Gutiérrez, Á., Campo, A., Dorigo, M., Amor, D., Magdalena, L., & Monasterio-Huelin, F. (2008). An open localization and local communication embodied sensor. *Sensors*, *8*(11), 7545–7563.

Humphrey, C., Henk, C., Sewell, G., Williams, B., & Adams, J. (2007). Assessing the scalability of a multiple robot interface. *Proceedings of the 2nd ACM/IEEE international conference on human–robot interaction (HRI 2007)* (pp. 239–246). Piscataway, NJ: IEEE Press.

Itoh, K., Miwa, H., Nukariya, Y., Zecca, M., Takanobu, H., Roccella, S., et al. (2006). Development of a bioinstrumentation system in the interaction between a human and a robot. In *IEEE/RSJ international conference on intelligent robots and systems* (pp. 2620–2625). IEEE.

Kolling, A., Sycara, K., Nunnally, S., & Lewis, M. (2013). Human swarm interaction: An experimental study of two types of interaction with foraging swarms. *Journal of Human–Robot Interaction*, *2*(2), 103–128.

Kolling, A., Walker, P., Chakraborty, N., Sycara, K., & Lewis, M. (2016). Human interaction with robot swarms: A survey. *IEEE Transaction on Human–Machine Systems*, *46*(1), 9–26.

Kulic, D., Croft, E. (2005). Anxiety detection during human–robot interaction. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS 2005)* (pp. 616–621). Los Alamitos: IEEE Computer Society Press.

Kulic, D., & Croft, E. (2007). Affective state estimation for human–robot interaction. *IEEE Transactions on Robotics*, *23*(5), 991–1000.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. H. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). New York: Ablex.

Mehrabian, A. (1996). Pleasure–arousal–dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, *14*(4), 261–292.

Mondada, F., Bonani, M., Raemy, X., Pugh, J., Cianci, C., Klaptocz, A., et al. (2009). The E-Puck, a robot designed for education in engineering. In *Proceedings of the 9th conference on autonomous robot systems and competitions* (pp. 59–65). Castelo Branco: Instituto Politècnico de Castelo Branco.

Nagavalli, S., Chien, S., Lewis, M., Chakraborty, N., & Sycara, K. (2015). Bounds of neglect benevolence in input timing for human interaction with robotic swarms. In *Proceedings of ACM/IEEE international conference on human–robot interaction* (pp. 197–204). New York, NY: ACM.

Nagi, J., Giusti, A., Gambardella, L., & Di Caro, G. A. (2014). Human–swarm interaction using spatial gestures. In *Proceedings of the 27th IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 3834–3841). Washington: IEEE Computer Society.

Nagi, J., Ngo, H., Gambardella, L., & Di Caro, G. A. (2015). Wisdom of the swarm for cooperative-decision making in human–swarm interaction. In *Proceedings of the IEEE international conference on robotics and automation (ICRA)*.

Novak, D., Ziherl, J., Olenšek, A., Milavec, M., Podobnik, J., Mihelj, M., et al. (2010). Psychophysiological responses to robotic rehabilitation tasks in stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *18*(4), 351–361.

Nunnally, S., Walker, P., Lewis, M., Kolling, A., Chakraborty, N., & Sycara, K. (2012). Connectivity differences between human operators of swarms and bandwidth limitations. In *Proceedings of the third international conference on swarm, evolutionary, and memetic computing* (pp. 713–720). Berlin: Springer.

Pendleton, B., & Goodrich, M. (2013). Scalable human interaction with robotic swarms. In *Proceedings of the AIAA Infotech@Aerospace conference* (pp. 633–645). Reston, VA: American Institute of Aeronautics and Astronautics.

Podevijn, G., O'Grady, R., Nashed, Y., & Dorigo, M. (2013). Gesturing at subswarms: Towards direct human control of robot swarms. In A, Natraj, S. Cameron, C. Melhuish, & M. Witkowski (Eds.), *Towards autonomous robotic systems—14th annual conference, TAROS 2013, Lecture notes in computer science* (Vol. 8069, pp. 390–403). Berlin: Springer.

Prinzel, L., Freeman, F., Scerbo, M., Mikulka, P., & Pope, A. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International Journal of Aviation Psychology*, *10*(4), 393–410.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(4), 601–614.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

Rani, P., & Sarkar, N. (2005). Making robots emotion-sensitive—Preliminary experiments and results. In *IEEE international workshop on robot and human interactive communication* (pp. 1–6). Piscataway: IEEE.

Rani, P., Sarkar, N., Smith, C. A., & Kirby, L. (2004). Anxiety detecting robotic system-towards implicit human–robot collaboration. *Robotica*, *22*(01), 85–95.

Salvendy, G. (2012). *Handbook of human factors and ergonomics*. Hoboken, NJ: Wiley.

Setter, T., Fouraker, A., Kawashima, H., & Egerstedt, M. (2015). Haptic interactions with multi-robot swarms using manipulability. *Journal of Human–Robot Interaction*, *4*(1), 60–74.

Swangnetr, M., & Kaber, D. (2013). 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.

Swangnetr, M., Zhu, B., Taylor, K., & Kaber, D. (2010). 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* (pp. 349–353). London: Sage.

Tiberio, L., Cesta, A., Cortellessa, G., Padua, L., & Pellegrino, A. (2012). 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)* (pp 833–838). New York: IEEE Press.

Ting, C., Mahfouf, M., Nassef, A., Linkens, D., Panoutsos, G., Nickel, P., et al. (2010). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *40*(2), 251–262.

Tullis, T., & Albert, W. (2008). *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. San Francisco, CA: Morgan Kaufmann.

Velagapudi, P., Scerri, P., Sycara, K., Wang, H., Lewis, M., & Wang, J. (2008). Scaling effects in multi-robot control. In *Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS 2008)* (pp. 2121–2126). Los Alamitos: IEEE Computer Society Press.

Zhang, T., Kaber, D., Zhu, B., Swangnetr, M., Mosaly, P., & Hodge, L. (2010). Service robot feature design effects on user perceptions and emotional responses. *Intelligent service robotics*, *3*(2), 73–88.