Eindhoven University of Technology

MASTER

The moderating effect of biofeedback on the relation between physiological synchrony and empathy & team performance

van Laar, C.K.

Award date:
2019

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. Apr. 2025

The moderating effect of Biofeedback on the relation between Physiological Synchrony and Empathy & Team Performance

Koen C.K. van Laar

0870217

In partial fulfilment of the requirements for the degree of

Master of Science

Human Technology Interaction

Eindhoven University of Technology

Master Thesis

April 2019

Supervisors:

prof.dr. J.H.D.M. Westerink	Philips Research / Eindhoven University of Technology
prof.dr. W. A. IJsselsteijn	Eindhoven University of Technology
MSc M. A. Feijt	Eindhoven University of Technology

Assessor:

dr. L. M. Huiberts	Eindhoven University of Technology

Abstract

Physiological Synchrony (PS) is the extent to which physiological processes of at least two persons fluctuate in synchrony over time (e.g. heartrate going up and down in a similar pattern over time). Research has shown that there is a positive relation between the level of PS and empathy and team performance. However, only little research is conducted on the effect of feedback of PS on empathy and team performance. This research is aimed at replicating previous findings on these relations as well as studying the effect of feedback of PS on empathy and team performance.

A laboratory study is conducted in which seventy dyads (groups of two participants) performed two discussion tasks (desert survival and lost at sea) while their heartrate was being measured, so that a real-time measure of PS could be calculated and given as feedback to participants. A within-subject design was performed to study the relation between PS and perceived empathy and team performance, while a between-subjects design was conducted to study the effect of feedback: only half of the dyads was provided with feedback of their level of PS. Perceived empathy was measured with a questionnaire and a score for team performance was obtained based on performance on the tasks.

No statistically significant correlations were found for the relation between PS and perceived empathy or team performance. Also, results showed no effect of PS feedback on perceived empathy or team performance. These results are in contradiction to earlier research that did find statistically significant effects and can partly be attributed to choices in the experimental design that are different from the previous studies.

Suggestions for future studies are 1) to study the effect of using different physiological measurements on PS, 2) to develop standard measurements for PS to compare studies meaningfully, and 3) investigate the potential use of tools using PS as a means to increase intimacy in therapeutic settings.

Keywords: physiological synchrony, perceived empathy, team performance, biofeedback

Contents

1. Introduction

People like to be part of a group and coordinate actions with the people around them, as can be derived from examples from various cultures and groups of people (Wiltermuth & Heath, 2009). Examples of such synchronized actions are soldiers marching together, or tribe members engaging in a ritual dance, but also non-physical acts such as choirs or football hooligans singing together. More day-to-day interpersonal synchronicity[1] that occurs frequently are bodily, vocal and postural mimicry of nature (Hatfield, Rapson, Ye, Decety, & Ickes, 2009): This can be mimicking facial expressions or posture by taking a seat synchronously and crossing the same leg. Research by Condon and Ogston (1966) has shown that this mimicry occurs without conscious control. They did so by studying the time it took for participants to match a partner's posture and found that the speed in which these imitations were executed was too high to be initialized by conscious thought.

While this behavioral matching is already intriguing, synchronicity can also manifest itself at an even more fundamental level: human physiology has been shown to exhibit substantial synchronicity between persons (DiMascio, Boyd, & Greenblatt, 1957; Levenson & Gottman, 1983). DiMascio et al. (1957) studied neurotic patients undergoing long-term psychotherapy and found that patient's and therapist's heartrates moved in similar directions as the levels of tension in the interview varied but moved in opposite directions when the patient expressed antagonism. This phenomenon is referred to as Physiological Synchrony (PS) and is defined by Palumbo et al. (2016, p.100) as "*any interdependent or associated activity identified in the physiological processes of two or more persons*." These interdependent or associated physiological activities can be quantified for example by computing correlations of physiological parameters such as heartrate. PS has been found to be related to several psychological constructs of which empathy and team performance are two of the most studied. Feelings of empathy and empathic accuracy tend to increase when PS increases (Hove & Risen, 2009; Marci & Orr, 2006; Marci, Orr, & Ham, 2007; Adler, 2002). PS has also been positively related to team performance in several studies (Elkins et al., 2009; Wiltermuth & Heath, 2015; Valdesolo et al., 2010; Henning, Boucsein, & Gil, 2001; Henning & Korbelak, 2005; Henning, Armstead, & Ferris, 2009; Ferrer & Helm, 2012) although this positive relation has not always been found (Järvelä et al., 2014; Strang et al., 2014).

Even though PS has been studied relatively intensively (see Palumbo et al., 2016 for an overview) the role that providing feedback on physiological measures might play (e.g. heart rate), has not received much attention yet. Research shows that perceiving someone's physiology can increase feelings of intimacy (Janssen et al., 2010), which indicates that receiving information on physiology affects psychological states. Also, providing feedback on PS has been shown to influence perceived empathy (Okel, 2018). Even more interesting, the level of PS provided in feedback was shown to have an effect; when people were given feedback that indicated their PS with a partner was high, they rated perceived empathy as higher compared to when feedback of low PS was given. However, the feedback provided in Okel's study was controlled by the experimenter and thus fake. Therefore, to provide more conclusive evidence for the effect of PS feedback on perceived empathy, it is important to study whether this effect can still be found when real feedback on PS is given.

[1] In this report, the words synchronicity and synchrony are used interchangeably.

In the current study, the relation between PS and empathy and team performance will be examined. Secondly, the moderating effect of providing feedback of PS on these relations will be investigated. Heart rate will be the physiological parameter to measure PS.

The remainder of the introduction is organized as follows: first the concept PS will be discussed in more depth together with explanations and skepticism towards it. Following this, empathy and team performance will be discussed together with argumentation for the hypotheses regarding the relation between PS and these concepts. Finally, feedback on physiological signals will be discussed with special focus on PS feedback.

1.1 Physiological Synchrony

Despite the increasing amount of research aimed at understanding the synchronization of physiology and its relations with other psychological constructs, this research area is still fragmented (Kleinbub, 2017). This becomes apparent when looked at the various terms used for PS, including physiological concordance (Marci, Ham, Moran, & Orr, 2007), physiological linkage (Reed, Randall, Post, & Butler, 2013), and physiological coherence (Chanel, Kivikangas, & Ravaja, 2012). Palumbo et al. (2016) use the overarching term interpersonal physiology. However, the current study focuses specifically on the synchronicity of physiology and therefore the term physiological synchrony (PS) is used.

The part of the nervous system that is responsible for the variations in physiology that will be measured in this study is called the Autonomic Nervous System (ANS) (Cacioppo, Tassinary, & Berntson, 2007). The ANS is part of the Peripheral Nervous System together with the Somatic Nervous System. Whereas the SNS is largely under conscious control, the ANS is largely outside conscious control. Due to our inability to consciously control the ANS, studying the ANS offers great opportunities to study human behavior more objectively and more unobtrusively than other measurement techniques such as questionnaires or interviews.

The ANS in turn consists of the Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS). These systems control respectively the “fight and flight” response and the “rest and digest” process (Palumbo et al., 2016). Many organs are influenced by both the SNS and PNS. For example, heartrate tends to increase during moments of both anger and joy (Cacioppo, Tassinary, & Berntson, 2007). Furthermore, relations can be context- dependent, meaning that in some situations, happiness leads to an increased heartrate, but not in others. Despite the limitations imposed by these restrictions, psychophysiology offers useful tools to study the mind through bodily responses that cannot be consciously controlled. Several processes have been put forward to explain how these automatically occurring physiological reactions lead to the existence of PS, which will be described in the next paragraph.

1.2 Explanations and Skepticism on PS

1.2.1 Influences of task and environment

Despite the interesting research opportunities provided by the concept PS, some skepticism exists on its legitimacy. Strang et al. (2014) conducted an experiment involving a cooperative team task from which they concluded that observed levels of PS were driven by the features of the team-task environment, and that PS was likely influenced by similar task demands and interpersonal coordination dynamics that were not ‘unique’ to a particular team. From this, they concluded that environmental influences cause identical bodily reactions in individuals rather than individuals have a ‘special connection’ with each other that leads to a certain level of synchrony. This did not yet exclude the possibility that PS can be higher in dyads with a strong social connection compared to dyads with a weaker social connection. However, Strang et al. (2014) also created random virtual dyads of participants who performed the experiment and found that the level of PS of these virtual dyads was not significantly different from the level of PS of real dyads. These results suggest that either no significant social connection was established between members of real dyads or that there was no effect of unique teams on the level of PS.

On the other hand, several studies do show support for the existence of PS. Liu, Zhou and Palumbo (2016) showed that participants seated back to back had significantly less PS compared to participants sitting face to face, suggesting that synchrony is likely due to interpersonal processes rather than shared experiences of shared environmental demands. Others tested whether data from randomly paired individuals who underwent the same task conditions showed the same level of PS as pairs of individuals who did the task together; they showed significantly greater synchrony when participants are together (Marci et al, 2007; Reed et al 2013). Also, in the study by Elkins et al. (2009), PS existed even though physical demands in their experiment differed between participants. Combining these results, Palumbo and et al. (2016) argue that PS goes beyond metabolic demands and can be influenced by psychosocial processes.

1.2.2 Fragmented results in physiological parameters

A different point of criticism stems from the fact that many studies measure multiple physiological parameters and multiple dependent variables simultaneously, while only rarely researchers find an effect for all of them. For example, Henning and Korbelak (2005) used two team performance measurements and only found an effect for the relation between physiological synchrony and one team performance measurement. Other studies measure several physiological parameters and team performance variables but only find an effect between one team performance variable per physiological measure (Elkins et al., 2009; Henning et al., 2001). Reed et al. (2013) measured blood pressure, inter-beat-interval, and skin conductance, and only found an effect for mean blood pressure. By also considering the effects that were not found, the evidence for relation between PS and team performance is weakened somewhat.

1.2.3 A possible explanation of PS: Mutual Social Tracking

Not much discussion is given to the underlying processes of PS, but one pragmatic explanation for the cause of PS comes from the concept of mutual social tracking that is part of the cybernetic model of behavior developed by Smith and Smith (1987). The Cybernetic model of behavior argues that by controlling motor behavior one can control other internal physiology behaviors. Take physical exercise as an example; exercising leads to higher respiration volume and heart rate. These cannot easily be controlled mentally, but by letting the body perform demanding physical work, the heartrate and respiration frequency increase. The same principle can be applied to social interactions. This is referred to as mutual social tracking as is well explained by Henning et al. (2009). Consider two people having a conversation; person 1 is actively listening to words being spoken by person 2 and provides sensory feedback to person 2 by means of motor control (i.e. facial expressions, posture etc.). Person 2 uses this feedback to dynamically adjust her speaking style to retain the interest of the first person. These continuous and coordinated small adjustments of both persons represent mutual social tracking (Henning et al., 2009). This will have reciprocal effects on their physiological states and these would be expected to be synchronous. Empirical evidence for this model comes from several studies (Levenson &

Gottman, 1983, Levenson & Ruef, 1992) that show higher synchrony of heart rate during intense social interaction.

However, objections can also be made against the cybernetic model of behavior as the most appropriate model to offer perspective on the underlying mechanisms of PS. It can be argued that the example of physical exercise is inappropriate, as this kind of behavior is often impossible to conduct in normal situations and can therefore not be used to attempt to obtain some control over the ANS. Also, research has shown that socio-visual contact does not influence the level of PS as one would expect based on the explanation of mutual social tracking (Henning et al., 2009). Another study showed that even when people are sitting next to each other while blindfolded, some level of PS was obtained (Ferrer & Helm, 2013). The authors mention the mere presence of a partner as a possible cause for this. Despite these counterarguments, mutual social tracking does seem to be a reasonable explanation for at least part of PS.

1.2.4 Concluding remarks regarding PS

The inconsistency in finding effects may not be surprising as this research area is relatively new, but it does add to the skepticism towards physiological synchrony. However, Palumbo et al. (2016) argue that although no definitive patterns have yet emerged, relatively consistent findings of PS across physiological measures suggest that synchrony is a robust phenomenon identifiable through numerous methods.

As mentioned before, several studies have shown positive relations between PS and empathy and team performance. In the following paragraphs, the concepts empathy and team performance will be discussed together with their relation with PS.

1.3 Empathy

Since the term empathy was first coined by Tithener more than 100 years ago (Wispé, 1986), it has been widely studied (Cuff et al., 2015). It is considered to be one of our most important social processes (Janssen, 2012). However, empathy has been defined and used in so many ways that Cuff et al. (2015) wrote a literature review for the sole purpose of suggesting one unified overarching definition. The definition they proposed is as follows:

"Empathy is an emotional response (affective), dependent upon the interaction between trait capacities and state influences. Empathetic processes are automatically elicited but are also shaped by top-down processes. The resulting emotion is similar to one's perception (directly experienced or imagined) and understanding (cognitive empathy) of the stimulus emotion, with recognition that the source of the emotion is not one's own." (p.150)

According to this definition, to be empathetic, one not only needs to experience another one's emotion but also understand it and be aware that it is not one's own emotion. Arguably, what could be emphasized more here is the response to another's emotion. Hatfield, Rapson and Le (2009) mention that a 'socially beneficial' intention to respond compassionately to a person's distress is also needed.

This corresponds with the three-component framework of empathy based on psychology and neuroscience proposed by Janssen (2012). These components are cognitive empathy, emotional convergence and empathic responding. Even though Cuff et al. (2015) made the wide variety in definitions of empathy apparent, Janssen (2012) argues that there exists considerable agreement among psychologists and neuroscientists on certain aspects of empathy. In general, the consensus is that empathy consist of the following three components: a) recognizing someone else's emotional state and understanding what it means (cognitive empathy), b) connecting and coupling of the emotion (emotional convergence), and c) responding socially appropriately to this received emotion (empathic responding).

Next to breaking empathy down to these three components, empathy can be studied in a dispositional or situational approach (Janssen, 2012), where dispositional empathy is referred to as the individual differences between people in their ability to be empathetic and, on the other hand, situational empathy is about the experienced empathy at specific moments or during specific interactions (Janssen, 2012). Simplified, one could say that dispositional empathy is how empathetic a person is as a personal characteristic, while situational empathy is how empathetic a person feels at a certain moment. The current study concerns itself with the situational aspect of empathy that can change based on environmental context and individual inner state. A type of situational empathy is perceived empathy, which is the level to which one feels understood and empathized with by another person (Cramer & Jowett, 2010). This is different from other empathy measures that usually focus on the empathetic state or trait of the person itself. A questionnaire for measuring perceived empathy has been developed by Okel (2018) which is based on a questionnaire originally devised for measuring perceived empathy in a salesperson-customer relation (Plank, Minton, & Reid, 1996).

As has been noted earlier, PS has been studied quite a lot in relation to empathy. These studies often investigated patient-therapist or parent-child relationships (Palumbo et al., 2016). One of the reasons for this is that empathy is a crucial concept in psychological therapy (Mohr, 1995) and the development of children (Feldman, 2012). According to Mohr (1995), the absence of perceived empathy is one of the best predictors of poor outcomes in psychotherapy. Regarding the parent-child relation, Feldman (2012) states that when a child is 3-4 months old, parent-child synchrony has been shown to predict infant's capabilities to empathize with others later in life.

Early research on empathy in relation to PS was conducted by Levenson and Gottman (1992) with their research on marital interaction in which they found that detection accuracy of negative emotions in partners was highest when PS of skin conductance was high. This relation was not found for three other physiological measurements (heart rate, finger pulse amplitude and somatic activity). Nevertheless, these results suggest that capability to emphasize with someone is related to the level of PS experienced. Further support for this relation comes from studies done by Marci and colleagues (Marci & Orr, 2006; Marci et al., 2007) where they found that PS in skin conductance between therapist and client was significantly lower in an emotionally distant condition where the therapist on purpose ignored verbal and emotional cues, compared with an emotionally neutral condition where the therapist did react to social cues. Perceived empathy was therefore related to the level of PS. This relation was again found in study by Messina et al. (2013) in their experiment that involved a fabricated therapist-patient role playing setting where volunteer students pretended to be a patient and had to discuss a personal problem with a "listener" that was either a real therapist, a psychologist or a non-therapist. Results showed a clear relationship between PS measured by skin conductance and listener's empathy based on scores given by the pseudo-patients.

Based on results found in previous studies on the relation between PS and empathy, the following hypothesis is formulated:

H1: PS and perceived empathy are positively related.

1.4 Team Performance

Research suggests that PS can be used as a predictor of team performance (e.g. Elkins et al., 2009; Henning, Boucsein, & Gil, 2001; Henning & Korbelak, 2005; Mitkidis, Roeptstorff & Wallot, 2015). This is valuable since it can be complicated to assess team work and team coordination in an objective manner (Henning et al., 2001). Physiology measurements can be said to be more objective than traditional psychological measurements because they do not suffer from the subjectiveness that accompanies introspection which is needed for questionnaire measurements. Having said that, physiological measurements are context dependent and leave room for multiple interpretations of the data. Nonetheless, measuring physiology does add a new dimension to analyze human behavior which can be used to guide human factors design and the evaluation of sociotechnical systems (Henning et al., 2001).

Henning et al. (2001) conducted an experiment where dyads played a custom-made game while visual contact was manipulated and found that, although socio-visual contact did not influence PS, PS was predictive of improved performance in terms of task completion time. Another study on the relation between teamwork and PS conducted by Elkins et al. (2009), analyzed data from an earlier experiment where teams of marines had to “clear” rooms by shooting opposing combatant forces in a military setting used for practicing the clearance of buildings to simulate fights in urban areas. Results showed that greater PS was associated with better performance, although they only compared physiological measurements between 2 members of each team. Further evidence for the positive relation between PS team performance comes from Montague, Xu, and Chiou (2014) who studied dyads playing a game where one person had control and the second person was a passive adviser: under varied task difficulties different measures of PS were significantly related to teammates performance. A study by Walker et al. (2013) showed that 10% of the variation of team performance on a processing plant simulation could be explained by PS.

However, a positive relation between PS and team performance is not always found as is characterized by a field study by Henning et al. (2009) where the physiology of a group of graduates was measured during a six-month period and where results showed barely any statistically significant results. This could suggest that PS as predictor of team performance outside lab settings is more complex. Järvelä et al. (2014) conducted an experiment to study whether the nature of conflict influences the level of PS. Participants in dyads played a video game in four conditions: 1) together against a computer, 2) together but individual score was counted, 3) versus each other with a computer as teammate and 4) versus each other without a computer as teammate. Results showed no difference in PS across conditions while a difference would be expected if PS is a predictor of team performance.

Considering the results of the mentioned studies one can say that despite the seemingly overwhelming evidence on the existence of the relation between PS and team performance, some studies indicate that there are several factors influencing the strength of this relation (Järvelä et al., 2014; Strang et al., 2014). Nevertheless, overall results in the literature suggest that there is indeed a positive relation between PS and team performance (Palumbo et al., 2016), which leads to the second hypothesis being formulated as follows:

H2: PS is positively related to team performance.

1.5 Feedback on Physiology

The homepage of the website of The Association for Applied Psychophysiology and Biofeedback defines Biofeedback as follows:

"Biofeedback is a process that enables an individual to learn how to change physiological activity for the purpose of improving health and performance. Precise instruments measure physiological activity such as brainwaves, heart function, breathing, muscle activity, and skin conductance. These instruments rapidly and accurately 'feed back' information to the user. The presentation of this information – often in conjunction with changes in thinking, emotions and behavior – supports desired physiological changes. Over time, these changes can endure without continued use of an instrument."

This elaborate and therefore somewhat unsatisfying definition of biofeedback was the result of a year of debating by a special task force commissioned with the sole goal of developing an updated official definition of biofeedback (Schwartz, 2008). This definition includes a specific purpose for biofeedback by saying it should aim to improve health and performance, while this might not always be the case. This could be restricting the use of the term biofeedback and therefore a more compact definition for biofeedback is found in a paper by van Dijk et al. (2008). The definition distilled from that paper reads: Biofeedback can be seen as measuring internal signals (e.g. heart rate) and presenting them via some external channel (e.g. monitor showing a number).

Biofeedback has been found to be efficacious in treating problems ranging from headaches, anxiety, and motion sickness to attention deficit hyperactivity disorder (Yucha & Montgomery, 2008). It has also been studied in relation to self-reported stress (van Dijk et al., 2008) where it was shown to increase self-awareness.

It is yet unclear what the effects of exposure to one's own physiological measurements are (van Dijk et al., 2015), let alone what these effects are in in combination with measurements of a partner. Having said that, there has been research conducted towards understanding and applying physiology feedback (i.e. feedback containing information about physiology) in social interactions: Merrill and Cheshire (2016) conducted a survey-based vignette experiment to investigate the social interpretation of heart rate in a conflicting and non-conflicting hypothetical situation in which participants read a story of someone apologizing for running late for an appointment and received information about that person's heart rate. The results indicated that a normal heartrate (rather than an elevated heartrate) leads to negative trust-related assessments of the person running late. Although this study does not include synchronicity of physiology in any form, this finding does show that providing information on physiology influences our judgements of others. More research on physiology feedback is conducted by Janssen et al. (2010) who studied the possibilities of heartbeat communication as an intimate cue. Results showed that heartbeat perception influences social behavior in a comparable manner as other intimate signals such as gaze and interpersonal distance. Participants who could hear the heartbeat of another person used this intimate signal by keeping a greater interpersonal distance. These results suggest that perceiving feedback on someone's physiology does not only influence our judgements of others, but also influence our own behavior. In a study by Chanel et al. (2010), the effect of displaying someone else's heart rate while watching video clips was analyzed. Results showed that displaying heart rates of participants increased co-presence. This provides further support that perceiving another person's heartbeat influences how we perceive one another. Lastly, a study by Okel (2018) showed that participants who received (fake) feedback indicating a high level of PS, rated perceived empathy higher compared to participants who received feedback of a low level of PS. This feedback was shown as a colored border around a video of an emotional clip from a movie.

Based on these previous studies, it is expected that providing people with feedback on the level of PS with a partner will influence their perception of each other in terms of perceived empathy. Furthermore, the level of PS in the provided feedback is expected to influence perceived empathy and team performance. Higher levels of PS people might cause people to think that they are indeed in sync with each other, leading to a greater level of perceived empathy. This is expected to work the other way too: low levels of PS could have a negative effect on the level of perceived empathy. To summarize these thoughts, the following hypotheses are formulated:

H3a) Feedback of PS leads to a stronger relation between PS and empathy and team performance than no feedback.
H3b) Feedback of high PS leads to higher empathy or team performance than no feedback whereas feedback of lower PS leads to lower empathy and team performance than no feedback.

An experiment has been designed to find answers to these hypotheses which is described in the next chapter.

2. Method

2.1 Design

This study includes a lab experiment in a room simulating a living room in which participants performed two discussion tasks while their heartrate was being measured. A within-dyads design was used to measure the relation between PS, perceived empathy, team performance. A between dyads design was used to measure the potential effect of feedback of PS on the relation between PS and perceived empathy. One group received feedback on their level of PS during the second discussion task whereas the other group received no feedback.

2.2 Participants

Participants were recruited from the JFS database from the Human Technology Interaction Group at the Technical University in Eindhoven. Seventy dyads participated in the experiment, thus 140 participants in total. Their age ranged from 18 to 81 years old (M_{age}=31.05, SD_{age} = 16.67). Sixty females participated (M_{age} = 29.49, SD_{age} = 14,95) and eighty males (M_{age} = 32.27, SD_{age} = 17,91). Using age as a distinction variable, the sample could clearly be split into young students and older, mostly retired, people. Ninety-five of the participants were younger than 30 years old and were mostly students while twenty-six were more than sixty years old. There were 11 female dyads, 21 male dyads and 38 mixed dyads. Out of the 70 dyads, 52 dyads indicated they had never seen the other participant before, 6 dyads knew each other superficially and 12 dyads knew each other very well (i.e. either married or best friends).

In the invitation mail, participants were told that the experiment included having a discussion with a second participant and that this conversation could be in English depending on whether both participants were able to speak Dutch. Participants who had any cardiovascular disorders could not participate as this could lead to disturbed ECG signals not suitable for analysis. Participants received ten or twelve euros for participation depending on whether they received travel-time-compensation or not. Also, some participants received course credits instead of money.

2.3 Experimental setting

The experiment took place at the Technical University Eindhoven in the USE-lab which is intended to resemble a standard living room. The participants were seated at a dinner-like table where they half faced each other, and half faced a computer monitor placed on the table. Figure 1 shows how the participants were seated during the experiment. The setup was constructed in such a way to make sure the participants would be able to have visual contact during their discussion and at the same time be able to receive the feedback on PS on the monitor in front of them. In Figure 1, the participants appear slightly rigid, possibly caused by the experimenter taking the picture, but during the discussion, participants were mostly relaxed and at ease.

Figure 1: Picture of experimental setting. Two participants sitting across each other and the monitor on which the task is shown and feedback on their PS (red in this case).

2.4 Tasks

Participants performed two discussion tasks in the experiment, called the "Desert Survival Task" and the "Lost at Sea Task". In both these tasks, participants were supposedly lost in either the desert or at sea and had to rank fifteen items (water, knife, raincoat etc.) based on the importance for them to survive these situations. These tasks were chosen because they provide an incentive for people to have a natural discussion and therefore stimulate social interaction which is required to have a meaningful measure of perceived empathy. First, participants made a list individually using pen and paper. After this, they compared their lists, discussed their rationale for their specific order and created a list together on the monitor by simply dragging and dropping images of the items in a certain order. There was a time limit of three minutes for the individual part and twelve minutes for the collaboration part. The Desert survival task has previously been used in research on social presence including psychological constructs such as empathy (Biocca, Harms, & Gregg, 2001), computer mediated cooperation (Bradner, & Mark, 2002) and group behavior (Staples & Zhao, 2006).

During the experiment it was aimed to have the discussions of all dyads of equal length. Participant were shown a timer on the screen that indicated how much time was left for the discussion counting down from 12 minutes. However, nearly all dyads finished the discussion before time ran out. If participants were finished discussing before the 8th minute, the experimenter would come out and say: *"Have you thought about the secondary uses of the items? Perhaps that will make you change your order"*. By prolonging the discussion like this, nearly all discussions lasted between eight and ten minutes.

2.5 Measurements

2.5.1 Questionnaires

Perceived empathy

To measure to what extent participants felt emphasized with and understood by the other participant, seven questions on perceived empathy were asked. These questions have previously been used by Okel (2018) which in turn based them on a questionnaire originally developed for measuring perceived empathy in a salesperson-customer relation (Plank, Minton, & Reid, 1996). The internal validity is considered sufficient as the Cronbach's alpha is high: α =0.84.

Social flow

Flow is defined by Walker (2010) as: *"an intrinsically rewarding highly absorbing state in which people lose a sense of time and the awareness of self."* (p. 3). Early research on flow saw it as an individual and not a group phenomenon, while later research noted that perhaps the most enjoyable flow experiences occur during social interactions (Walker, 2010). Social flow is referred to when flow is experienced in a social context where people have interdependent relations such as in a football team or a jazz band performing a jam session. No previous studies have been found focusing on the relation between social flow and PS and therefore, including social flow in this study is somewhat exploratory in nature. Nevertheless, it intuitively makes sense that social flow and PS are positively related, since experiencing social flow requires a certain level of interconnectedness with another person, based on the definition of social flow by Walker (2010).

To measure social flow, six questions on a Likert scale from one to seven were asked. Together these questions aimed at obtaining a measure of the extent participants felt immersed in the task and were in a state of flow together with the other participant. These questions are based on a study on social flow in a collaborative mobile learning setting (Ryu & Parsons, 2012). The Cronbach's Alpha of this measurement was high: α =0.84.

IOS

Previous research on the effect of PS feedback showed an effect of (fake) PS feedback of the level of PS on IOS (Okel, 2018); high PS feedback was related to higher IOS scores compared to low PS feedback. To study whether this effect will also be found for real PS feedback, the same measurement will be taken. The Inclusion of Other in Self is a single-item, pictorial measure of closeness devised by Aron, Aron & Smollan (1992) and has been widely used. It consists of seven pictures that differ to the extent to which the self and the other person are overlapping (see Appendix A).

2.5.2 Physiology

Heart rate. The heart rate of two participants is measured simultaneously using two Mobi-8 devices with ECG electrodes with a sampling frequency of 1024 Hz. To measure the ECG signal, three sticker electrodes, two for the bipolar input and one for the participant's ground, are placed according to the standard II lead positioning.

Heart rate synchronization. To provide feedback on the level of heart rate synchronization, the level of heartrate synchronization is calculated in real time. Heart rate synchronization is defined as the correlation between the participant's inter-beat-intervals (IBI's) over a period of 60 seconds. Formula 1 shows how this correlation is calculated.

$$r_t = \frac{\sum(IBI_{1k} * IBI_{2k}) - \frac{1}{N} * \sum IBI_{1k} * \sum IBI_{2k}}{\sqrt{(\sum IBI_{1k}^2 - \frac{1}{N} * (\sum IBI_{1k})^2) * (\sum IBI_{2k}^2 - \frac{1}{N} * (\sum IBI_{2k})^2)}}$$

Formula 1: Heart rate synchronization formula

Where IBI_1 and IBI_2 are the values of measured inter-beat-intervals for participant 1 and 2 at the sample moments indicated by the index *k*. N stands for the window size, in this case the number of IBI's taken into the calculation for a single correlation measurement. Each second, the latest IBI value is sampled from each person. This is done to prevent a time-lag between measurements of a person with a higher IBI and a person with lower IBI's, but it does mean that a subsample of IBI's is taken to measure the correlation between IBI's of participant 1 and 2.

This formula is implemented in HRSync, which is a home-made application created by the HTI group of Eindhoven University of Technology (Boschman, 2018). HRSync detects QRS complexes from ECG data received from Mobi-8 devices and provides feedback on the correlation coefficient between the IBI's of the two participants measured over the last N seconds, sampling the last measured IBI each second.

Outlier Detection

To prevent outliers having a large influence on the PS level, an impulse filter that is implemented in HRSync is applied to the incoming IBI data (Boschman, 2018). These outliers can be caused by bodily movements or temporal signal loss and can have substantial influence on the calculated correlation value. The following test statistic is used to account for this:

$D(n) = \frac{|x(n) - x_m|}{1.483 * med\{|x(n) - x_m|\}}$ (McNames, Thong, & Aboy, 2004).

Formula 2: Outlier detection formula implemented in HRSync.

Where med{.} is the median operator applied of the entire window of 60 seconds and thus specifies the median difference between a certain IBI and the median IBI in the window-size of 60 seconds. x_m is the median value of the signal over the 60 seconds period x_m = med{x(n)}.

An outlier is detected when a data point exceeds the user-specified threshold, in this case when *D*(n)>4. If this happens, a new value is calculated and inserted instead of the actual detected IBI value. The formula for this new value is as follows:

$$S_i\,(n) = med\{\, x(n+m) : |m| \leq \frac{w_m - 1}{2}\}$$

Formula 3: Outlier criteria formula implemented in HRSync.

Where w_m is the window length of the segment centered about n that the median operator is applied to. For this study a user specified threshold of 4 is used and a window length of 5 samples. Thus, when D(n) becomes greater than 4 for a specific data point n, a new value S_i is estimated and inserted instead. S_i is the median IBI value of the two preceding and two succeeding IBI's of IBI *x*(n) (Boschman, 2018).

2.5.3 Team performance

To calculate team performance on the two survival tasks, the order of items made by participants is compared to the 'optimal order' created by experts on survival situations. In the case of the Lost at Sea task, the experts are the US coastguard and regarding the Desert Task, the optimal order is created based on over 2000 actual cases in which men and woman lived or died depending upon the survival decision they made (Pond, Lafferty, & Eady, 1974). The final score was derived by computing the absolute difference between the participants order and the optimal order. Thus, for each item the difference in rank was calculated and summed up. Subsequently, this absolute difference between the two rankings was subtracted from 100 to obtain a measure where high scores reflect high-quality solutions as can be seen in Formula 4.

$$\text{Score} = 100 - \left(\sum_{15}^{1} P_i - C_i \right)$$

Formula 4: Team performance calculation.

Score is the team performance measurement, P_i the rank of the item in the participants list and C_i the rank of the item in the correct list. This scoring procedure has been used in previous studies using the Desert survival task (Thompson & Coovert, 2003; Rogelberg & O'Connor, 1998).

2.6 Physiology Feedback

Feedback on the level of heart rate synchronization was provided in the form of a colored border around the task on the monitor. The order of the items for the collaboration task had to be made on the screen to maximize the amount of time participants looked at the feedback shown on the monitor. A simple program was created that would receive a value from HRSync and convert it to a window colour to provide real time feedback in an easily interpretable manner. The correlation value obtained from the HRSync program was converted to either green, yellow or red depending on the value. A small pilot study was done to find appropriate values to make sure participants would not only see one color for the whole duration of the task, but instead see some variation in colors that would be more informative. Green was used for a correlation larger than 0.2, yellow for correlation between 0.2 and 0.1, and red for correlation lower than 0.1. A new correlation value would be sent from HRSync to the color program only once every five seconds to prevent a constant change of color that might lead to annoyance. However, at the end of the experiment, some participants mentioned that they still found the changing of colors distracting, but a sampling frequency lower than five seconds was judged to lead to a too large reduction in accuracy of feedback. The colors used are similar as the ones used in the study of Okel (2018).

2.7 Procedure

This section describes, step by step, the procedure of the experiment; a visual overview can be seen in Figure 2 together with approximations for the duration of the separate steps.

A dyad of two participants was required to be present to run the experiment. If only one participant showed up, the experimenter tried to quickly find a second person close to the experiment room who was willing to participate. When two participants were present, they were welcomed into experiment room and asked to fill in a form of informed consent prior to the experiment. Then participants were guided to their seats where they were given instructions how to attach the electrodes for the ECG measurements. This usually involved checking whether the signal was correct a few times and readjusting the electrodes or firmly pressing the electrodes on the skin so that they would not fall off during the experiment. When all was set and ready, participants were first shown a short 3-minute video of fish in an aquarium for a baseline measurement as well as a means to calm the participants such that they were in the same mental state. During the baseline measurement and subsequent survival tasks, the experimenter would be seated in a different room with the door open so that he could hear when participants were done discussing.

After 3 minutes, the experimenter would come back in and explain the first task. To control for the effect that slight differences in the tasks could have, half of the dyads would do the Desert task first and the other half would do the Sea task first. Participants first got 3 minutes to make an order of items individually after which they got 12 minutes to discuss and create an order of items together. The individual part had to be written down on paper whereas the list made together was done on the monitor using only a mouse to drag and drop pictures. When dyads were finished after at least eight minutes of discussion, the experimenter would come back in and give the first set of questionnaires. No feedback was given yet during the first task for neither condition. Following the first set of questionnaires, the second survival task was given for which the same procedure was applied where only half of the dyads received feedback on the level of PS. The only difference was that a verbal explanation was given on the changing colors on the monitor to the dyads who received physiology feedback during the discussion.

The explanation was as follows: *"During this task, you will see changes of color in the background. The colors are red, yellow or green. These colors represent the extent to which your heartrates are synchronized. Red indicates a low level of synchrony, yellow an intermediate level and green represents a high level of synchronization of your heartrates. When the heartrate of person A and person B increase and decrease together at similar moments in time, a high level of synchrony is obtained. Lastly, it is not your task to aim for a specific color, you should just do the task while you are able to see the colors changing."*

After the explanation, the experimenter left the room again and came back when the discussion was finished to bring the second set of questionnaires. After this, the experiment was finished, and a debriefing was given while participants removed the electrodes and signed for receiving their reward. If there was enough time left, participants could read the 'optimal order' of the survival tasks to get some idea on how well they did, while the room was made ready for the next dyad.

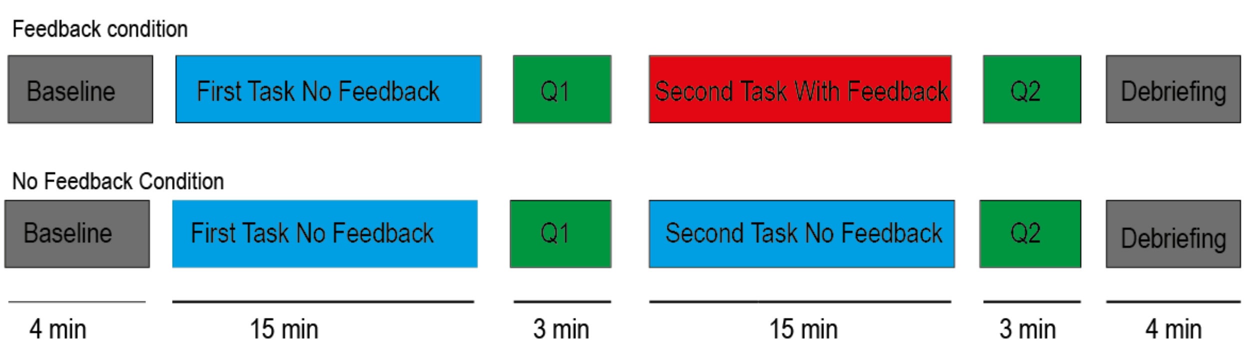


Figure 2: Schematic overview experiment procedure.

2.8 Statistical Analyses

2.8.1 Preprocessing

Physiological Synchrony

To obtain one measure of PS per dyad per session[2] (baseline, first task and second task), an average is taken of all synchrony values during each session. The first 60 values are ignored as these values capture information prior to the session since a correlation is calculated over a period of the preceding 60 seconds.

Outlier Detection Accuracy

In this section, a short description is given on the process for checking the accuracy of the outlier detective implemented in HRSync. This testing of accuracy is done by quantifying the amount of IBI's that have been interpolated in the HRSync program. Using the raw ECG data obtained from running the experiment in this study, a second set of IBI's is calculated using the program RRDetect (Boschman, 2017) in which a QRS detection method is implemented (Hamilton, 2002). These new IBI's are compared to the IBI's used in the HRSync program for the first 10 dyads to obtain a rough estimation for the number of interpolated IBI's. On average, less than 1% of the IBI's used in HRSync have been interpolated (see Appendix B). Furthermore, after manual inspection of the replaced IBI's in the ECG signal, the replacements seem to be appropriate, as often sections with poor signal quality were found at those specific points in the ECG data. These noisy sections might be the result from artefacts such as movement by the participant or possibly disturbances in the signal transmission from the Mobi device to the laptop. Appendix C gives an example how an unusual IBI can be detected by anomalies in the ECG signal.

Even though only a small portion of all the IBI's has been interpolated, the heart rate synchrony formula (Formula 1) is also implemented to calculate a second synchrony score using the IBI's detected by RRDetect to investigate the potential difference of using slightly different IBI's. Again, this was only done for the first ten dyads. On average, there is an absolute difference in correlation of 0.07 between the correlation metric based on IBI's from HRSync compared to IBI's from RRDetect, which is deemed as appropriately small (see table 2 in Appendix B).

Inter-Beat-Interval check

Even though the program HRSync already has an outlier detection mechanism build into it as explained in section 2.5.2, an extra check has been done to ensure no anomalies regarding IBI values occurred. This extra check has been performed after HRSync smoothed out potential anomalies in IBI values by taking the average over a segment with window size of five. According to criteria specified by Overbeek, Boxel and Westerink (2012), IBI's smaller than 0.4 or larger than 1.4 are inspected as well as IBI's being 0.5 smaller or 1.5 larger than the running average of the ten preceding IBI' values. No abnormal IBI's were detected, suggesting that the inherent outlier detection method in HRSync functions well.

[2] The word session refers to a sub-part of the experiment; i.e. the baseline measurement, first task measurement and second task measurement.

Creation dependent variables

This study uses a dyad of two participants as a single datapoint, meaning that a single value for each dyad is calculated based on two inputs. To compute a single score per dyad for perceived empathy, IOS and social flow, a simple average is calculated based on the scores of two individuals forming one dyad. Although this averaging is not ideal since it ignores potential large differences between participants, due to absence of a better combination metric, forthcoming analyses are performed using the average scores. Therefore, each time the variables perceived empathy, IOS and social flow are mentioned, it is referred to as the average score of a dyad.

Feedback

A post hoc analysis on the heart rate synchrony data, obtained from the experiment in this study, was conducted to obtain the amount of time each color was shown as feedback to participants. A small issue had to be resolved as feedback was given once each five seconds while a heart rate synchrony correlation was calculated each second, making it difficult to find out which value was taken to provide feedback. To provide a reliable measure of this, five separate counts of color distributions were calculated, each with a step size of 5 seconds but with a different starting point such that each possibility is covered. For each dyad and for each color, the average percentage of each color shown is taken of the five separate counts. Red (*M*=63.7%, *SD*=1.81%, *Min*=28%, *Max*=93%) clearly was shown most of the time, followed by green (*M*=14.9%, *SD*=11.72%, *Min*=6.30%, *Max*=37.5%), and participants saw yellow the least (*M*=21.4%, *SD*=13.5%, *Min*=0.19%, *Max*=56%). To get a single feedback score for each dyad, the percentage of green is subtracted from the percentage red.

In order to answer the research question on the potential effect of feedback of PS on the dependent variables (perceived empathy, team performance, IOS and social flow), two groups are created; a high feedback group and a low feedback group. The distinction between high and low feedback is made with a bi-split at the median of the feedback score variable, such that group sizes are equal, which is shown in Figure 3. Because this Figure 3 shows that this split is not optimal due to the absence of two distinct lumps, three other splits have been made, but all lead to the same results when studying the effect of feedback of PS on the dependent variables as described in the third section of this chapter (see Appendix E).

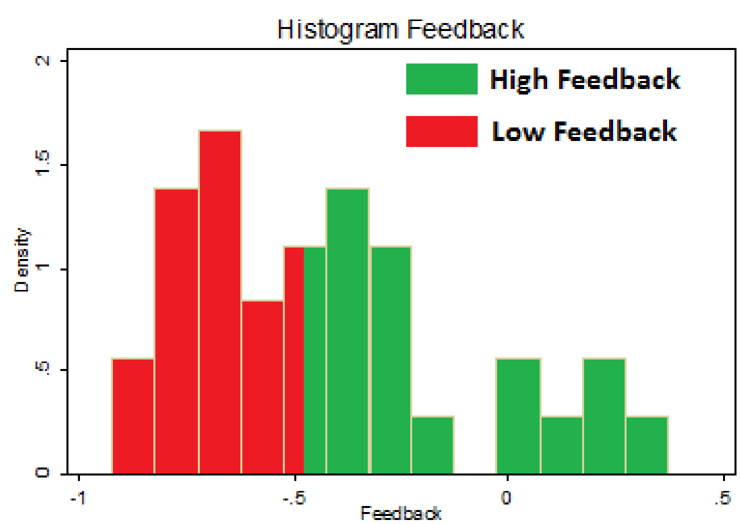


Figure 3: Histogram of feedback score to decide on position of the bi-split (-0.45).

Figure 4 shows the distribution of colors shown as feedback in terms of percentages for the high and low feedback group. While even the high feedback group still gets to see red most of the time, there is a clear difference in distribution of colors between the high and low feedback group.

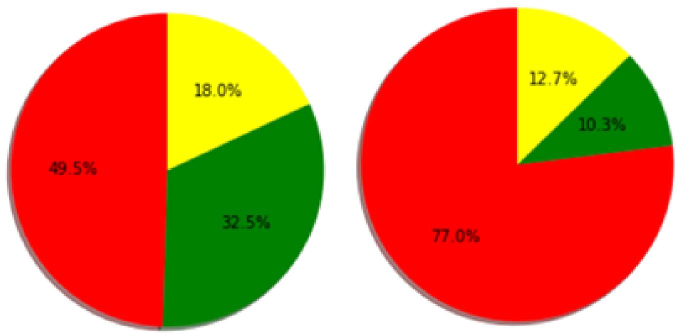


Figure 4: Distribution of color shown in feedback for the high feedback group (left) and low feedback group (right).

Nevertheless, this is a much smaller difference in distribution of colors shown than was the case in Okel's (2018) study as can be seen in table 1.

Table 1:

Overview of the distribution of colors shown as feedback in percentages.

	High feedback group			Low Feedback group		
	Red	Yellow*	Green	Red	Yellow*	Green
Fake Feedback (Okel)	0	24	76	76	24	0
Real Feedback (Current study)	49.5	18	32.5	77	12.7	10.3

**Okel (2018) used orange as the color for intermediate levels of PS. By accident this study used yellow instead.*

2.8.2 Analysis

Four Pearson Correlation tests will be done to answer the first research question on the relation between PS and the dependent variables: perceived empathy, team performance, IOS and social flow.

Four two-way repeated measures ANOVA's will be conducted to answer the second research question on the moderating effect of feedback on the relation between PS the dependent variables. Factors that will be used are session (baseline, first task, and second task) and feedback (no feedback, low PS feedback and high PS feedback).

3. Results

Due to malfunctioning of the heartrate sensors, no data was collected for four dyads, resulting in a dataset of 66 dyads.

In the first section of this chapter, the descriptive of the dependent and independent variables will be discussed. Following this, in section 3.2 and 3.3 the research questions will be answered using the statistical tests mentioned in section 2.8.2 and finally a short section on people's experiences with the feedback mechanism is given in section 3.4.

3.1 Descriptive measurements

3.1.1 Physiological Synchrony

A one-way repeated measures ANOVA was run to investigate differences in PS between the three sessions (i.e. baseline, first session and second session). Results show no statistically significant effect of session on PS; $F(2,128) = 0.19$, $p = 0.82$. Shapiro-Wilk tests for PS across all sessions show that normality could not be rejected (p-values of respectively p=0.98, p=0.21, p=0.71). Homogeneity of variances was rejected indicated with $p< 0.05$ after performing Levene's test. However, ANOVA F tests are generally robust to violations of the equal variance assumption when group sizes are equal (Rogan & Keselman, 1977), so violation of this last assumption is deemed as of no big importance. Figure 5 shows per session the average level of PS together with the 95% confidence interval.

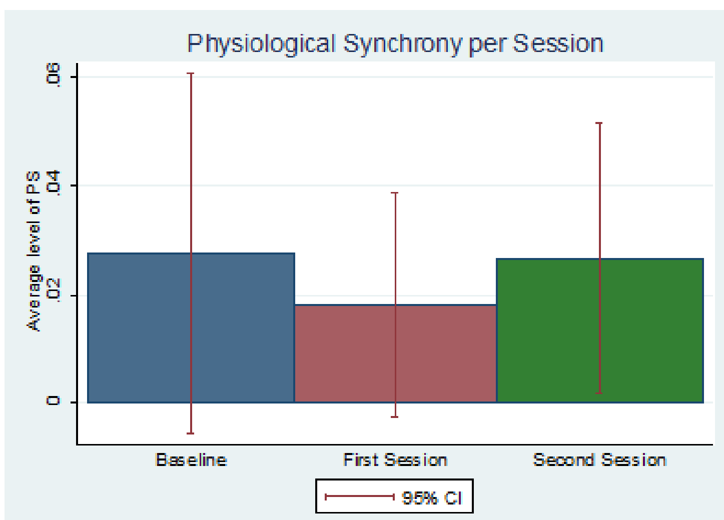


Figure 5: Mean Physiological Synchrony over the three sessions.

Pearson correlation tests were performed to study the relation of PS between sessions within dyads. While no statistically significant correlations were found between the baseline measurement and either of the two sessions, there is a statistically significant correlation between PS during the first and second session ($r(66)=0.26$, p=0.047), suggesting a relatively stable level of PS within a dyad when having a discussion. Next to this, to assess whether a general pattern occurred during the session, a plot showing the development of PS over time has been made that includes data from all dyads, which can be seen in two graphs in Appendix D. Eyeballing these graphs, makes it clear that there is no general pattern, which is also supported by the average level of PS during the session that fluctuates around zero.

3.1.2. Team performance

To test whether the separate tasks (Desert task and Sea task) were equal in terms of difficulty, a paired samples t-test was done on the scores for the Desert Task (M_d = 29.97, SD_d = 11.29) and the Sea task (M_s =38.71, SD_s = 12.53). The null-hypotheses of equal task difficulty was rejected ($t(69) = 5.07$, $p<0.001$). Thus, dyads scored significantly better on the sea task compared to the desert task.

While this inequality in task difficulty is not ideal, it is not problematic as long as there is no difference in scores within the tasks themselves depending on which task was done first. Meaning that scores on the desert task should not be different depending on whether the desert task was done first or second. Likewise, for the sea task, there should be no difference between scores on the sea task whether it was done first or second. To check for this, two independent t-test were done. For the desert task it made no statistically significant difference whether the task was performed first (M_{df} = 31.00, SD_{df} =10.92) or second (M_{ds} = 29.00, SD_{ds} = 11.69) ($t(68) = -0.74$, $p = 0.46$). Also, for the Sea task it did not matter if the Sea task was done first (M_{sf} = 38.42, SD_{sf} = 11.14) or second (M_{ss} = 39.03, SD_{ss} = 14.03) ($t(68) = -0.20$, $p = 0.84$).

Next to this, team performance did not improve during the experiment. There was no statistically significant difference between team performance during the first session (M_f= 34.81, SD_f=11.57) compared to the second session (Ms=33.87, SD_s=13.75) ($t(69)=0.47$, $p=0.64$). In some more detail; team performance increased when the desert task was done prior to the sea task and decreased when the sea task was performed prior to the desert task. This was to be expected as the average score on the sea task was significantly higher than on the desert task. Figure 6 shows the scores for the tasks as well as an average together with a 95% confidence interval for the two sessions.

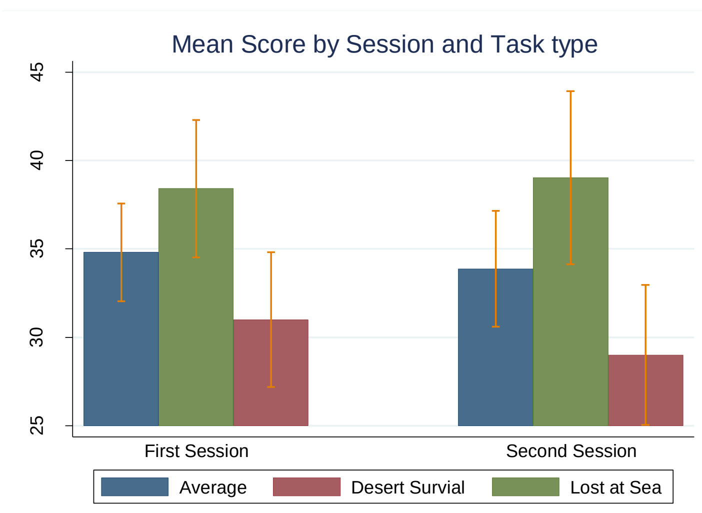


Figure 6: Mean Team Performance per task type and session.

3.1.3 Perceived Empathy

No statistically significant difference was found in perceived empathy during the first session (M_f = 5.41, SD_f=0.74) compared to the second session (M_s = 5.17, SD_s = 0.78) (t(69) = 1.20, p = 0.23). Figure 7 shows this in a graph along with the 95% confidence interval.

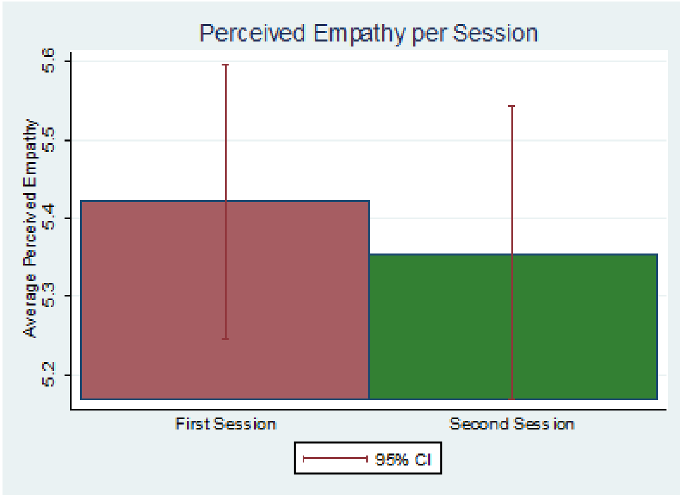


Figure 7: Mean Perceived Empathy for each session

3.1.4 IOS

Using a dependent t-test, no statistically significant difference is found in IOS between the first (M=4.82, SD=1.28) and second session (M=4.74, SD=1.57) (t(67)=0.69, p=0.49). Figure 8 shows the average level of IOS for the two sessions together with the 95% confidence intervals.

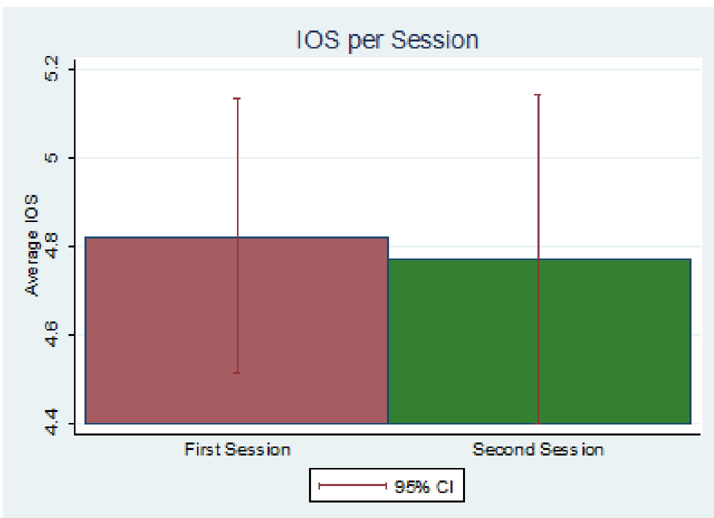


Figure 8: Mean IOS per session.

3.1.5 Social Flow

Using a dependent t-test, results show no difference of social flow between the first (*M* = 5.92, *SD* = 0.60) and second session (*M* = 5.88, *SD* =0.63) (t(69)= 1.15, *p*=0.26), this is visualized in Figure 9.

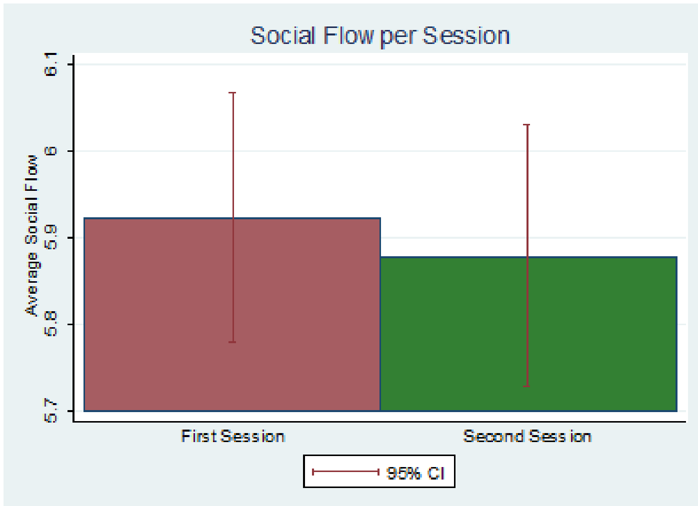


Figure 9: Mean Social Flow per session.

3.1.6. Checking influence of control variables

Since there were no strict requirements for participation in this experiment, control tests were conducted to investigate the potential confounding effects of three variables: 1) dyad composition in terms of gender, 2) whether participants performed the discussion tasks in their native language or not and 3) the difference in age between participants forming one dyad. A one-way repeated ANOVA was conducted for the effect of gender composition where gender composition was split into three categories (male-male, female-female, male-female). For the effect of language, a one-way repeated measures ANOVA was conducted where language was split into two categories (first language or not). Lastly, for the effect of the difference on age, also a one-way repeated measures ANOVA was conducted. After inspecting a histogram of age difference, it was decided to split the variable age difference into two groups, depending on whether the difference in age was larger or smaller than ten years. To control for the increased type 1 error by doing so many tests, a Bonferroni correction was applied. Details of these tests are summarized and can be found in Appendix F. The results only show a statistically significant of native language on team performance; team performance is better when participants spoke their native language (*M*=37.46, *SD*=9.69) compared to when they spoke in their non-native language (*M*=30.42 *SD*=7.76) (*t*(68)=-3.29 , *p*=0.002). Aside from this, no statistically significant effects are found of these control variables on the dependent variables: perceived empathy, team performance, PS, IOS, and social flow.

3.2 Correlations Physiological Synchrony and Dependent Variables

To answer the first research question on the relation between PS and perceived empathy and team performance, several Pearson Correlation tests have been conducted. The following sections discuss the relation between PS and the dependent variables (perceived empathy, team performance, IOS, and social flow) separately.

3.2.1 Physiological Synchrony and Perceived Empathy

No significant correlation was found between PS and perceived empathy during the first session ($r(67)=0.04$, $p=0.72$) nor during the second session ($r(67)=0.18$, $p=0.13$). Figure 10 visualizes these results and it can easily be seen that the lines are almost horizontal, indicating no statistically significant correlations.

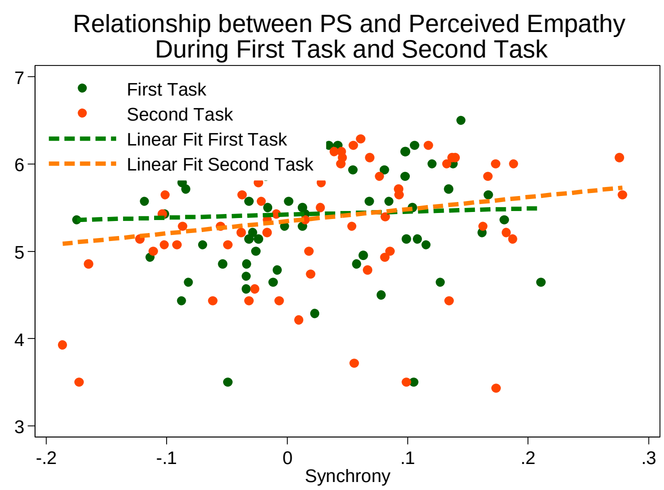


Figure 10: Relation between Physiological Synchrony and Perceived Empathy for session 1 (green) and session 2 (orange).

Figure 11 shows a scatter plot with a linear fit for the relation between PS and perceived empathy during the second session for three different types of feedback groups: 1) no feedback, 2) low feedback and 3) high feedback. Pearson Correlation tests are conducted to study the relation between PS and perceived empathy for each feedback group. From the graph it can be seen that there is even a negative trend between PS and perceived empathy for the High feedback group ($r(17)=-0.19$, $p=0.47$) although this is statistically non-significant. The no feedback group and the low feedback group show a positive relation between PS and perceived empathy but are also non-significant with statistics respectively: ($r(34)=0.20$, $p=0.26$) and ($r(19)=0.41$, $p=0.09$). Using the *cortesti* module (Hervi, 2000) in Stata, pairwise comparisons between these correlations were made, resulting in no statistically significant differences between the mentioned correlations; high feedback vs no feedback (*z-score*=-1.22, $p=0.22$), high feedback vs low feedback (*z-score*=-1.72, $p=0.09$) and no feedback vs low feedback (*z-score* = -0.076, $p=0.45$).

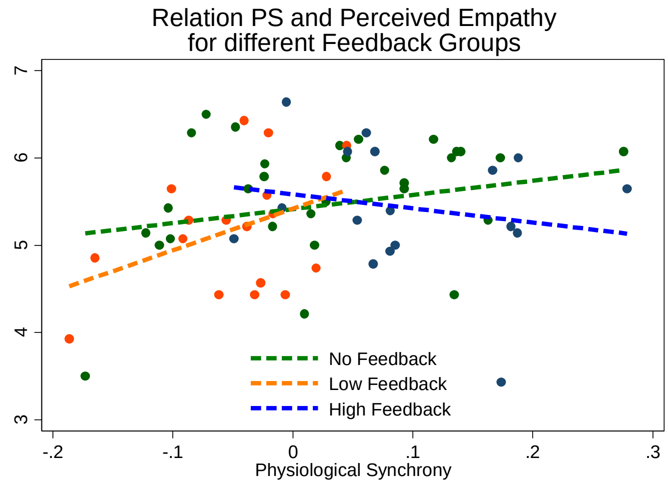


Figure 11: Relation between Physiological Synchrony and Perceived empathy during session 2 for the no feedback group (green), low feedback group (orange) and the high feedback group (blue).

3.2.2 Physiological Synchrony and Team Performance

No statistically significant correlation was found between the PS and team performance during the first session (r(67)=0.-0.02, p=0.87) nor was there one during the second session (r(67)=0.10, p=0.43). Figure 12 visualizes these correlations.

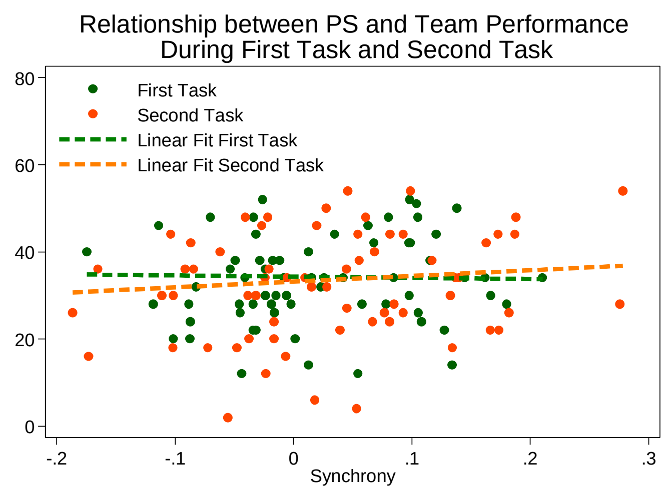


Figure 12: Relation between Physiological Synchrony and Team Performance for session 1 (green) and session 2 (orange).

3.2.3 Physiological Synchrony and IOS

No significant correlation was found between the PS and IOS during the first session ($r(65)=0.05$, $p=0.68$) nor was there a statistically significant effect during the second session ($r(67)=-0.03$, $p=0.79$). This is visualized in Figure 13.

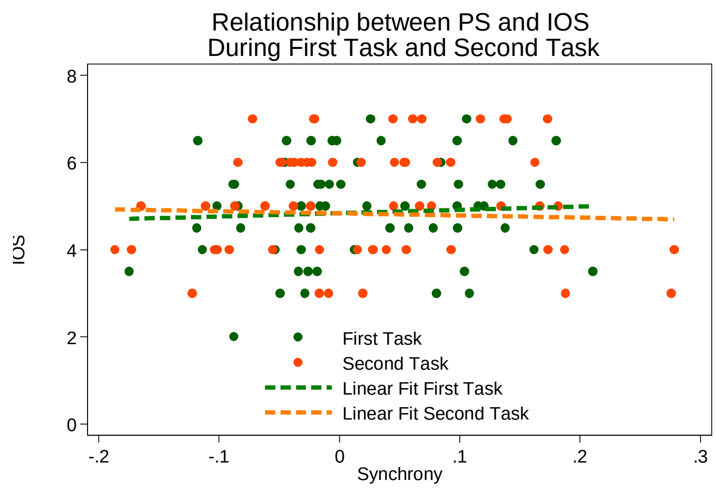


Figure 13: Relation between Physiological Synchrony and IOS for session 1 (green) and session 2 (orange).

Figure 14 shows a scatter plot in which the relation between PS and IOS during the second session is shown for the three different feedback groups: 1) no feedback, 2) low feedback and 3) high feedback. There is a statistically non-significant negative relation between PS and IOS for the high feedback group ($r(17)=-0.33$, $p=0.20$). Both for the no feedback group and the low feedback group there is a slight positive relation but also statistically non-significant; respectively ($r(32)=0.067$, $p=0.72$) and ($r(18)=0.022$, $p=0.93$).

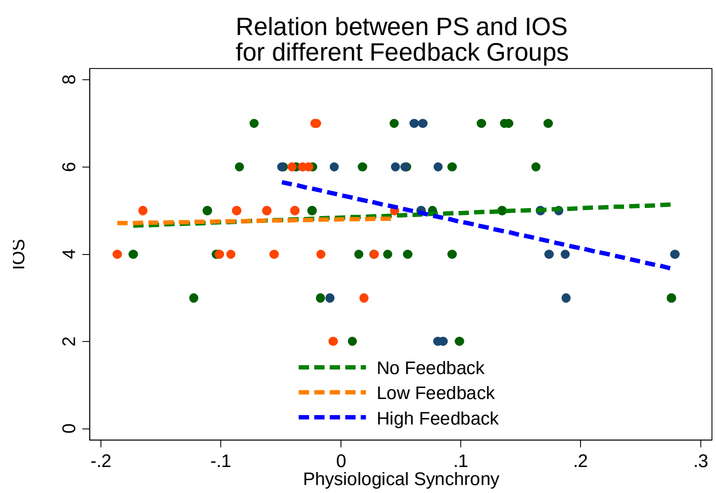


Figure 14: Relation between Physiological Synchrony and IOS during session 2 for the no feedback group (green), low feedback group (orange) and the high feedback group (blue).

3.2.4 Physiological Synchrony and Social Flow

No significant correlation was found between the PS and social flow during the first session ($r(67)=0.03$, $p=0.84$) but there was a statistically significant relation during the second session ($r(67)=0.25$, $p=0.04$). However, considering the many tests performed, a stricter p-value than 0.05 might be more appropriate to reduce the changes to type 1 errors. These results are again visualized in Figure 15.

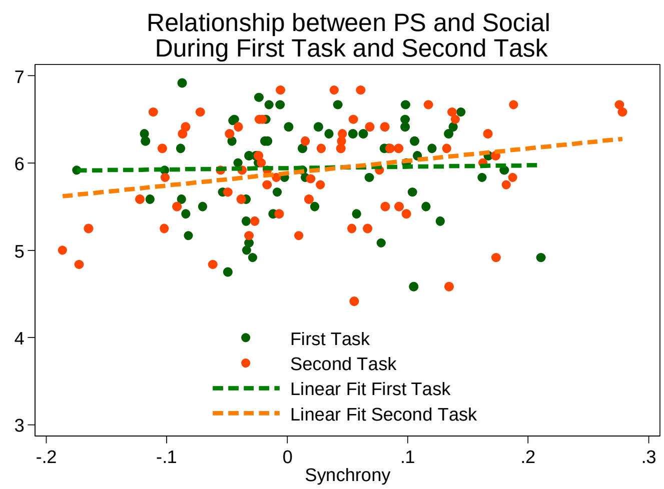


Figure 15: Relation between Physiological Synchrony and Social Flow for session 1 (green) and session 2 (orange).

3.3 Effect of showing Physiology Feedback

To study the effect of showing feedback of the level of PS, four two-way repeated measures mixed ANOVA's are conducted. Session and feedback group are taken as factors to study their effect on perceived empathy, team performance, PS, IOS and social flow. Session has two levels (task 1 and task 2) and feedback consists of three levels (no feedback, low feedback, and high feedback).

3.3.1 Effect of Feedback on Perceived Empathy

A two-way (2 by 3) repeated measure mixed ANOVA was conducted to study the effect of feedback, session and their interaction effect of on perceived empathy. No significant effect was found for the main effects of session ($F(1, 67)=1.66$, $p=0.20$) and feedback ($F(2, 67)=0.44$, $p=0.64$) nor for the interaction effect of session and feedback ($F(2,67)=0.55$, $p=0.58$). Figure 16 shows the level of perceived empathy for each session and level of feedback received. Even though it is uncommon to compare the perceived empathy for the low feedback group between session 1 and session 2 because no interaction was found, a paired sample t-test was conducted and showed no significant difference in perceived empathy for the low feedback group in the first session ($M=5.33$, $SD=0.64$) and second session ($M=5.17$, $SD=0.69$) ($t(18)1.46$, $p=0.16$).

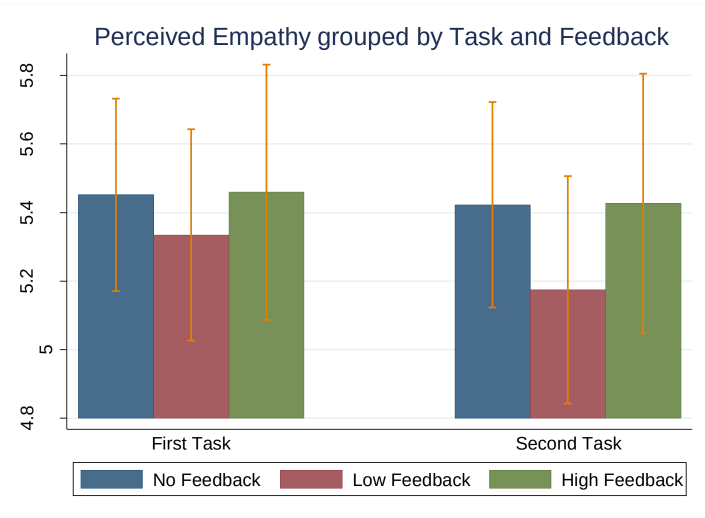


Figure 16: Mean Perceived Empathy per session per feedback group. Note that during the first task, none of the groups received feedback.

3.3.2 Effect of Feedback on Team Performance

No statistically significant effects were found for session (F(1, 67)=0.34, p=0.56), level of feedback (F(2,67)=2.42, p=0.10) and the interaction effect of session and feedback (F(2,67)=0.16, p=0.85) on team performance. Figure 17 shows the overview of team performance for the different sessions and feedback groups.

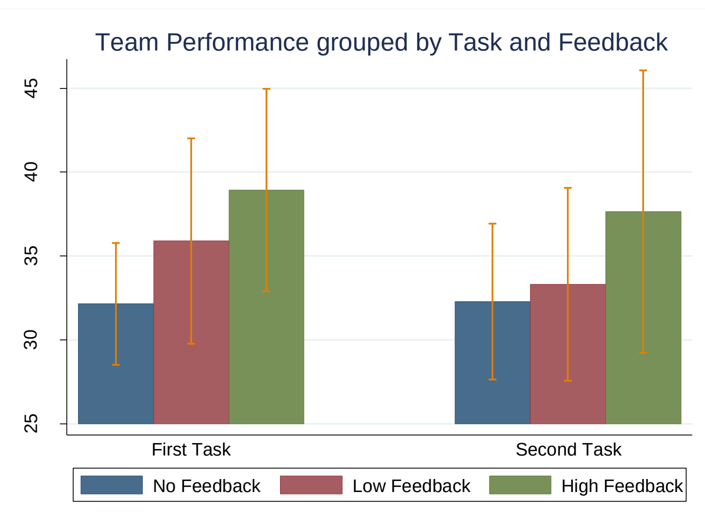


Figure 17: Mean Team Performance per session per feedback group.

3.3.3 Effect of Feedback on Physiological Synchrony

There was a statistically significant main effect of the level of feedback on the level of PS (F(2, 67)=7.23, p=0.0015). Post hoc comparisons show that the high feedback group has a statistically higher level of PS (M=0.061, SD=0.061) compared to the low feedback group (M=-0.024, *SD*=0.057) (*t*(32)=-4.18, *p*<0.001). However, this is not surprising as the level of feedback is directly based on the level of PS and it would be highly troublesome if there was not a main effect of the level of feedback on the level of PS. More surprisingly, the no feedback group has a statistically higher level of PS (*M*=0.028, *SD*=0.078) compared to the low feedback group (*M*=-0.024, *SD*=0.057) (*t*(47)=2.42, *p*=0.02). No statistically significant effects of session (*F*(2,67)=0.16, *p*=0.84) or interaction between session and level of feedback (*F*(4,67)=1.59, *p*=0.18) was found. Figure 18 shows an overview of the level of PS over the two sessions and levels of feedback.

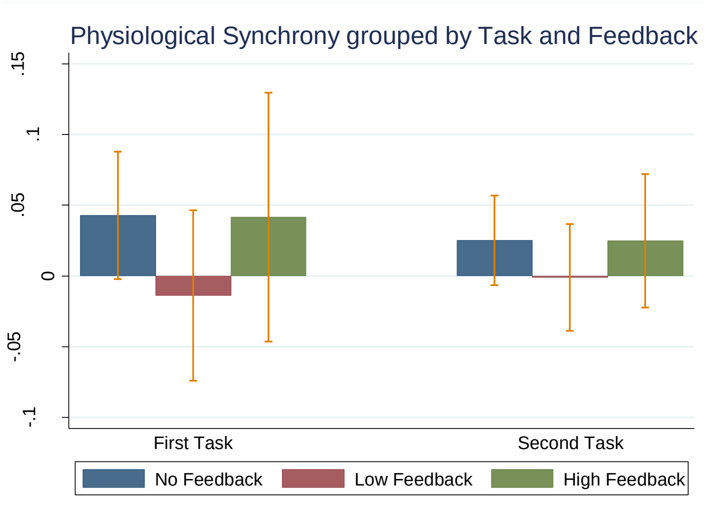


Figure 18: Mean Physiological Synchrony per session per feedback group.

3.3.4 Effect of Feedback on IOS

No statistically significant effects were found for the main effects of session (*F*(1, 67)=1.25, *p*=0.27) and feedback (*F*(2, 67)=0.11, *p*=0.90) nor for the interaction effect of session and feedback (*F*(2, 67)=1.11, *p*=0.33) on IOS. Figure 19 shows these results in a bar chart.

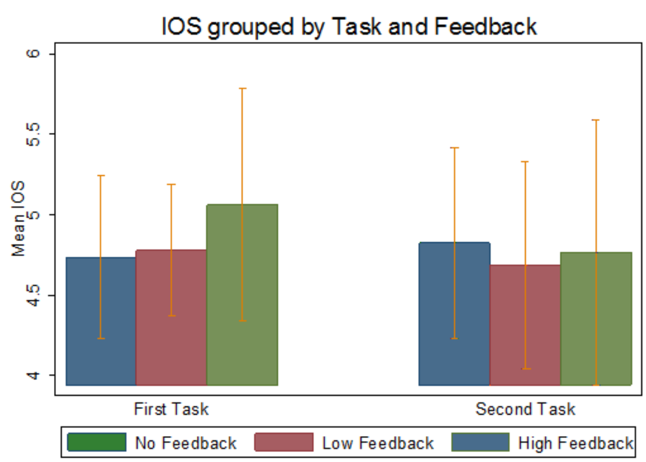


Figure 19: Mean IOS per session per feedback group.

3.3.5 Effect of Feedback on Social Flow

No statistically significant effects were found for the main effects of session ($F(1, 67)=1.13$, $p=0.29$) and feedback ($F(2, 67)=0.80$, $p=0.45$) and the interaction effect of session and feedback ($F(2, 67)=0.97$, $p=0.39$) on social flow. Figure 20 shows these results in a bar chart.

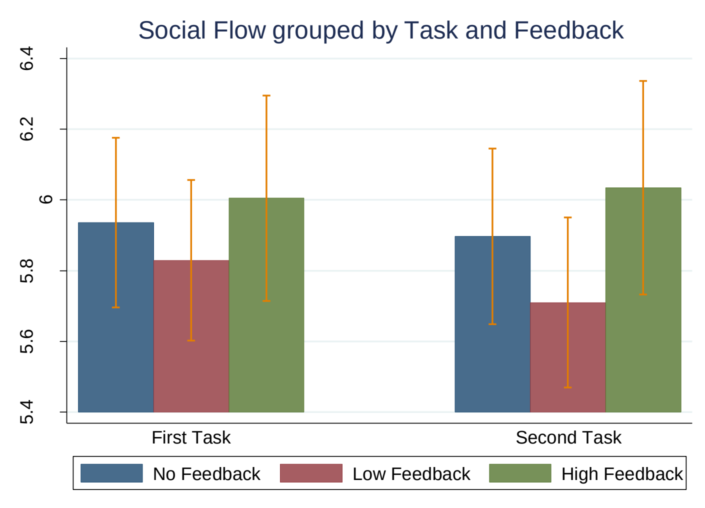


Figure 20: Mean Social Flow per session per feedback group.

3.4 Participant’s experiences on receiving feedback

When in the feedback condition, participant’s opinion was asked about receiving feedback on the synchrony of their heartrate and their partner’s heartrate. 58 participants answered the open question, while twelve participants left the answer box blank.

28 participants (48%) found the feedback distracting but also interesting. This is well summarized in one participant’s answer: *“It was very informative but also very distracting. When the color changed, our focus would shift to the screen instead of the discussion.”* (dyad: 65) Eighteen (30%) participants believed to some extent that their heart rate synchronization did include information about the nature of their discussion. An example of such a perception is: *“It was informative, but a bit distracting. You start to think: Why is the color red when are discussing so well?”* (dyad: 30). Unfortunately, there were also thirteen people (22%) that indicated they did not pay much attention to the feedback as they were mostly concerned with the survival task. One participant mentioned the following: *“The feedback was interesting, but I was too immersed in the task to pay attention to it all the time”* (dyad: 10). Next to this, 15 (26%) participants mentioned that they did not see the usefulness of the feedback as characterized with the following sentence: *“It was interesting to see if we are in sync. It was not distracting, but also not really informative of useful”* (dyad: 57). These percentages do not add up to 100 since participants could mention more than one category.

In general, attention given to the feedback on the monitor seemed to differ quite a lot between participants. Although no quantitative data is obtained on time spent watching the feedback, the experimenter’s perception was that especially older participants did not seem to care much about the feedback whereas the younger participants were more interested. On the other hand, the fact that many participants reported that they experienced receiving feedback as distracting or interesting or both, could suggest that feedback was noticed. Therefore, it is not clear whether this feedback mechanism captured the participant's attention enough for feedback to have an effect.

4. Discussion

The aim of this study was two-fold: First, to replicate findings of previous research that found positive relations between PS and perceived empathy and team performance and secondly, to study the potential moderating effect of feedback on the relation between PS and perceived empathy. No statistically significant correlations were found between PS and perceived empathy and team performance and no effect of providing feedback has been found. The remainder of this discussion includes in the following order: a more detailed discussion of the results, the limitations of this research, the implications of this study, and finally a general conclusion.

4.1 Relation between Physiological Synchrony and other dependent variables

Although the results mainly describe a positive relation between PS and the dependent variables (perceived empathy, team performance, IOS and social flow) these effects are non-significant. This is not in line with most of the findings in previous studies and the first hypothesis of this research that described the expectation of finding a positive relation between PS and perceived empathy and team performance. In the sections below, the relation between PS and dependent variables will be discussed separately.

PS and Perceived Empathy

There was no statistically significant correlation between PS and perceived empathy during the first session ($r(67) = 0.04$, $p = 0.72$) nor was there one during the second session ($r(67) = 0.18$, $p = 0.13$). This is not in line with previous research (Levenson & Gottman, 1992; Marci & Orr, 2006; Marci et al., 2007; Messina et al., 2013) that did find a relation between PS and (perceived) empathy.

Part of the discrepancy between these finding in literature and our results, might be caused by taking heartrate as the physiological measure, whereas skin conductance might have been a better choice as physiological measure. While there are studies done on PS using heart rate as physiological measure (Elkins et al., 2009; Henning & Korbelak, 2005; Henning et al., 2009), most studies that focus specifically on the relation between PS and empathy used skin conductance instead (Marci & Orr; Marci et al., 2007; Messina et al., 2013). Levenson & Gottman used a combination of several physiological measures including heart rate and skin conductance but did not specify how this combined metric was established. Marci et al. (2007) mention that after reviewing the literature on PS, they found that skin conductance was consistently the most sensitive physiologic measure of emotional and empathetic responsiveness. A potential reason for this it that, while most organs are influenced by both the SNS and PNS, sweat glands, which are responsible for the underlying mechanism of skin conductance, are an exception to this and are only influenced by the SNS. This is seen as a large advantage because sweat glands are not influenced by the PNS and other neurohormonal influences that can confound interpretation of measures of other physiologic variables (Marci et al., 2007). This is important, because physiological measures and psychological states are rarely mapped as a one on one relation. For example, heartrate tends to increase both during moments of anger and joy (Cacioppo, Tassinary, & Berntson, 2007).

A second reason might lie in the fact that the task was not ‘strong’ enough in eliciting empathetic emotions because in comparison by previous studies, the survival tasks in this study did not include the explicit discussion of emotions whereas some of the previous studies are based on therapeutic sessions where feelings are more the focus of the conversation (Marci & Orr, 2006; Marci et al., 2007; Messina et al., 2013).

PS and Team Performance

There was also no statistically significant correlation between PS and team performance during the first session ($r(67) = -0.02$, $p = 0.84$) nor during the second session ($r(67) = 0.10$, $p = 0.43$). The above explanation on using heartrate as the physiological measure for a lack of finding statistically significant results, does not hold for the absence for the relation between PS and team performance since there are several studies using heart rate as physiological measure that did find a statistically significant effect (Elkins et al., 2009; Henning, Boucsein, & Gil, 2001; Henning & Korbelak, 2005; Mitkidis, Roeptstorff & Wallot, 2015). However, one of the reasons for not finding an effect might instead have to do with the choice of the team performance measurement, which was the difference in order of items between the dyad’s list and the ‘professional’s’ list. While it can be argued that having a meaningful discussion is part of performing well as a team, the effect on the team score by being able to discuss well, might be much smaller in comparison of having knowledge on survival situations. To illustrate this, consider the following an example: Imagine two people not knowing anything about survival situations but who are able to cooperate very well and listen to each other’s arguments. Compare them with a dyad consisting of a bully that knows a lot of survival situations and an average person who has average knowledge of survival situations. Although, the first dyad would seem to be able to cooperate far better than the second dyad, just because one person in the second team knows a lot about survival situations, their list will probably match the professional’s list better and thus leading to a better team performance score. Therefore, it is argued that a more suitable task would require a higher level of coordination between team members to make the team performance score more dependent on the interconnectedness of the team members and less dependent on the individual knowledge and skills.

PS and IOS

No statistically significant relation has been found between PS and IOS during the first session ($r(67) = 0.05$, $p = 0.68$) nor during the second session ($r(67) = -0.03$, $p = 0.79$). No previous research was found that studied the relation between PS and IOS, making it impossible to compare these results with previous findings.

PS and Social Flow

There was no statistically significant correlation between PS and social flow during the first session ($r(67) = 0.03$, $p = 0.84$) but there was one during the second session ($r(67) = 0.25$, $p = 0.04$). However, due to the fact that eight correlation tests were conducted, a Bonferroni correction was applied to compensate for the increased chance of making a type-1 error, thus weakening statistical support for this relation. Also, when comparing to the effect of session 1 it seems unlikely that there is a true effect. While social flow and PS might seem to be intuitively related, no previous studies were found that study this specific relation.

4.2 Moderating Effect of Physiology Feedback

It was expected to find a moderating effect of feedback of PS on the relation between PS and perceived empathy. This was not the case, as results of a two-way repeated measures ANOVA showed no main effect of feedback (no feedback, low feedback and high feedback) on perceived empathy ($F(2, 67) = 0.44$, p=0.64). This was unexpected, since results from an experiment by Okel (2018) showed a significant difference in perceived empathy when participants received feedback of high levels of PS ($M = 3.17$, $SD = 0.51$, CI[2.99, 3.35]) compared to feedback of low levels of PS ($M = 2.57$, $SD = 0.58$, $CI = [2.36, 2.77]$) ($t(62.9) = -4.54$, $p < 0.001$, $d = 1.12$).

However, there are several potential reasons why this study did not replicate the findings from Okel's research. Firstly, the feedback in Okel's study was fake, whereas feedback in this study was real. This is a major difference, as the fake feedback was far more polarized than the real feedback; in this experiment, the participants in the high feedback group were presented with the color red 50% of the time, whereas this percentage was 0% in Okel's study.

A second reason for not replicating Okel's findings might lie in the difference in task between the two experiments. In Okel's study, participants watched a video that was explicitly chosen for its capacity to evoke empathetic emotions, whereas for this study, as task was chosen based on its capacity of involving social interaction and as well as providing a possibility to measure a team performance score.

Next to the potential difference of evoking empathetic emotions, the tasks most likely also differ in their extent to which feedback was noticed. In the experiment of the current study, participants had to divide their attention between their partner and the monitor on which pictures for the task could be seen, whereas in Okel's experiment, participants only had to watch the monitor. It could very well be that participants were more aware of the feedback in Okel's experiment than they were in this experiment.

Together, these differences might explain at least a part of the discrepancy between the results of Okel (2018) and the current study.

4.3 Limitations

This study suffers from several limitations that are elaborated on in this section. These are: 1) no consistent length of task duration, 2) feedback not being prominent enough, 3) the threshold values of the feedback, and 4) the diversity of the participants sample.

Since the aim of this study was to measure team performance as well as perceived empathy, a means to calculate a team score was required. The difficulty of choosing a time limit for this task was to balance the amount of time in such a way that dyads had enough time to complete the whole list while at the same time ensuring that there was not too much of a discrepancy of time used between the dyads. Based on a small pilot study, it was decided that participants would get twelve minutes of discussion time to create the list. However, during the experiment, some dyads were already finished after six minutes whereas others used all the time that was given to them. Even though dyads who were finished quickly, were urged by the experimenter to take some more time, there was still a considerable difference in discussion length, making it less valid to compare team performance scores.

Secondly, even though measures were taken to make the feedback clearly visible for participants by requiring performing the task on the monitor where the feedback was shown, some participants still reported that they were not giving much attention towards the feedback. One participant mentioned the following: *"I was not really looking at the screen, the task was too much fun"*. However, other participants reported being distracted by the feedback which suggest that they did perceive the feedback. Therefore, it is not clear whether the feedback mechanism used in this study did stand out enough.

The third point concerns the arbitrariness of the thresholds that determine whether feedback shown, indicated a low, intermediate or high level of PS. While this is not a clear limitation, it is an important factor to keep in mind when designing future studies to make sure results can be compared between studies.

Due to the relatively large required sample size, it was decided to not impose severe participation restrictions, which resulted in a relatively diverse sample population. While a high diversity in the participants sample is usually the opposite of a limitation since it provides stronger claims that effects found can be extended to the broader population, it is argued that for this experiment it poses a limitation by introducing factors that could possible influence PS. Not to exaggerate the diverseness of the sample, it has to be mentioned that the large majority of participants were highly educated students from the same university. However, compared to other research on PS, there were considerable differences between dyads in terms of, gender, familiarity, age difference, and cultural background. For example, several studies only used couples (Ferrer & Helm, 2013; Levenson & Gottman, 1983; Liu et al., 2016), specifically focused on therapist-client relationships (Marci & Orr, 2006; Marci et al., 2007; Robinson, Herman & Kaplan, 1982) or specifically focused on only male students (Li & Liu, 2016). In hindsight, even though control checks showed no effects of these control variables, it might have been better to include a more homogeneous sample population.

4.4 Implications and future research

Despite the previously mentioned limitations, the results of this study have several implications of which the first is the complexness of studying the relation between PS and empathy and team performance. When doing research on physiology, two important decisions with regards to the experimental design have to made by the researchers for various factors of which the effects are not well known yet are: the choice of physiological measure and the task during which measurements are taken. A second implication is that the effect of feedback of PS is not so straightforward as initially seemed according to Okel's results, because using real feedback instead of fake feedback leads to a more nuanced situation with less clear differences between dyads regarding the feedback they receive.

Nevertheless, the literature suggests that feedback on physiology does influence our perceptions, because creating awareness of one's physiology can increase the intensity of emotional experience (Janssen et al., 2010). For instance, results from a study by Herbert et al. (2010) showed greater cardiac awareness (how well an individual perceived his or her own heartbeats) is related to greater responsivity of the autonomic nervous system during situations evoking autonomic reactivity. Also, an early study on heartbeat perception showed that giving false feedback on heartbeats, influences perception of emotional states (Valins, 1966). This finding received support from a study by van Dijk et al. (2015) on stress perception in which participant's rating of stress corresponded better with their heartrate when they received physiology feedback. However, in a follow up study, this effect was contributed to an increased awareness of participants' physiology instead of the informational properties of the feedback. Simply put, reminding people to listen their bodily signals influences self-reported stress in such as way it corresponds more to their heartrate.

Continuing this line of arguing it could be that receiving feedback on the level of PS with another person could lead to a higher awareness of the relation with the other person. An oversimplified example of this is a person receiving information about an increase of heart rate when talking to a person and interpreting this as follows: "My heart rate went up and I am not feeling angry, therefore I must like this person". However, this reasoning is highly speculative and, apart from the study by Okel (2018), no empirical evidence has been found that there is an effect of feedback, let alone for why feedback has an effect. Also, results by van Dijk et al. (2015) suggested that it is the increased awareness caused by providing physiology feedback and not the absolute feedback itself that changes how people think about their psychological state. Therefore, when feedback indicates that PS is low, people might use this information to listen more carefully to their body without immediately interpreting this low PS as reason for having a bond or 'click' with someone. Therefore, the effect of showing real feedback on PS should be subject of further investigation, not just on the extent to which it occurs by also on the underlying mechanism why feedback would have an effect.

A second subject for future research is to establish standard measures of PS to be able to compare studies more meaningfully. Currently it is unclear what is defined as a low or high level of PS and studies differ in the calculation of a PS score. Also, as Kleinbub (2017) mentioned, another issue is the many degrees of freedom that multivariate time-series analysis imposes; window size, lag interval size and correction for outliers can lead to a wide range of possibilities that make it possible to 'tweak' the data until results are in line with the original hypotheses. Furthermore, Kleinbub (2017) argues that even though most papers since 2006 reported using the same setting at the study of Marci & Orr (2006), Marci & Orr themselves did not provide a rationale for choosing their parameters. While copying the same set of parameter settings makes it possible to compare these studies, other researchers make the important notion that "*blindly adopting the same settings might lead to skewed conclusions and hinders moving from exploratory studies to confirmatory studies"* (Kleinbub, 2017, p.7). Therefore, there must be a balance between, on the one hand, using the same parameter settings such that studies can be compared, while at the same time making sure that these parameter settings make sense and do not lead to skewed conclusions.

The last implication that will be discussed, concerns the application of technology using PS in therapeutic settings. A qualitative study on the perceived drivers and barriers to adopting technology by psychologists in their line of work with patients, showed that one of the perceived advantages of adopting technology would be the increased intimacy of the therapeutic relationship (Feijt, de Kort, Bongers, & IJsselsteijn, 2018). An example of a technology that could be developed would for instance inform the therapist about the level of PS between the therapist and client such that the therapist has an indicator of the intimacy according to physiological signals. While Kirk (2007) argues that empathy is not the same an intimacy, as intimacy requires intentionality that is not by definition required for empathy (Kirk, 2007), intimacy and empathy are closely related as being able to understand a person's feeling is needed to form an intimate relationship. Having said this, it should be noted that the findings from this research are not in favor of providing evidence that the development of tools based on PS as an important indicator of empathy are helpful in increasing intimacy between therapists and clients. Nevertheless, research on the relation between PS and psychological constructs such as empathy and intimacy should continue as it could provide great prospects for developing tools that might be used to improve therapeutic sessions.

4.5 Conclusion

This research studied the relation between PS and perceived empathy and team performance as well as the effect of feedback of PS. As a measure of PS, the correlation between IBI's of participants was used. Results showed no statistically significant correlation between PS and perceived empathy nor between PS and team performance. Also, when feedback of PS was given, perceived empathy did not differ significantly between dyads with a low level of PS compared to a high level of PS. Therefore, there was no statistically significant effect found for feedback of PS on perceived empathy and team performance.

Because these findings are not in line with results from previous studies, several limitations of this study have been discussed together with differences in experimental design of this study and previous ones. The most important differences are the choice of physiological measure, which often is skin conductance instead of heartrate, and choice of task, which might have been too weak with regards to establishing empathetic feelings and dependency on interconnectedness of dyad members.

Future studies should focus on establishing standard methods for measuring PS to make sure meaningful comparisons can be made between studies. Also, despite having found no significant results on the relation between PS and perceived empathy nor of feedback of PS, research should be aimed at studying the potential of using tools based on PS that can be helpful in increasing feelings of intimacy in therapeutic settings.

References

Adler, H. M. (2002). The sociophysiology of caring in the doctor-patient relationship. *Journal of general internal medicine, 17*(11), 883-890.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology, 63*(4), 596-612. doi:10.1037/0022-3514.63.4.596

Babiloni, F., Astolfi, L., Cincotti, F., Mattia, D., Tocci, A., & Fallani, F. D. (2007). Cortical activity and connectivity of human brain during prisoner's dilemma: an EEG hyperscanning study. *Engineering in Medicine an Biology Society* (pp. 4953-4956). EMBS.

Barone, D. F., Hutchings, P. S., Kimmel, H. J., Traub, H. L., Cooper, J. T., & Marshall, C. M. (2005). Increasing empathetic accuracy through practice and feedback in a clinical interviewing course. *Journal of Social and Clinical Psychology, 24*(2), 156-171. doi:10.1521/jscp.24.2.156.62275

Biocca, F., Harms, C., & Gregg, J. (2001). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. *4th annual international workshop on presence*, (pp. 1-9). Philadelphia.

Boschman, M. (2017). RR Detect User Manual. Eindhoven University of Technology Human Technology Interaction.

Boschman, M. C. (2018). *HRSync User Manual.* Eindhoven: Technische Universiteit Eindhoven Human Interaction Group.

Bradner, E., & Mark, G. (2002). Why distance matter: effects on cooperation, persuasion and deception. *ACM*, (pp. 226-235). New Orleands.

Cacioppo, J. T., Tassanary, L. G., & Berntson, G. G. (2007). *Handbook of Psychophysiology.* Cambridge: Cambridge University Press.

Chanel, G., & Mühl, C. (2015). Connecting brains and bodies: applying physiological computing to support social interaction,. *Interacting with Computers, 27*(5), 534-550. doi:10.1093/iwc/iwv013

Chanel, G., Kivikangas, J. M., & Ravaja, N. (2012). Physiological compliance for social gaming analysis: Cooperative versus competitive play. *Interacting with computers, 24*(4), 306-316. doi:10.1016/j.intcom.2012.04.012

Chanel, G., Pelli, S., Ravaja, N., & Kuikkaniemi, K. (2010). Social interaction using mobile devives and biofeedback: effects on presence, attraction and emotions. *Workshop on Multiuser and Social Biosignal Adaptive Games and Playful Applications, In conjunction with the Fun and Games Conference 2010.*

Condon, W. S., & Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease, 143*, 338-347. doi:10.1097/00005053-196610000-00005

Cramer, D., & Jowett, S. (2010). Perceived empathy, accurate empathy and relationship satisfaction in heteosexual couples. *Journal of Social and Personal Relationships, 27*(3), 327-349. doi:10.1177/0265407509348384

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: a review of the concept. *Emotion review, 8*(2), 144-153. doi:10.1177/1754073914558466

DiMascio, A., Boyd, R. W., & Greenblatt, M. (1957). Physiological Correlates of tension and Antagonism During Psychotherapy: A study of Interpersonal Physiology. *Psychosomatic medicine, 19*(2), 99-104. doi:10.1097/00006842-195703000-00002

Elkins, A. N., Muth, E. R., Hoover, A. W., Walker, A. D., Carpenter, T. L., & Switzer, F. S. (2009). Physiological compliance and team performance. *Applied ergonomics, 40*(6), 997-1003. doi:10.1016/j.apergo.2009.02.002

Feijt, M. A., de Kort, A. W., Bongers, I. M., & IJsselsteijn, W. A. (2018). Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. *Journal of Medical Internet Research, 20*(4), e153. doi:10.2196/jmir.9485

Feldman, R. (2012). Parent-infant synchrony: A biobehavioral model of mutual influences in the formation of affiliative bonds. *Monographs of the Society for Research in Child Development., 77*(2), 42-51. doi:10.1111/j.1540-5834.2011.00660.x

Ferrer, E., & Helm, J. L. (2013). Dynamical systems modeling of physiological coregulation in the dyadic interactions. *International Journal of Psychophysiology, 88*(3), 296-308. doi:10.1016/j.ijpsycho.2012.10.013

Gächter , S., Starmer, C., & Tufano, F. (2015). Measuring the closeness of relationships: a comprehensive evaluation of the Inclusion of the Other in the Self scale. *PloS One, 10*(6). doi:10.1371/journal.pone.0129478

Hamilton, P. S. (2002). *Open Source ECG Analysis Software Documentation.* Retrieved from EP Limited: http://www.eplimited.com/osea13.pdf.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional Contagion.* Paris: Cambridge University Press. doi:10.1111/1467-8721.ep10770953

Hatfield, E., Rapson, R. L., Ye, Y. C., Decety, J., & Ickes, W. (2009). *The social neuroscience of empathy.* Boston: MIT Press. doi:10.7551/mitpress/9780262012973.001.0001

Henning, R. A., & Korbelak, K. T. (2005). Socio-psychophysiological compliance as a predictor of future team performance. *Psychologia, 48*, 84-92. doi:10.2117/psysoc.2005.84

Henning, R. A., Armsteads, A. G., & Ferris, J. K. (2009). Social psychophysiological compliance in a four-person research team. *Applied Ergonomics, 40*, 1004-1010. doi:10.1016/j.apergo.2009.04.009

Henning, R. A., Boucsein, W., & Gil, M. C. (2001). Social-physiological compliance as a determinant of team performance. *International Journal of Psychophysiology, 40*, 221-232. doi:10.1016/S0167-8760(00)00190-2

Hervi, M. C. (2000, Nov 7). Cortesti: Stata module to tset equality of two correlation coeficients, Statistical Software Components S407302. Boston, New York, USA. Retrieved from https://ideas.repec.org/c/boc/bocode/s407302.html

Hodges, S. D., Kiel, K. J., Kramer, A. D., Veach, D., & Villanueva, B. R. (2010). Giving Birth to Empathy: The Effects of Similair Experience on Empatic Accuracy, Empathic Concern, and Perceived Empathy. *Personality and Social Psychology Bulletin, 36*(3), 398-409. doi:10.1177/0146167029350326

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition, 27*(6), 949-960. doi:10.1521/soco.2009.27.6.949

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organisations: From input-process-output models to IMOI models. *Annu. Rev. Psychology, 56*, 517-543. doi:10.1146/annurev.psych.56.091103.070250

Janssen, J. H. (2012). A three-component framework for empathetic technologies to augment human interaction. *Journal on Multimodal User Interfaces, 6*(3-4), 143-161. doi:10.1007/s12193-012-0097-5

Janssen, J. H., Bailenson, J. N., IJsselsteijn, W. A., & Westerink, J. H. (2010). Intimate heartbeats: Opportunities for affective communication technology. *IEEE Trans. Affective Computing, 1*(2), 72-80. doi: 10.1109/T-AFFC.2010.13

Järvelä, S., Kivikangas, J. M., Kätsyri, J., & Ravaja, N. (2014). Physiological Linkage of Dyadic Gaming Experience. *Journal on Multimodal User Interfaces, 45*(1), 24-40. doi:10.1177/1046878113513080

Kaplan, H. B., Burch, N. R., Bloom, S. W., & Edelberg, R. (1963). Affective orientation and physiological activity (GSR) in small peer groups. *Psychomatic medicine, 25*(3), 245-252. doi:10.1097/00006842-196305000-00006

Kirk, T. W. (2007). Beyond empathy: clinical intimacy in nursing practice. *Nursing Philosophy, 8*(4), 233-243. doi:10.1111/j.1466-769X.2007.00318.x

Kleinbub, J. R. (2017). State of the Art of Interpersonal Physiology in Psyhoctherapy: A systematic Review. *Frontiers in Psychology, 8*, 1-10. doi:10.3389/fpsyg.2017.02053

Kozlowski, S. W., & Bell, B. S. (2003). Work groups and teams in organizations,. *Handbook of psychology: Industrial and organizational psychology, 12*, 333-375. doi:10.1002/9781118133880.hop212017

Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: physiological linkage and affective exchange. *Journal of personality and social psychology, 45*(3), 587-597. doi:10.1037/0022-3514.45.3.587

Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of personality and social psychology, 63*(2), 234-246. doi:10.1037/0022-3514.63.2.234

Li, B., & Liu, B. (2016). *Improving team performance by synchronizing heartbeats.* Human Technology Interaction. Eindhoven: Einhoven University of Technology.

Liu, S., Zhou, Y., Palumbo, R., & Wang, J. L. (2016). Dynamic correlaion: A new method for quantifying synchrony with multivariate intensive longitudinal data. *Psychological methods, 21*(3), 291-332. doi:10.1037/met0000071

Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathetic accuracy in a clinically relevant setting. *Journal of personality and social psychology, 68*(5), 854-896. doi:10.1037/0022-3514.68.5.854

Marci, C. D., & Orr, S. P. (2006). The Effect of Emotional Distance on Psychophysiologic Concordance and Perceived Empathy Between Patient and Interviewer. *Applied Psychophysiology and Biofeedback, 31*(2), 115-128. doi:10.1007/s10484-006-9008-4

Marci, C. D., Ham, J., Moran, E., & Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *The Journal of nervous and mental disease, 195*(2), 103-111. doi:10.1097/01.nmd.0000253731.71025.fc

Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of management, 34*(3), 410-476. doi:10.1177/0149206308316061

McNames, J., Thong, T., & Aboy, M. (2004). Impulse Rejection FIlter for Artifact Removal in Spectral Analysis of Biomdedial Signals. *Proceedings of the 26th Annual International Conference of the IEEEE EMBS* (pp. 145-148). San Fransisco: IEEE. doi:10.1109/IEMBS.2004.1403112

Merrill, N., & Cheshire, C. (2016). Habits of the Heart(rate): Social interpretation of Biosignals in Two Interaction Contexts. *Proceedings of the 19th international conference on supporting group work.* (pp. 31-38). ACM. doi:10.1145/2957276.2957313

Messina, I., Palmieri, A., Sambino, M., Kleinbub, J. R., Voci, A., & Calco, V. (2013). Somatic underpinnings of perceived empathy: The importance of psychotherapy training. *Psychotherapy Research, 23*(2), 169-177. doi:10.1080/10503307.2012.748940

Mitikidis, P., McGraw, J. J., Roepstorff, A., & Wallot, S. (2015). Building trust: Heart rate synchrony and arousal during joint action increased by public goods game. *Physiology & Behavior, 149*, 101-106. doi:10.1016/j.physbeh.2015.05.033

Mohr, D. C. (1995). Negative outcome in psychotherapy: A critical review. *Clinical psychology: Science and practice, 2*(1), 1-27. doi:10.1111/j.1468-2850.1995.tb00022.x

Montague, E., Xu, J., & Chiou, E. (2014). Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encountres. *IEEE transactions of human-machine systems, 44*(5), 614-625. doi:10.1109/THMS.2014.2325859

Okel, S. (2018). *The Effect of Biofeedback on Perceived Empathy and Social Connectedness during social interaction.* Eindhoven: Eindhoven University of Technology.

Overbeek, T. J., Boxtel, A., & Westerink, J. H. (2012). Respiratory sinus arrhythmia responses to induced emotional states: Effects of RSA indices, emotion induction method, age and sex. *Biological Psychology, 91*(1), 128-141. doi:10.1016/j.biopsycho.2012.05.011

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Kiu, S., & Goodwin, M. S. (2016). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review, 21*(2), 99-141. doi:10.1177/1088868316628405

Pesarin, A., Cristani, M., Murino, V., & Vinciarelli, A. (2012). Conversation analysis at work: detection of conflict in competitive discussion through semi-automatic turn-orgnization analysis. *Cognitive processing, 13*(2), 533-540. doi:10.1007/s10339-011-0417-9

Reed, R. G., Randall, A. K., Post, J. H., & Butler, E. A. (2013). Partner influence and in-phase versus anti-phase physiological linkage in romantic couples. *International Journal of psychophysiology, 88*(3), 309-316. doi:10.1016/j.ijpsycho.2012.08.009

Robinson, J. W., Herman, A., & Kaplan, B. J. (1982). Autonomic responses correlate with counselor-client empathy. *Journal of Counseling Psychology, 29*, 195-198. doi:10.1037/0022-0167.29.2.195

Rogan, J. C., & Keselman, H. J. (1977). Is the ANOVA F-test robust to variance heterogeneity when sample sizes are equal? *American Educational Research Journal, 14*(4), 493-498. doi:10.3102/00028312014004493

Rogelberg, S. G., & O'Connor, M. S. (1998). Extending the Stepladder Technique: An Examination of Self-Paces Stepladder Groups. *Group Dynamics: Theory, Research, and Practice, 2*(2), 82-91. doi:10.1037/1089-2699.2.2.82

Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Mast, M. S., & Gatica-Perez, D. (2013). Emergent leaders through looking and speaking from audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces, 1*(2), 39-53. doi:10.1007/s12193-012-0101-0

Schwartz, M. S. (2008). A new improved universally accepted official definition of biofeedback: Where did it come from? Why? Who did it? Who is it for? What's Next? *Biofeedback, 38*(3), 88-90. doi:10.5298/1081-5937-38.3.88

Smith, T. J., & Smith, K. U. (1987). Feedback-control mechanisms of human behavior. In G. Salvendy, *Handbook of Human Factors* (pp. 251-293). New York: Wiley.

Staples, D. S., & Zhao, L. (2006). The effects of cultural diversity in virtual teams versus face-to-face teams. *Group decision and negotiation, 15*(4), 389-406. doi: 10.1007/s10726-006-9042-x

Stinson, L., & Ickes, W. (1992). Empathetic accuracy in the interactions of male friends versus male strangers. *Journal of personality and social psychology, 62*(5), 787-797. doi:10.1037/0022-3514.62.5.787

Strang, A. J., Funke, G. J., Russel, S. M., Dukes, A. W., & Middendorf, M. S. (2014). Physio-behavioral coupling in a cooperative team task: Contributors and relations. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 1-13. doi:10.1037/a0033125

Thompson, L. F., & Coovert, M. D. (2003). Teamwork Online: The Effect of Computer Conferencing on Perceived Confusion, Satisfcation and Postdiscussion Accuracy. *Group Dynamics: Theory, Research and Practice, 7*(2), 135-151. doi:10.1037/1089-2699.7.2.135

Valdesolo, P., Ouyang, J., & DeSteno, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology, 46*(4), 693-695. doi:10.1016/j.jesp.2010.03.004

Valins, S. (1966). Cognitive effects of false heart-rate feedback. *Journal of personality and social psychology, 4*(4), 400-408. doi:10.1037/h0023791

van Dijk, E. T., Westerink, J. H., Beute, F., & IJsselsteijn, W. A. (2015). In sync: the effect of physiology feedback on the match between heart rate and self reported stress. *BioMed research international*. doi:10.1155/2015/134606

Walker, C. J. (2010). Experiencing flow: Is doing it together better than doing it alone? *Journal of Positive Psychology, 5(*1), 3-11. doi:10.1080/17439760903271116.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological science, 20*(1), 1-5. doi:10.1111/j.1467-9280.2008.02253.x

Wispé, L. (1986). The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of personality and social psyhology, 50*(2), 314-321. doi:10.1037/0022-3514.50.2.314

Yucha, C., & Montgomery, D. (2008). *Evidence-Based Practice in Biofeedback and Neurofeedback.* Wheat Ridge, Colorado, USA: Association for Applied Psychophysiology and Biofeedback.

Appendix A: Inclusion of Other in Self

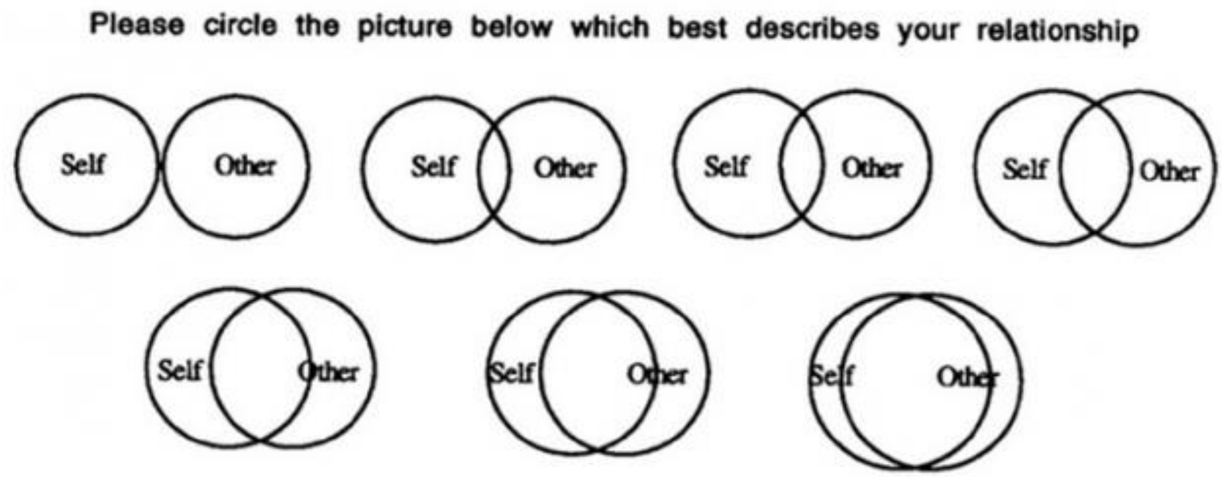


Figure 21: Question to measure IOS (Aron et al., 1992).

Appendix B: Percentages IBI's Changed

Table 1a

Percentage IBI's changed according to the inherent outlier detection in HRSync.

Dyad	Baseline Percentage ibis changed	First Task Percentage ibis changed	Second Task Percentage ibis changed
1	0,00%	0,20%	0,19%
2	0,00%	0,20%	0,00%
3	0,00%	0,68%	0,00%
4	0,00%	0,68%	0,00%
5	0,00%	0,52%	1,09%
6	0,00%	2,97%	3,64%
7	0,00%	0,00%	0,00%
8	0,00%	0,27%	0,00%
9	0,00%	0,00%	0,24%
10	0,00%	0,00%	0,00%
11	0,00%	0,17%	0,75%
12	3,85%	0,00%	0,00%
13	0,93%	0,76%	0,43%
14	0,00%	0,00%	0,00%
15	5,71%	0,00%	0,21%
16	2,86%	0,54%	0,00%
17	5,21%	3,44%	1,01%
	0,00%	5,35%	3,77%
	Average: 1,03%	Average: 0,88%	Average: 0,63%

Table 2b

Difference PS score for IBI obtained from HRSync Compared to IBI obtained from RR Detect.

	Abs Diff Baseline	Abs Diff First Task	Abs Diff Second Task
1	0,07	0,04	0,00
2	0,32	0,15	0,00
3	0,09	0,11	0,06
4	0,08	0,12	0,04
5	0,07	0,03	0,01
6	0,09	0,03	0,05
7	0,03	0,01	0,03
8	0,05	0,06	0,05
9	0,05	0,06	0,21
	Average: 0,09	Average: 0,07	Average: 0,05

Appendix C: Example anomaly IBI and ECG signal

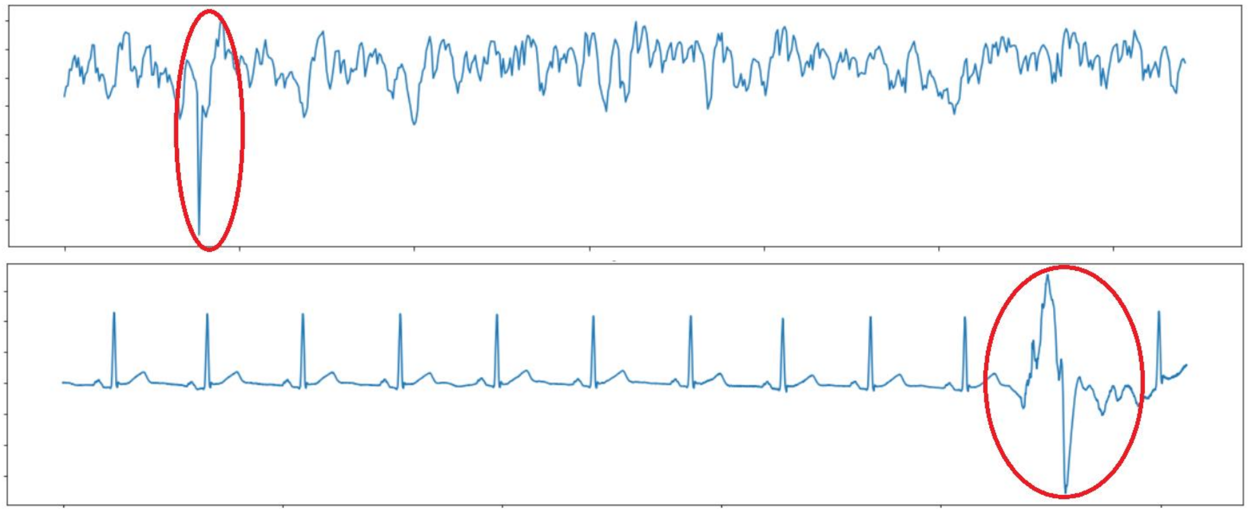


Figure 22: Example of an anomaly in IBI signal (top graph) being reflected in original ECG signal (bottom graph).

Appendix D: Average HR Synchrony of all dyads of the first task

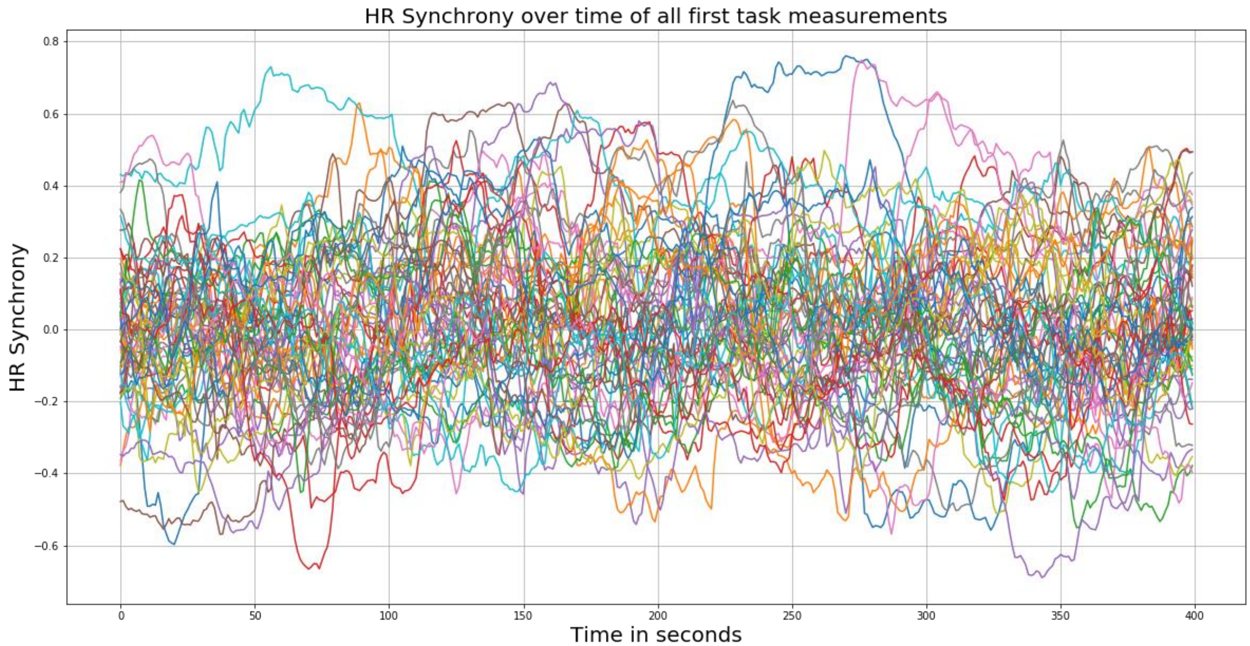


Figure 23: Physiological synchrony values of the first session from all dyads combined to detect a potential general pattern.

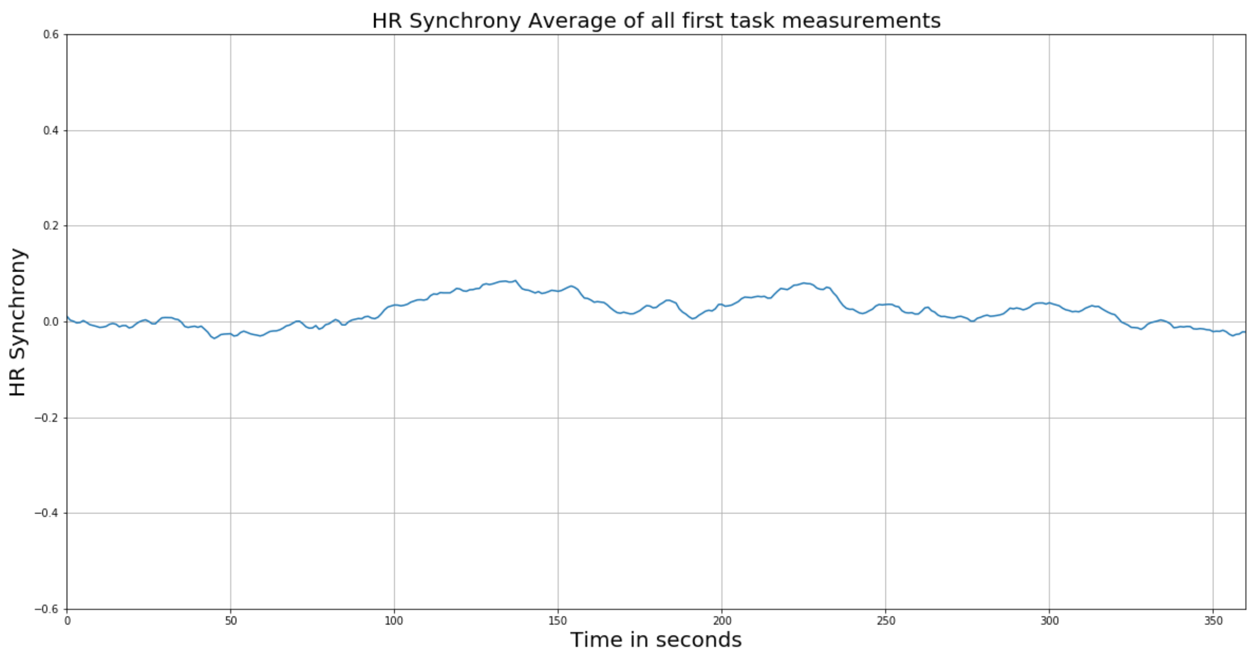


Figure 24: Average physiological synchrony value of all dyads during the first session.

Appendix E: Effect of Feedback with other splits on Feedback-Score

Split at 0 with no exclusions

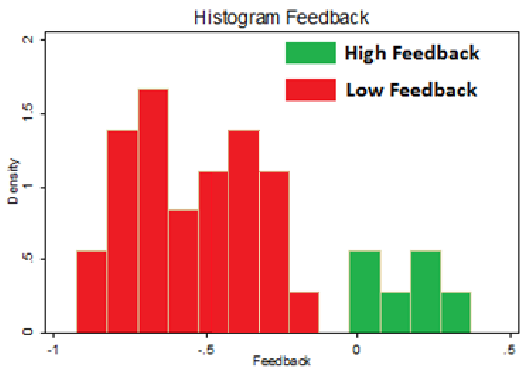
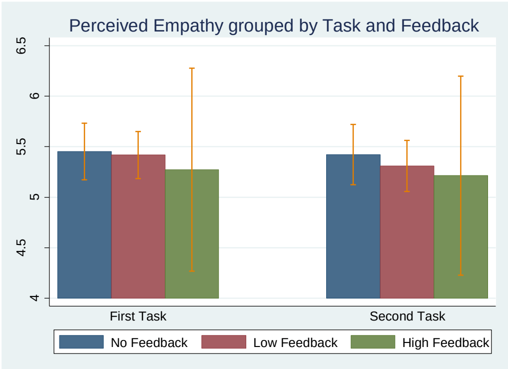
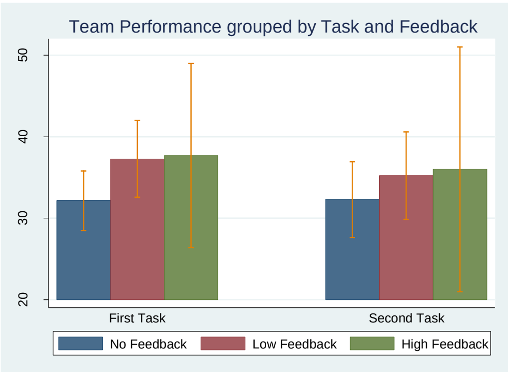
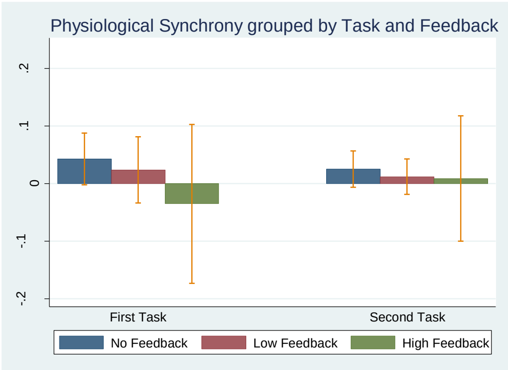


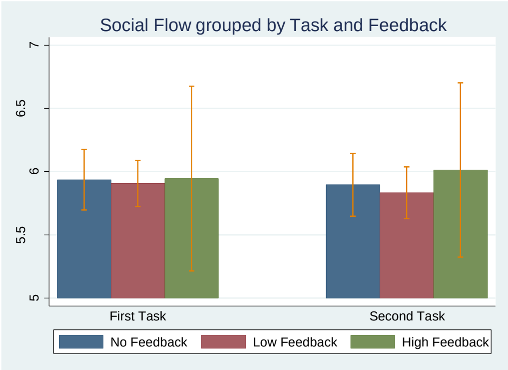
Table 2:

ANOVA results of session and feedback for split at 0 no exclusions			
	Main effect session	Main effect feedback	Interaction session-feedback
Perceived Empathy	F(1,67)=0.79 p=0.37	F(2,67)=0.21 p=0.81	F(2,67) =0.23 p=0.79
Team Performance	F(1,67)=0.19 p=0.66	F(2,67)=1.69 p=0.19	F(2,67) =0.14 p=0.87
PS	F(1,67)=4.12 p=0.019	F(2,67)=1.96 p=0.15	F(2,67)=5.06 p=0.0008
Social Flow	F(1,67)=0.07 p=0.79	F(2,67)=0.10 p=0.90	F(2,67)=0.49 p=0.61









Split a 0 excluding the high feedback group

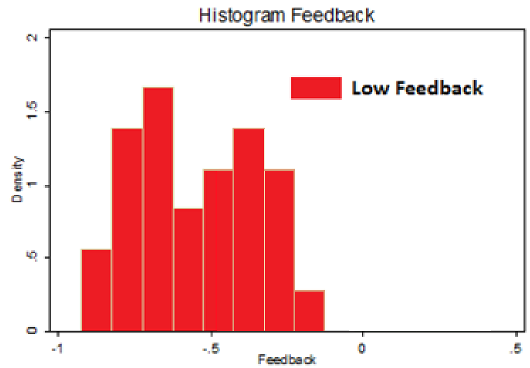
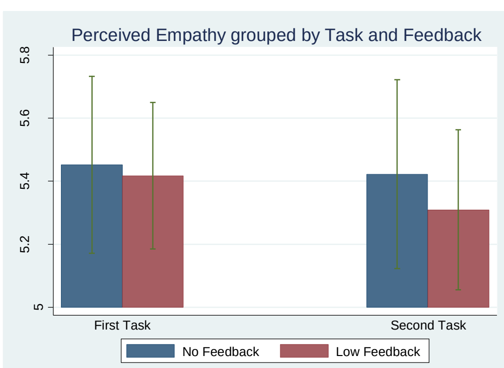
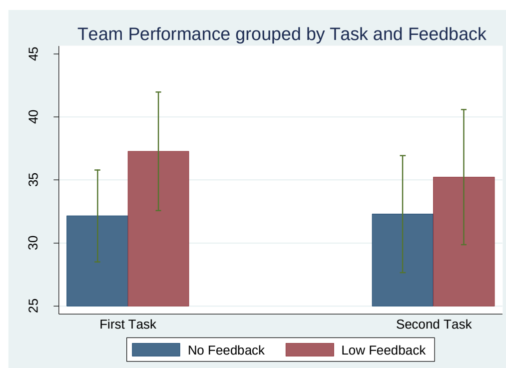


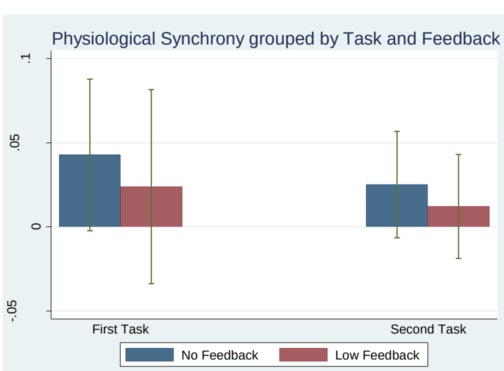
Table 3:

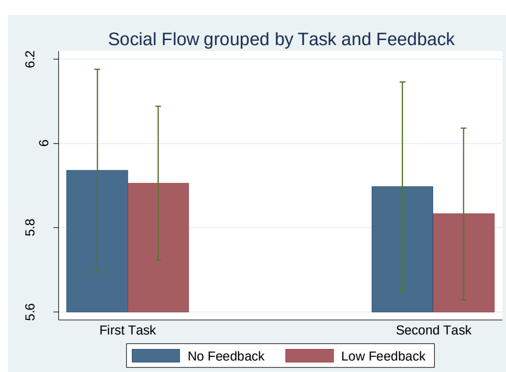
ANOVA results of session and feedback for split at 0 excluding high feedback group

	Main effect session	Main effect feedback	Interaction session-feedback
Perceived Empathy	$F(1, 62)=1.35$ $p=0.25$	$F(1, 62)=0.17$ $p=0.68$	$F(1, 62)=0.44$ $p=0.51$
Team Performance	$F(1, 62)=0.19$ $p=0.66$	$F(1, 62)=2.91$ $p=0.093$	$F(1, 62)=0.26$ $p=0.61$
PS	$F(1, 62)=1.12$ $p=0.33$	$F(1, 62)=2.13$ $p=0.15$	$F(1, 62)=0.46$ $p=0.63$
Social Flow	$F(1, 62)=2.01$ $p=0.16$	$F(1, 62)=0.10$ $p=0.75$	$F(1, 62)=0.19$ $p=0.67$









Split at -0.45 excluding middle data points

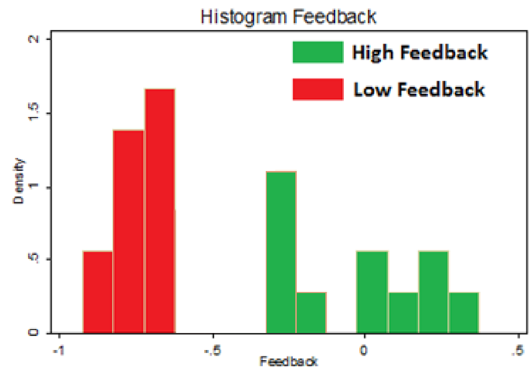
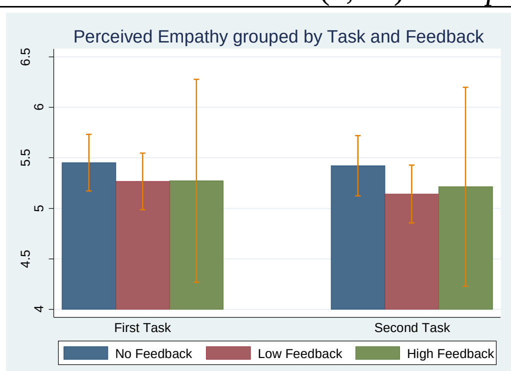
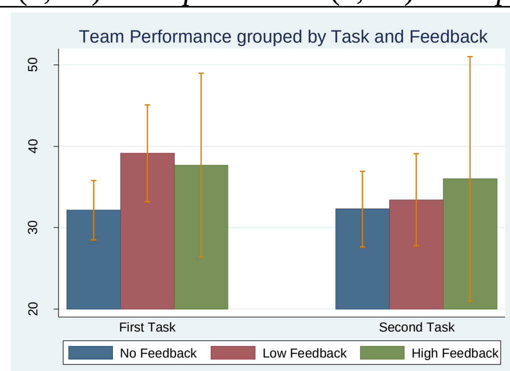


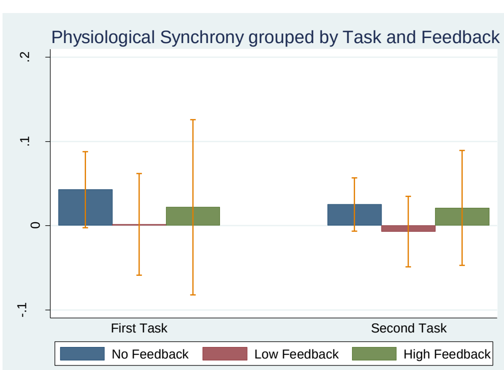
Table 4

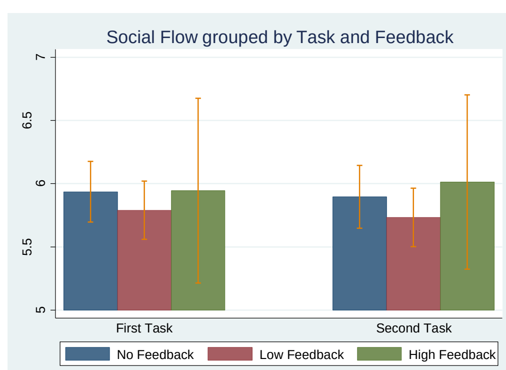
ANOVA results of session and feedback for split at -0.45 excluding middle pairs

	Main effect session	Main effect feedback	Interaction session-feedback
Perceived Empathy	*F*(1, 67)=0.97 *p*=0.33	*F*(2, 67)=0.68 *p*=0.50	*F*(2, 67)=0.30 *p*=0.74
Team Performance	*F*(1, 67)=0.81 *p*=0.37	*F*(2, 67)=1.56 *p*=0.22	*F*(2, 67)=0.84 *p*=0.44
PS	*F*(1, 67)=3.52 *p*=0.033	*F*(2, 67)=2.48 *p*=0.093	*F*(2, 67)=5.37 *p*=0.0006
Social Flow	*F*(1, 67)=0.03 *p*=0.87	*F*(2, 67)=0.51 *p*=0.60	*F*(2, 67)=0.39 *p*=0.68









Appendix F: Checking control variables

Dyad Gender composition: There are three dyad composition in terms of gender: 1) male-male, 2) female-female and 3) female-male.
Difference in age: The difference in age between participants within a dyad in categorized in two groups; a difference smaller than 10 years or larger than 10 years.
Native language: Language consist of two groups, either participants were able to talk in their native language or not.

Table 5

Effect of control variables on dependent variables analyzed with ANOVA's

	Dyad Gender Composition	Age Difference	Native Language
Perceived Empathy	$F(2, 67) = 2.00$, $p=0.14$	$F(1, 66)=0.97$, $p=0.33$	$F(1, 68)=3.07$, $p=0.08$
Team Performance	$F(2, 67) = 1.09$, $p=0.34$	$F(1, 66)=0.15$, $p=0.70$	**$F(1, 68)=10.83$, $p=0.002$***
Physiological Synchrony	$F(2, 65)=0.29$, $p=0.74$	$F(1, 64)=0.23$, $p=0.64$	$F(1, 66)=1.13$, $p=0.29$
Social Flow	$F(2, 67)=0.72$, $p=0.49$	$F(1, 66)=2.62$, $p=0.11$	$F(1, 68)=1.97$, $p=0.17$

*Figure 25 shows the direction of this effect: Team Performance is higher for participants who spoke in their Native language compared to participants who spoke in their non-native language.

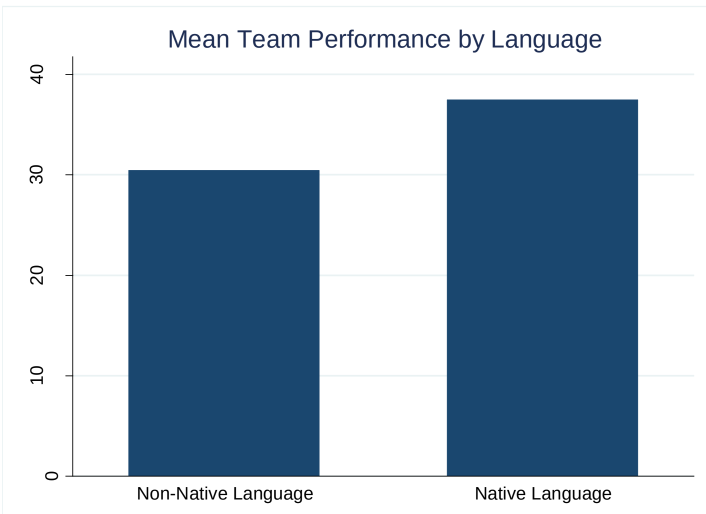


Figure 25: Dyads performed better on the survival tasks when they spoke in their native language