Physiological synchrony in the context of cooperation: Theoretical and methodological considerations

Behrens, F.

Physiological Synchrony in the Context of Cooperation:

Theoretical and Methodological Considerations

Friederike Behrens

Cover by: Evy van Berlo & Friederike Behrens

Layout and typesetting: Michał Sławiński, thesisprint.eu

ISBN: 978-90-9033657-2

Physiological Synchrony in the Context of Cooperation: Theoretical and Methodological Considerations

Proefschrift

ter verkrijging van de graad van Doctor aan de

Universiteit Leiden,

op gezag van Rector Magnificus Prof. mr. C.J.J.M. Stolker,

en volgens besluit van het College voor Promoties

te verdedigen op

woensdag 28 oktober 2020 om 11:15 uur

door

Friederike Behrens

geboren te Bad Oldesloe (Duitsland)

in 1989

Promotor: *Prof. dr. Bernhard Hommel*

Co-promotor: *Dr. Mariska E. Kret*

Doctorate Committee: *Prof. dr. Carsten K.W. de Dreu, Leiden University*

Prof. dr. Natalie Sebanz, Central European University

Prof. dr. Michiel P.M. Westenberg, Leiden University

Acknowledgements

First of all, I would like to thank Mariska for being such an amazing supervisor. From Day 1, you have guided me through this journey, gave me the freedom to make this thesis my own and supported me wherever you could, professionally and personally. I would also like to thank Bernhard for being my promotor.

A special thanks to my paranymphs Eliska and Evy. Eliska, we complemented each other in every aspect throughout the PhD trajectory and made an unbeatable team. Evy, I am still amazed how you can turn every sentence of my self-doubt and self-criticism into something positive. A big thanks also for helping me with the cover!

Iliana, you were there at one of the deepest moments of my PhD (I only say: Driverslab, Elio, no data) and always helped wherever you could! Joyce, you have been an incredible help, practically and mentally. I also want to thank all students for helping me with the data collection and recruiting participants. I am very grateful to all the participants who build the foundation of this thesis.

I have been surrounded by a fantastic lab. Julia, your fresh enthusiasm and passion are contagious and I will miss your inspirational quotes in the morning! Yena, your fabulous cooking skills and talent for singing karaoke are priceless. I want to thank Chris, Evania, Fabiola, Katharina, Maarten, Tom, Tonko, Wouter and all other members of the CoPAN crew for wonderful and fun times together.

Thank you, Maureen, Elio and Evert, for implementing my (sometimes chaotic) thoughts into concrete lab set-ups. To all other colleagues, thank you for the interesting talks and fun times during lunch, drinks and group outings! Steve, thank you for the warm welcome in your lab and for sharing your statistical expertise. Bobby, thank you for all your support and for showing me the American "cuisine" of the best pizzas, burgers and cupcakes in town.

Outside the work environment, I have been surrounded by fantastic friends, who celebrated my achievements and distracted me during stressful times. Dana, thank you for being the best roomie ever! Thank you, Alex, Annika, Arnaud, Charlotte, Cynthia, Erik, Maaike (and of course Sjoerdje), Marcella, Mark, Niek, René, Rianka, Roderick, Ruben, Sanne, and Wouter, for the needed distraction during boardgame-nights, barbeques and parties. Fritzi, Emily & Julia, Friede & Mitra, Anna-Katharina, Julia, Louise, Miriam, Svenja, and Ronja, thank you for staying my friends, even though we cannot see each other that often anymore. Alex, Katja and Nadine, thank you for making my studies such an unforgettable experience.

Tjardo, you followed the same path as I did, recognizing every phase of the emotional rollercoaster of my PhD. Thank you for your endless patience, support and love during the past years. I am grateful for officially calling you, Arja, Bert, Carla, Chris, Erik, Gerda, Jeroen and Roberto, my second family. A special thanks also to Joke and Marijn for being the best parents-in-law!

Moreover, I am very grateful to my close and distant family for supporting me. I would have never obtained my PhD without my parents and sister. Mama, you are my emotional rock and I treasure your honesty, comfort and support you have always given me. Papa, you are not much of a talker when it comes to emotions, but I can feel your endless love in every small and big act(ion) you have done for me. Juli, I can always count on you, no matter what. Thank you for being the best sister! Marcel, I am grateful that you have been in our family for such a long time. I cannot wait to watch Lina grow up; I am already the proudest aunt on earth!

Of course, it is impossible to mention and recall all the people who have contributed to this thesis directly or indirectly. I would like to thank all of them.

Friederike Behrens
September 22nd, 2020
Groningen

Table of Contents

Chapter 1

General introduction

Humans donate thousands of dollars to strangers they will never meet, start societal and political movements to fight climate change, share their possessions with people in need, and build enormous constructions from pyramids to skyscrapers. These are all examples of prosocial behavior, an umbrella term referring to actions that are intended to help others (Batson & Powell, 2003). In the current dissertation, I first take a look at what this broad term means in terms of testing it in the lab and then focus on cooperation and its link with how we nonverbally communicate with each other.

What makes cooperation particularly interesting is that humans are unique in the complexity, scale, and frequency of working together with other individuals. One key ingredient for cooperation to flourish is nonverbal communication which allows us to distinguish cooperators from defectors (Boone & Buck, 2003; Frank, 1988). Research has shown that such communication is a back-and-forth process where people tend to automatically and unconsciously synchronize the nonverbal signals they receive from their interaction partner (Chartrand & Bargh, 1999; Kret, 2015; Levenson & Gottman, 1983; Prochazkova et al., 2018). Preliminary evidence suggests that such synchrony affects cooperative behavior. However, theoretical and methodological questions remain to understand such link, which is the focus of my dissertation.

The aim of the current dissertation is to investigate how nonverbal communication between individuals affects cooperative success and how it can be best investigated in the lab. I shed light on these questions in four chapters. In Chapter 2, I zoom out from cooperation and investigate how different measures of prosocial behavior, some of which we use in the following chapters, relate to one another. In the next chapters, we zoom back in on cooperation and investigate the question of what makes cooperation successful. In Chapter 3, I test the effect of face-to-face contact on cooperation. Face contact allows for nonverbal communication and therefore potentially fosters cooperation. Another factor people rely on when making the decision to work together is past experiences with the interaction partner. In Chapter 3, I investigate how these two sources of information are integrated to make a cooperative decision. In Chapter 4, I follow up on the beneficial effect of face contact on cooperation and investigate the link between synchrony and cooperative success as a potential mechanism of such beneficial effect. In Chapter 5, I zoom in on how to optimize the statistical quantification of synchrony. Specifically, I develop guidelines on how to apply a statistical method to different physiological measures. In Chapter 6, I finalize the dissertation with a general discussion. In the following, I will provide a general overview and introduction of the key questions addressed in this dissertation.

The main focus of my dissertation is on how humans cooperate. In Chapter 2, I start by zooming out and investigate how two of the cooperation games used in Chapter 3 and 4 can be placed into the context of the overarching umbrella term of prosocial behavior by making comparisons with four other games. Prosocial behavior refers to “a broad range of actions intended to benefit one or more people other than oneself— behaviors such as helping, comforting, sharing, and cooperating” (Batson & Powell, 2003, p. 463). Previous studies have shown that people behave similarly across a range of tasks. Such studies have focused on anonymous, one-shot economic games. In Chapter 2, I extend these findings by investigating whether the consistent behavior observed in these economic games generalizes to more naturalistic, interactive games. Specifically, I compare six different paradigms: three variants of a social dilemma game, and three naturalistic games (an egg-hunt game, a puzzle game, and a hidden-profile game where people needed to exchange information in order to reach a common goal). This comparison can shed light on questions such as: How robust are the behavioral consistencies across economic games when looking at more ecologically valid games? Can we use the different games interchangeably or does the choice of a paradigm crucially affect the behavior we measure and the conclusions we draw? Can we generalize findings from one task to another?

Besides these general questions of how these six games compare to each other, I was specifically interested in two of the six tasks, which are both measures of cooperative behavior and on which I zoom in during later chapters. In the comparison study, I aimed to verify that they indeed capture comparable behavioral tendencies. The two games were the classical Prisoner's Dilemma game and an extended version of it where the binary choice to cooperate or defect was extended to a six-option scale. The principles stayed the same which were as follows: Participants can choose between maximizing one's own or the collective outcome by deciding to cooperate or defect. To induce such conflict, two premises hold for the payoff structure: (1) a person receives the highest individual outcome for choosing to defect independent of whether the other player chooses to defect or cooperate; (2) if both participants choose to cooperate, they will receive higher joint outcomes than if they both defect (Dawes, 1980). This most common version of the social dilemma game has been used to measure cooperative behavior for decades, a popularity that can be devoted to its simplicity in tapping into complex motives, emotions, and cognition. I used this game to measure cooperative behavior in Chapter 3. In the Chapter 4, I extended the payoff structure from a binary to a six-option scale in order to measure the *degree* of cooperation rather than the binary decision to cooperate or not. A positive moderate correlation between these two versions of the social dilemma game in the first study supported the claim that they measure similar behavioral tendencies. After taking a methodological perspective on *how* to measure cooperation, the next question to be addressed in Chapter 3 is "*what* makes cooperation successful?".

Given the widespread potential of successful cooperation from building an IKEA wardrobe to international collaborations in research, businesses, and politics, it is crucial to understand which factors contribute to its success. When looking at real-life examples, one aspect that stands out is that people fly around the world to cooperate. In other words, despite the great technological advances in phone calling and video chatting, people still prefer to meet face-to-face. Is it worth the effort? Previous studies suggest that people are indeed more willing and more successful in cooperating when they face each other compared to when they write messages, call via the phone, or interact with a human-like avatar (Balliet, 2010; Bohnet & Frey, 1999; Drolet & Morris, 2000; Frohlich & Oppenheimer, 1998; Kiesler, Sproull, & Waters, 1996). The reason for such boost in successful cooperation has been attributed to the possibility to exchange nonverbal signals that give away the intentions of the other person. I will elaborate on this topic below.

Another factor that crucially determines cooperative behavior is the knowledge we have about the interaction partner, a factor I explore in Chapter 3. Outside the lab, cooperation often occurs between people who have some information about that person's previous behavior either from personal experiences or via gossip or other indirect sources. From an evolutionary perspective, exchanging information about other's behavior has been suggested to be the driving factor that allowed humans to live in large-scaled groups and succeed in forming alliances with a large number of individuals (Dunbar, 2004). If you know that a person has been cooperative in the past, that person is more likely to act selflessly in the future. Likewise, in the case of the social dilemma game, the outcome of one individual depends on the decision of the other, so knowing what the other person chose in previous rounds can help predict future decisions. Not surprisingly then, research has shown that people show higher cooperation rates when they receive direct feedback about each other's choices when playing multiple rounds (Bixenstine & Wilson, 1963; Jorgenson & Papciak, 1981; Monterosso, Ainslie, Pamela Toppi Mullen, & Gault, 2002; Tedeschi, Lesnick, & Gahagan, 1968). In Chapter 3, I investigated how humans integrate the two sources of information: nonverbal signals and explicit information. Do people rely more on nonverbal information when no explicit information is available? To answer this question, two individuals played the Prisoner's Dilemma game in a dyadic interaction setting where they sometimes played face-to-face and sometimes with a visual cover between them preventing nonverbal communication. Additionally, dyads received either no, correct, or random feedback about each other's decisions. This mixed design allowed us to deepen our understanding of the beneficial effect of face contact on cooperation and how it operates in the face of the less uncertain, but sometimes false information from the explicit feedback. From establishing the beneficial effect of face contact on cooperation in Chapter 3, I aimed to investigate its underlying mechanisms in Chapter 4.

What is it about face contact that makes people more cooperative? In a classic study conducted by Axelrod and Hamilton (1981), they showed that cooperation can evolve under three conditions: (1) individuals are likely to meet again, (2) cooperators can be distinguished from defectors, and (3) the fruits of cooperation can be harvested by other cooperators. Focusing on the second condition, how can we distinguish cooperators from defectors? Humans use nonverbal dynamic signals that help us identify the intentions and emotions of others[1]. Over thousands of years humans have developed a unique signaling system that started as simple ritualized acts and that has developed into a multilevel, fine-grained system of nonverbal signals and cues (Boone & Buck, 2003). Such system is the foundation of (nonverbal) communication which "involves a pair of behaviors–a signal and a response–that are functionally interdependent" (Scott-Phillips, Blythe, Gardner, & West, 2012, p. 1943). There is a debate about whether such behavior is restricted to socially shared, intentional signals or whether they also include spontaneous, nonvoluntary and non-intentional expressions (Buck & VanLear, 2002; Ekman, 1997; Gibbs & Van Orden, 2003). In this dissertation, I focus on physiological responses within social interactions which cannot be controlled and therefore not expressed intentionally, favoring the proposition to include both spontaneous and symbolic signals in nonverbal communication.

[1] The signaling system communicates both intentions and emotions. Albeit different concepts, here they are strongly linked as making cooperative decisions *is* emotional. If a person decides to cooperate, she might feel fear that the other person will exploit her. If the other person indeed exploits her selfless act, she might feel anger or disappointment. Such emotional responses can in turn influence the other person's decision. It is therefore difficult to disentangle intentions from emotions which is why we treat them similarly in the present context (Van Kleef, 2010).

One body part that is particularly salient in the signaling system is the face. The substantial amount of fine muscles, the hairless skin, and the high contrast between the sclera and the iris in the eyes offer a unique landscape that allows for an enormous variety of fine-tuned expressions (Kret, 2015). Such variation and nuances in expressions facilitate and enrich the communication of our intentions and emotions, and thereby help us distinguish cooperators from defectors. Tweaking the facial expressions and other cues in computer tasks and observing individuals' expressions while making prosocial decisions have identified a range of signals that are considered communicating selfless intentions such as smiling (Krumhuber et al., 2007; Reed, Zeglen, & Schmidt, 2012), pupil dilation (Kret et al., 2015), blushing (Dijk, Koenig, Ketelaar, & de Jong, 2011), and eye contact (Kleinke, 1986). However, there is not a single expression or a fixed combination of expressions that reliably reflect prosocial intentions as the interpretation of nonverbal signals is highly context-dependent (Barrett, Mesquita, & Gendron, 2011; De Melo, Carnevale, Read, & Gratch, 2014). Although we cannot pinpoint to a specific set of expressions used to communicate our intentions, it is well-established that face contact boosts cooperative behavior by being able to exchange prosocial intentions through nonverbal signals.

The emotional expressions and other nonverbal indices of cooperative intent which we perceive from our interaction partners influence the social decisions that are being made during cooperative endeavors, partly because they impact our own emotions and cognition (Prochazkova & Kret, 2017). As illustrated by Loewenstein and Lerner (2003), our decisions are influenced by the expected outcome of our decision and its associated emotions. Such anticipatory changes in affect influence our immediate inner state which guides our decisions. Damasio, Everitt, and Bishop (1996) referred to such an internal signaling system as "somatic markers" that unconsciously and automatically influence our decisions. The focus from a communicative (explicit) to an internal (implicit) signaling system has great implications for studying social decision-making as it opens a new layer of cues that are evident *within* a person such as changes in arousal levels as measured by skin conductance responses and heartrate changes. The integration of the two sources of information from signals of oneself and the other person is particularly important for the topic I will introduce next.

The majority of research on how we perceive and express our intentions has been focused on intrapersonal processes in computerized paradigms (Kret et al., 2015; Krumhuber et al., 2007; Scharlemann, Eckel, Kacelnik, & Wilson, 2001). Although such an approach provides a controlled setting that allows researchers to disentangle the many factors at play in social decision-making, it neglects the two-directional back-and-forth interplay between two individuals engaging in an actual interaction. Acknowledging such interplay has unraveled a new layer of interpersonal processes where people have been shown to mimic or synchronize each other's explicit and implicit emotional expressions[2]. Such synchronization has been shown to be a multifaceted phenomenon occurring on the behavioral, physiological, and neural level impacting a broad range of interper-

[2] The words mimicry and synchrony are often used interchangeably in the broader context. In the case of physiological responses, we use the term physiological synchrony rather than mimicry as this term has been prominently used in this context (e.g., Prochazkova & Kret, 2017). Researchers have distinguished the two terms based on time lags between responses (Rennung & Göritz, 2016). In our analyses, we took a data-driven approach and included aligned responses with and without time lags.

sonal processes such as cooperative success between strangers, marital satisfaction in couples, mother-child relationships, and therapeutic outcomes (Chartrand & Bargh, 1999; Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Levenson & Gottman, 1983; Prochazkova et al., 2018; Ramseyer & Tschacher, 2011). In their perception-action model, Preston and de Waal (2002) proposed that synchrony forms the basis of emotional contagion which is proposed to be the most basic manifestation of empathy and provides the fundament for higher-order cognitive empathy and prosocial behavior. Hatfield, Cacioppo, and Rapson (1993) described emotional contagion as follows: "by attending to this stream of tiny moment-to-moment reactions, people can and do 'feel themselves into' the emotional landscapes inhabited by their partners." (p.96). Such landscape includes the sensory, motor, physiological, and emotional state of the partner which is in line with the notion that emotional responses constitute behavioral, physiological, and cognitive components that activate each other (Wood, Rychlowska, Korb, & Niedenthal, 2016). This implies that people will only feel the same emotional experience if synchrony emerges on most of these levels. Successfully aligning emotionally with another person then helps to recognize and understand the other person's emotional state and subsequently respond appropriately (e.g., show empathy and/or helping behavior towards a distressed person; Preston & de Waal, 2002). From a developmental perspective, when language is yet to develop in infants and communication with the caregiver is mostly nonverbal, imitation constitutes an innate and automatic learning process to develop emotion regulation abilities, learn about the dangers in the environment, and acquire increasingly more complex social abilities (Preston & de Waal, 2002). As such abilities become more and more automatic, synchrony has been suggested to mostly serve affiliative purposes (Lakin & Chartrand, 2003). Thus, the link between aligning nonverbal signals and social decision-making provides a potential mechanism for explaining the beneficial effects of face contact on cooperation observed in Chapter 3.

In Chapter 4 of this dissertation, I examine this potential link by investigating whether physiological synchrony is positively related to cooperative success in a dyadic interaction study. The setting was similar to the study presented in Chapter 3 with the addition of measuring skin conductance level and heartrate responses throughout the experiment. In the literature, two lines of research have emerged by either manipulating synchrony or the prosocial setting. The former has concentrated on motor and vocalization synchrony asking people to dance, tap, or sing together and investigate how prosocial behavior changes between synchronized and non-synchronized conditions. Two recent meta-analyses showed that being in sync has a medium-sized positive effect on prosocial behavior (Mogan, Fischer, & Bulbulia, 2017; Rennung & Göritz, 2016). In the context of physiological synchrony, manipulating the level of synchrony in, for example, heartrate or skin conductance responses is less straightforward, which is why research has focused on manipulating the cooperative setting and investigating its effect on synchrony. For example, people were asked to play a computer game with or against another player (Chanel, Kivikangas, & Ravaja, 2012; Järvelä, Kivikangas, Kätsyri, & Ravaja, 2014) or to build something together (Mitkidis, McGraw, Roepstorff, & Wallot, 2015; Mønster, Håkonsson, Eskildsen, & Wallot, 2016). Although there were some inconsistencies with regard to which measures exactly played a role, these studies generally support a link between physiological synchrony and cooperation.

Synchronization between individuals and its effect on social processes has been observed at different levels. In our study, we focused on *physiological* synchrony for three reasons: first, this type of synchrony and its effect on prosocial behavior is less understood than other forms of synchrony. Although there is preliminary evidence that physiological synchrony plays a role in cooperative decision-making, the findings are equivocal (Järvelä et al., 2014; Mitkidis et al., 2015; Mønster et al., 2016; Vanutelli, Gatti, Angioletti, & Balconi, 2017). Second, as mentioned above, emotional experiences constitute a multifaceted combination of behavioral, physiological, and cognitive components, making physiological changes a crucial part of the experience. Likewise, "feeling into" the emotional state of another person eases the synchronization of these responses (Prochazkova & Kret, 2017). In other words, to most optimally experience the emotional state of another person, synchronizing on an arousal level is essential. And as described above, such changes have been shown to influence our decision-making (Bechara, Damasio, Tranel, & Damasio, 1997; Crone, Somsen, Van Beek, & Van Der Molen, 2004). Third, while motor movements such as gestures and facial expressions can be consciously controlled, physiological responses are difficult to control. Therefore, physiological responses and their synchronization between interaction partners might provide more genuine information about their relationship. In line with this, Prochazkova and her colleagues demonstrated that attraction between individuals on a blind-date was positively associated with the level of physiological synchrony, but not with the mimicry of explicit signals such as gestures and facial expressions (Prochazkova, Sjak-Shie, Behrens, Lindh, & Kret, 2019). In sum, given the lack of conclusive results regarding the role of specific physiological measures affecting cooperation, their importance in emotional states, and its elusive nature, we focused on physiological synchrony.

There is a range of physiological measures that has been shown to synchronize between two individuals and in the current thesis I concentrated on two measures most often used in previous studies: heartrate and skin conductance level (Palumbo et al., 2017). These measures reflect activity in the autonomic nervous system (ANS). This system is part of the peripheral nervous system (as opposed to the central nervous system) and its function is to maintain homeostasis and adapt our body to changes in the environment. The ANS is an integral component of emotional experiences and has been shown to influence cognitive processes, among others social decision-making (Kreibig, 2010). The ANS is divided into two antagonistic, yet intertwined branches referred to as the sympathetic and the parasympathetic nervous system. The former prepares the body for a fight-or-flight response and activation of this system causes the pupils to dilate, the heart to beat faster, and the hands to sweat. The latter response is measured by changes in skin conductance which is elevated with sweat. Skin conductance responses have been associated with a range of processes such as activation, attention, and significance or affective intensity of a stimulus (Dawson, Schell, & Filion, 2000). The parasympathetic nervous system is also referred to as the "rest and digest" system pinpointing to its role in relaxation and recovery from the elevated activity of the sympathetic branch. Biologically, activation of the parasympathetic nervous system constricts the pupils, decreases heartrate and activates, among others, the digestion processes. A measure of parasympathetic nervous system activity is the Respiratory Sinus Arrhythmia (RSA) which reflects the high-frequency component of the general heartrate measure and reflects the respiratory cycle. Chapter 4 focuses on the global heartrate which is controlled by both sympathetic and parasympathetic nervous system activity and is therefore less specific in identifying distinct processes in the body than the skin conductance measure. Nonetheless, heartrate has been shown to influence psychological processes such as decision-making and emotional processing (Crone et al., 2004; Kreibig, 2010). With regard to the link between synchrony and cooperation, both measures have been shown to play a role, however, the findings were equivocal between studies. Most showed effects in one of the measures (Mitkidis et al., 2015; Mønster et al., 2016; Vanutelli et al., 2017), whereas others did not observe any effects (Järvelä et al., 2014). Therefore, by combining skin conductance level with heartrate measures, we could shed more light on the inconsistency in the literature and the role of the two branches of the ANS in social decision-making.

While conducting the study presented in the Chapter 4, I encountered a methodological challenge: how can we accurately quantify physiological synchrony? Despite the popularity across different disciplines to understand the phenomenon of interpersonal synchrony, no standardized guidelines have been developed on how to properly capture the dynamics between two individuals statistically. The method that I rendered most appropriate for our research question and type of data was the Windowed Cross-Correlation analysis (WCC; Boker, Xu, Rotondo, & King, 2002). The method incorporates two features that allows for dynamic changes over time: first, by segmenting the time series into smaller, overlapping windows and calculating the cross-correlation for each window, the strength of synchrony (i.e., the correlation estimate) can change throughout an interaction. This is important as two individuals most certainly do not establish the same degree of synchrony throughout an interaction, but rather show moments of weak and strong synchrony. Second, there is great intra- and interpersonal variation in the pace of (physiological) responses introducing varying time delays between the responses of two individuals. The method accounts for such variations by shifting the windows of the interacting people away from each other with increasing delays. Thus, this method offers a neat way to compute a quantification of the overall strength of synchrony throughout a conversation.

The challenge lies in tailoring the WCC analysis to the characteristics of the signal of interest by specifying parameters. There is great variation between studies on which parameters are used and researchers have proposed that the choice of parameters is arbitrary and does not change the relative results (McAssey, Helm, Hsieh, Sbarra, & Ferrer, 2013). For the study presented in Chapter 4, I applied a similar approach and based the choice of the parameters mostly on previous literature and the biological nature of the signals of interest. However, during a research visit to the lab of Steven Boker, who developed the statistical method, I had the opportunity to shed light on this issue in a data-driven manner, a project that is presented in Chapter 5. Particularly, I systematically investigated the influence of the parameters and developed guidelines for the best parameter configurations for four different physiological measures: heartrate, skin conductance level, pupil size, and facial expressions (the left zygomaticus major, a muscle associated with smiling). Such guidelines can guide other researchers to make informed choices about which parameters to use, thereby increasing the comparability between studies and contributing to solving the inconsistencies in findings between studies.

Chapter Overview

This dissertation is based on four empirical stand-alone research articles that are presented in Chapters 2 to 5. They build upon each other by holding a magnifying glass over one aspect of the previous study. However, as they are written as independent research articles, they contain some theoretical overlap. In the following, I will give a short overview of the studies presented in each chapter.

In **Chapter 2**, I investigated how six different prosocial behavior tasks relate to one another. To that end, 74 participants played all tasks in a within-subject design with three different partners. The games have been used previously to measure prosocial behavior and include three variants of the social dilemma game, an egg-hunt game, a puzzle, and a communication task. By means of a Principal Component Analysis, I examined whether these games measure similar behavioral tendencies (i.e., prosociality). Two of the examined games were used in Chapter 3 and 4 to measure cooperative behavior.

In **Chapter 3**, I investigated the effect of face contact and feedback on cooperative behavior. People played multiple rounds of the Prisoner's Dilemma game in a dyadic interaction setting ($N = 116$). In a mixed design I manipulated whether people could see each other or not (within-subject manipulation) and whether they received correct, random, or no feedback (between-subject manipulation). I was particularly interested in the interaction between the two manipulations, investigating whether the effect of face-to-face contact on cooperative behavior was moderated by how much information people received about each other's past behavior.

In **Chapter 4**, the aim was to dive deeper into the beneficial effect of face contact on cooperation. Specifically, I investigated whether physiological synchrony functions as a potential mechanism of such beneficial effect. To investigate this, I used a similar set-up to the previous study ($N = 152$). Additionally, participants' physiological responses by means of their skin conductance level and heartrate responses were measured throughout the experiment. I hypothesized that physiological synchrony would be higher when people could face each other compared to when they could not. Most importantly, I expected physiological synchrony to predict cooperative success in a dyad, particularly when participants interacted face-to-face.

In **Chapter 5**, I focused on the methodological challenge of properly quantifying synchrony. In particular, I advanced an existing analysis, the Windowed Cross-Correlation analysis, that has been used to measure synchrony by refining it to four physiological responses. The data I used for this methodological study was from a dyadic interaction study where people engaged in storytelling while their heartrate, skin conductance level, pupil size, and facial expressions (i.e., smiling) were measured ($N = 68$). I elaborated the analysis by investigating its sensitivity to discriminate the original dyads from dyads who participated in the same experiment, but never interacted. Such distinction is particularly important to draw conclusions about that synchrony is the result of interpersonal processes rather than artifacts deriving from the recurrent nature of the signals or the same structure of the experiment across dyads. Based on these outcomes, I could provide recommendations on how to tailor the analysis to each physiological measure.

In **Chapter 6**, I close the dissertation with a general discussion, where I highlight and integrate the key findings from the different chapters. I also pinpoint to new questions that this dissertation has provoked and propose ideas for future studies. I finalize the chapter and the dissertation with concluding remarks.

Finally, I want to emphasize that the empirical studies are the result of the collaboration with my co-authors as acknowledged by including them in the author list of each study and by writing these chapters using plural personal pronouns. However, Chapter 1 and 6 are based on my own thoughts which is why I use singular personal pronouns in these chapters.

Chapter 2

Under the umbrella of prosocial behavior: A critical comparison of paradigms

Abstract

Despite the discontent, cruelty, and warfare that fill the daily news, people show tremendous capacities to help and cooperate with others. Prosocial behavior is used as an umbrella term capturing the diversity of selfless acts. As such, researchers have developed a variety of tasks and it is crucial to verify that they measure the same underlying construct of prosocial behavior. Previous studies have focused on comparing anonymous, one-shot economic games providing evidence for behavioral consistency across games. The current study extends these findings by (i) comparing both repeated economic and naturalistic interactive games in a within-subject design, and (ii) letting participants play in face-to-face dyadic settings. In total, 74 participants completed six tasks: three variants of a social dilemma game, an Egg Hunt game measuring helping behavior, a group decision-making paradigm requiring communication skills, and a Tangram game where participants solved puzzles together. A Principal Component Analysis revealed that two components best describe the behavior in these tasks. The three social dilemma games loaded on the first component, termed "social dilemma games". These games were distinct from the interactive games and the helping and decision-making tasks loaded on the second component, termed "naturalistic games". The Tangram game was unrelated to all other games. These findings suggest that the behavioral consistency observed in economic games has its limits to generalize to other types of tasks and emphasizes the importance of choosing the appropriate (combination of) paradigms to measure prosocial behavior. Theoretical and methodological differences between tasks are discussed to explain these findings.

Based on: Behrens, F., & Kret, M.E. (2020). Under the umbrella of prosocial behavior – a critical comparison of paradigms. *PsyArXiv*.

Introduction

Prosocial behavior is one of the pillars of human society. Thousands of dollars are donated every day to strangers the donor will never see, leaders from 195 countries agreed on the Paris Agreement to fight climate change, and researchers from around the world form collaborations to advance our understanding of the human mind. Prosocial behavior is used as an umbrella term incorporating a "broad range of actions intended to benefit one or more people other than oneself— behaviors such as helping, comforting, sharing, and cooperating" (Batson & Powell, 2003, p. 463). In order to understand how prosocial behavior arises, it is of pivotal importance to elucidate its underlying mechanisms, detect individual differences, and highlight situations where it is common as well as those where it is a rarity. To that extent, previous researchers have developed multiple paradigms to investigate prosocial behavior. One question that arises given the variety of tasks and the broad definition of the term "prosocial behavior" is whether behavioral tendencies are consistent across these paradigms. The few studies that have addressed this question generally observed similar behavior across tasks, reflecting a person's overarching, temporally stable prosocial preferences (Blanco, Engelmann, & Normann, 2010; Böckler, Tusche, & Singer, 2016; Epstein, Peysakhovich, & Rand, 2016; Peysakhovich, Nowak, & Rand, 2014; Volk, Thöni, & Ruigrok, 2012; Yamagishi et al., 2013). However, these studies have mostly concentrated on anonymous, one-shot economic games, as discussed below. Although such games provide valuable insights in controlled laboratory settings, whether similar patterns emerge in more naturalistic settings remains less understood. Taking a step in this direction, the current study aims to extend the previous findings by considering a more heterogeneous set of paradigms including both economic games and three more naturalistic tasks, and by playing iterated games in face-to-face dyadic interaction settings.

In the fields of economics, psychology, political sciences and biology it is common to measure prosocial behavior by means of social dilemma games. A strength is that the social dilemmas are, for the neurotypical population, easy to comprehend, yet tap into complex motives, cognition and emotions. These games are designed to reflect real-life decision-making problems with conflicting self- or collective interests. Specifically, such games have been developed to measure different subtypes of prosocial behavior, including but not limited to cooperation (e.g., Prisoner's Dilemma game and Public Good game), trust (e.g., Trust game), generosity (e.g., Dictator game). Although they are designed to measure different aspects of prosocial behavior, behavior is believed to be driven by a person's general prosocial preference or "phenotype" (Balliet, Parks, & Joireman, 2009; Peysakhovich et al., 2014; Poncela-Casasnovas et al., 2016). In line with this, previous studies have shown that participants behave quite consistently across a variety of tasks. For example, Yamagishi et al. (2013) compared five economic games (two variants of the Prisoner's Dilemma game, Trust game, Dictator game, and a Faith game) and showed that behavior across these games were correlated (except for the Dictator game). Similarly, Peysakhovich and colleagues (2014) observed that participants who cooperated in the Public Good game were also more likely to allocate more resources to their partner during the Trust game and the Dictator game. Importantly, although these cooperation games showed moderate correlations, suggesting they tap into a similar construct, they were distinct from norm-based games such as the Ultimatum game and the Punishment game. In other words, participants who cooperated in one game, were more likely to cooperate in another, but were not necessarily inclined to punish others for non-cooperative decisions. This distinction was further supported by Böckler et al. (2016), who compared 14 computerized tasks including economic games, hypothetical distribution tasks, and self-reported measures. Applying a Principal Component Analysis, the authors showed that games measuring altruistic-motivated prosocial behavior, such as donating and helping, clustered together; however, were distinct from norm-based punishment games and self-reported prosocial behavior. In sum, these studies support the notion that social dilemma games measuring different aspects of prosocial behavior generally tap into similar overarching behavioral tendencies.

The question remains whether such clustered behavior is also evident when using more heterogeneous paradigms. Böckler and her colleagues (2016) have made an attempt in this direction by including a broader range of paradigms other than economic games. However, these games were still based on straightforward pay-off structures where people indicate their decisions by pressing keys in an isolated room. Given that prosocial behavior is social by nature, it is crucial to investigate whether the above described findings also apply to situations where people interact with each other in a face-to-face setting and play games with an uncertain pay-off structure.

To that end, in the current study we compared behavior in six prosocial behavior games: three variants of a social dilemma game and three more naturalistic games, where people were asked to build puzzles together (Vink, Hasselman, Cillessen, Wijnants, & Bosman, 2018), communicate and exchange information (Nevicka, Ten Velden, de Hoogh, & van Vianen, 2011), and help each other in collecting eggs (McClung, Placì, Bangerter, Clément, & Bshary, 2017). With regard to the social dilemma games, we tested three variants of the Prisoner's Dilemma game: (1) the classical variant with two response options (cooperate or defect); (2) an extended version of the Prisoner's Dilemma game, where the pay-off structure was extended from a 2×2 to a 6×6 matrix; and (3) an adjusted version developed to test children and chimpanzees, where participants can decide to pull a rope (i.e., cooperate) or not (i.e., defect) (Sánchez-Amaro, Duguid, Call, & Tomasello, 2019). Including these three variants allowed us to investigate whether participants based their decision on the same principles across games despite changes in the scale (six versus two response options) and the way of indicating a decision (key press versus rope pull).

The three additional games are less commonly used paradigms to measure prosocial behavior, yet tap into related processes. First, the Tangram game requires participants to build puzzles together (Vink et al., 2018). The more efficiently participants work together, the more puzzles they can complete. In contrast to other games described in this study, this task required people to physically work together on one problem such that participants could complement their actions to complete the puzzles. The second paradigm was the Hidden Profile game which was originally designed to study group decision-making processes rather than prosocial behavior (Nevicka et al., 2011; Stasser & Titus, 1985). In this game, participants are asked to find the most suitable job candidate, by sharing information. Crucially, performance during this game depends on individuals' skills to cooperate and share partly *unique* information (Wittenbaum, Hollingshead, & Botero, 2004). If participants only consider shared information, they are led to an incorrect decision. However, when all unshared information is exchanged, the correct "hidden" decision is revealed. Thus, this task measures how well people communicate with each other and reach the goal of finding the best job candidate. The third game was the Egg Hunt game measuring helping behavior (McClung et al., 2017). Specifically, the behavior of interest was how much effort participants put into helping others collect their eggs at the cost of the limited time to collect their own. This task still loosely incorporates the idea of the social dilemma because a person can choose between maximizing his or her own rewards at the expense of the joint outcome. However, this structure was neither directly visible nor highlighted in the instructions and helping behavior could develop in a natural environment.

By incorporating a variety of tasks ranging from different versions of the classical controlled social dilemma paradigms, to interaction tasks reflecting more natural and less abstract situations, we aimed to answer the questions: How robust is the consistent behavioral tendency observed across different games in previous studies? Are some games more related to each other than other games? We expected similar behavior among the three variants of the social dilemma game, because they are based on the same principles and similar pay-off structures. With regard to the three naturalistic games, the expectations were less straightforward as to how they would correlate with each other and with the social dilemma games. The social dilemma games incorporate clear response options and participants could choose to act prosocially or not. In contrast, in the three naturalistic games the options were less clear-cut and the *willingness* to act prosocially was constraint by the *ability* to do so. For example, the classical Prisoner's Dilemma game requires a certain amount of cognitive abilities to understand the abstract payoff structure. The Tangram game depends on spatial skills and the Hidden Profile game addresses communication abilities. Differences in these abilities could undermine the correlation in behavior between the games. On the other hand, as described above, prosocial preferences have been demonstrated to show across a range of tasks and subtypes of prosocial behavior have been shown to relate to similar underlying motivational tendencies. For example, prosocial people as classified by their social value orientation have been shown to cooperate more in the social dilemma game and show more helping behavior than proself individuals (C. Boone, Declerck, & Kiyonari, 2010; Van Lange, Bekkers, Schuyt, & Van Vugt, 2007). Similarly, social value orientation has also been related to motivational processes in the Hidden Profile game (De Dreu, Nijstad, & Van Knippenberg, 2008). Furthermore, behavior in economic games have been shown to translate to behavior outside the lab, suggesting that behavior should also translate to the more naturalistic tasks used in the current study (Benz & Meier, 2008; Böckler et al., 2016; Franzen & Pointner, 2013). These arguments suggest that participants would show similar behavioral tendencies across all six games.

Method

Participants

In total, 74 individuals participated in this study, completing six separate experimental tasks (M_{age} = 22.05, SD_{age} = 2.55, $Range_{age}$ = 18–31 years). The subject-to-items ratio (13:1) was good which has been shown to be important when performing a Principal Component Analysis (Osborne & Costello, 2004). Participants were recruited via an online recruitment system, flyers in the University building, or personal contacts. More demographical information about the participants are listed in Appendix A1 (see Table A.S1). Participants received written and oral instructions in Dutch or English. All participants provided written informed consent after receiving information about the study and prior to the start of the experiment. Due to participants arriving late or leaving early during the experiment and due to technical issues, the number of participants per game differed between 53 and 73. The study was approved by the Psychology Research Ethics Committee of Leiden University (CEP19–0318/223).

Design

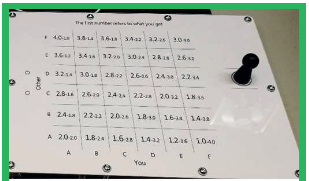
The study was a within-subject design with all participants playing each of the six games (with the exception of those arriving late/leaving early). All games were played in dyads. To minimize the influence of the partner on a person's prosocial behavior, the dyadic composition changed after two games. The two games that a dyad played together as well as the order of the games was counterbalanced between groups. The only restriction was that the Prisoner's Dilemma game was played as one of the first two games and the extended Prisoner's Dilemma game as one of the last two games. This restriction was introduced for two reasons: (1) the extended Prisoner's Dilemma game builds on the original Prisoner's Dilemma game, making it easier to instruct participants on the more complex version. (2) Compared to the other tasks, the instructions were more complex, making these two games cognitively more demanding. Therefore, having a break in between would increase participants' level of concentration.

Material

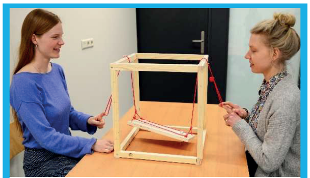
All participants played six games with three different partners (Figure 1). They were not allowed to talk about the (strategy of the) games, but could chat about task-unrelated topics (monitored by the experimenter who was either involved in administering the task or stood close to the participants). This restriction was not applied during the Hidden Profile game, as the task was to have a discussion.

Prisoner's Dilemma game. In this two-choices game, participants can choose to cooperate or defect with their partner and the amount of points a person receives depends on the own and the partner's choice. When both players cooperate, the outcome is higher compared to when both defect and gives the highest joint outcome. However, independent of what the partner chooses, the higher individual outcome is always achieved by choosing to defect (Table 1). A player receives the lowest amount of points when s/he cooperates and the other player defects. Thus, a dilemma is created between maximizing the individual or the joint outcome. In the current study, participants played three practice and ten experimental trials. Practice trials were included to ensure that participants understood and were familiarized with the game. These trials were not included in the analysis. Auditory pre-recorded instructions were provided via headphones (for a similar approach, see Behrens & Kret, 2019). The sequence of a trial was as follows: Participants were asked to look at each other (participants could only see each other's face). After four seconds, participants were asked to look down and indicate their decision (i.e., cooperate or defect) on a keyboard. The response window was three seconds. Afterwards, they heard that they both had made a choice and the next trial began. Participants did not receive any feedback about their decisions during the game. The instructions were phrased to choose between option A and option B, avoiding any references to cooperation and defection. The face-to-face interval was introduced to allow nonverbal communication which helps participants to read the intentions of their partner (Behrens & Kret, 2019). The measure of interest was the proportion of decisions to cooperate (i.e., number of decisions to cooperate divided by the total number of trials). Two participants were excluded because they had more than four out of ten missing values.

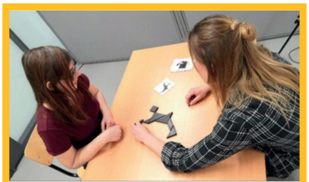
Prisoner's Dilemma
- Choose to cooperate or defect
- Before each decision, participants look at each other
- Dichotomous scale (1 = cooperate)
- 10 trials
- Duration: 5 min
- MoI: % cooperative decisions



Extended Prisoner's Dilemma
- Extends Prisoner's Dilemma to a 6 x 6 payoff structure
- 6-point scale (6 = fully cooperate)
- 10 trials
- Duration: 5 min
- MoI: mean of points (between 1 – 6) / 6

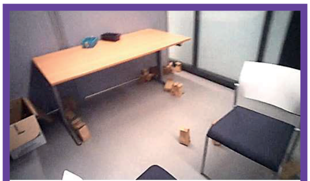


Rope Pull
- Based on the principles of the Prisoner's Dilemma
- Pull (=cooperate) or not (=defect)
- Dichotomous scale (1 = cooperate)
- 10 trials
- Duration: 3 min
- MoI: % cooperative decisions



Tangram Puzzle
- Complete as many puzzles as possible
- Two conditions: complete puzzles cooperatively and alone
- Duration: 2 x 5 min
- MoI: completed puzzles (coop) / completed puzzles (total)

Hidden Profile
- Find best job candidate with different characteristics
- Un-/Shared information in profiles
- Combined information reveals best candidate
- Duration: 7 min (5 min discussion)
- MoI: % unique information exchanged



Egg Hunt
- Participants are rewarded for different colors of eggs
- Collecting eggs is time-costly
- Do participants help each other?
- Duration: 7 min (5 min egg hunting)
- MoI: % helping behavior when finding egg from other participant

Figure 1. Information about the set-up of the six prosocial behavior games played by each participant; MoI = Measure of Interest included in the analysis.

Table 1
The payoff matrix of the Prisoner's Dilemma game

	Other	
You	C	D
C	3–3	1–4
D	4–1	2–2

Note. The first number refers to the points earned by "You".

Extended Prisoner's Dilemma game. This game extends the original Prisoner's Dilemma game from a 2 × 2 to a 6 × 6 payoff structure and was adopted from Behrens et al. (2019). Thus, the response options changed from a dichotomous choice (i.e., cooperate or defect) to a six-point scale ranging from option A (= fully defect) to option F (= fully cooperate). Apart from this adjustment in the payoff structure, the procedure was the same as for the classical Prisoner's Dilemma game described above. The measure of interest was the mean willingness to cooperate on the 6-point scale (1 = fully defect [option A] to 6 = fully cooperate [option F]). To make the measure comparable to the other games, we transformed the mean value to a proportion value.

Rope Pull game. This game was adopted from Sánchez-Amaro and colleagues (2019) who used this game to measure cooperation in chimpanzees and children. The game is based on the principles of the Prisoner's Dilemma game, but instead of pressing keys, participants could pull a rope or not. When both participants pulled the rope (i.e., mutual cooperation), a tray with two rewards was lifted and both participants received one reward each. If only one participant pulled (i.e., unilateral cooperation), the tray was lifted on that participant's side and the two rewards rolled to the side of the other participant. If nobody pulled for 15 seconds (i.e., mutual defection), no one received a reward. Hence, the principle is similar to the Prisoner's Dilemma game in that mutual cooperation is more beneficial for both participants than mutual defection and that with unilateral cooperation the defector earns more points than through mutual cooperation. However, the game differs such that mutual defection and unilateral cooperation result in the same outcome (no reward). Compared to the payoff structure of the current Prisoner's Dilemma game another difference was that mutual cooperation led to the same reward than a turn-taking strategy (unilateral cooperation by one participant and then by the other; Sánchez-Amaro et al., 2019). Participants first read instructions about the game and subsequently the experimenter showed the four possible outcomes to the participants with the apparatus. Afterwards, participants played ten trials and the experimenter recorded the outcome of each trial on a sheet of paper. As was the case for all other tasks as well, the game was phrased in a neutral way without referring to cooperation and defection. The measure of interest was the proportion of times a participant cooperated out of the total number of trials. The willingness to cooperate was based on the reward distribution such that a person cooperated if both participants received one reward each, or when the other participant received both rewards.

Tangram game. In this game, participants are asked to make predefined figures with a number of puzzle pieces that have different shapes (e.g. triangle, square; Vink et al., 2018). In the current study, the same seven pieces were used to make different figures. Participants were asked to complete as many figures as possible within five minutes. They were only allowed to continue with the next figure after completing the previous one which was checked by the experimenter. Participants played two rounds, once individually and once together. The individual condition served as a baseline measure of their skill in performing the task. The measure of interest was the difference in completed puzzles between the cooperative and individual condition (completed puzzles in the cooperative condition minus completed puzzles in the individual condition). Additionally, we explored different measures of prosocial behavior on the dyadic level which are presented in Appendix A3. We made three sets of figures with eighteen figures each that were individually printed in black with only the outline of the figure visible. Which set was given in the individual and cooperative condition was counterbalanced. The order of the figures within each set was kept constant. The difficulty of the figures was based on the performance of the experimenters prior to data collection. For some figures, the outline of one or two puzzle pieces were shown in a different color to make it easier.

Hidden Profile game. This game was adopted from Nevicka and colleagues (2011) and was originally designed to measure group decision-making processes. Here, participants need to find the most suitable candidate for a job. Each participant receives a profile with information about the three potential candidates. What participants do not know is that some information is shared between them, whereas other information is unique for each participant. Based on the (incomplete) information of each profile, a suboptimal candidate stands out; however, after combining all information from the different profiles, another candidate is more suited (has the most positive and least negative characteristics). Thus, the true best candidate is "hidden" in the unshared information distributed among the profiles. In the current study, we used the profiles from Nevicka and colleagues (2011) where participants needed to find the best candidate for a secret agent position with validated positive, neutral and negative characteristics. To adjust the original three-player game to the dyadic setting of the current study, we excluded one characteristic per candidate. Both profiles (one for each participant) consisted of nine characteristics for each candidate, of which six were shared among the profiles. Eight additional characteristics per candidate were evenly distributed between the two profiles, that is, each profile included four unique characteristics per candidate. In the shared information, candidate A had three neutral and three negative characteristics, candidate B had six positive characteristics and candidate C had three positive and three negative characteristics. Based on this information, candidate B would be the preferred candidate. In the unshared information, candidate A had eight positive characteristics, candidate B had two neutral and six negative characteristics, and candidate C had two positive and six neutral characteristics. Therefore, after combining the information candidate A was most suited. Examples of the characteristics were: "can read code language" (positive), "is 180 cm tall" (neutral), and "is afraid of heights" (negative). During the game, participants first read the profiles and made an individual decision about which candidate they would choose.

Subsequently, participants had a five-minutes discussion and then made a joint decision about the best candidate. The performance of a dyad was operationalized by the number of unshared characteristics that were exchanged during the discussion. To measure this, participants filled out a checklist with the characteristics of the candidates and indicated whether each of the items described the candidates. A characteristic was considered "exchanged" if the participant who did not have that characteristic in his/her profile indicated that it belonged to the corresponding candidate. The measure of interest was the proportion of exchanged characteristics in relation to the total number of unshared information of a participant's profile.

Egg Hunt game. The Egg Hunt game was adopted from McClung and colleagues (2017) who used this game to measure helping behavior. The idea is that participants are assigned different colors of eggs and earn rewards for each egg of their color they collect during the egg hunt. However, there is not enough time to collect all own eggs and collecting an egg is time-costly. Thus, the behavior of interest is what participants decide to do when they find an egg of the partner's color: Does the participant invest the time to help the partner by collecting the eggs or not? In the current study, 90 eggs were wrapped in paper sandwich bags: 18 orange, 18 pink, and 54 green eggs. One participant was rewarded for all collected pink eggs, the other participant for all orange eggs. The green eggs were not rewarded. The wrapped eggs were placed in a room and participants could simultaneously hunt for the eggs. For each egg, participants first had to unwrap the paper bag, take out the egg to see the color and then (i) put the egg back in the bag, (ii) put the egg next to the bag, or (iii) take the egg, run around a chair twice (there were two chairs in the room) and then put the egg into one of two baskets (one basket for the orange eggs and another basket for the pink eggs). Participants were rewarded for the eggs in their basket with chocolate (see below). The game lasted five minutes. During the game, both participants wore eye-tracking glasses (Tobii Pro Glasses 2) recording their behavior. The behavior was coded afterwards following McClung and colleagues' (2017) scoring scheme. Specifically, when an egg of the other participant was found, the behavior was classified into three categories: (1) costly helping: the egg is collected or passed on to the other player; (2) no costly helping: the egg is left visible to the other participant; (3) neglect: the egg is put back in the paper bag. We also added a fourth category representing the "competitive" behavior in other games: (4) active hiding: the paper bag with the egg is actively hidden. The latter behavior was, however, not evident in any video recordings. The measure of interest was the proportion of helping behavior (costly and no costly helping) from the times that an egg with the color assigned to the other participant was found. Unfortunately, some behavior could not be classified because either the participant did not look down to see the egg or the color could not be clearly identified from the video. Participants with ten or more unidentified behaviors were excluded from the analysis. Data of 53 participants could be included in the analysis.

Questionnaire. At the end of the experiment, participants filled out a questionnaire where they were asked to indicate their gender, age, nationality, highest completed education level, and the number of siblings (for descriptive statistics, see Table A.S1). Furthermore, participants indicated how much they knew and liked the three different partners on a scale from 0 – 100. Descriptive statistics of these scales are presented in Appendix A2 (see Table A.S2 and Figure

A.S1). We also present the Spearman's rho correlation between the liking rating and prosocial behavior in each game (see Table A.S3). Finally, participants were asked about what they thought the purpose of the study was.

Procedure

A group of four same-sex participants was invited into the lab. Upon arrival, participants were separated, read the information about the study and gave informed consent. They were also given a colored wristband which was used to form the dyads. Which participant received which color was based on the time of arrival (the first individual had the green wristband, the second the orange, the third the purple and the fourth the blue). When there were no more questions and all participants were ready, the first two games started. All games were played in a big lab with a separation wall in the middle so that the two dyads could play games simultaneously without seeing each other. The first two dyads were formed with participants wearing the green and orange (Dyad 1) and the purple and blue (Dyad 2) wristbands. One of the dyads started with the Prisoner's Dilemma game, the other dyad started with another game. Which dyad began with the Prisoner's Dilemma game was counterbalanced, and the second game was changed between groups. After finishing the games, the dyads switched places and played the game the other dyad just played. Then, the formation of the dyads changed (Dyad 3: green + purple; Dyad 4: orange + blue) and the next round of two games were played. In the last round, the last two dyads were formed (Dyad 5: green + blue; Dyad 6: orange + purple) and the extended Prisoner's Dilemma game and the last remaining game were played. Thus, all participants played six games with three different partners. Because the order of the games was changed between the groups and therefore the combination of games that were played within dyads, order effects were negligible and the dependency between observations on a dyadic level was minimalized (e.g., five out of 19 groups played the Prisoner's Dilemma game paired with the Egg Hunt game in one round). After that, participants filled out the questionnaire (see Materials). Finally, participants received as many M&Ms as they earned points during the Egg Hunt, Rope Pull, and Tangram games (Egg Hunt: the number of eggs collected of the participant's color; Rope Pull: number of rewards received; Tangram: number of completed puzzles in both conditions). Depending on their preference, participants were additionally paid with course credits or money and thanked for participating.

Statistical Analysis

For each game, one measure of interest was calculated per participant as described in the Material section. In the first step, descriptive statistics of each game are presented. Second, correlation coefficients and 95%-confidence intervals for each combination of games are reported. Given that most measures were not normally distributed, we report Spearman's rho correlations. In the third step, we conducted a Principal Component Analysis (PCA) to investigate which tasks could be described as measuring similar behavioral tendencies. The following settings were used: To determine the number of components, the parallel analysis was applied where a component was selected when the eigenvalue of that component was larger than the parallel average random eigenvalue. For the rotation method, the oblique rotation method *promax* was chosen, because this method allows the components to be correlated. In an additional step, we further explored differences between the social dilemma games by looking at the absolute cooperation rates both on the individual level (i.e., a person's willingness to cooperate) and on the dyadic level (i.e., mutual cooperation). These results are reported in Appendix A4. A significance level of $\alpha = 0.05$ was used and analyses were performed in JASP 0.10 (JASP-Team, 2019) and SPSS (Version 25).

Results

Descriptive statistics

Descriptive statistics of each prosocial behavior measure are presented in Table 2 and displayed in Figure 2. Except for the Tangram game, the prosocial behavior is operationalized as proportions. For the Tangram game, the difference in completed figures between the cooperative and individual conditions was calculated. A positive value therefore corresponds to more completed figures in the cooperative condition.

The highest cooperation rate was observed in the Rope Pull game with almost .85. The minimum was 0.4 meaning that participants cooperated at least four out of ten times. This was not the case in the two Prisoner's Dilemma games where some people also defected at all times. The cooperation rates of the two Prisoner's Dilemma games and the Egg Hunt game were in a similar range (around .60). The lowest rate was observed in the Hidden Profile game with only .33.

Table 2

Descriptive statistics of the proportion of prosocial behavior for the six games

Game	Mean	Median	SD	MAD	Maximum	Minimum	Range	N	Missing
Prisoner's Dilemma	.61	.60	.33	.30	1.00	.00	1.00	70	4
Extended Prisoner's Dilemma	.69	.68	.26	.23	1.00	.17	.83	70	4
Rope Pull	.84	.90	.19	.10	1.00	.40	.60	72	2
Tangram	0.10	0.00	1.80	1.00	5.00	-5.00	10.00	73	1
Hidden Profile	.33	.33	.29	.25	1.00	.00	1.00	70	4
Egg Hunt	.60	.86	.45	.14	1.00	.00	1.00	53	21

Note. SD = standard deviation; *MAD* = median absolute deviation; *N* = sample size.

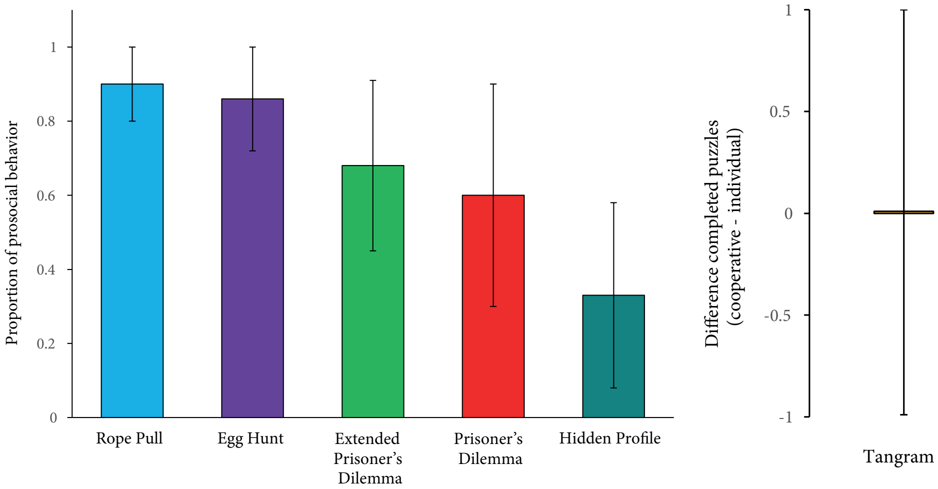


Figure 2. Median (± Median Absolute Deviation) of the proportion of prosocial behavior per game (left). The Tangram game (right) is displayed separately because it is based on a scale between -5 and 5, whereas all other measures are based on proportions.

Correlations and Principal Component Analysis

The bivariate correlation coefficients of each combination of games are displayed in Table 3 and Figure 3. The strongest correlation was evident between the Prisoner's Dilemma game and its extended version. Both versions were also positively correlated with the outcome of the Rope Pull game. Thus, if a person was willing to cooperate in the Prisoner's Dilemma game, s/he was also likely to cooperate in the extended version of the Prisoner's Dilemma game and the Rope Pull game. The three games were not significantly correlated with the more naturalistic games. The correlations between the latter games were non-significant.

The Principal Component Analysis confirmed the pattern seen in the correlation matrix and revealed that the six games could be best represented by two components. The component loadings are depicted in Table 4. The Prisoner's Dilemma game, its extended version, and the Rope Pull game loaded positively on the first component. As the three games are all variants of a social dilemma game, we referred to this component as the "social dilemma games" component. The Egg Hunt game loaded positively and the Hidden Profile game loaded negatively on the second component. To reflect the distinction between the economic games of the first component and the more naturalistic games of the second component, we called the latter component "naturalistic games". The two components were slightly correlated (.03).

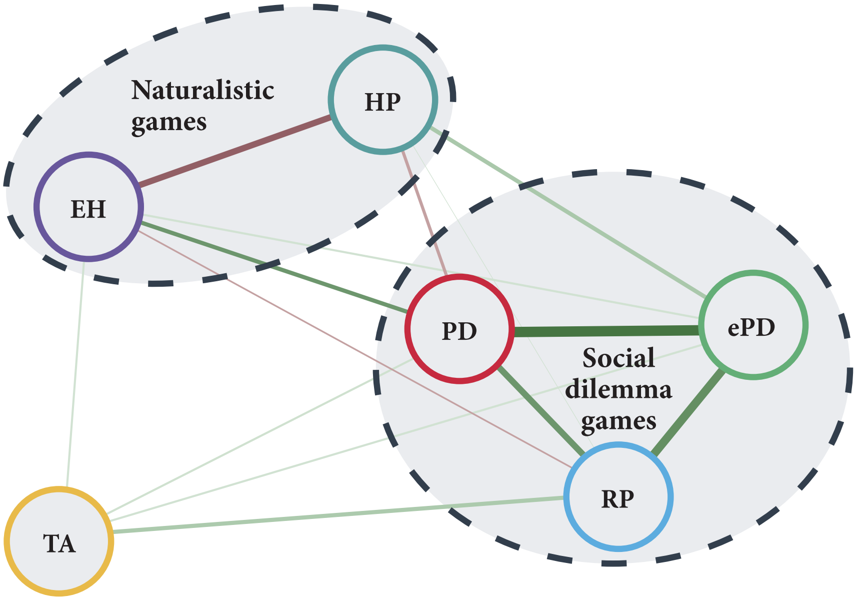


Figure 3. Visualization of the correlations between the six cooperation tasks. The thickness of the lines represents the strength of the correlation. Green lines reflect positive correlations, red lines negative ones. EG = Egg Hunt; HP = Hidden Profile; PD = Prisoner's Dilemma; ePD = extended Prisoner's Dilemma; TA = Tangram; RP = Rope Pull.

Table 3

Correlation matrix of the six games with the Spearman's rho coefficients, the corresponding 95% confidence interval and the sample size of each pair

		Prisoner's Dilemma	Extended Prisoner's Dilemma	Rope Pull	Tangram	Hidden Profile	Egg Hunt
Prisoner's Dilemma	r_s	1	**.44*****	**.31***	.08	-.15	.22
	95%-CI		.22-.62	.08-.51	-.16-.31	-.38-.09	-.06-.47
	p-value		< .001	.010	.517	.222	.117
	N	70	66	68	69	67	52
Extended Prisoner's Dilemma	r_s		1	**.34****	.05	.09	.06
	95%-CI			.11-.53	-.19-.28	-.16-.32	-.22-.33
	p-value			.004	.709	.487	.676
	N		70	69	70	67	50
Rope Pull	r_s			1	.13	.02	-.05
	95%-CI				-.11-.35	-.22-.25	-.32-.23
	p-value				.287	.881	.712
	N			72	71	69	51
Tangram	r_s				1	-.00	.04
	95%-CI					-.23-.23	-.24-.31
	p-value						.785
	N				73	70	52
Hidden Profile	r_s					1	-.24
	95%-CI						-.49-.04
	p-value						.088
	N					70	51
Egg Hunt	r_s						1
	95%-CI						
	p-value						
	N						53

Note. r_s = Spearman's rho correlation coefficient; *CI* = confidence interval; *N* = sample size;
* p < .05; ** p < .01; *** p < .001.

Table 4

Component Loadings of the Principal Component Analysis

	PC 1	PC 2	Uniqueness
Extended Prisoner's Dilemma	0.80	.	0.36
Egg Hunt	.	0.77	0.40
Hidden Profile	.	-0.77	0.41
Prisoner's Dilemma	0.72	.	0.34
Rope Pull	0.69	.	0.48
Tangram	.	.	0.96

Note. Applied rotation method is promax.

Discussion

Prosociality is central to humanity's unique capacity for large-scale cooperation. Experimental paradigms that measure prosocial behavior can help us understand how it emerges, as it allows researchers to investigate its contextual boundaries, zoom in on individual differences, and take factors such as previous experiences and costs into account. Researchers have designed multiple paradigms that tap into prosocial behavior. A crucial question is whether such paradigms measure a similar construct, that is, a person's general tendency to act prosocially, or whether the different paradigms tap into distinct subcomponents of prosociality. Previous studies have shown that behavioral tendencies are fairly consistent across economic games and translate to prosocial behavior outside the lab (e.g., voluntarily filling out a feedback form, Peysakhovich et al., 2014); sending back a "misdirected letter" enclosing money, (Franzen & Pointner, 2013; Stoop, 2014; Blanco et al., 2010; Böckler et al., 2016; Epstein et al., 2016; Yamagishi et al., 2013). The current study extends these findings by investigating both economic games and more naturalistic games. Contrary to the previous studies, in all six games, participants directly interacted with their game partners, better simulating real-life interactions. Investigating whether previous findings in anonymous, one-shot economic games also translate to more ecologically valid settings scrutinizes the robustness of the previously observed consistent behavior across paradigms. Given that participants engaged in face-to-face interactions in all games, the effect of such interaction on prosocial behavior was assumed to be constant across games. Consistent with previous findings, we observed that behavior in the three variants of a social dilemma game positively correlated. However, such consistency did not generalize to the three naturalistic games as evident by negligible correlations between the economic and naturalistic games. This pattern was also apparent in a Principal Component Analysis (PCA) showing that the series of tasks split into two components, which we dub "social dilemma games" and "naturalistic games". The three variants of the social dilemma game loaded positively onto the first component, whereas the Egg Hunt game loaded positively and the Hidden Profile game loaded negatively onto the latter component. The Tangram game was not related to any of the other games. In the section below, we will discuss these results in terms of theoretical and methodological considerations.

The "social dilemma games" component of the PCA showed that behavior in the three different variations of the social dilemma game was positively related (the classical and extended Prisoner's Dilemma game and the Rope Pull game). If participants cooperated in one of these games, they were more likely to do so in the other two as well. This was expected as all three games were designed to measure cooperative behavior. Specifically, they use similar outcome structures incorporating the essential ingredients to induce the dilemma between self- and collective interests: mutual cooperation is more beneficial for the joint outcome than mutual defection, and unilateral cooperation is more beneficial for a defector than mutual defection. The games differed with respect to the response scale (dichotomous versus 6-point scale) and the way of making a decision (pulling a rope versus pressing a button). The clustering of the three variants of the social dilemma game suggest that they are robust against these differences and capture people's general tendency to cooperate.

The behavior shown in the social dilemma games was distinct from the behavior displayed in the three naturalistic games. In other words, the two types of tasks did not measure the same underlying prosocial preferences. This discrepancy could result from theoretical differences, suggesting that the tasks measure distinct subcomponents of prosocial behavior. It could also be that methodological issues contributed to these findings as tasks differed regarding the level of feedback and clarity of the response options and their consequences. In the following, we will discuss these possible explanations in more detail.

The distinction between the social dilemma games and the more naturalistic games might be explained from a theoretical point of view such that they capture distinct subcomponents of prosocial behavior. For example, while social dilemma games are designed to measure cooperative behavior, the Egg Hunt game measures helping behavior. The lack of a relationship in behavior between these tasks might be the result of conceptual differences: While helping behavior is one-directional with one person helping to attain the goal of another person, cooperative behavior is bidirectional and implies interdependence, that is, the success of cooperation depends on two or more people working together towards a common goal (Penner, Dovidio, Piliavin, & Schroeder, 2005). Thus, while the conflict between self- and collective interests is inherent in cooperation, it is less salient in one-directional helping. A person might therefore be willing to cooperate with another person to achieve a common goal, but not necessarily help to achieve another person's goal. However, although conceptual differences are evident, such distinction has not been shown to elicit distinct behavior. For example, Böckler and colleagues (2016) showed that people who act prosocially during economic games also showed more helping behavior in the Zurich Prosocial Game (ZPG, Leiberg, Klimecki, & Singer, 2011). Furthermore, behavior in other games where the outcome depended on the other player (i.e., the Public Good and Trust Game) have been shown to correlate with behavior in the Dictator Game where the outcome of the dictator is independent of the other player (Peysakhovich et al., 2014; Yamagishi et al., 2013). As emphasized by Peysakhovich and colleagues (2014), these findings do not imply though that behavior across these games are driven by the same underlying motivation. For example, some games might be driven by reciprocity, while others might be driven by equality and altruistic preferences. Nevertheless, these studies suggest that despite differences in the interdependence of individuals that are inherent to the games, this likely does not explain the distinct behavior observed in the current study. Thus, the question remains why we did not observe consistent behavior across tasks. We now turn to possible methodological considerations that might shed more light onto this question.

The aim of the current study was to extend previous studies by incorporating more naturalistic games that reflect a range of situations also encountered in real-life interactions. Stepping away from the controlled context of economic games increases the ecological validity of the tasks, but simultaneously introduces additional factors that might influence the behavior of interest. Methodological differences between tasks such as the level of feedback and transparency of the response options and their consequences might have therefore contributed to the results observed in the current study.

First of all, the level of feedback differed such that participants received no feedback about each other's decisions during the classical and extended Prisoner's Dilemma games, whereas in all other games, individuals knew how prosocially their partner acted. Feedback has been shown to increase cooperation as it provides valuable information about a person's future decisions (Behrens & Kret, 2019; Jorgenson & Papciak, 1981; Monterosso et al., 2002). In line with this, we observed that participants cooperated more during the Rope Pull game, where participants received immediate feedback about the other person's move, compared to the classical and extended Prisoner's Dilemma game, where no feedback was provided (see Appendix A4). However, despite the overall increase in cooperation rates, the relationship between the games still showed that people behaved consistently between the games. In other words, a prosocial person still acted more prosocially than a less prosocial person despite the additive effect of feedback. Thus, although feedback influenced the overall level of prosocial behavior, our results suggest that this factor did not substantially affect the relationship between the games.

Second, a crucial difference between games was how obvious the response options and their consequences were for each game. In the social dilemma games, the response options were described in a pay-off structure and a participant knew that she could either cooperate or defect depending on her own preferences. On the contrary, in the Egg Hunt game participants were not explicitly informed that they could help each other and in the Hidden Profile game it was not mentioned that some, but not all of the information that both partners received was identical (see also McClung et al., 2017). Therefore, in the social dilemma games people could make a weighted, informed choice about whether they wanted to cooperate or not; whereas in the Egg Hunt and Hidden Profile game, people needed to discover the possibility of helping and sharing unique information first before they could work together. In other words, the degree of prosocial behavior displayed in these two games might have been undermined by whether people discovered *how* they could behave prosocially. However, as described above in the case of the effect of feedback, such difference in explicitly informing about the response options could affect the overall level of prosocial behavior without affecting the relationship between games. Although it would fit the pattern that behavior in the Egg Hunt and Hidden Profile game loaded on the same component, but were distinct from the other tasks, future studies are needed to investigate this explanation more directly.

Although behavior was related in the Hidden Profile and Egg Hunt game, we were surprised to observe that the Egg Hunt loaded positively and the Hidden Profile game negatively on the "naturalistic games" components which was consistent with a marginally significant negative correlation between the two games: a person who was more helpful in the Egg Hunt game shared less information in the Hidden Profile game. One possible explanation might be attributed to differences in the underlying motives causing the behavior. While the motivation for engaging in helping behavior in the Egg Hunt game is likely to be prosocially-driven, this is not necessarily the case for information exchange in the Hidden Profile game. Although it is assumed that working together on a common goal stimulates information exchange (De Dreu et al., 2008; Toma & Butera, 2009), research has shown that people generally stick with their own à priori decision (Wittenbaum et al., 2004). This bias not only motivates people to share as many characteristics of their own profile to convince others of their preferred candidate, but also influences individuals' encoding and retrieval of information from others' that is inconsistent with their own preference (De Dreu et al., 2008). The bias can be driven by competitive motives, where people want others to adopt their opinion, or by cooperative motives, where they are genuinely convinced that the own preferred candidate serves the group interest best (Wittenbaum et al., 2004). A previous study observed more information exchange in a cooperative compared to a competitive condition suggesting that information sharing is driven by cooperative motives (Toma & Butera, 2009). However, this study investigated the relationship between cooperation and information sharing on the contextual rather than individual difference level. Specifically, they manipulated the group's goal by emphasizing that the individuals should come to a joint decision (cooperative condition) or by encouraging group members to be the first to make a decision (competitive condition). Crucially, participants knew that some information was unshared and that unshared information was more important than shared information for the decision process. This allowed participants to strategically withhold information, directly linking sharing and withholding information to cooperative and competitive motives, respectively. In contrast, in the current study, participants were not informed that some information was unshared and more important. Therefore, deliberately withholding information did not function as the "competitive alternative" to information sharing. We do not know what motivated people in the current study, but given that helping behavior in the Egg Hunt game measured opposing behavioral tendencies in the Hidden Profile game (as evident by the negative correlation), it might be argued that behavior in the latter reflected competitive motives. However, this is speculative and future studies should use additional measures to understand people's underlying motives. For example, a person's social value orientation has been shown to relate to a variety of tasks (Balliet et al., 2009; Behrens & Kret, 2019; Böckler et al., 2016; C. Boone et al., 2010). Quantified by hypothetical distributions of resources between oneself and a hypothetical partner, SVO indicates the extent to which people take into account the welfare of another person when distributing resources between oneself and that other person (van Lange, 1999). Based on the distribution, participants are classified as generally being proself or prosocial (Van Lange, De Bruin, Otten, & Joireman, 1997). Thus, this measure could help shed light on why people share information in the Hidden Profile game.

To conclude, the second component of the PCA in our study indicates that the Egg Hunt game and the Hidden Profile game measure opposing tendencies–with the former reflecting helping behavior and the latter potentially indicating a person's competitive motives.

Finally, we observed that behavior during the Tangram game was not related to behavior during any of the other games. The game is designed to measure cooperation, but possibly, methodological differences with studies using the same game might explain its lack of correlation with the other games. Cooperation was operationalized as the difference in completed puzzles between the cooperative and individual condition. While we accounted for individual differences in spatial ability necessary to complete the task within the individual condition, two issues arose from the comparison of the two conditions. First, because people were randomly paired, their spatial skills were not matched, which is particularly relevant to the performance in this game. Consequently, if a person who performed poorly individually was paired with a skilled person, the skilled individual would inevitably complete most of the puzzles in the cooperative condition. As a result, the cooperation rate is overestimated for the poorly performing person and underestimated for the skilled one. Second, for all puzzles, the difficulty level was relatively high from the beginning. Although the first puzzles included highlighted pieces that made it easier to detect their individual shapes, this might not have been enough to prevent floor effects. Moreover, in order to equalize the duration of the different games in this study, participants had less time to complete the Tangram game than in two earlier studies (Saleem, Anderson, & Barlett, 2015; Vink et al., 2018). In sum, the random matching of participants with different skills and the difficulty of the task might explain why cooperative behavior in our Tangram game was not related to the behavior in the other games. Performance during the Tangram game might therefore be a reflection of participants' spatial abilities rather than prosocial tendencies, but future studies need to verify this presumption.

In sum, we extend previous findings by showing that not only different economic games, but also changes within the same economic game elicit similar prosocial behavior. In other words, as long as the principle of the game (i.e., the pay-off structure) stays similar, methodological changes in the response scale and the way people give their response still allows researchers to measure similar prosocial tendencies by these variants. Furthermore, the aim of the current study was to investigate whether the consistent behavioral tendency observed in economic games would generalize to more naturalistic games. The short answer is: no. Our results revealed that the social dilemma games measured different behavioral tendencies than the naturalistic games. Our discussion suggests that methodological differences in, for example, the clarity on *how* to act prosocially might explain such distinction between the economic and naturalistic games. This explanation is, however, speculative and further research is needed to draw strong conclusions about what drove the current findings. Economic games are designed to create conflicts of interests within individuals and changing the payoff structure determines the nature of these conflicts. Such experimental control combined with an extensive body of literature has portrayed a rather detailed picture of the underlying motives of choosing one over another option in these games.

However, for the more naturalistic, less often used games such as the Egg Hunt and Tangram game, the motivation behind people's behavior is more ambiguous and the effects of methodological changes are less well known. This does not mean that economic games measure prosocial behavior better. In fact, we encourage researchers to include naturalistic games to investigate whether previous findings generalize to these games. However, researchers need to be aware of potential differences between games and we advise to combine naturalistic games with other paradigms to tap into potentially different aspects of prosocial behavior and to integrate the findings with existing literature. Our current study takes a first step in this direction and sheds light onto the generalizability of prosocial behavior as measured by different paradigms.

Chapter 3

The interplay between face-to-face contact and feedback on cooperation during real-life interactions

Abstract

Cooperation forms the basis of our society and becomes increasingly essential during times of globalization. However, despite technological developments people still prefer to meet face-to-face, which has been shown to foster cooperation. However, what is still unclear is how this beneficial effect depends on what people know about their interaction partner. To examine this question, 58 dyads played an iterated Prisoner's Dilemma game, sometimes facing each other, sometimes without face contact. Additionally, explicit feedback regarding their decisions was manipulated between dyads. The results revealed that participants were more cooperative when they saw each other compared to when they could not, and when receiving reliable compared to unreliable or no feedback. Contradicting our hypothesis that participants would rely more on nonverbal communication in the absence of explicit information, we observed that the two sources of information operated independently on cooperative behavior. Interestingly, although individuals mostly relied on explicit information if available, participants still cooperated more after their partner defected with face-to-face contact compared to no face-to-face contact. The results of our study have implications for real-life interactions, suggesting that face-to-face contact has beneficial effects on prosocial behavior even if people cannot verify whether their selfless acts are being reciprocated.

Based on: Behrens, F., & Kret, M.E. (2019). The Interplay Between Face-To-Face Contact and Feedback on Cooperation During Real-Life Interactions. *Journal of Nonverbal Behavior, 43* (4), 513–528.

Introduction

Cooperation is an important foundation of social group life and essential in diverse activities ranging from riding a tandem to raising a child. On the one hand, relatively recent technological developments such as the Internet, social media and Skype, allow cooperation on a heretofore unknown scale; researchers separated by an ocean can conduct research together and discuss findings via video chat. On the other hand, more and more people express their concerns; in public transport, travelers hardly interact anymore, but are sucked up by their phones. The big question is whether and to what extent face-to-face contact fosters cooperative endeavors and helps society flourish. To that extent, the current study investigates people's cooperative tendencies towards others when partners can see each other or not, and when they have true, unreliable, or no information about their partners' previous cooperative decisions.

Face-to-face contact may be an important predictor for cooperation. Indeed, using economic games, previous research has shown that with the introduction of more artificial forms of communication, cooperation declines. For instance, cooperation drops in the context of writing messages, having telephone conversations and interacting with human-like avatars compared to face-to-face encounters with real people (Balliet, 2010; Bohnet & Frey, 1999; Drolet & Morris, 2000; Frohlich & Oppenheimer, 1998; Kiesler et al., 1996). Possibly, during face-to-face interactions, social information including body language, facial expressions, and eye gaze, shape our expectations about the partner's intentions and therewith decreases the risk of being exploited (Balliet, 2010; R. T. Boone & Buck, 2003). Furthermore, face-to-face contact makes social norms more salient, and hence, boosts cooperation (Bohnet & Frey, 1999). While these studies have often investigated the effect of face-to-face contact by looking at "communication" in a broader sense including both verbal and nonverbal communication (Balliet, 2010; Bicchieri & Lev-on, 2007; Brosig, Weimann, & Ockenfels, 2003; Jorgenson & Papciak, 1981; Sprecher, 2014), the current study focused on nonverbal communication only, refining its contribution to the overall beneficial effect of face-to-face contact. Verbal communication allows people to explicitly exchange information, discuss strategies and agree on future steps in the game. Nonverbal information, such as facial expressions, eye gaze, and pupil dilation, on the other hand, are more subtle, but still carry rich and genuine information that we use to express and interpret other's intentions, which consequently influences our (prosocial) behavior (Adolphs & Tusche, 2017; R. T. Boone & Buck, 2003; R. H. Frank, Gilovich, & Regan, 1993b; Jahng, Kralik, Hwang, & Jeong, 2017; Kret, 2015; Kret et al., 2015; Myllyneva & Hietanen, 2015; Prochazkova & Kret, 2017). Until now, studying the effect of dynamic nonverbal communication on cooperation in natural dyadic interactions has been largely neglected. Only one study conducted by Jahng et al. (2017) used a similar set-up where people were restricted to use nonverbal communication only. The authors reported that seeing each other increased mutual cooperation compared to when participants could not see each other. In sum, the literature shows that face-to-face contact is likely to be important for cooperation in different contexts.

Another factor that people generally take into account when considering cooperation, is knowledge about the person and how cooperative s/he has been in the past. Cooperation is a vulnerable act associated with the risk of being exploited. Knowing that the interaction partner has cooperated before lowers this risk. Research using economic game paradigms clearly shows that this kind of knowledge modulates the outcomes of such games (Bixenstine & Wilson, 1963; Jorgenson & Papciak, 1981; Monterosso et al., 2002; Tedeschi et al., 1968). What is less clear, is how explicit knowledge about a partner's previous decisions is integrated with the nonverbal signals that are being transmitted. In other words, how strongly do people rely on face-to-face contact compared to past behavior? The current study addresses this question for the first time by manipulating the visibility of nonverbal information and knowledge about past behavior.

In the current study, two naïve participants played an iterated Prisoner's Dilemma game while they could either see each other or not, and where they received feedback about their partner's decisions that was either true, unreliable or absent (no feedback). We had three main hypotheses. First and foremost, we expected participants to be more willing to cooperate when facing each other, allowing for the implicit transmission of nonverbal signals, compared to when a visual barrier blocked the view of one another. Second, we hypothesized that cooperation would be influenced by the type of feedback participants received. Specifically, we expected individuals to cooperate on the largest scale when receiving correct feedback about the partner's decisions. In this experimental condition, the predictability of the partner's next choice was highest and the risk of being exploited lowest. We further expected that a participant's decision was influenced by the partner's latest decision, and more likely being the same than dissimilar. Third, investigating the possible interplay between feedback and face-to-face contact, we hypothesized that the advantage of face-to-face contact would be most pronounced when participants received no feedback at all, as they would have to rely on nonverbal signals exclusively.

For exploratory reasons, we investigated potential effects of individual differences in social value orientation, emotion recognition ability, social anxiety, and empathy, as previous research suggests that these might modulate cooperative decisions (Bogaert, Boone, & Declerck, 2008; Doesum, Van Lange, & Van Lange, 2013; Eisenberg & Miller, 1987; Sylwester, Lyons, Buchanan, Nettle, & Roberts, 2012; Wehebrink, Koelkebeck, Piest, De Dreu, & Kret, 2018) or the effect of face-to-face contact on cooperation (Adolphs, Sears, & Piven, 2001; Emonds, Declerck, Boone, Vandervliet, & Parizel, 2011; Kret, Stekelenburg, De Gelder, & Roelofs, 2017; Pierce, 2009).

Method

Participants

In total, 116 individuals ($M_{age} = 21.05$, $SD_{age} = 2.49$) participated in the study with 72 (62.1%) females. They were randomly paired to form 58 same-sex dyads, ensuring that they did not know each other before. One dyad was excluded from the analysis due to missing data for more than 50% of the trials in the face-to-face condition. Participants received either course credits or a monetary reward of €7 per hour. In addition, all participants had the chance to win between €0.5 and €2.0 extra based on their performance during the Prisoner's Dilemma game (no deception). Participation took two hours, which included completing questionnaires at home and in the lab and playing the Prisoner's Dilemma game. Upon arrival at the lab, participants gave informed consent to participate in the study, which was approved by the Ethics Committee of Leiden University (CEP16–0314/131). They received full debriefing afterwards.

Materials

Two participants played two rounds of 50 Prisoner's Dilemma games, measuring cooperative behavior. This game provides two choice alternatives, where the performance of one player depends on both one's own and the other person's choices. Specifically, both can choose to cooperate (C) or defect (D) during each trial. When both cooperate (CC), the incentive is larger compared to when both defect (DD). However, when one player defects whereas the other cooperates (DC), the former gets the highest possible incentive while the latter receives the lowest. In that way, a conflict emerges between self- and collective-interests because the joint outcome is larger when both players cooperate, while the trade-off for each individual is larger when defecting (DC > CC > DD > CD from the perspective of Player 1). The payoff structure for the current study was as follows: DC = 4, CC = 3, DD = 2, CD = 1 (see Table 1; Balliet & Van Lange, 2013). Following standard procedures and in order to avoid confounds through suggestive formulations, the game was phrased as choosing between options A or B rather than between cooperation and defection.

Study Design

The study used a mixed design with one between-subject (feedback) and one within-subject variable (nonverbal communication). The dependent variables were the willingness to cooperate (0 = defect, 1 = cooperate) and the dyad's joint outcome, i.e., mutual cooperation (CC), mutual defection (DD) and one-sided cooperation (CD/DC).

Table 1
The payoff matrix of the Prisoner's Dilemma game

	Other	
You	C	D
C	3–3	1–4
D	4–1	2–2

Note. The first number refers to the points earned by "You".

Feedback was manipulated between dyads with a third of the dyads receiving no feedback about the other person's decision, a third receiving correct feedback, and a third receiving random feedback (50% correct, 50% incorrect). Participants in the latter condition were not informed that the feedback was random. Furthermore, the possibility for participants to use nonverbal communication was manipulated within dyads. Participants played the game twice, once where they faced each other, allowing for nonverbal communication (face-to-face condition), and once while a visual cover was placed in between them, constraining nonverbal communication (face-blocked condition).

Procedure

Before coming to the lab, participants received information about the study and a link to fill out an online questionnaire consisting of the Liebowitz Social Anxiety Scale (LSAS; Beard et al., 2011), the Interpersonal Reactivity Index (IRI; Davis, 1980) and nine brief, decomposed games to measure their social value orientation (SVO; Van Lange, de Bruin, Otten, & Joireman, 1997). When the two participants forming one dyad arrived at the lab, they were immediately separated to avoid interactions before the experiment started. In different rooms, they gave informed consent and filled out the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) and completed the Reading the Mind in the Eyes task (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). After filling out these questionnaires, they read the instructions of the Prisoner's Dilemma game and answered verification questions to make sure that they understood the game correctly. The questionnaires were not the main focus of our study, which is the reason why we present the descriptive statistics and additional analyses regarding their relation with the experiment in Appendix B1.

When both participants were ready to start the experiment, they were asked to sit at a table facing each other. Participants could not see each other's responses (button presses), as a visual cover was placed in the middle, such that participants could only see each other's faces, but not their bodies. At the beginning of the experiment when participants sat at the table and during the no-face condition, there was an extra visual cover on top of the other one, such that participants could not see each other at all (see Figure 1). The payoff matrix was placed in front of them on the table to reduce mental effort trying to remember the payoff structure (see Table 1; Balliet & Van Lange, 2013). The experiment started with five practice trials. When no errors were made, the real experiment began, consisting of 45 testing trials. In order to keep the experiment controlled and standardized, an audio file was played so that participants heard all instructions via their headphones. The sequence of events per trial was as follows: First, both participants were instructed to look at each other ("Look at each other" or: "Look at the cross in front of you"). This interval allowed participants to make a decision and to "read" the other person's mind, and decide whether he/she would cooperate or defect. After four seconds, they were told to look down and make a decision (cooperate or defect), by pressing the corresponding button on the keyboard (the corresponding keys were marked with stickers saying "A" for option A (cooperate) or "B" for option B (defect) ("Look down at the table. After the beep, choose as fast as possible between option A and B. Keep looking down"). Subsequently, they were instructed to indicate what they thought the other person had decided to choose (using the same keys as for their own choice) ("Indicate after the beep what you think the other person chose, option A or B. Keep looking down"). In the correct/random feedback conditions, the points each player earned were subsequently communicated via the headphones (e.g., "Player 1 receives 1 point, player 2 receives 4 points"). In the correct feedback condition, the feedback reflected the actual responses. In the random feedback condition, the points were correct 50% of the time and 50% incorrect. In the "no feedback" condition, they were only informed that both players had made a decision without any information about the points earned ("Both of you have made a decision"). The next trial started when the auditory instruction to look at each other was given. No information was given about the cumulative performance of the players during the experiment.

Before and after the testing trials, participants rated their experience of the interaction regarding their level of connection, awkwardness, and shyness towards the other person on visual analog scales (VAS). After the first session of 50 trials (5 practice, 45 testing trials), participants had to change their sitting position so that they could not (could) face each other (the order of starting in the face-to-face or no-face condition was counterbalanced) and played the game again for 50 trials (5 practice, 45 testing trials). After the practice and testing trials, participants filled out the same VAS as in the first session. The descriptive statistics of the VAS at the four time points are provided in Appendix B4 (see Table B.S4).

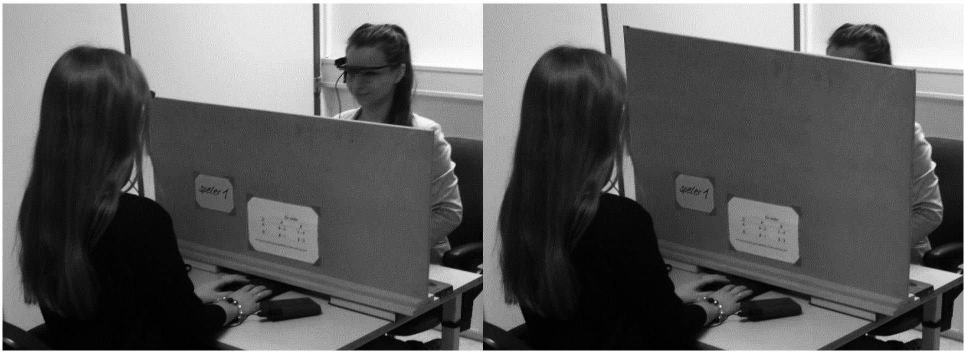


Figure 1. Set-up of the experiment for the face-to-face condition (left) and the face-blocked condition (right).

At the end of the experiment (after the second session), participants filled out the DFI (Desire for Future Interaction Scale, see Table B.S4) and answered questions about their insight and experience of the interaction with questions including "How much could you 'read' the other person's intentions?", "Which information did you use to 'read' the other person's intentions?", and "How much did the feedback influence your choices?". In the two feedback conditions, a manipulation check verified that participants believed that they received correct feedback (participants in the random feedback condition were not informed that the feedback was incorrect 50% of the time). Finally, the two individuals were separated again to give them their monetary reward based on their performance on a randomly selected trial (DC = €2, CC = €1.5, DD = €1, CD = €0.5 from the perspective of Player 1).

Data Analyses

We performed two types of analyses, one with the willingness to cooperate (per person) and one with the joint outcome (per dyad) as the dependent variable. We made this distinction because we aimed to understand the effects of nonverbal communication and explicit information on the personal level and how that translates into successful cooperation. For the willingness of cooperation, we performed binary logistic mixed model analyses with cooperation (0 = defect, 1 = cooperate) as the dependent variable. For the joint outcome, we used a multinomial mixed model analysis with mutual defection as the reference. This results in binary estimations of the likelihood of achieving mutual cooperation (CC) over mutual defection (DD) and the likelihood of reaching one-sided cooperation (CD/DC) over mutual defection (DD). To account for the differences between dyads and interdependence within dyads, Dyad and Dyad * Player were added as random intercept effects. As predictor variables, we added face condition (0 = face-blocked, 1 = face-to-face) as a within-dyad factor and feedback type condition (0 = no, 1 = correct, 2 = random) as a between-dyad factor. On a trial-by-trial basis, we coded the feedback that participants received in the previous trial (previous feedback, 0 = defect, 1 = cooperate). Participants did not directly hear whether the partner defected or cooperated but were informed about the amount of points each player received, from which they could deduce the decision the partner had made (e.g., when player 1 heard "player 1 received 4 points and player 2 received 1 point", s/he knew that the partner [player 2] cooperated). Keep in mind that in the random feedback condition participants sometimes received correct feedback and sometimes incorrect feedback. Both players received either correct or incorrect feedback on the same trials. Given that players knew what they chose themselves, the incorrect feedback had to be tailored to their choice, only changing the feedback about the other player's decision. In the correct feedback condition, both players always heard the same audio clips.

In an exploratory analysis, we investigated the relation between cooperative behavior, face-to-face contact and personality traits. In addition, we appended an analysis investigating whether people were able to predict each other's decisions based on nonverbal cues, an ability that has been reported previously (Lewkowicz, Quesque, Coello, & Delevoye-Turrell, 2015; Sparks, Burleigh, & Barclay, 2016; Verplaetse, Vanneste, & Braeckman, 2007). A more detailed description and the results of these analyses can be found in Appendix B1 and B2. Given the number of analyses we ran, increasing the chance of Type I errors, we decided to lower the significance level to $\alpha = 0.005$ for the exploratory analysis section. For the confirmatory analyses, the significance level was $\alpha = 0.05$.

Results

The mean cooperation rates and joint outcomes per condition are presented in Table 2 and 3, respectively. First, our hypothesis that participants would cooperate more in the face-to-face compared to face-blocked condition was confirmed, $B = .17$, $SE = .05$, CI (.06, .27), $OR = 1.19$, $p = .002$. This effect also translated into more successful cooperation considering the joint outcome with people successfully cooperating more (compared to mutually defected) in the face-to-face condition compared to the face-blocked condition, $B = .32$, $SE = .10$, CI (.12, .51), $OR = 1.38$, $p = .001$. Participants also chose one-sided cooperation over mutual defection more in the face-to-face compared to face-blocked condition, $B = .19$, $SE = .08$, CI (.03, .35), $OR = 1.21$, $p = .018$. In Appendix B3 we describe additional effects regarding the order of the sessions.

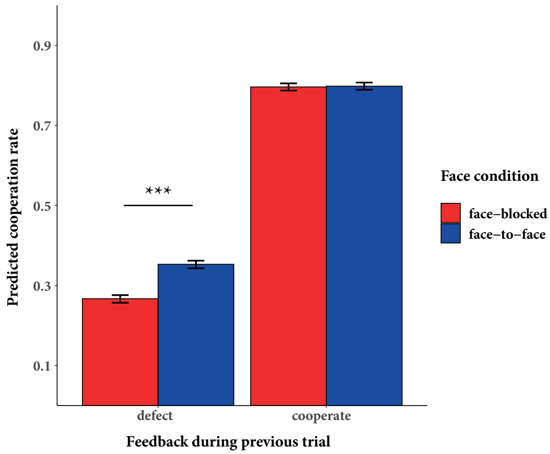


Figure 2. Predicted mean cooperation rate in the face-blocked and face-to-face condition moderated by what people heard what their partner chose in the previous trial. Error bars represent 95%-confidence intervals. (*** $p < 0.001$).

Regarding our second hypothesis, we observed a significant effect of feedback type, $F(2, 10226) = 8.08$, $p < 0.001$: participants were more willing to cooperate in the correct compared to the unreliable feedback condition, $B = 2.88$, $SE = 0.72$, CI (1.47, 4.29), $OR = 17.81$, $p < .001$, and the no feedback condition, $B = 1.64$, $SE = 0.72$, CI (0.23, 3.04), $OR = 5.16$, $p = .023$. The difference between no and unreliable feedback was not significant ($p = .075$). Focusing on the dyads that did receive feedback about their partner's previous decisions, we investigated whether the feedback they received during the previous trial impacted on participant's cooperative behavior during the subsequent trial. As participants did not know whether the feedback they received was correct or not, we only investigated the effect of what they actually heard, independent of whether this was correct or not[1]. The results reveal that participants indeed relied on what they learned from the explicit feedback when deciding on their next choice, with a greater willingness to cooperate after hearing their partner cooperated versus defected in the previous trial, $B = 1.53$, $SE = .07$, CI (1.40, 1.67), $OR = 4.62$, $p < .001$.

[1] There was a significant interaction between feedback type and previous feedback ($p < .001$) driven by extreme cooperation rate values, i.e., by a ceiling effect. After excluding extreme participants with a cooperation rate higher than .95, the interaction became non-significant suggesting that the willingness to cooperate based on what they heard did not differ between the correct and random feedback condition; $p = .120$ (number of excluded trials: 6463–4638 = 1825)

With respect to our third hypothesis that people would rely more on nonverbal communication during their decision to cooperate when no explicit information was available (i.e. in the no feedback condition), the interaction between feedback type and face condition, $F(2, 10223) = 3.16$, $p = .043$, was not meaningful as it was entirely driven by dyads with extreme cooperation rates. After excluding the highest cooperation rates (mean cooperation rate of 0.99 or higher), the effect was rendered insignificant, indicating that the interaction was driven by a ceiling effect, $F(2, 7441) = 1.36$, $p = .256$. With respect to the content of the feedback participants received in the previous trial, we observed an interaction effect with face condition, $F(1, 6459) = 14.0$, $p < .001$ (see Figure 2), with significantly more choices to cooperate in the face-to-face compared to face-blocked condition when participants heard that their partner defected in the previous round, $B = .43$, $SE = .10$, CI (.22, .63), $OR = 1.54$, $p < .001$. However, there was no such difference when participants heard that their partner cooperated during the previous trial ($p = .527$). Disentangling the interaction from another perspective, although participants' willingness to cooperate decreased after hearing that their partner defected in the previous round, this decrease was less pronounced when players saw each other compared to when they did not, as was evident by a smaller beta estimate in the face-to-face condition, $B = 1.17$, $SE = .10$, CI (0.98, 1.36), $OR = 3.22$, $p < .001$, compared to the face-blocked condition, $B = 1.61$, $SE = .10$, CI (1.41, 1.81), $OR = 5.00$, $p < .001$.

Next, we investigated whether the benefits of explicit feedback on partners' willingness to cooperate also translated into greater cooperative success. This was indeed the case, with more mutual cooperation, $B = 1.80$, $SE = .13$, CI (1.54, 2.06), $OR = 6.05$, $p < .001$, and more one-sided cooperation, $B = 0.82$, $SE = 0.11$, CI (0.61, 2.03), $OR = 2.27$, $p < .001$, compared to mutual defection when participants heard their partner cooperated rather than defected in the previous trial. This effect was independent of the face condition the dyads were in ($p = .052$).

Table 2

Overview of the means and their standard errors of the cooperation rates for the (a) main effects of Face and Feedback condition and (b) their interaction effect

a)

Main effect	Condition	Mean (SD)
Face condition	face-to-face	.64 (.03) ***
	face-blocked	.62 (.03) ***
Feedback condition	no	.60 (.04) *
	correct	.79 (.03) ***
	random	.48 (.02)

Note. * $p < .05$; ** $p < .005$, $p < .001$, one-sided one-sample t-tests (m = 0.5, level of chance).

b)

		Face condition	
		face-to-face	**face-blocked**
Feedback condition	no	.60 (.06)	.61 (.06)
	correct	.80 (.04) ***	.78 (.05) ***
	random	.51 (.03)	.46 (.03)

Note. * $p < .05$; ** $p < .005$, $p < .001$, one-sided one-sample t-tests (m = 0.5, level of chance).

Table 3

Proportions (mean & standard errors) of the joint outcomes (CC, DD, CD/DC) per condition

		Joint outcome		
Feedback condition	**Face condition**	**CC**	**DD**	**CD/DC**
no	face-to-face	.35 (.009)	.17 (.005)	.48 (.007)
	face-blocked	.37 (.009)	.19 (.006)	.44 (.007)
correct	face-to-face	.72 (.009)	.13 (.005)	.16 (.005)
	face-blocked	.69 (.009)	.16 (.006)	.15 (.004)
random	face-to-face	.25 (.004)	.28 (.004)	.48 (.003)
	face-blocked	.19 (.003)	.30 (.003)	.51 (.002)

Note. The standard errors of the means are calculated per dyad per face condition.

Discussion

In the current study, we investigated the joint effects of face-to-face contact and knowledge about partners' previous behavior on the willingness to cooperate and on cooperative success. Our key results are threefold and show that first, face-to-face contact stimulates people's willingness to cooperate, even when their partner defected earlier. This positive effect also translates into more successful cooperation (cooperative decisions in both players). Second, participants are most cooperative when receiving reliable feedback about their partner's behavior compared to unreliable or no feedback, reciprocating their partner's past behavior when that information is available. Third, the benefit of face-to-face contact operates independently of whether people have knowledge about their partner's previous behavior or not. In other words, the positive effects of face-to-face contact and knowledge about a partner's previous behavior on cooperation are additive rather than interdependent. These results, along with other findings, will be discussed in detail in the sections below.

The first key result, that face-to-face contact promotes cooperative behavior and translates into successful joint cooperation, replicates previous studies that have shown the beneficial effects in different contexts (Balliet, 2010; Bohnet & Frey, 1999; Frohlich & Oppenheimer, 1998; Jahng et al., 2017; Sally, 1995). One study conducted by Jahng et al. (2017) used a similar set-up as the current study where dyads played multiple rounds of the iterative Prisoner's Dilemma game while looking at each other or not. In line with our findings, their study showed that cooperation was more successful when participants looked at each other compared to when a visual barrier was placed between them. In their study, participants always received feedback about each other's decisions. Interestingly, in the reliable feedback condition of our study, mutual cooperation rates were much higher in both the face-to-face and face-blocked condition (72% and 69%, respectively) compared to their study (around 35% and 20%, respectively). This difference might be attributed to the differences in the sample population: participants in Jahng et al.'s study (2017) were male, Korean students from the local university, whereas we tested both male and female, mostly Dutch psychology students. Although males and people from collectivistic cultures have been shown to be more cooperative than females and people from individualistic cultures (Balliet, Li, Macfarlan, & Van Vugt, 2011; Parks & Vu, 1994), including participants with a broader range of backgrounds might have led participants to be less cooperative in the sample from Jahng et al. (2017) compared to the psychology students of our study (Frank, Gilovich, & Regan, 1993a). Specifically, individuals from a beta-science background (included in Jahng and colleagues' study [2017] , but not in our study) might be more sensitive to the mathematical advantage of choosing to defect (players will always receive higher rewards when defecting independent of what their partner chooses). The implications of this could be important in different settings, but future studies are needed to directly compare different groups of people and to make valid statements about the effects of sample populations on the cooperative behavior in dynamic social interactions. In sum, the current study replicates previous studies supporting the beneficial effect of face-to-face contact and nonverbal communication in particular on cooperation in dyadic interactions. Our study extends previous works in various ways, which will be discussed in the next section.

Besides manipulating the access to nonverbal information, we also varied the degree of feedback participants received about their partner's behavior, providing reliable, unreliable, or no feedback after each decision. In line with previous studies (which did not manipulate face contact), our second key finding shows that cooperation is higher when receiving reliable compared to unreliable or no feedback (Jorgenson & Papciak, 1981; Monterosso et al., 2002; Pillutla & Chen, 1999). In real-life situations, we often have information about the past behavior of our interaction partners from previous experiences or through gossip with a third person. Based on this information we can predict our partner's future behavior and promote cooperation by encouraging others to reciprocate one's own prosocial behavior. Feedback provides a way to control and verify these predictions, infer a partner's strategies directly and unambiguously and eases the adjustment of own behaviors accordingly (Jorgenson & Papciak, 1981). One strategy that is often adopted in social dilemma games such as the Prisoner's Dilemma game is that people reciprocate the decisions the partner has made, a finding that is also supported in the current study: people were more willing to cooperate when their partner cooperated, but also tended to reciprocate a selfish decision (Axelrod & Hamilton, 1981; Fehr & Fischbacher, 2004; Rilling et al., 2008). Given that mutual defection is less favorable than mutual cooperation (in the current study receiving 2 versus 3 points, respectively), the latter becomes more beneficial, resulting in more cooperation. On the other hand, when feedback is not provided, people cannot be "caught" violating the social norm of reciprocating cooperation and consequently decreasing the incentive to make prosocial decisions (Biel & Thøgersen, 2007; Fehr & Fischbacher, 2004). In a similar vein, when feedback is provided that is sometimes correct and sometimes incorrect (as in the unreliable feedback condition in our study), participants appear overall more selfish, evoking more mutual defection following the reciprocity strategy. This is indeed what we observe: cooperation drops substantially in the unreliable feedback condition and participants reciprocate their partner's decisions independent of whether the feedback is correct or not. Hence, receiving feedback does not stimulate cooperation per se, but rather provokes reciprocity promoting cooperation only if the prosocial effort is returned.

Apart from the independent effects of face-to-face contact and knowledge about the partner on tendencies to cooperate and on cooperative success, we were particularly interested in their putative combined effect. Both sources of information can be used to predict the partner's next decision, which reduces the risk of being exploited when cooperating. Consequently, one of our predictions was that if only one of these sources is available, people would rely more on information from that source. Specifically, we hypothesized that we would find a greater benefit of face-to-face contact on cooperative behavior when no explicit information was provided compared to when such explicit information was available. In contrast to our hypothesis, the results showed that the "boost" in cooperation when facing each other was independent of whether and what type of explicit feedback participants received (correct or unreliable feedback). In other words, the beneficial effects of implicit and explicit information on cooperative decision-making were additive. On the other hand, the effect of the *content* of the feedback on cooperation was moderated by whether people could face each other or not. Interestingly, specifically when face-to-face contact was allowed, people cooperated more often despite a selfish partner. Cooperation is often seen as the social norm in social dilemma games and it has been suggested that the more intimate the interaction, the stronger social norms are activated (Bohnet & Frey, 1999). Hence, people might be more "forgiving" when facing their partner when he/she defects and therefore encourage the defecting partner to return to cooperation by opting for a cooperative decision themselves. Our study, however, does not allow for strong conclusions about the motivation to cooperate in response to a partner's defection, which is something future studies might want to look into specifically.

To our knowledge, only one other study has investigated the effects of communication and feedback before. In contrast to our findings, Jorgenson and Papciak (1981) observed that whether people were allowed to communicate or not, altered the effect of feedback on cooperation. Specifically, receiving feedback fostered cooperation but only if people could discuss their strategies and outcomes with each other. However, there were three essential methodological differences compared to our study that might explain the discrepancy in findings: First, the authors investigated the effect of *verbal* rather than nonverbal communication, either allowing participants to discuss their strategies for the game or not. Although face-to-face contact was not prohibited, the seating arrangements kept nonverbal communication to a minimum in both conditions (Jorgenson & Papciak, 1981). Hence, the type of communication was qualitatively different as nonverbal communication does not allow individuals to discuss strategies and comes with more uncertainty about the prediction of other's intentions. Research has shown that people are rather good at detecting emotions from nonverbal sources (Ekman & Friesen, 1971; Elfenbein & Ambady, 2002), but reading another person's intentions is more complex and difficult (Bonnefon, Hopfensitz, & De Neys, 2017). For that reason, people might be less likely to form and rely on predictions about the future decisions of a partner. On the other hand, verbal promises are less open to interpretation making predictions easier. The cost of this ease is that promises can easily be broken (i.e., lying), which is why feedback is especially important in these situations. As a consequence, the difference between receiving feedback or not might be more pronounced when communicating verbally compared to nonverbally.

Another important difference is that in Jorgenson and Papciak's study (1981) participants played in groups of four players, while our interactions consisted of only two players. Research has shown that trust, a premise of cooperative behavior, is enhanced among dyadic interactions compared to larger groups (Lev-on, Chavez, & Bicchieri, 2010). On the other hand, in a meta-analysis, Balliet (2010) reported a positive effect of group size on the relation between *verbal* communication and cooperation suggesting that the effect of communication becomes stronger in larger groups. These studies suggest that people might generally be more willing to cooperate in dyadic interactions, but that face-to-face interaction has a larger beneficial effect on cooperation when playing in larger groups. It might be that verbal communication is more important to coordinate the behavior and discuss strategies among members of a larger group (Jorgenson & Papciak, 1981), whereas nonverbal communication becomes more important in smaller groups because the transmission of nonverbal signals is a back-and-forth interplay between two people (Kret, 2015; Prochazkova & Kret, 2017). The exact consequences of group size on cooperative behavior are hence unclear and future studies would need to investigate the relation between group size, (nonverbal and verbal) communication and cooperation. In sum, the fact that Jorgenson and Papciak (1981) observed an interplay between communication and feedback, while we did not, might be explained by methodological differences, in particular, the type of communication and the group size. Future studies are needed to address these differences and scrutinize their impact on the relation between communication, feedback, and cooperation.

There are a few limitations in this study that should be considered. First, due to very high cooperation rates in the reliable feedback condition, we found a ceiling effect for the interaction effect between the feedback and face conditions. Future studies using a different payoff structure might want to investigate whether an interaction effect can be detected without such a ceiling effect and therefore show that the two sources of information might influence each other. Second, our results revealed that the order of the face condition influenced the effect of face-to-face contact on cooperative behavior (see Appendix B3). Future studies should investigate whether this is a true finding and if so, what can explain such an effect. It might be that the "social connection" between the two participants can only be established when the initial contact is face-to-face and that cooperation declines as face-to-face contact is not possible anymore. However, this order effect is confounded by a general "time" effect. Research has shown that cooperation declines with an increasing number of trials (Bó, 2005) and therefore the decrease in cooperation might have occurred with or without face-to-face contact. Future studies should include conditions where dyads cooperate only in the face-to-face or in the face-blocked condition over the two sessions to disentangle the order effect of the manipulation and a general time effect.

The current study has been the first to investigate the interplay between nonverbal and explicit information, raising many open questions for future studies to investigate. First, although various studies have provided evidence for the beneficial effect of face-to-face contact on prosocial behavior, the motivation and underlying mechanisms are less known. What exactly is it that we rely on when making our decisions and what is it about the face that makes us behave more socially? Which nonverbal signals do we pick up and take into account when making our social decisions? Although more and more studies address these questions, it is still an understudied topic and future studies using eye-tracking and physiological measures are highly needed to address these questions. Second, in an exploratory analysis (see Appendix B1 for details) we observed that the way face-to-face contact affects a person's cooperative behavior depends on that person's level of empathy, prosociality, social anxiety and emotion recognition skills. Future studies are needed to verify these findings and draw stronger conclusions about such effects. Related to this, we observed different levels of cooperation rates compared to another study with a similar set-up. One explanation might be that differences in the sample population might have caused such deviations. Future studies are needed to investigate whether the strength of the effect of face-to-face contact might differ between cultures, work disciplines, and other contexts. This might be particularly interesting in light of globalization as people are more and more often asked to work with people from different backgrounds who might benefit from different environments.

In conclusion, the findings of the current study emphasize the need to study social phenomena during real-life interactions and investigate the interaction and complexity of information derived from different sources. Fortunately, real-life interaction studies using realistic, ecologically valid contexts are on the rise and scientists increasingly realize that knowledge about social cognition cannot be merely based on studies lacking in actual social interactions. In the current study, we highlight the power face-to-face contact has on social decision-making and the remarkable willingness of people to cooperate with strangers even though they have no previous experiences or knowledge of that person.

Chapter 4

Physiological synchrony is associated with cooperative success in real-life interactions

Abstract

Cooperation is pivotal for society to flourish. To foster cooperation, humans express and read intentions via explicit signals and subtle reflections of arousal visible in the face. Evidence is accumulating that humans synchronize these nonverbal expressions and the physiological mechanisms underlying them, potentially influencing cooperation. The current study is designed to verify this putative linkage between synchrony and cooperation. To that end, 152 participants played the Prisoner's Dilemma game in a dyadic interaction setting, sometimes facing each other and sometimes not. Results showed that synchrony in both heart rate and skin conductance level emerged during face-to-face contact. However, only synchrony in skin conductance levels predicted cooperative success of dyads. Crucially, this positive linkage was strengthened when participants could see each other. These findings show the strong relationship between our bodily responses and social behavior, and emphasize the importance of studying social processes between rather than within individuals in real-life interactions.

Based on: Behrens, F., Snijdewint, J. A., Moulder, R. G., Prochazkova, E., Sjak-Shie, E. E., Boker, S. M., & Kret, M. E. (in press). Physiological synchrony is associated with cooperative success in real-life interactions. *Scientific Reports*.

Introduction

Cooperation is one of human society's core pillars, distinguishing us from other species in its scale and complexity (Bowles & Gintis, 2013). Despite countless examples of tremendous successes of people working together towards a common goal, there are as many examples where cooperation fails. An important question therefore is: How can cooperation be achieved? In order to be able to foster cooperation, we must first understand the mechanisms. The current study takes a step in that direction.

When making decisions, such as whether to cooperate or not, people rely on a variety of nonverbal expressions to communicate their own and predict others' intentions (Damasio et al., 1996; R. H. Frank, 1988). Cooperation is risky as individuals can take advantage of those investing time and resources, and nonverbal expressions reflecting a person's benign intents can help ensure cooperative success. Intriguingly, research has shown that emotional states tend to synchronize between interaction partners on several levels including the behavioral (Chartrand & Bargh, 1999), neural (Hasson et al., 2004), and physiological level (Fawcett, Wesevich, & Gredebäck, 2016; Levenson & Gottman, 1983). This is in line with the idea that emotional states are multidimensional constructs and that activation of one of these levels simultaneously activates the other levels (Wood et al., 2016). Although some of these emotion-induced changes cannot be observed by the naked eye directly, people perceive them indirectly through visual cues such as pupil size or a blush on the cheeks, and align their bodily responses accordingly (Prochazkova & Kret, 2017). Whether or not physiological synchrony is associated with cooperative decisions is a key question that has thus far remained unanswered.

Raising awareness of synchronized emotion states has had a vast impact on different disciplines with researchers investigating its clinical (Galazka et al., 2019), developmental (Fawcett, Arslan, Falck-Ytter, Roeyers, & Gredebäck, 2017), social (Tarr, Launay, & Dunbar, 2016a), evolutionary (Mancini, Ferrari, & Palagi, 2013), neural (Prochazkova et al., 2018), and cognitive (Kret et al., 2015) implications. It has been proposed that the *function* of this alignment is to infer the other person's emotions, to empathize, and to provide subsequent consolation, help, or other prosocial behavior (De Waal & Preston, 2017). Despite the clear predictions regarding the function of synchrony, studies have thus far only investigated the benefits of synchrony in artificial settings with either participants interacting with virtual characters on a computer screen (Kret & De Dreu, 2017), or two people interacting in cooperative compared to competitive contexts (Chanel et al., 2012). Thus far, no research has investigated the direct link between synchrony and subsequent cooperative decisions.

To what extent are synchrony and cooperative success linked? This pivotal question has never been directly addressed before. We aim to close this knowledge gap, focusing on physiological synchrony because it is implicit, hard to control or regulate, and is a crucial component of emotion processing (Critchley & Harrison, 2013; Kret, 2015). In psychology, the most commonly studied physiological responses are skin conductance level, a purely sympathetic nervous system response, and heart rate, which reflects both sympathetic and parasympathetic nervous system activity (Critchley & Harrison, 2013; Dawson et al., 2000). Previous research has shown that before people make a decision by, for instance, pressing a button in an experiment, that decision is already reflected in their physiology (Crone et al., 2004; Quesque, Behrens, & Kret, 2019). We here focus on these two measures, investigating whether they synchronize between interaction partners and if so, whether that relates to the cooperative success of a dyad.

To that end, participants played a modified iterated Prisoner's Dilemma game in dyads, sometimes facing each other (allowing for nonverbal communication), and sometimes with a visual cover between them (constraining nonverbal communication). Throughout the experiment, participants' heart rate and skin conductance levels were measured. The aim of the study was twofold: First, we aimed to confirm that physiological synchrony emerges during dyadic interactions. Second, we aimed to investigate whether synchrony is related to cooperative success and whether such a relationship was bound to interactions where partners could see each other.

Methods

Participants

In total, 152 individuals participated in the study (71% females, $M_{age} = 23$, $SD_{age} = 4.3$), who were recruited via the University online recruitment system (SONA) and by approaching people on University ground. By the time of data collection, we were not aware of methods to calculate a prior power analyses for hierarchical data structures. Instead, we based the sample size on our previous studies, where we used a very similar set-up (Behrens & Kret, 2019). Although recent advances would make it possible to conduct a post-hoc power analysis, we refrain from this as it has been suggested to greatly depend on the p-value of the observed effects [for a detailed explanation, see e.g., (Lenth, 2007; Plate, Borggreve, van Hillegersberg, & Peelen, 2019)]. Instead, we conducted a sensitivity analysis which has been recommended as a valid post-hoc method (Green & Macleod, 2016). In contrast to an a priori power analysis where the necessary sample size is calculated for a given power and effect size, the sensitivity analysis consists of simulation-based power analyses for different effect sizes with the fixed sample size of the study assuming that the effect sizes are the true population parameters. The results show that the minimum true effect that we can detect with a power of 80% and the sample size of our study ($N = 50$) is .70. The observed effect size of .86 is associated with a power of 89%, again assuming that the observed effect size reflects the true population effect size. Details on the sensitivity analysis and the associated power curve are described in Appendix C1.

A dyad consisted of two same-sex individuals who did not know each other ($N_{dyads} = 76$). The reason for using only same-sex dyads were that (i) factors such as sexual attraction could have influenced the level of synchrony in mixed-sex dyads (Prochazkova et al., 2019) and (ii) people have been shown to behave differently in social dilemma games when playing with their own compared to the other gender (Balliet et al., 2011). All participants had normal or corrected-to-normal vision wearing contact lenses (glasses were not compatible with the eye-tracking glasses, see below). They received either course credits or a monetary reward (8€) for participation and could earn an additional maximum of 2€ depending on their performance during the experiment (no deception). Informed consent was obtained from all participants (all participants were 18 years old or older). The study was approved by the Psychology Research Ethics Committee of Leiden University (CEP17–0113/18) and follows the relevant guidelines and regulations to conduct a study with human participants.

Missing data. For the behavioral data, three of the 152 participants (=76 dyads) were excluded because they had missing data for 30 or more out of 60 trials. For the physiological data, the decision to exclude data was based on the manual preprocessing of the data. Either the measurement of the physiological responses was erroneous in at least one of the two participants during the whole session or more than 70% of the responses were missing due to local measurement errors in the data. Based on these criteria, 14 dyads had to be excluded. The reason for such high rates of measurement errors is that we measured multiple physiological responses wirelessly and the recording devices would sometimes lose the signal during the experiment. In addition, the synchrony level was computed on the dyadic level, therefore we needed to exclude both participants if one of them had inaccurate measurements. Two additional dyads were excluded because they did not make any eye-contact during the face-to-face condition trials which was verified by means of eye-tracking glasses worn during the experiment. Ten additional dyads were excluded from only the skin conductance level analysis due to measurement errors. Thus, the heart rate analysis included 60 dyads and the skin conductance level analysis 50 dyads which lies in the upper range of sample sizes across studies investigating physiological synchrony (Palumbo et al., 2017). In addition, 29 single trials for the heart rate data and three single trials for the skin conductance level data were excluded.

Design

The objective of the study was to investigate whether cooperative success could be predicted based on the physiological synchrony between two individuals in a real-life interaction setting. To this end, two participants played a modified iterated Prisoner's Dilemma game while their heart rate and skin conductance responses were measured. A mixed-design study was conducted with one within-dyad (Face manipulation) and one between-dyad (Feedback manipulation) variable. In the latter manipulation, people received auditory feedback about their decision or not. However, this manipulation did not influence cooperation ($\chi^2(1) = 1.29$, $p = .256$), and was not the focus of this article. As such, the Feedback manipulation is not discussed and only included as a control variable in the analyses. Regarding the Face manipulation, participants could either see each other's faces (face-to-face condition) or they could not see each other (face-blocked condition). All dyads played a block of 30 rounds of the game in each condition with the order counterbalanced. The dependent variable was cooperative success which was measured by means of a modified version of the Prisoner's Dilemma game (see below). All dyads played 30 rounds of the game in both conditions with the order counterbalanced. During the whole experiment, participants' heart rate, skin conductance level and eye movements were measured.

Materials

Cooperation game. To measure cooperation, a modified version of the Prisoner's Dilemma game was used. The general idea of the game is that people can choose between two options (cooperate versus defect) that affect both a person's own and the partner's outcome. In particular, if both players cooperate (CC), each player receives more points compared to if both players defect (DD). If one player cooperates and the other defects, the latter receives the highest points possible, while the former receives the lowest points. Hence, the dilemma is to choose between maximizing the own outcome by defecting (which is more advantageous independent of the other player's choice) or maximizing the joint outcome by cooperating (the highest joint outcome is achieved when both players cooperate). In the current study, the idea of the game stayed the same, but people could choose between six instead of two options (option A-F) creating a cooperation scale (Table 1). For this purpose, we built two boards where participants could put a pawn on the response matrix to indicate their response. That response incorporated two choices: (1) the level of willingness to cooperate; moving from the left (option A) to the right (option F) on the x-axis, the willingness to cooperate increased with option A reflecting complete defection and option F reflecting complete cooperation; (2) what the participant thought the other person would choose on that trial; moving from the bottom (option A) to the top (option F) on the y-axis indicates that the participant expected the partner to cooperate more. Hence, the highlighted options in the four corners in Table 1 reflect the payoff structure of a traditional Prisoner's Dilemma game, but the extended matrix shows the innovative structure designed for the current experiment. We recently observed that behavior displayed in this extended version of the Prisoner's Dilemma game positively correlated with the behavior shown in the classical Prisoner's Dilemma game suggesting that they measure similar behavioral tendencies (Behrens & Kret, 2020).

Table 1

Payoff structure of the current study (bold numbers were not highlighted during the experiment)

	F	**4.0–1.0**	3.8–1.4	3.6–1.8	3.4–2.2	3.2–2.6	**3.0–3.0**
	E	3.6–1.2	3.4–1.6	3.2–2.0	3.0–2.4	2.8 -2.8	2.6–3.2
Other	D	3.2–1.4	3.0–1.8	2.8–2.2	2.6–2.6	2.4–3.0	2.2–3.4
	C	2.8–1.6	2.6–2.0	2.4–2.4	2.2–2.8	2.0–3.2	1.8–3.6
	B	2.4–1.8	2.2 -2.2	2.0–2.6	1.8–3.0	1.6–3.4	1.4–3.8
	A	**2.0–2.0**	1.8–2.4	1.6–2.8	1.4–3.2	1.2–3.6	**1.0–4.0**
		A	B	C	D	E	F
				You			

Note. The first number refers to the points earned by "You".

Physiological data acquisition and preparation. Throughout the experiment, four physiological responses were measured on both participants: heart rate (HR), skin conductance level (SCL), zygomaticus major (smiling muscle) and eye movements by means of electrocardiography (ECG), electrodermal activity (EDA), electromyography (EMG), and eye tracking glasses, respectively. The former three were recorded wirelessly with the MP150 BIOPAC data acquisition system and sampled at 2000 Hz. The EMG data contained many artifacts where the source could not be identified and the shape of the artifacts did not allow for clear distinction between artifacts and responses. Therefore, the facial expression data were not included in this paper.

For the analyses, the preprocessed heart rate and skin conductance level measures were down-sampled to 20 Hz. The software AcqKnowledge (AcqKnowledge v. 4.4; BIOPAC Systems Inc.) was used to record and sync the signals from the physiological signals, the event markers from E-Prime which was used to present the instructions and lock the behavioral responses, and markers sent by the eye tracking glasses.

Heart rate. To measure participants' heart rate, electrodes were attached on the left and right side of the abdomen and on thorax below the right collar bone. To process the data, an in-house developed software, PhysioData Toolbox (Sjak-Shie, 2017), was used offline. The signals were band-filtered with a cut-off of 1 Hz and 50 Hz. The R-peaks that were automatically detected by the software were afterwards visually inspected and manually corrected in case of missed or incorrect R-peaks. To still generate a smooth and continuous heart rate signal, inter-beat intervals (IBI) were linearly interpolated in these locations. Participants with less than 30% coverage of the sum of the IBIs relative to the duration of the time signal were excluded. The signal used for the analyses was heart rate which was measured in beats-per-minutes.

Skin conductance level. Two electrodes were attached on the intermediate phalanges of the index and ring finger of the non-dominant hand. To improve the quality of the signal, there was a time interval of around 15 minutes between the attachment of the electrodes and the beginning of the data collection. The skin conductance level measures were low-pass filtered with a cut-off of 5 Hz and subsequently visually inspected for artifacts using the PhysioData Toolbox (Sjak-Shie, 2017).

Eye movements. Participants were wearing Tobii Pro Glasses 2 to track their eye movement and to verify whether they were looking at each other during the face-to-face condition trials. Fixation points were manually coded in Tobii Lab Pro (version 1.64, 2017). Trials in which participants were not at least once looking at the face of the other person were excluded.

Procedure

Before participants came to the lab, they received information about the study and filled out three questionnaire about empathy (Interpersonal Relation Index; IRI; Davis, 1980), social anxiety (Liebowitz Social Anxiety Scale; LSAS; Beard et al., 2011), and social value orientation (SVO; Van Lange et al., 1997). Upon arrival at the lab, participants signed an informed consent in separate rooms and a female researcher attached the electrodes for measuring heart rate, skin conductance level, and facial expressions. Next, participants filled out the Positive And Negative Affect Scale

(PANAS; Watson et al., 1988) and read the instructions for the social dilemma game. Their understanding of the game was checked with multiple choice questions which were discussed in more detailed when answered incorrectly. Afterwards, both participants sat on a table in front of each other with a wooden board between them such that they could only see each other's faces. Finally, the eye tracking glasses were calibrated, the researcher left the room and started the experiment.

After three practice trials (face-to-face condition), participants played the game two times, 30 rounds in the face-to-face and face-blocked condition. The order of starting in one or the other condition was counterbalanced. To block nonverbal communication in the latter condition, a visual cover was placed on top of the wooden board. The sequence of the trial was as follows with auditory instructions given via speakers: First, participants were instructed to look at each other (look at the cross in front of them [drawn on the visual cover] in the face-blocked condition). After four seconds, they were asked to look down and make a decision. When both individuals made their decision, they either heard that they have both made a decision (no feedback condition) or heard how many points each player received based on their choices (feedback condition). As mentioned above, the role of feedback is not discussed here and only added as a control variable in the analyses.

After each session, participants filled out a visual analogue scale (VAS) about their current feelings and experiences. After the second session, participants were separated again in different rooms where they filled out the Desire for Future Interaction scale (DFI; Coyne, 1976) and read the debriefing form. Finally, they were paid and thanked for participation.

Statistical Analysis

During the study, different questionnaires about the participants' characteristics and current mood and experiences were measured as mentioned in the Procedure. These data were not the focus of the current article and are not discussed any further. In Appendix C2, we provide descriptive statistics of these questionnaires (see Table C.S1).

Behavioral data. We hypothesized that face contact would increase the joint outcome, i.e. cooperative success. Specifically, cooperative success was measured as the points both players earned together which ranged from 4.0 to 6.0 points. The Face condition variable was coded 0 = face-blocked condition and 1 = face-to-face condition. We conducted a multilevel linear regression analysis with dyads added as a random intercept effect. The inclusion of the random effect was verified by running an empty model consisting of the random effect only and calculating the intra-class correlation which quantifies how much dependency there is in the data. The significance level of .05 was applied. We report the f^2 as a measure of effect size which is classified as small at a value of 0.02, medium at a value of 0.15, and large at a value of 0.35 (Cohen, 1992; Lorah, 2018). Dyads with more than 50% missing data (more than 30 trials) were excluded.

Physiological data. We conducted a lagged windowed cross correlation analysis to quantify physiological synchrony for the heart rate and skin conductance level measures separately (Boker et al., 2002). The objective of this analysis is to calculate the strength of association between two time series while taking into account the non-stationarity of the signals and the lag between responses, that is, to consider the dynamics of a dyadic interaction. Non-stationarity is accounted for by breaking down the time series into smaller windows (in the current study, the size of the windows is 8 seconds) and calculating the cross-correlation of each segment, allowing the correlation to change throughout the time series. These overlapping window segments are moved along the time series in steps of two seconds starting from the beginning to the end of each Face condition (i.e., moving along the 30 trials per condition). In addition, for each window segment, the signals of the two participants are lagged in relation to one another (in the current study, up to a maximum of four seconds in steps of 100ms) allowing for differences in how fast people react to events and to one another (Boker et al., 2002). For each window segment, the maximum cross-correlation (called "peak cross-correlation") is detected across the different lags and subsequently, these maximum cross-correlations are averaged over all window segments within each Face condition. We therefore obtained a measure of the strength of synchrony for each Face condition per dyad. A more detailed description of the analysis can be found in Appendix C3.

Hypothesis testing. Based on the synchrony measures we conducted two analyses to (i) investigate whether synchrony is influenced by the face contact manipulations, and (ii) test whether the joint outcome can be predicted based on synchrony and on whether people could see each other or not. For both analyses, multilevel linear regression analyses were performed with the same procedure as for the behavioral data. Regarding the first part, Face condition was added as the predictor and the synchrony measure for heart rate and skin conductance level responses as the outcome variables. For the second part, we ran one model with cooperative success as the outcome variable and the main effects and two-way interaction effects of the synchrony measures and Face condition as the predictors. Additionally, we included Feedback (feedback = 1; no feedback = 0) as a control variable. To check that multicollinearity does not confound our results, we calculated the variance inflation factor (Sheather, 2009).

Results

Investigating the joint outcome, the results showed that the interaction effect between skin conductance level synchrony and Face condition significantly predicted cooperative success ($t(2882.33) = 3.24$, $p = .001$, $f^2 = .013$). As depicted in Panel B of Figure 1, the interaction shows a positive slope in the case of face-to-face interactions (beta coefficient = .86) and a flat (very slightly negative) slope in the face-blocked condition (beta coefficient = -.01). Thus, in line with our expectation, there was a positive relation between skin conductance level synchrony and cooperation when people could see each other, but not when they could not see each other. With regard to heart rate synchrony, results yielded no significant interaction effect with Face condition on cooperative success ($t(2861.92) = 0.86$, $p = .389$, $f^2 < .001$). The VIF values were all smaller than 1.75, which is lower than the cut-off value of 5, suggesting that multicollinearity did not influence our results (Sheather, 2009). The full model summary is shown in Table C.S2. In a post-hoc control analysis, we demonstrated that cooperative success could not be significantly explained by the two individuals' independent arousal levels ($ps > .10$) suggesting that the effects of the current study cannot be explained by the mere arousal responses of the two individuals (see Appendix C5 for more details and Table C.S3 for the model summary). The VIF values were all smaller than 3.15 suggesting that multicollinearity did not influence our results (Sheather, 2009).

Other findings underscored the importance of face contact. Regarding the behavioral responses, participants were more successful in cooperating when they faced each other as compared to when they did not ($M_{face} = 0.65$; $M_{face\text{-}blocked} = 0.59$; $t(3629.74) = 7.59$, $p < .001$, $f^2 = .02$; Figure C.S2) [for similar findings, see (Behrens & Kret, 2019; Kiesler et al., 1996)]. With respect to physiological synchrony, as predicted, face-to-face contact amplified the level of synchrony in heart rate and skin conductance level (HR: $t(59) = 3.76$, $p < .001$, $f^2 = .24$; SCL: $t(49) = 2.40$, $p = .020$, $f^2 = .12$). See Panel C of Figure 1 for the corresponding plots. Finally, in a control analysis, we compared the level of synchrony from the original dyads with newly generated, randomly matched dyads. Specifically, participants were paired with another partner than the one they had actually interacted with in the experiment. This analysis verified that the level of synchrony was due to the interaction rather than the experimental set-up of the study. For both heart rate and skin conductance level, the original dyads showed significantly higher Fisher-Z transformed correlations than the newly generated dyads (HR: $t(3622.7) = 8.06$, $p < .001$, $d = .27$; SCL: $t(3015.5) = 4.38$, $p < .001$, $d = .15$). In Appendix C6, we provide a more detailed description of the control analysis.

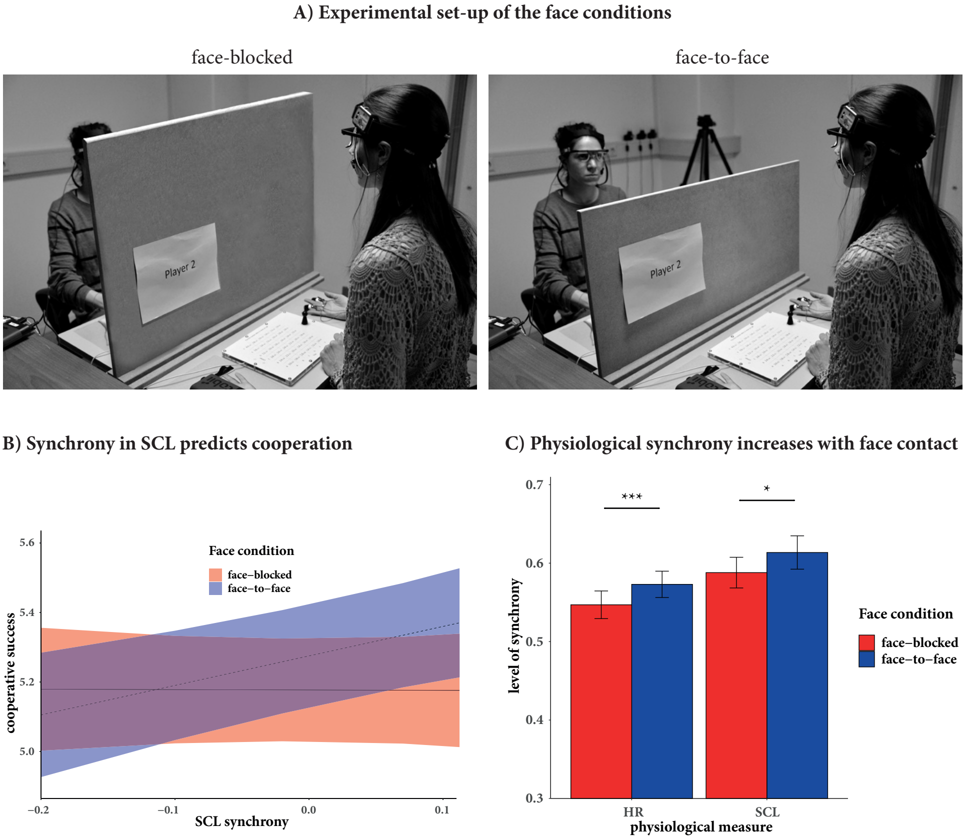


Figure 1. Experimental set-up and results. (A) Dyadic interaction in the face-to-face (left) and face-blocked (right) conditions. Inclusion of the two images was approved for publication by both individuals seen in the pictures. (B) Predicted values of cooperative success based on the interaction effect between synchrony in skin conductance level and Face condition. (C) Mean differences between the face-to-face (blue) and face-blocked condition (red) for heart rate and skin conductance level synchrony. The shaded areas in (B) and error bars in (C) represent 95%-confidence intervals. Physiological synchrony is measured by the mean windowed cross-correlation and is grand-mean centered for the analysis (see Methods for details). Cooperative success is measured by the joint outcome of a dyad per trial in the economic game (range: 4–6 points). HR = Heart rate; SCL = Skin conductance level. * $p < .05$; ** $p < .01$; *** $p < .001$.

Discussion

For thousands of generations, humans have cooperated with others on unprecedented scales, which has been essential for their survival (Bowles & Gintis, 2013). However, as is clear when opening the newspaper, cooperation also often fails. The core question is: what is the mechanism underlying successful cooperation? The current study gives insight into this question by investigating whether cooperative success is related to interaction partners detecting nonverbal signals reflective of physiological arousal, emotionally converging, and fostering mutual understanding and trust. Specifically, the aim of the current study was to investigate the linkage between physiological synchrony and cooperation. For the first time in the literature, we demonstrate that physiological synchrony is associated with cooperative success in real-life interactions. Importantly, this link is especially pronounced when people face each other, that is, when people are able to exchange nonverbal signals. Interestingly, these effects are only evident for skin conductance level synchrony, but not heart rate synchrony. Furthermore, both physiological synchrony and cooperative success are higher when people face each other, and synchrony levels are higher in real compared to artificially-generated dyads. These findings imply that people can detect subtle changes in another person's face, and react to these changes, which is positively associated with cooperation success. Physiological synchrony therefore acts as an unconscious mechanism that affects our behavior and improves the success of close social interactions.

Synchronization is observed on many different levels (Prochazkova & Kret, 2017), in infants (de Klerk, Hamilton, & Southgate, 2018; Fawcett et al., 2016), and in different species (Mancini et al., 2013). Theoretically, it has been proposed to make two interaction partners more similar, aligned, and easier to predict, which is why they are able to cooperate more efficiently (De Waal & Preston, 2017). By manipulating a cooperative versus competitive context, previous research showed increased heart rate synchrony (Mitkidis et al., 2015) and skin conductance synchrony (Vanutelli et al., 2017) in a cooperative compared to a competitive context. The current study builds on this work by showing that when people could decide themselves on a trial-by-trial basis whether they wanted to cooperate or not, these decisions were positively associated with the level of synchrony. This new approach better reflects natural situations where multiple small decisions are taken and thus shows the true relationship between synchronization and cooperative success.

Cooperation carries the risk of exploitation by non-cooperators, therefore being able to detect the integrity of another person's intent is crucial. These intentions are reflected in a variety of behavioral and physiological signals that are visible in the face (Wood et al., 2016). This is supported by the current finding that people were more successful when they played face-to-face compared to when they could not exchange nonverbal signals. We observed a similar effect in a previous, separate study where we used the same set-up, but a new sample of participants played the classical instead of the extended Prisoner's Dilemma game (Behrens & Kret, 2019). Here, we would like to note that in that study we also manipulated whether participants received feedback about each other's decisions or not and, contrarily to the current study, observed a positive effect of feedback. Two methodological differences might have contributed to such discrepancy: (i) the payoff structure was extended from a 2×2 to a 6×6 response matrix, and (ii) while in the previous study, participants first made a decision about whether they wanted to cooperate or not and subsequently indicated what they thought the other person chose; in the current study, the decision and prediction were combined into one response (i.e., participants place a pawn in the payoff matrix where the x-axis represents their own decision and the y-axis indicates their prediction about the other participant's decision). As these two factors are the most prominent changes to our previous study, we believe that they are likely candidates to explain the differences in findings. Coming back to the effect of face contact, people have been shown to be more willing to cooperate when they could talk face-to-face rather than write emails, again supporting the beneficial effect of nonverbal signals (Kiesler et al., 1996). Although the positive effect of face-to-face contact on cooperation is well documented, is it less clear what it is exactly that elicits such effect.

Behavioral signals such as facial expressions and eye gaze can provide valuable information about the intentions of others. however, these signals can in principle be consciously controlled and therefore faked and do not necessarily reflect a person's true intentions (R. H. Frank, 1988; Prochazkova et al., 2019). Physiological responses, on the other hand, are difficult to control and are indicative of social decision-making (Critchley & Harrison, 2013; Damasio et al., 1996). Synchronizing on the physiological level has been proposed to change the way Person A feels about and behaves towards Person B which is consequently reflected in signals visible to Person A (Prochazkova et al., 2019). Likewise, if the explicit signals do show benign intentions, such signals and their mimicry can influence autonomic responses and their synchrony implying a bi-directional interaction between autonomic cues and explicit signals. The influence of visible signals on the synchrony in heart rate and skin conductance level is supported by the current finding that people synchronized more when they interacted face-to-face compared to no face contact; visible signals could be exchanged in the former but not the latter condition. Thus, we argue that cooperation flourishes when people synchronize their autonomic responses because they align emotional states based on genuine emotional cues that are perceived by interaction partners.

The question remains which emotional cues the observer perceives to pick up the changes in heart rate and skin conductance level which can lead to synchrony in these measures. Besides pronounced signals such as facial expressions and eye gaze, other subtle, yet visible cues that are closely linked to changes in arousal are pupil dilation and blushing. It has been demonstrated that people can observe changes in blushing in another person's face and that blushing increases trust, a precursor of cooperation (Dijk et al., 2011; Voncken & Bögels, 2009). In addition, changes in pupil size have been specifically linked to changes in skin conductance level, but not in heart rate (Bradley, Miccoli, Escrig, & Lang, 2008). Again, people have been observed to be sensitive to these pupil size changes in another person (Behrens, Moulder, Boker, & Kret, 2020) and to show more trust towards people with dilated pupils (Kret et al., 2015). These studies suggest that visible physiological responses such as pupil dilation and blushing might constitute suitable candidates for emotional cues that people use to perceive and synchronize changes in arousal as reflected in heart rate and skin conductance level. However, future research is needed to draw strong conclusions about the underlying mechanisms of how physiological synchrony emerges.

Interestingly, we observed that only synchrony in skin conductance level, but not in heart rate affected cooperative success. Such specificity to the purely sympathetic response was not anticipated, but can potentially be explained from hindsight. Sympathetic synchrony has been shown to elicit perceived similarity between interaction partners (Danyluck & Page-Gould, 2019) and perceived similarity has been shown to foster cooperation (Kaufmann, 1967). Furthermore, the sympathetic changes in skin conductance level have been related to (disadvantageous) decision-making and emotion regulation (Crone et al., 2004; Werner, Duschek, & Schandry, 2009). Given the risk of being exploited during cooperation, one might need increased emotion regulation to control the urge to defect in order to successfully cooperate. "Clicking" with another person on the autonomic level might therefore be an essential component of cooperation. These suggestions are, however, speculative and future research is needed to draw strong conclusions about how different responses and their synchrony are integrated in affecting social decision-making.

Two crucial control analyses underscore that synchrony was more than the sum of the arousal responses of two individuals or an artifact of sharing the same environment (e.g., participating in the experiment, receiving the same instructions, etc.). First, it might be argued that if both participants cooperate, their skin conductance level will increase as a reflection of their own decision without any influence of the interaction partner. However, the fact that cooperative success could not be predicted based on participants skin conductance levels alone argues against such interpretation. Second, it might be argued that the increased synchrony levels observed in our study could be the result of a shared environment. However, this argument is confuted by the finding that synchrony was higher for people interacting with each other compared to dyads who shared the same environment, but never actually interacted. This strengthens the notion that synchrony elevated during the actual interaction rather than constitutes an artifact of being in the similar situation. Here, we would like to note that with "the similar situation" we refer to the broader situation such as participating in the same experiment and hearing the same instructions. What is not captured by the two control analyses is the influence of sharing the same specific experience of, for example, making the same decision at the same time. Such shared experience is by definition created when cooperation succeeds, as both individuals need to decide to cooperate. However, the same is true for situations where both participants decide to defect. An important question is therefore whether the link between cooperation and synchrony goes beyond the shared experience of choosing the same response option. In that case, we would expect higher levels of synchrony when both people cooperate compared to when they both defect. We tried to run an additional control analysis to test this hypothesis, however, due to the fact that the data incorporate twice the number of mutual cooperation trials compared to mutual defection, we were not able to perform a valid analysis. Future research is therefore needed to investigate the effect of sharing the same experience on the observed association between synchrony and cooperation. Besides this open question, based on the two control analyses that we did perform, we are confident that the measure of physiological synchrony is the result of a social interaction and that interpersonal rather than intrapersonal processes drive the link with cooperation in the current study.

At this point we would like to clarify that we do not make any claims about the direction of the observed effects. Although some models, such as the Perception Action Model (De Waal & Preston, 2017), suggest that synchrony drives social perception, it could also be a reflection of social processes. To the best of our knowledge, no studies have addressed this question directly.

The design of the current study, that is, people first look at each other before making the decision, is in line with the idea that synchrony drives cooperation. However, previous studies showing that manipulating a cooperative versus competitive context increased synchrony supports the opposite direction. Future studies should scrutinize the causal relation between synchrony and cooperation by manipulating both variables.

The current study has significant implications for studying the intricate dynamics of cooperation. We provide unique evidence that physiological synchrony plays a crucial role in how successful people cooperate. Studying cooperation in real-life interactions unfolded a new layer of communicative processes that is ignored when using computerized, one-person paradigms. This new layer incorporates how two bodies communicate on a subtle level that we are not aware of, yet that is related to how we behave towards other individuals. Shedding light onto what makes cooperation successful in healthy interactions can help us understand situations where human interactions fail. Conflict resolution, whether in a conversation, a company or an international collaboration, is dependent on moment-by-moment cooperative tendencies of its individuals. Such tendencies are by virtue reliant on human's ability to understand each other's emotions and on the capacity to balance their emotions with one another. Applying this to clinical populations, it has been suggested that the lack of interpersonal exchange of non-verbal signals underlies deficits evident in autism, social anxiety, and depression, insights that can advance new therapies in these populations (Galazka et al., 2019; Oberman, Winkielman, & Ramachandran, 2009). Our findings broaden our understanding of the role of synchrony in social behavior and add a hereto forth missing piece to the puzzle of understanding the link between cooperation and nonverbal communication.

Chapter 5

Quantifying physiological synchrony through windowed cross-correlation analysis: Statistical and theoretical considerations

Abstract

Interpersonal synchrony is a widely studied phenomenon. A great challenge is to statistically capture the dynamics of social interactions with fluctuating levels of synchrony and varying delays between responses of individuals. Windowed Cross-Correlation analysis accounts for both characteristics by segmenting the time series into smaller windows and shifting the segments of two interacting individuals away from each other up to a maximum lag. Despite evidence showing that these parameters affect the estimated synchrony level, there is a lack of guidelines on which parameter configurations to use. The current study aimed to close this knowledge gap by comparing the effect of different parameter configurations on two outcome criteria: (1) the ability to distinguish synchrony from pseudosynchrony by means of surrogate data analyses and (2) the sensitivity to detect change in synchrony as measured by the difference between two within-subject conditions. Focusing on physiological synchrony, we performed these analyses on heartrate, skin conductance level, pupil size, and facial expressions data. Results revealed that a range of parameters was able to discriminate synchrony from pseudosynchrony. Window size was more influential than the maximum lag with smaller window sizes showing better discrimination. No clear patterns emerged for the second criterion. Integrating the statistical findings and theoretical considerations regarding the physiological characteristics and biological boundaries of the signals, we provide recommendations for optimizing the parameter settings to the signal of interest.

Based on: Behrens, F., Moulder, R. G., Boker, S. M., & Kret, M. E. (2020). Quantifying Physiological Synchrony through Windowed Cross-Correlation Analysis: Statistical and Theoretical Considerations. *bioRxiv*.

Introduction

During social interactions, humans tend to synchronize on different levels: They mimic postures (Ramseyer & Tschacher, 2011), facial expressions (Chartrand & Bargh, 1999) and align their level of physiological arousal (Feldman, Magori-Cohen, Galili, Singer, & Louzoun, 2011; Levenson & Gottman, 1983; Prochazkova et al., 2018). Although this synchrony comes naturally and without effort, it is a great challenge for social scientists to measure it statistically. The current paper addresses this issue and proposes a Windowed Cross-Correlation (WCC) analysis to investigate the dynamic changes in heartrate, skin conductance level, pupil size, and facial expression. Recommendations are provided on which parameter configurations to use to quantify synchrony of these four responses.

Synchrony is a multifaceted phenomenon evident on the behavioral, physiological, and neural level. Not surprisingly then, the causes and consequences of synchrony have been studied in a broad range of contexts investigating the dynamic nature of social interactions from clinical (Galazka et al., 2019; Wehebrink et al., 2018), developmental (de Klerk et al., 2018; Shih, Quiñones-Camacho, Karan, & Davis, 2019), evolutionary (Mancini et al., 2013; Palagi, Leone, Mancini, & Ferrari, 2009), neural (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Prochazkova et al., 2018), social (Behrens et al., 2019; Tarr, Launay, & Dunbar, 2016b), and cognitive (Kret, Fischer, & De Dreu, 2015; Kret & De Dreu, 2017) perspectives. Such fascination across disciplines has revealed the far-reaching scope of synchrony: it has been demonstrated in different species, it occurs from birth on, and it influences a variety of interpersonal processes such as marital quality, cooperative success between strangers and outcomes of therapeutic interactions (Behrens et al., 2019; Feldman et al., 2011; Kret, Tomonaga, & Matsuzawa, 2014; Levenson & Gottman, 1983; Ramseyer & Tschacher, 2011). Because of these implications and this wide interest, it is of particular importance to establish solid statistical methods to quantify synchrony.

A variety of methods have been proposed in previous literature to quantify synchrony including correlations, regressions, structural equation models and recurrence quantification analyses. These approaches differ in their assumptions, their operationalization of synchrony, and the type of synchrony they measure (for reviews, see Gates & Liu, 2016; McAssey et al., 2013; Schoenherr et al., 2018; Thorson, West, & Mendes, 2017). In the current article, we focus on continuous time series measures in dyads. For this type of data, it is important that the method captures responses that happen "in sync" (e.g., two individuals react simultaneously to an external event), but also responses that occur with a small time delay (e.g., one individual responds to another or at a different pace). Furthermore, the method needs to allow for changes in the level of synchrony as it will vary depending on the events happening in a conversation with moments of stronger and weaker synchrony. Moreover, we focus on the strength rather than the frequency of synchrony. Some methods first specify intervals of synchrony and subsequently compute the frequency of these intervals within a time series (Altmann, 2011). This method is particularly interesting for movement synchrony where people can either move or not. In the current study, on the other hand, we concentrate on physiological measures that constantly change, therefore categorizing intervals into synchronous and non-synchronous segments is difficult. Instead, we are interested in obtaining a global estimate of the strength of synchrony in a conversation. A method that fulfills these different criteria is Windowed Cross-Correlation (WCC) analysis, the focus of the current study (Boker et al., 2002).

WCC analysis offers a neat method to account for dynamic changes in synchrony (Boker et al., 2002). This is achieved by extending a classical cross-correlation estimate by two aspects: windows and lags. Specifically, rather than calculating a correlation coefficient over the whole time series, the signals are broken into smaller overlapping segments or windows. Changes in synchronization can be captured because the degree to which two signals co-vary is estimated for each window separately. The lag is introduced to account for differences in the pace of individuals' responses to one another and to track the follow-lead relationship between them. It might be that at some point Person A responds to Person B and a moment later the pattern is reversed. Consequently, allowing for varying time lags can account for such dynamics. Although this method offers an advanced way to quantify synchrony in naturalistic settings, it does not come without a challenge: parameters need to be specified to tailor the analysis to the signal of interest. In the original paper by Boker and colleagues (2002), the authors advised on parameters using data from motor movements. To this date, there are no guidelines on which parameter settings are most suitable for physiological measures. The goal of the current paper is to close that knowledge gap.

WCC analysis requires the specification of four parameters that tailors the method to the signal of interest: window size, maximum lag, window increment, and lag increment (see Figure 1). Carefully choosing the right parameter settings is crucial, because these settings can substantially affect the outcome of the WCC analysis (Schoenherr et al., 2018). First, the window size determines the number of observations (i.e., data points) in each sliding window across the time series. The window should be small enough to be sensitive to changes in the degree of synchronization and the lead-follow relationship between individuals. Disregarding fluctuations within a large window might undermine the strength of association at certain moments. Here, the biological nature of the signal of interest and its time course are of particular importance. A relatively slow signal such as skin conductance requires a longer window than a fast signal such as facial expressions. Moreover, the window segments need to be small enough such that the assumption of stationarity is likely to hold (Boker et al., 2002). However, if the window size is too small, there are not enough data points left to provide reliable estimates of the relationship between the two segments. Whereas 50–70 values have been proposed as sufficient (Cappella, 1996), more recent work performing Monte-Carlo simulations recommends 65 to 250 values, depending on the strength of the correlation (Schönbrodt & Perugini, 2013). Given the high sampling rates incorporated in many psychophysiological measurement devices, this range should be fairly easy to accomplish, if the window size is not overly small. Decisions on the window size should be based on both statistical and theoretical considerations.

Second, the maximum lag indicates the maximum number of observations one window is shifted in relation to the other window and consequently determines the maximum lag two events are still considered reactions to one another. For example, if the maximum lag is three seconds, then if Person A smiles two seconds later in response to Person B, this would be captured with the three second window. However, if that smile occurs four seconds later, it would not be considered a response to the smile of the other person anymore. If the maximum lag is too long, synchrony might be attributed to two unrelated events. However, if the maximum lag is chosen too small, then important delayed responses between two individuals are missed. Previous research suggests that the maximum lag between responses impacts on synchrony. Specifically, it has been shown that skin conductance responses within, but not beyond seven seconds correlate with the empathetic relationship between counselors and clients (Robinson, Herman, & Kaplan, 1982). The authors did not, however, directly compare whether the shorter latency could predict the relationship better than the longer latency. Additionally, although this study provides an indication that the maximum lag indeed matters, the categorization of latencies (responses between 0 and 7 sec compared to responses between 7 and 40 sec) does not allow for fine-grained conclusions about which maximum lag is optimal. To our knowledge, this is the only study investigating the impact of the maximum lag on synchrony. Thus, a systematic comparison of different maximum lags is needed to make well-informed decisions on this parameter.

Third, the window increment determines the size of the steps (i.e., the number of observations) when moving from one window segment to the next. If the increment is one, then the window is moved by one data point. If the window increment is the same size as the window size or greater, then adjacent windows are non-overlapping. Similarly, the fourth parameter, the lag increment, indicates how big the steps are between time lags. Both increment parameters regulate the resolution in terms of time lag and elapsed time. Ideally, the increment should be kept as small as possible to ensure the best resolution. However, at some point the estimates will stabilize and the limited additional information that can be added by increasing the resolution is not worth the increased computational time. Comparing it to sampling rates, if one aims to measure heartrate changes, a sampling rate of 1000 Hz gives a smooth signal. Increasing the sampling rate to 2000 Hz adds little information because the heartrate does not change this fast resulting in very similar heartrate signals using both sampling rates. Similarly, increasing the resolution of the increment of the moving windows and lags will eventually stabilize around a correlation estimate. The size of the increment will, of course, also depend on the sampling rate which represents the lower bound of possible increments. Therefore, setting the increment parameters for the windows and lags is a question of balancing the benefit of a better resolution and the drawback of increased computational time.

In order to determine the best parameter configurations, we used two criteria. The first criterion was the ability to discriminate synchrony from pseudosynchrony. Pseudosynchrony has been defined as “the amount of apparent and spurious synchrony between two individuals not engaged in information exchange with one another” (Moulder et al., 2018, p. 2). The reason for spurious synchrony is that the signals of interest are restricted in their patterns and how they can behave across contexts. For example, heartrate is constantly changing, decreasing and increasing depending on the person’s inner state and environmental circumstances (i.e., participating in a study with the same procedure across dyads). However, the changes stay in a certain range causing recursiveness and commonality within and between heartrate measures. As a consequence, to determine whether synchrony exists between two time series, the null hypothesis is not zero as for standard null-hypothesis testing, but rather a fundamental value due to the similarities between the biological time series. It is therefore necessary to find an appropriate comparison between the level of synchrony of individuals engaging in an interaction and the level of synchrony that occurs due to the nature of the signals. One way to account for pseudosynchrony is to perform a surrogate data analysis (Moulder et al., 2018). The idea is that the original time series is compared to the same time series where synchrony is destroyed while keeping all other properties constant. Specifically, the synchrony level from the original dyads engaging in an interaction is compared to the synchrony levels from newly generated dyads that never actually interacted. To generate these dyads, the time series from each participant is coupled with every other participant. That way it can be tested whether being in an interaction adds something over and beyond being in the same situation and investigating the same physiological measure. Therefore, being able to distinguish synchrony from pseudosynchrony offers an ideal criterion to test whether some parameter configurations are more sensitive to this distinction.

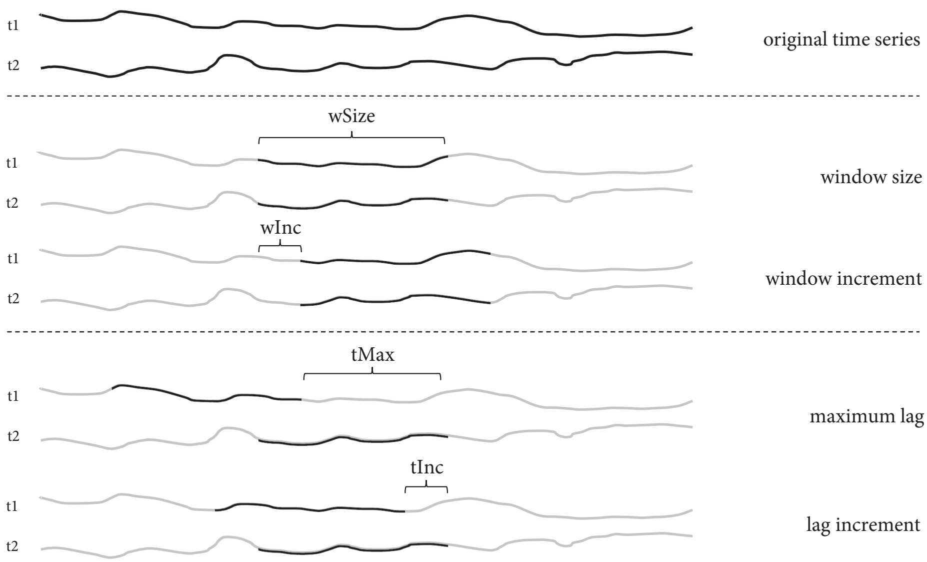


Figure 1. Schematic outline of the four parameters that are specified in the WCC analysis: window size (*wSize*), window increment (*wInc*), maximum lag (*tMax*), and lag increment (*tInc*). The abbreviations *tMax* and *tInc* originate from using "tau" (τ) to refer to the lags in the cross-correlation equation (see Equation 1).

The second criterion that is essential when it comes to research on synchrony is to be able to detect changes in synchrony. To study the underlying mechanisms of synchrony, its boundary conditions and individual differences, researchers are often interested in how synchrony changes in relation to experimental manipulations. For example, in a previous study, we observed that physiological synchrony promoted cooperative success, but only when partners could see each other and not when a cover prevented eye contact (our manipulation) (Behrens et al., 2019).

Another study investigated the effect of emotional salience during storytelling on pupil mimicry and showed that physiological coupling between the speaker and the listener was stronger during emotionally intense moments compared to less salient moments (Kang & Wheatley, 2017). Storytelling is particularly interesting because it is a uniquely human and universal activity creating social bonds between people (Smith et al., 2017). In Kang and Wheatley's (2017) study, listeners watched videos of speakers telling the story and therefore did not engage in an actual conversation. However, direct face-to-face interactions has been shown to affect synchrony levels (Behrens et al., 2019). Therefore, in the current study, two individuals engaged in face-to-face storytelling and completed baseline measures, silent moments of eye-contact. In line with the findings by Kang and Wheatley (2017), we expected higher levels of synchrony when people engaged in storytelling compared to the baseline measure. Ideally, the analysis that measures synchrony is sensitive to detect changes in synchrony between the two (within-subject) conditions.

The aim of the current study was to determine the best parameter configurations for the WCC analysis applied to different common physiological measures. The two criteria we used to decide on these configurations are (i) the ability to distinguish synchrony from pseudosynchrony and (ii) the sensibility to detect *changes* in synchrony (i.e., distinguish between two conditions). The reason to include two criteria is to investigate whether the purpose of the study (i.e., detect synchrony or change in synchrony) influences which parameters configurations are most suitable. We tested these criteria on data from dyadic interactions where two individuals told each other four stories. During the interaction, their heartrate, skin conductance level, pupil size, and contractions of the left zygomaticus major (a muscle associated with smiling) were measured. For a range of window sizes and maximum lags that were tailored to each signal, we calculated a measure of distance for the comparison (i) between the original dyads and newly generated surrogate dyads, and (ii) between intervals of storytelling and baseline measures in the original dyads. The window and lag increments were not systematically compared, but were adjusted as a function of the window size and maximum lag, respectively. Based on the outcome of these comparisons, we provide recommendations on which parameter configurations are best for detecting synchrony and change in synchrony for the four physiological measures. With these recommendations, we hope to help other researchers to make well-informed decisions in applying the WCC analysis and to increase the comparability of findings across studies.

Method

Participants

In total, 34 same-sex dyads participated in the study of which six dyads had to be excluded due to technical problems (dyads included in analysis: Female = 22 [78%]; M_{age} = 22.79; SD_{age} = 3.23; Dutch = 17 [30%]). Participants were recruited via the Leiden University online recruitment system, flyers distributed around the university building, and through personal contacts. In the latter case, participants were tested by a researcher they did not know. Individuals had normal or corrected-to-normal vision wearing contact lenses. Glasses were not compatible with the eye-tracking glasses worn during the experiment. The duration of the study was about one hour and participants received two course credits or 6€, and chocolate for compensation. The study was approved by the local Psychology Ethics Committee of Leiden University (CEP19–0313/208).

Design

The design of the study is outlined in Figure 2. The study consisted of two parts. First, participants completed a breathing exercise where they were instructed to look at each other and synchronize their breathing. Second, participants engaged in storytelling with each participant telling a neutral and a positive story while the other participant was listening. Thus, participants told four stories in total with story 1 and story 3 always being told by participant 1 (sitting on the left side) and story 2 and story 4 being told by participant 2 (sitting on the right side). Story 1 & story 2 and story 3 & story 4 were of the same valence, with the order of starting with the neutral or positive story being counterbalanced between dyads. The breathing and storytelling parts were both preceded by a 2-min baseline measure where participants were instructed to relax and look at each other. After the second baseline measure and after each story, participants filled out the Positive And Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) to measure their current affect. Also, they rated each story with regard to its valence and intensity on a scale from 0 to 10. The PANAS and the story ratings are not discussed any further, but the descriptive statistics are provided in Appendix D2 (see Table D.S1).

Procedure

Upon arrival at the lab, participants were separated, received information about the study, and gave informed consent for participation. Afterwards, electrodes were attached to the torso, fingers, and face as preparation for the measurement of ECG, EDA, and EMG activity, respectively. Specifically, three electrodes were attached on the left and right side of the abdomen and on the thorax below the right collar bone to measure heartrate; two electrodes were attached to the non-dominant hand on the intermediate phalanges of the index and ring finger to measure skin conductance level; and three electrodes were attached to the left face on the zygomaticus major and behind the ear to measure facial expressions. The MP160 BIOPAC data acquisition system was used to record these measures at a sampling rate of 2000 Hz. After the preparation, participants filled out the Interpersonal Reactivity Inventory (IRI; Davis, 1980) and the Five Facet Mindfulness Questionnaire (FFMQ; Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006) online. The descriptive statistics of both questionnaires can be found in Appendix D2 (see Table D.S1). Next, participants were seated on the same table and participants were asked to wear the eye-tracking device Tobii Pro Glasses 2 which were subsequently calibrated. Afterwards, the experimenters left the room and started the recordings of the physiological measures and the pre-recorded instructions that were provided via speakers. The experiment started with a 2-min baseline measure where participants were instructed to relax and look at each other (Baseline 1). Afterwards, the breathing exercise started where participants were again asked to look at each other, but this time synchronize their breathing for two minutes (not discussed in the current study). After this first part of the experiment, participants had time to think of a neutral and positive personal story. When they were ready to begin, another 2-min baseline (Baseline 2) was taken and participants filled out the first PANAS which was provided on the table. Then Participant 1 (the individual at the left side of the table) started with the first story. Participants were instructed to talk for at least three minutes till they heard a beep and were requested to finish up. Afterwards, both participants filled out the PANAS and rated the story based on its valence and intensity on a scale between 0 and 10. Then the next story began. Participants took turns in telling them and filled out the PANAS and the rating after each story. At the end, participants put all filled out papers in an envelope, read the debriefing, and the experimenters removed the electrodes. Finally, individuals were paid and thanked for participation.

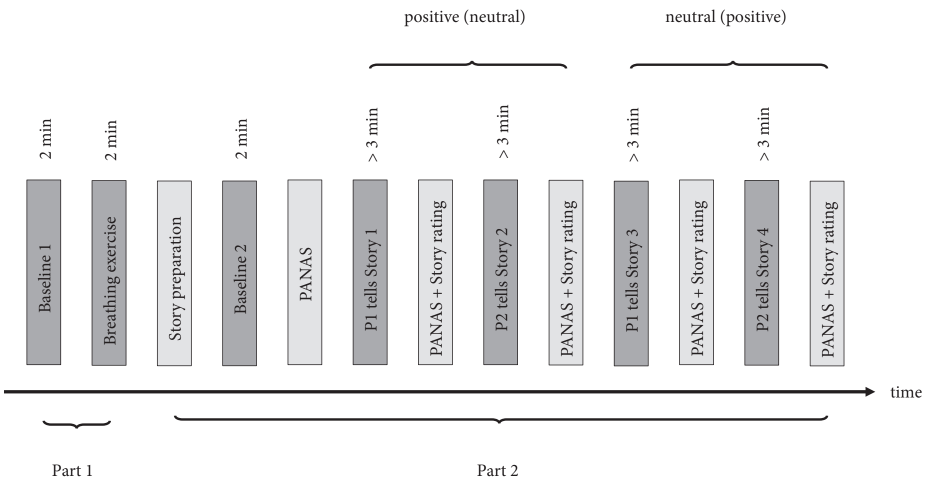


Figure 2. The time course of the study. The study was divided into two parts: breathing exercise (Part 1) and storytelling (Part 2). During the dark grey epochs, people interacted with each other; during the light grey epochs, they prepared the storytelling and filled out questionnaires; P1/P2 = Participant 1 and 2; PANAS = Positive And Negative Affect Schedule; Story 1 & 2 and Story 3 & 4 were of the same valence (positive or neutral); the order of starting with the positive or neutral story was counter balanced between dyads.

Preprocessing of the physiological measures

The physiological measures were pre-processed offline with the PhysioData Toolbox (Sjak-Shie, 2017). The heartrate data were preprocessed applying a band-filter between 1Hz and 50Hz. R-peaks were detected and transformed to inter-beat intervals (IBI) and subsequently to heartrate (bpm) values. The skin conductance signal was low-pass filtered with a cut-off of 5Hz. The EMG signal was preprocessed with a low-pass FIR filter of 28Hz and a high-pass FIR filter of 500Hz and a Notch-filter of 50Hz. The rectified signal was subsequently smoothed with a Boxcar filter of 100ms. The pupil size data were preprocessed in multiple stages according to recommended guidelines described elsewhere (Kret & Sjak-Shie, 2018). After applying the filters, each signal was visually inspected and if necessary, manually corrected. If missing or incorrect intervals were manually detected, the signals were linearly interpolated. Finally, all signals were down-sampled to 20Hz.

Windowed Cross-Correlation analysis

Two challenges in analyzing physiological responses between two individuals include i) to statistically represent the dynamics of an interaction and ii) to quantify the associated patterns that might vary in the strength of association and the timing of the responses. Windowed Cross-Correlation (WCC) analysis offers a method that addresses both challenges. Specifically, the two time series are broken into smaller, overlapping windows before the correlation is estimated for each window. This way, the strength of association can vary between these windows accounting for the non-stationarity of the signals. The overlap between windows assures that strong synchronization that occurs at the edge of non-overlapping adjacent segments is not missed. Additionally, for each window, the two segments are lagged away from each other up to a maximum lag such that the segment of either participant 1 or participant 2 precedes the other participant's segment in time. This way the method accounts for the (varying) delay between two responses. This generates a result matrix r with correlations for the different segments and time lags defined as

$$\boldsymbol{r(Wx, Wy) = \frac{1}{T_W}\sum_{t=1}^{T_w}\frac{(Wx_t - \overline{Wx})(Wy_t - \overline{Wy})}{sd(Wx)sd(Wy)}} \tag{1}$$

Where $\boldsymbol{T_w}$ is the total amount of observations (i.e., data points) in each window $\boldsymbol{Wx}$ and $\boldsymbol{Wy}$ consisting of observations $\boldsymbol{Wx_t}$ and $\boldsymbol{Wy_t}$ where $\boldsymbol{t} \in \{1, \ldots, \boldsymbol{T_w}\}$, Wx and Wy are the means of the observations in each window, and $\boldsymbol{sd(Wx)}$ and $\boldsymbol{sd(Wy)}$ the standard deviations of each window. In the result matrix, each row represents one window, while each column represents one lag. Because the first window needs to lag segments up to the maximum lag and because the window includes more than one data point, the number of rows is given by (*N* – *wSize* – *tMax*) / *wInc*. Dividing by *wInc* accounts for how many observations are skipped between one window and the next one. For example, if the window increment is one, then the number of rows of the result matrix will be equal to the number of observations of the time series (after accounting for the window size and maximum lag as just described). But if the increment is 10, then the steps are bigger between the windows, reducing the number of segments needed to cover the whole time series and therefore decreasing the number of rows in the result matrix. The number of columns in the result matrix is $(tMax * 2)/tInc + 1$ because the segments are shifted such that first Participant 1 and then Participant 2 precedes the other participant up to the maximum lag (i.e., twice the *tMax*). The *tInc* accounts for the size of the steps between two lags. The extra column (+1) represents the case where the lag is zero.

Peak picking. Following the WCC analysis, Boker et al. (2002) developed the so-called peak-picking algorithm where the maximum correlation across different lags is determined for each window (i.e., the maximum correlation per row of the result matrix). The maximum correlation should be preceded and succeeded by lower correlation values. For example, if Participant 1 synchronizes with Participant 2 at a lag of 1 second, then the correlation should be highest (i.e., peak) at that time lag and the correlation should be lower at both lag .5 and 1.5 seconds. This "peak" criterion is implemented to ensure that individuals indeed react to one another. If both individuals did nothing, they both would show more or less flat lines in their physiological responses and the correlation between their signals would be high for all lags. Requiring a peak in the correlation across lags prevents such events from being termed "synchrony". The peak-picking algorithm outputs a matrix with the maximum ("peak") correlation and its corresponding time lag for each window. In a last step, a summary statistic is computed by calculating the mean of the maximum correlations. This measure provides an indication of the overall level of synchrony between the two time series.

Choosing values for parameter configurations

5

As mentioned above, there are four parameters that need to be specified: window size, window increment, maximum lag, and lag increment. The window size (*wSize*) determines how long each window is, the window increment (*wInc*) indicates the size of the steps between two adjacent (overlapping) windows, the maximum lag (*tMax*) regulates how far the segments of the two time series are shifted away from each other, and the lag increment (*tInc*) determines the size of the steps with which the segments are shifted.

To choose the range of values we considered for the window size and maximum lag parameters, we employed a bottom-up approach by running preliminary WCC analyses on the whole time series (including all data of the study). Inspecting the result matrix plots, we examined the patterns seen in these plots. Examples of a "good" and "bad" parameter configurations are shown in Figure 3. Good parameter configurations show sharp contrasts between regions of high and low synchrony. The bad choices show a more smoothed image and thus less contrast between these regions, making differences more difficult to detect.

With regard to the maximum lag, we examined the plots inspecting whether the peak correlations fell within the range of lags or whether they fell outside the plots (not shown in Figure 3). For reasons of simplicity, the range of maximum lags was equal to the range of window sizes. In addition to the visual inspection, we ensured that the range of parameters included the parameters previously used in the literature. Finally, the minimum value for the window size was set to 3 sec to include at least 60 data points (20Hz sampling rate) per window size which is in line with previous guidelines for reliably estimating correlation coefficients (Schoeneberger, 2016). The window size and maximum lag parameters chosen for each physiological measure are listed in Table 1. For the window and lag increment parameters, we used 1/10th of the window size and the maximum lag, respectively.

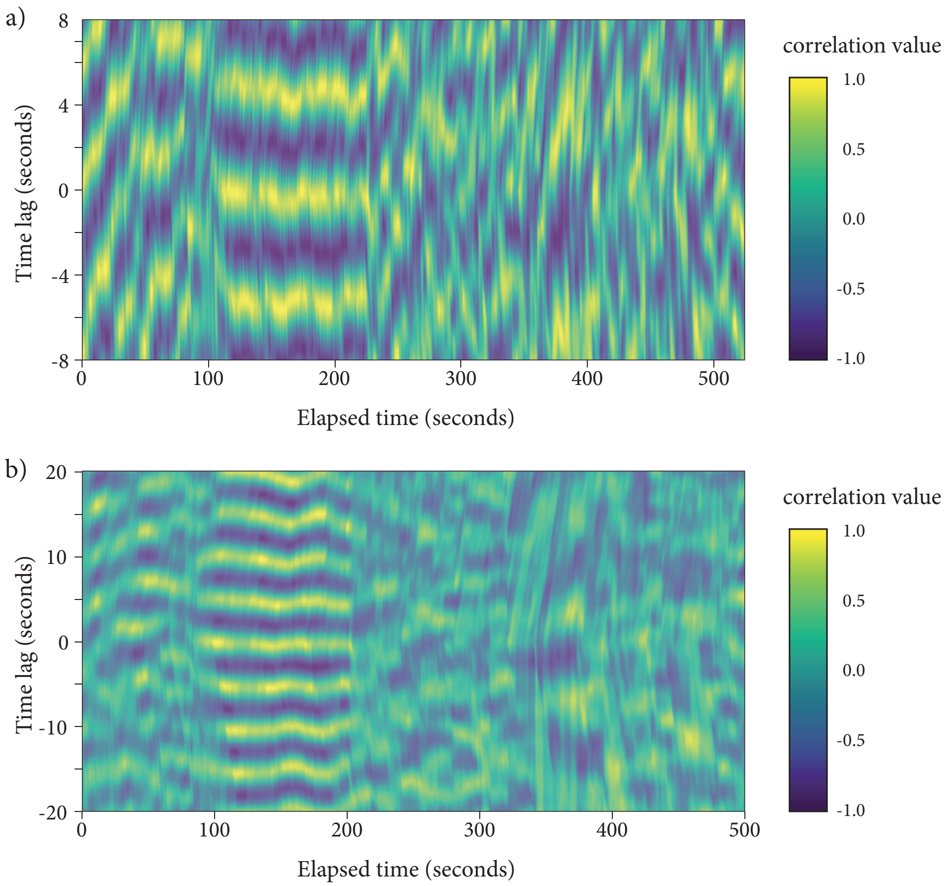


Figure 3. Examples of WCC analysis plots using heartrate data and a window size and a maximum lag of (a) 8 sec and (b) 20 sec, representing a “good” and “bad” example of parameter settings, respectively. Between around 100 and 200 seconds, people engage in a breathing exercise where they breathe synchronously which is reflected in the steadily high correlations around the time lag of zero.

Table 1
Window size and maximum lag parameters used for each physiological measure

Signal	Window size	Maximum lag
Heartrate	4 – 12 sec in steps of ½ sec	4 – 12 sec in steps of ½ sec
Skin conductance level	5 – 25 sec in steps of 1 sec	5 – 25 sec in steps of 1 sec
Pupil size	3 – 9 sec in steps of ½ sec	3 – 9 sec in steps of ½ sec
Facial expression	3 – 9 sec in steps of ½ sec	3 – 9 sec in steps of ½ sec

Note. The window and lag increments were equal to 1/10th of the window size and the maximum lag, respectively.

Choosing the best parameter settings

We conducted the WCC and peak-picking analyses for all combinations of the window size and maximum lag parameters with their corresponding increments as described in the previous section. For each parameter configuration, we calculated the mean peak correlation across window segments per dyad as the measure of synchrony. To determine the best parameter configurations for each physiological measure we used two criteria: (i) the ability to discriminate synchrony from pseudosynchrony, and (ii) the ability to detect change in synchrony. For the first criterion, we compared the original dyads consisting of the individuals who in fact interacted with each other during the experiment with the surrogate dyads consisting of all possible combinations of pairing individuals who did not interact during the experiment. If being in the specific social interaction evoked synchrony above and beyond the synchrony evoked by the fact of being in *any* actual interaction, synchrony levels are expected to be higher in the original compared to the surrogate dyads. Therefore, we calculated the mean peak correlation for both the original and the surrogate dyads and investigated whether specific parameter configurations were more sensitive to detect the difference between synchrony (original dyads) and pseudosynchrony (surrogate dyads). Sensitivity was quantified by the t-statistics of an independent t-test between the mean estimates of the two groups. A positive t-statistic indicates that the true dyads show higher levels of synchrony than the surrogate dyads. To determine the best parameter configuration, we located which configuration generated the largest t-statistic and inspected the pattern in changes of t-statistics across parameter configurations. Note that we used the t-statistic as a measure of distance between the two group means without running hypothesis testing (i.e., decide on whether the distance is significant or not). We therefore interpret the t-statistics in relative rather than absolute terms and do not draw any conclusions about whether the differences reveal significant results or not. The analysis was conducted with the data from the first baseline measure (see Figure 1). To investigate whether the results of this analysis would replicate, we additionally conducted the same analysis again with data from the second baseline measure.

For the second criterion, that is, which parameter configurations are most sensitive to detect change in synchrony, we concentrated on the original dyads and investigated which parameter configurations generated the biggest difference between two conditions of the experiment. We used the t-statistic based on a paired t-test as a measure of distance between the mean estimates of the two conditions. A positive t-statistic indicates higher levels of synchrony during storytelling than baseline. Similar to the first criterion, we identified the largest t-statistic and inspected the pattern in changes of t-statistics across parameter configurations. We also ran the analysis twice. First, we compared story 1 and story 3 with the two baseline measures. Second, we compared story 2 and story 4 with the two baseline measures (see Appendix D1 for the reasoning behind the choice of these comparisons). To keep the length of the stories equal, we only used the first three minutes of each story. This way, both comparisons included a positive and a neutral story (a preliminary analysis yielded no differences between the positive and negative stories). The only difference was that in the first analysis, Participant 1 told the stories and in the replication analysis, Participant 2 told the stories. Being Participant 1 or 2 was based on the participant number and therefore should not have had any systematic impact on the synchrony level between the two individuals. Therefore, we could investigate whether specific parameter configurations were more sensitive than others to detect differences in synchrony levels when people just looked at each other compared to when they engaged in storytelling.

Results

Synchrony versus pseudosynchrony

Heartrate. There was a range of positive t-statistics indicating that multiple parameter configurations could differentiate between the original and the surrogate dyads (Figure 4a). The best discrimination (maximum t-statistic = 28.32) was evident for the smallest window size (4 sec) and a maximum lag of 7.5 sec (the most yellow combination in Figure 5a). When mapping the t-statistics distribution onto the parameter configuration space, a clear pattern emerged: the smaller the window size, the larger the t-statistics. This pattern was evident by the gradual changes in coloring from blue to yellow in Figure 5a when moving down the y-axis (i.e., moving from large to small window sizes). When the window size became too large, the synchrony level dropped in the original dyads such that it became lower than the synchrony level apparent in the surrogate dyads (especially, when the maximum lag was small; dark blue coloring in Figure 5a).

The maximum lag was less influential on differentiating between original and surrogate dyads than the window size, yet not trivial. The maximum t-statistic was evident for a maximum lag of 7.5 sec. The optimal maximum lag was therefore around twice the optimal window size (4 sec). Increasing or decreasing the maximum lag reduced the sensitivity to distinguish between the original and surrogate dyads as indicated by less yellow colors when moving left or right on the x-axis in Figure 5a. The replication analysis using data from the second baseline measure revealed similar results to the primary analysis and is depicted in Figure D.S1a-D.S2a. The maximum t-statistic (35.23) for a window size of 4 sec was replicated. The maximum lag differed slightly by 1.5 sec showing the highest t-statistic at 9 sec. However, the pattern was comparable with smaller window sizes and maximum lags around twice the window sizes yielding the largest difference between the original and surrogate dyads. In conclusion, if the aim of the study is to verify whether synchrony evolved as a result of interpersonal processes during a conversation above and beyond the shared environment of two participant, the range of parameters able to detect that difference is rather wide. In general, we recommend using a small window size for heartrate synchrony. Regarding the maximum lag, the choice of parameters is less influential, however, we recommend using a maximum lag that is around twice the window size.

Skin conductance level. As with heartrate synchrony, there was a range of parameter configurations with a positive t-statistic that was sensitive to distinguish the original from the surrogate dyads (see Figure 4b). The largest t-statistic of 37.71 was observed for a window size of 6 sec and a maximum lag of 24 sec (see Figure 5b). Similar to the heartrate data, the smaller the window size, the greater the distance in estimated means between the original and surrogate dyads. Also, the outcome flipped with higher synchrony levels for the surrogate compared to the original dyads when the window size was too large paired with a small maximum lag. In contrast to heartrate, the discriminative ability steadily increased when the small window size was combined with an increasingly larger maximum lag (around four times the window size). In the replication analysis, the same pattern emerged as in the primary analysis: the greatest discrimination was seen for a small window size and a large maximum lag (see Figure D.S1b-D.S2b). The largest t-statistic (48.71) was observed for a window size of 5 sec and a maximum lag of 21 sec. Again, when the window size became too large paired with smaller maximum lags, the analysis would estimate higher synchrony levels for the surrogate compared to the original dyads. Based on these results, we recommend using a small window size and a large maximum lag that is around four times the window size.

5

Pupil size. The number of positive t-statistics depicted in Figure 4c indicates that there was a range of parameter configurations that could differentiate synchrony from pseudosynchrony. The maximum t-statistic of 16.12 was associated with a window size of 3 sec and a maximum lag of 9 sec. The general pattern as for the other measures was observed: the smaller the window size, the greater the difference between the original and surrogate dyads (see Figure 5c). Again, when the window size became too large, the estimates of synchrony level would become larger for the surrogate compared to the original dyads. With respect to the maximum lag, it was less influential than the window size, but showed a slight tendency to larger maximum lags. A similar pattern was observed for the replication analysis with a maximum t-statistic (18.04) evident for a window size of 3 sec and a maximum lag of 6.5 sec (see Figure D.S1c-D.S2c). In conclusion, smaller window sizes were more sensitive to distinguishing synchrony from pseudosynchrony in pupil size data. The maximum lag did not have as much of an impact, but should be set to two to three times the window size.

Facial expression. All t-statistics were positive indicating that the level of synchrony was higher for the original compared to the surrogate dyads for all parameter configurations. However, compared to the other three measures, the distribution showed less variance with t-statistics ranging from 1.68 to 5.14 (see Figure 4d). The latter was observed for a window size of 3 sec and a maximum lag of 5 sec. As shown in Figure 5d, the same pattern as for the other three measures emerged: the smaller the window size, the better the original dyads could be distinguished from the surrogate dyads. Furthermore, the maximum lag did not have a great impact on the discriminative ability, but the largest t-statistic was observed at almost twice the window size (5 sec). For the replication analysis, a similar pattern was observed (see Figure D.S1d-D.S2d) with a slightly wider range of t-statistics (maximum = 7.05; minimum = -.16). The maximum t-statistic was associated with a window size of 3 sec and a maximum lag of 8.5 sec. Again, the smaller the window size, the greater the difference between synchrony and pseudosynchrony with limited impact of the maximum lag. In conclusion, we recommend using a small window size and a maximum lag that is two to three times the window size.

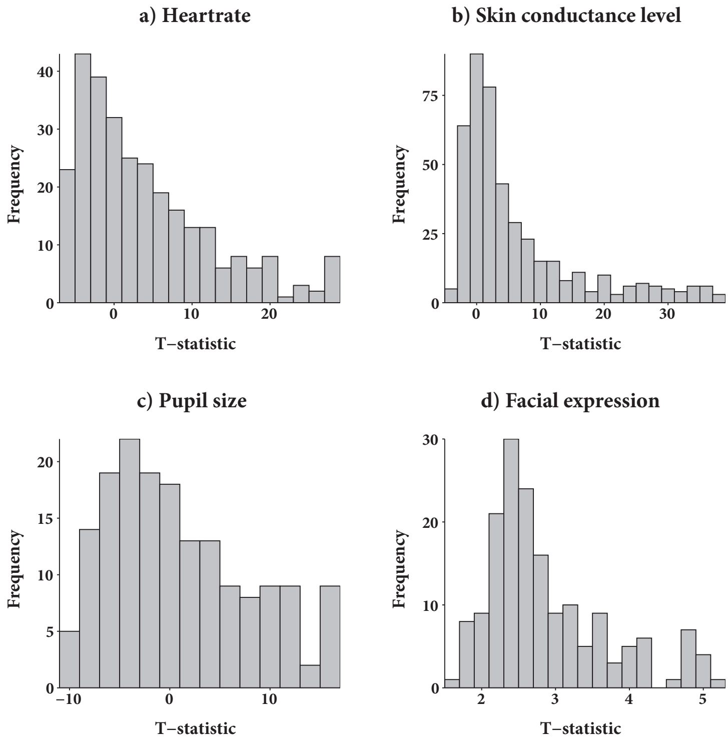


Figure 4. Distribution of t-statistics of the comparison between the original and surrogate dyads for each physiological measure. A positive value indicates higher synchrony level in the original compared to the surrogate dyads. Each data point represents one parameter configuration. For the analyses, data from the first baseline measure were used.

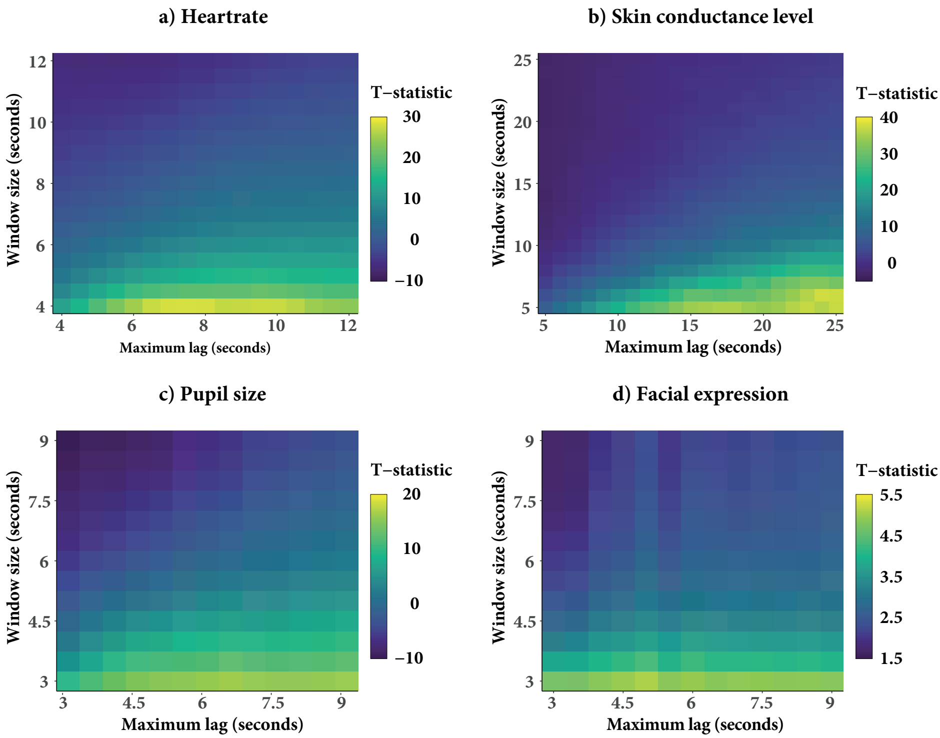


Figure 5. Distribution of the t-statistics of the comparison between the originate and surrogate dyads for all parameter configurations and each physiological measure. The color coding runs from the lowest (blue) to the highest (yellow) t-statistic. A positive t-statistic indicates that the original dyads showed higher synchrony levels than the surrogate dyads. The more yellow, the better the discrimination between the original and surrogate dyads. Data from the first baseline measure were used. Notice that the scaling of the axes and the color coding are adjusted to each physiological measure to increase comparability between parameters.

Change in synchrony

Heartrate. The largest absolute t-statistic was negative indicating that synchrony levels were higher during baseline compared to during storytelling (see Figure 6a). The highest absolute t-statistic of 4.86 was observed when the window size was set to 4 sec. Similar to the first comparison analysis, smaller window sizes could discriminate the two conditions better than large window sizes (see Figure 7a). Also, the maximum lag was less influential than the window size parameter, but the best outcome was observed for the smallest maximum lag of 4 sec. The absolute t-statistic steadily decreased with increasing maximum lags. For the replication analysis, the results were similar to the primary analysis, with smaller window sizes showing the greatest discriminative power between the conditions (see Figure D.S3a-D.S4a). Specifically, the largest absolute t-statistic was again observed for a window size of 4 sec. The maximum lag increased from 4 to 7 sec in the replication analysis with only slight changes across maximum lags. Therefore, based on both analyses the conclusion is: if the aim is to distinguish synchrony levels in heartrate responses between two (within-subject) conditions, the smaller the window size, the better. The maximum lag is less influential, but should be equal to or twice the window size.

Skin conductance level. All t-statistics were negative indicating that the level of synchrony was higher during the baseline measures compared to during storytelling (see Figure 6b). The highest absolute t-statistic of 4.37 was observed for a window size of 18 sec and a maximum lag of 25 sec. Interestingly, the previous pattern of smaller window sizes showing greater t-statistics was not evident (see Figure 7b). In fact, although there seemed to be a weak tendency for absolute t-statistics to become larger with larger window sizes and larger maximum lags, the pattern was rather weak. In addition, the difference between t-statistics was small ranging from -1.61 to -4.37. For the replication study, the range was also rather narrow from -.19 to -2.56 (see Figure D.S3b-D.S4b). The maximum absolute t-statistic was observed for a window size of 5 sec and a maximum lag of 12 sec, deviating substantially from the primary analysis. Although the general pattern (i.e., the smaller window size, the higher the t-statistic) was observed to a stronger degree compared to the primary analysis, it was still weak. In conclusion, given the lack of clear patterns in the parameter configuration space and considerable discrepancies in the results between the primary and replication analyses, we cannot draw strong conclusions about which parameter configuration is best to distinguish between two conditions when looking at skin conductance level synchrony.

Pupil size. For this measure, the parameter configurations strongly influenced whether synchrony levels were higher during baseline or storytelling (see Figure 7c). Generally, if both the window size and the maximum lag were small, synchrony levels were higher during storytelling; if the window size and maximum lag were large, synchrony levels were higher during the baseline measures. Specifically, the largest positive t-statistic of 1.72 (storytelling showed more synchrony) was observed for a window size of 3.5 sec and a maximum lag of 3 sec. However, the largest absolute t-statistic of 2.07 (baseline showed more synchrony) was associated with a window size of 8.5 sec and a maximum lag of 9 sec. A similar, but weaker pattern was evident for the replication analysis (see Figure D.S3c-D.S4c). The window sizes and maximum lags associated with the largest (absolute) t-statistic were the same as for the primary analysis. Given the ambiguity across parameters, we refrain from providing any recommendations about the best parameter configurations when the aim is to detect change in pupil size synchrony between conditions and instead caution that parameter choices can have a large effect on the outcome of this type of study.

Facial expressions. All t-statistics were positive indicating that the level of synchrony was higher during storytelling compared to baseline (see Figure 6d). The largest t-statistic of 3.99 was evident for a window size of 3 sec (see Figure 7d). Albeit weak, the general pattern emerged with larger t-statistics being associated with smaller windows sizes. The maximum lag associated with the biggest difference between conditions was 3.5 sec, but the differences across lags were trivial. The replication analysis revealed similar results with the largest t-statistic (4.53) observed for a window size of 3 sec (see Figure D.S3d-D.S4d). The maximum lag of 9 sec deviated from the primary analysis, however, the differences across the maximum lags were again rather small. To conclude, if the aim is to detect change in synchrony between two conditions in facial expressions, then the window size should be set to a small value. The effect of the maximum lag was negligible, however, to be consistent with the other measures, we recommend a maximum lag twice the window size.

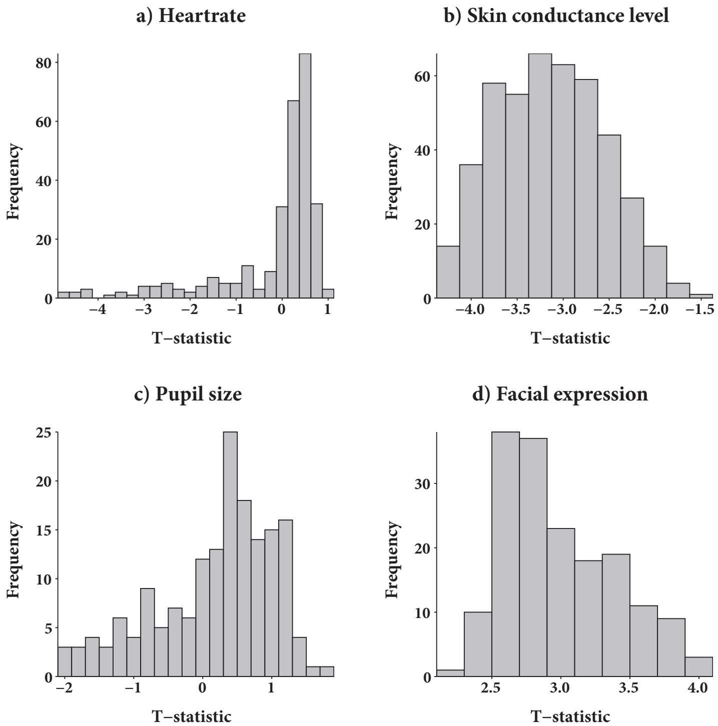


Figure 6. Distribution of t-statistics of the comparison between storytelling and baseline for each physiological measure. A positive value indicates higher synchrony levels during storytelling compared to baseline. Each data point represents one parameter configuration.

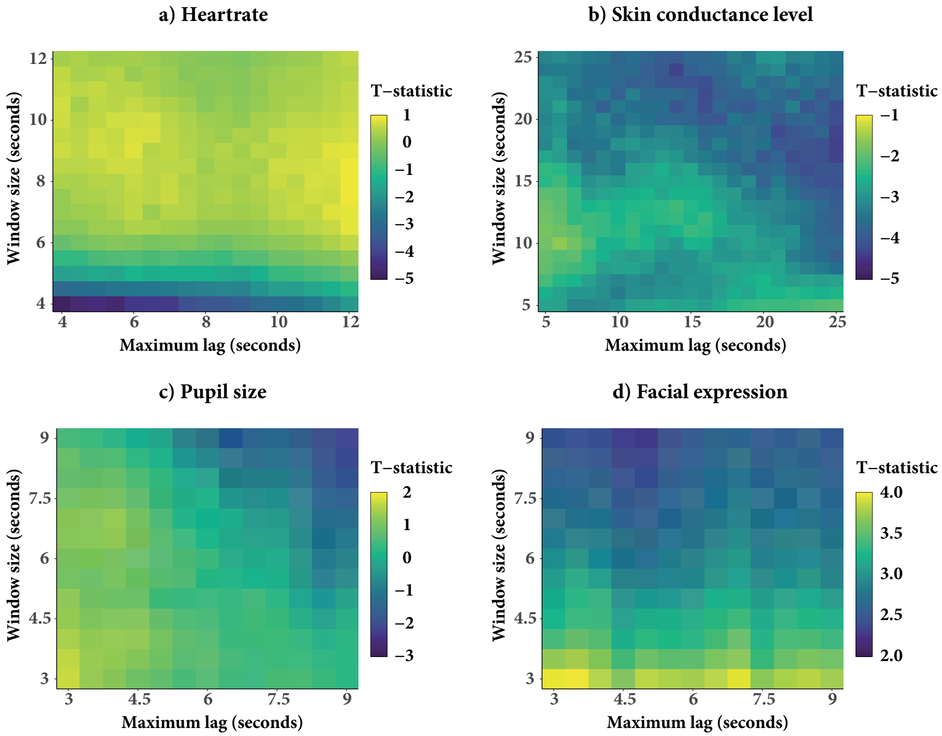


Figure 7. Distribution of t-statistics of the comparison between storytelling and baseline of all parameter configurations for each physiological measure. The color coding runs from the lowest (blue) to the highest (yellow) t-statistic. A positive t-statistic indicates that the level of synchrony was higher during storytelling than during baseline. Analysis was based on data from both baseline measures and the first and third stories. Notice that the scaling of the axes and the color coding are adjusted to each physiological measure to increase comparability between parameters. Also, the highest t-statistic was not always the highest absolute value with the latter value being discussed in the result section. However, the general idea of greater (absolute) t-statistics indicating better discrimination between the two conditions remains.

Discussion

The phenomenon that people synchronize each other's emotional expressions and physiological states has intrigued researchers in many different disciplines. Studying this phenomenon comes with the challenge of statistically capturing the dynamic nature of a social interaction. Over the years, several methods have been developed that address this dynamic to different degrees and in different ways. One such method is the Windowed Cross-Correlation analysis (Boker et al., 2002). It accounts for changes in the strength of synchrony throughout an interaction and in the different paces in which people respond. The method requires researchers to specify parameters that allow us to tailor the method to the signal of interest. Albeit increasing the method's flexibility, there is a lack of guidelines on which parameters to use for which signal, which can have an impact on the outcome of the analysis. The aim of the current study was to statistically determine the most suitable parameter settings applied to four different physiological measures: heartrate, skin conductance level, pupil size, and activity of the zygomaticus major muscle (associated with smiling). To that end, we systematically investigated the influence of a range of parameter configurations on two criteria: i) the ability to distinguish synchrony from pseudosynchrony, and ii) the sensitivity to detect change in synchrony (i.e., distinguish two within-subject conditions).

Regarding the first criterion, the results revealed that a wide range of parameter configurations could distinguish between the original dyads and dyads that participated in the study, but never engaged in an actual interaction (i.e., surrogate dyads). Additionally, a general pattern across all physiological measures emerged: the smaller the window size, the better the discriminative ability tear apart the original dyads from the surrogate dyads. In contrast, if the window size became too large, the estimated level in true dyads dropped to such an extent that it became lower than the synchrony level estimated in the surrogate dyads. With respect to the second parameter, the maximum lag was generally larger than the corresponding window size. How much larger differed between physiological measures: the optimal maximum lag was two, four, and two to three times the window size for heartrate, skin conductance level, pupil size and facial expressions, respectively.

Regarding the second criterion, that is, the sensitivity to detect change in synchrony, the results were less clear cut. Here, we compared two baseline measures where people looked at each other in silent with periods where participants engaged in storytelling. For heartrate and facial expressions, the general pattern was visible with better discriminative ability between storytelling and baseline with smaller window sizes. For facial expressions, this pattern was, however, weak at best. Interestingly, differences across measures emerged of whether synchrony levels were higher during storytelling or baseline. (Almost) all parameter configurations for the heartrate and skin conductance level measures indicated higher levels of synchrony during baseline. For pupil size, both patterns emerged with small window sizes and maximum lags showing more synchrony during storytelling, whereas large window sizes and maximum lags revealed more synchrony during baseline. For facial expressions, storytelling showed higher levels of synchrony for all parameter configurations. Other than these differences between measures, the range of t-statistics within each measure was considerably smaller than for the surrogate data analysis, suggesting less sensitivity to parameter choice. In the following, we will discuss our findings in depth and integrate them with theoretical considerations. In Table 2, we summarize the global recommendations on determining the parameter configurations. We hope that these guidelines provide researchers with information that assist them to make well-informed decisions about the optimal parameters for their WCC analysis.

Table 2

Summary of global recommendations per parameter of the WCC analysis

Parameter	Recommendations
Window size	• Lower bound: large enough to capture meaningful information and variance within the signal of interest • Upper bound: the response duration of the signal of interest; assumption of stationarity is met
Maximum lag	• Lower bound: at least as long as the window size • Upper bound: at most twice as long as the window size
Window and lag increment	• Lower bound: 1 datapoint • Upper bound: same as the window size / maximum lag • Balance computational time and resolution: 1•5% of the window size / maximum lag

Synchrony versus pseudosynchrony–Window size

We observed that a wide range of window sizes was able to distinguish between synchrony and pseudosynchrony. However, in general smaller window sizes performed better. However, if the window size became too large, synchrony levels dropped to the extent that the levels became lower for the true dyads than the surrogate dyads. How can this general pattern across measures be explained? To understand it, let us quickly recapture the purpose of the surrogate data analysis. As introduced above, the aim is to destroy any synchrony that is the result of interpersonal processes while preserving all other statistical properties by generating new dyads that participated in the study, but never actually interacted. This way we *know* that the null hypothesis that there is no synchrony between participants is true. As the null hypothesis will be true independent of the parameter configurations, the distribution of cross-correlations stays constant across all parameters. In contrast, for the original dyads, synchrony does emerge, which we expected based on prior research. During a dynamic interaction, there are moments when dyads are in sync, but also out of sync (Boker & Rotondo, 2002). If the window size becomes too large, both moments of synchrony and "anti-synchrony" are likely to be included into one window segment, substantially reducing the strength of synchrony. This causes a drop in overall synchrony that can be lower than in the surrogate dyads with no synchrony at all (i.e., no synchrony and no "anti-synchrony"). On the other hand, decreasing the window size decreases the variance within a window causing the overall synchrony level to increase. Specifically, as seen in Equation 1, the cross-correlation is calculated by dividing the distance between each datapoint and the mean of the window segment by its standard deviation. The smaller the window size, the less change for variation to occur within a window (i.e., the smaller the standard deviation), which causes the correlation to increase. Thus, while the distribution of correlation estimates stays constant for the surrogate dyads, the estimates for the original dyads increase with smaller window sizes. Consequently, the distance between the mean of these two groups becomes increasingly larger, causing the general pattern we see across the physiological measures. This pattern is therefore an intrinsic characteristic of the way the cross-correlation is estimated applying to all types of time series.

The question then arises whether steadily decreasing the window size will also steadily increase the ability to distinguish synchrony from pseudosynchrony. The short answer is no. Imagine the extreme case, where the window size consists of two datapoints. These two datapoints hold very little information and would only allow possible correlation values of -1 and 1. This reduces the sensitivity for measuring synchrony and therefore for distinguishing synchrony from pseudosynchrony. Consequently, somewhere between a window size containing two datapoints and the smallest window sizes we examined, there will be a turning point, where the two types of dyads will become distinguishable.

Although statistically possible, making the window size as small as possible (but above the turning point) is not advisable for two reasons: (1) a sufficient number of data points are needed to reliably estimate correlation coefficients (Schönbrodt & Perugini, 2013), and (2) the window should capture a meaningful response. As outlined earlier, in order to reliably estimate a correlation coefficient, a recent study showed that 65 to 250 datapoints are necessary depending on the strength of the correlation. With a sampling rate of 20Hz across all measures, we therefore used a window size of at least 3 seconds (60 datapoints). If researchers want to further decrease the window size, they should increase their sampling rate accordingly. With that said, a window size must include responses constricted by a meaningful upper and lower bound.

In the current study, we narrowed the possible values for the window size parameter by showing a range of parameters that qualify as potentially suitable parameters. How can researchers choose between these options? To answer this question, let us go back to the aim of cutting the time series into segments in the first place, namely, reducing the non-stationarity in the signals. A stationary signal has constant statistical properties with, among others, a constant mean and standard deviation within that signal. In a dynamic interaction, the strength of synchrony (our statistical property of interest) will vary between moments of strong and weak synchrony. The window size needs to be small enough such that the synchrony level stays constant within that window. Determining how small the window must be, depends on the nature of the signal and is contained by an upper and lower bound.

Imagine smiles of two interaction partners are coded during a conversation such that a person either smiles or not. If the two participants smile at the same time for the same duration, there will be perfect synchrony between them for the entire duration of the two smiles (given an appropriate correlation measure for categorical variables). In this case, the window size could be as large as the duration of the smile because the level of synchrony is constant during that interval. However, if the smiling response occurs in a real conversation and is measured continuously reflected in the activity of the zygomaticus major (as in the current study), there are variations in latency, magnitude and duration of the smiles within and between individuals. In this case, the level of synchrony is likely to change even within the window that would have been categorized as a "smile" in the artificial categorical scenario just described. For example, one person might show a long, pronounced smile, while the other person might smile later and and for a shorter length of time. Then the synchrony would only occur during the short time where both people smile simultaneously. Therefore, the window size should be smaller than the duration of a "typical" smile to capture these variations. More specifically, we recommend a window size that is at most half the response duration, such that at least two estimates of the level of synchrony will be computed for that response capturing changes in synchrony that are twice the speed of the overall response.

Other than the upper bound for window size being smaller than the response duration of interest, there is a lower bound as well. In particular, the window size should be large enough to capture meaningful variations within a response. For example, if the signal of interest is skin conductance level, a window size of 1 second would contain straight lines in most windows. This produces extreme cross-correlations without capturing meaningful changes in the signal. On the other hand, applying the same window size to facial expressions might be considered a medium to large window given that a smile has been shown to last 500ms to 4 seconds (Frank, Ekman, & Friesen, 1993). Both the upper and lower bound therefore determine the potential values for the window size.

When talking about "the duration of a response" we realize that this can be difficult to define as physiological measures show great variations within and between individuals. In the section "physiological boundaries" below, we provide an overview of the "typical" temporal characteristics of each physiological measure realizing that this overview is far from being exhaustive. It is beyond the scope of the current paper to provide concrete guidelines for this matter and it is up to the researcher to decide on which responses she is interested in. As the most suitable (range of) window size(s) likely differs across situations and conditions, choosing a window size should be seen as a hypothesis that is tested, namely, that responses synchronize that are equal to or longer than the window size chosen. Although faster responses are still included in the window segments, they are likely to be averaged out and changes in the faster responses will be reduced.

If the researcher has no strong a priori hypotheses, multiple window sizes can be tested across a range of possible values taking a data-driven bottom-up approach to determine the best parameter choice. Obviously, it is not realistic that researchers perform such elaborated analyses as in the current study, however, comparing two to three potential values can shed light on the rate at which synchrony occurs in a particular context. Of course, it is unlikely that people will synchronize on one specific response duration only, so one would expect more similar results for window sizes closer together. However, referring to "skin conductance synchrony" based on one parameter setting is likely an overgeneralization and needs more detailed investigation.

To conclude, the results of the current study indicate that a range of window sizes is able to detect synchrony that occurs as a result of interpersonal processes with a preference for shorter window sizes. From a theoretical perspective, the range of potential window sizes is contained by (i) an upper bound defined by the length of the duration of the responses under investigation and (ii) a lower bound defined by sufficient variation within the window. Rather than searching for that one most suitable parameter for each physiological measure, choosing a window size should be seen as a hypothesis being tested. Importantly, researchers need to be specific about what aspect of a signal they investigate which should be clearly stated in both their hypothesis and conclusions.

Synchrony versus pseudosynchrony: maximum lag

Our results revealed that the maximum lag was less influential than the window size, yet not trivial. In contrast to the window size, the optimal maximum lag differed between the physiological measures. For heartrate, pupil size, and facial expression, the optimal maximum lag was around 5–10 seconds. Skin conductance level deviated from the other measures with the optimal parameter being around 20–25 seconds. This is consistent with the fact that skin conductance level is a considerably slower signal compared to the others. However, it contrasts the finding reported by Robinson and colleagues (1982) who showed that synchrony in skin conductance response within, but not outside the range of 7 sec was associated with the empathetic relationship between therapists and clients. Such discrepancy can be explained by the fact that while these authors concentrated on the phasic response, we have focused on the tonic, slower responses. This underscores the importance of being specific about what aspect of a signal the researcher is interested in and shows again the importance of the theoretical consideration for choosing the parameter configurations for the WCC analysis. In the following, we aim to provide the reader with a sense of how the maximum lag influences the analysis.

Essentially, the maximum lag indicates how far responses between participants can lie apart that can still be considered a response to one another. Thus, similar to choosing the window size, the maximum lag considerably depends on the interest of the researcher. Given their link, it seems reasonable to choose the maximum lag in relation to the window size. In line with our findings, we recommend using a maximum lag that is equal to or twice the size of the window. For simplicity, let us assume that stationarity is met for the length of a full response, all response cycles have the same length and the window segments start at the beginning of a new response. If the maximum lag is the same length as the window size, the window segment of Person A will be shifted away from the segment of Person B (and vice versa) until the two segments succeed one another with no overlap in time. When Person A, now later in time, shows a response, then Person B reacts *right after* the response of Person A. Thus, over the range of all considered lags, synchrony can happen between people being in sync (lag = 0) and people responding to each other in direct succession. In a similar vein, setting the maximum lag to twice the window size means that there can be up to a full response duration between the responses of the interacting individuals. For example, imagine the measure of interest is facial activity and the window size is 2 seconds. If the maximum lag is 4 seconds, then two smiles that occur simultaneously up to the situation that they are 4 seconds apart from each other are considered synchronized responses. The latter situation seems still reasonable in the context of a real conversation, yet on the upper limit. Therefore, expanding the maximum lag to 6 seconds likely increases the chance of linking two unrelated events to one another. The decision to set the maximum lag equal to or twice the window size depends on the researcher's preference of what she considers reasonable in the context of interest. In a controlled environment with straightforward, stereotypical displays of emotions, a person should react rapidly and a smaller maximum lag might be sufficient. However, in a natural interaction where ambiguous expressions and verbal conversations require more elaborated processing, a response might take longer and therefore a larger maximum lag might be appropriate. In addition, the latency of a response itself is important, especially in relation to the response duration. For example, if a response is expected to be initiated rapidly, but last relatively long, a small maximum lag is sufficient. However, if the latency of a response is long and the duration of the response short, then a longer maximum lag is required. In sum, as a general rule of thumb we recommend a maximum lag of at least equal to and at most twice the size of the window size.

We would like to point out that the results considerably deviated for the skin conductance level. While the three other measures showed the largest discrimination between synchrony and pseudosynchrony for a maximum lag that was about twice the window size, for skin conductance level it was four times (around 20–25 seconds). As described above, this is consistent with the fact that skin conductance level is a substantially slower response compared to the other signals. One might therefore argue that the associated window size of 5 seconds might be too small capturing mostly responses with little meaningful variation. Increasing the window size might therefore be advisable, which then align with our recommendation of choosing a maximum lag that is at most twice the window size. In conclusion, our findings revealed that from a statistical point of view, the maximum lag is less influential than the window size. Nevertheless, this does not safeguard the researcher from using any parameter and tailoring it to the nature of the signal of interest is essential. Here, we have provided more information about the meaning of the maximum lag and recommended to specify the maximum lag equal to or twice the window size.

Window and lag increments

In the current study, we have adjusted the increments such that the windows and lags moved by 10% of the window size and maximum lag, respectively. This was a choice of practicality, reducing the computational time in light of the huge amount of analyses run while keeping the resolution sufficiently high. As already mentioned at the beginning of the paper, both parameters determine the resolution of the changes occurring between window segments and lags. Ideally, the increments should be as small as possible (i.e., 1 data point). However, the increments heavily influence the computational time which is why researchers might want to increase these parameters. Nevertheless, the increments should never be set higher than the window size and maximum lag themselves. In case the lag increment is equal to the maximum lag, three situations are analyzed: people responding in sync (lag = 0), Person A responds to Person B with a delay of the maximum lag, and Person B responds to Person A with a delay of the maximum lag. For the window size, two adjacent segments would not overlap. If the increment would be greater than the window size, there would be a gap between two adjacent segments. This is problematic because moments of synchrony occurring during that gap are missed. Generally, unless researchers are specifically interested in one particular time lag, we recommend keeping the increment small in relation to the window size. Using the 10%-rule of thumb was fine for the current study, however, we needed to account for an enormous amount of analyses. We believe that reducing the percentage to 1 to 5% offers a good balance between analysis sensitivity and computational time.

Change in synchrony

Besides the ability to detect synchrony, we also investigated the effect of the parameter configurations on the sensitivity to detect change in synchrony. The results were less clear-cut here. While for heartrate and facial expression synchrony, the general pattern of smaller window sizes increasing the discrimination ability was (weakly) apparent, it was not observed for skin conductance level and pupil size. Additionally, the primary and replication analyses sometimes showed large deviations. For example, for the skin conductance level, the greatest differences between conditions was apparent for a window size of 5 seconds in the primary analysis and 18 seconds in the replication analysis. On top of that, there were differences between measures and parameters in whether synchrony levels were higher during storytelling or baseline. In particular, for heartrate and skin conductance synchrony, (almost) all parameter configurations suggested higher levels of synchrony during baseline, whereas the reverse was evident for facial expressions. Such discrepancy might be explained by the function of the signal. Facial expressions are displayed for communicative purposes which is particularly important during storytelling where people react to one another more than during silent moments of eye-contact during baseline. While arousal levels also play a crucial role during conversations, during the baseline measure people could concentrate on each other nonverbally and were not “disturbed” by engaging in conversations, overall leading to higher synchrony during baseline. On top of that, the baseline condition consisted of two baselines measures with the second being preceded by the breathing exercise where participants were instructed to synchronize their breathing. This might have influenced the second baseline measure leading to higher overall synchrony levels. In general, given the lack of clear patterns and inconsistencies between the primary and replication analyses, we refrain from giving recommendations for parameter configurations based on these results.

The inconclusiveness of the results might be attributed to two potential explanations: (1) the difference between the two conditions was negligible and the sensitivity to detect such small differences was barely affected by the parameters; (2) there were differences between the two conditions, but the method was not sensitive to detect them. In support of the first explanation, in two previous studies, we have used parameters included in the current analysis with which we were able to detect differences in within-subject conditions and could link it to interpersonal outcomes (Behrens et al., 2019; Prochazkova, Sjak-Shie, Behrens, Lindh, & Kret, 2019). The method therefore has been shown to be sensitive in other contexts. However, future studies are needed to address this question using either simulated data or more extreme conditions where the difference is more pronounced and possible differences between parameters are more likely to show.

Physiological boundaries

Every physiological measure has its temporal characteristics and we will give a short overview for each of the four measures considered in the current study. First, the time course of heartrate is controlled by several physiological processes that can operate to varying degrees depending on the situation and psychological process of interest. Generally, parasympathetic nervous system activity slows the heartrate down, while sympathetic activity increases heartrate. While parasympathetic activity is associated with fast changes in heartrate and is predominantly related to breathing (changes within millisecond to second range), sympathetic activity takes more time to show and is attributed to changes in arousal levels (changes within second range) (Berntson, Cacioppo, & Quigley, 1991). The pace of the heart can change on a beat-by-beat interval and the peak of heartrate acceleration has been shown to occur within the first 4 seconds (Critchley et al., 2005). The duration of a response to an external event (e.g., stimulus presentation) usually takes around 5–8 seconds, although full recovery from stressful events can take several minutes (Berntson et al., 1991; Bradley, Codispoti, Cuthbert, & Lang, 2001; Bradley et al., 2008; McAssey et al., 2013).

Skin conductance measures are indications of arousal resulting from sympathetic nervous system activity and are divided into tonic (skin conductance level) and phasic (skin conductance response) components. The tonic activity consists of gradual changes over time that vary considerably between and within individuals. It decreases during rest and increases more quickly in response to new events (Dawson et al., 2000). The phasic activity, the high-frequency component of the skin conductance measure, is faster than the tonic response and reflects responses directly linked to an external or internal event. The latency of a response is usually between 1–3 seconds and the time to reach the peak amplitude takes between 1–4 seconds. The duration of a full response from stimulus presentation to 50% recovery of the amplitude after the response peak varies between 4 to 16 seconds (Dawson et al., 2000). This is consistent with a power spectral analysis showing that the sympathetic activity is reflected in frequencies between .045–.25 Hz, corresponding to response durations of 22 and 4 seconds, respectively (Posada-Quintero et al., 2016).

Changes in pupil size can result from both parasympathetic and sympathetic activity. Specifically, pupil constriction is mainly controlled by parasympathetic activity, whereas pupil dilation is an indication of sympathetic activity. Pupil size changes in response to light are rapid showing a constriction response 200ms after turning on the light (Mathôt, 2018). Pupil size changes in response to psychosensory processes are slower and vary with, among others, mental effort and saliency of the stimulus (for a review, see Beatty & Lucero-Wagoner, 2000). The typical response is characterized by an initial constriction in response to the stimulus and subsequently, a more pronounced dilation of the pupil with a peak after 2 to 3 seconds and a total response duration of 4 to 6 seconds (Bradley et al., 2008; Kret et al., 2015; Oliva & Anikin, 2018).

Facial expressions consist of changes in facial muscles such as the zygomaticus major, associated with smiling, and the corrugator supercilii, associated with frowning. The duration of a facial response depends on whether researchers investigate subtle, rapid changes or full-blown smiles in a natural conversation. For example, a facial response can occur as fast as 200–300ms in response to stereotypical, controlled stimuli (Achaibou, Pourtois, Schwartz, & Vuilleumier, 2008). In a more natural setting, Frank, Ekman, and Friesen (1993) showed that a Duchenne smile of enjoyment lasts between 500ms to 4 seconds. Accordingly, response windows used in previous studies greatly differ ranging from 1.4 – 5 seconds after stimulus onset showing static images (Achaibou et al., 2008; Drimalla, Landwehr, Hess, & Dziobek, 2019; Lang, Greenwald, Bradley, & Hamm, 1993), to 15 second intervals investigating facial activity in real-life interactions (Hess & Bourgeois, 2010). This section gives a brief glimpse into the temporal characteristics of the physiological measures we have focused on in this paper. However, we would like to emphasize that this overview is far from being exhaustive and researchers need more elaborated knowledge to make well-informed decisions about the signal of interest.

Limitations

There are a few limitations that we would like to point out. First, in the current study we concentrated on the window size and maximum lag parameters, while setting the window and lag increments to 10%. A systematic investigation of the effect of changing these parameters is needed. As mentioned earlier, estimations of the level of synchrony will stabilize with smaller increments such that decreasing the increments even further will add little information at the cost of extra computational time. Although we propose to set the increments to 1– 5% of the window size and maximum lag, this suggestion is not based on statistical analyses and future research is needed to determine the optimal balance between sensitivity and computational time. Second, the general guidelines we propose in Table 2 may not be generalizable to other measures of synchrony and may not be applicable to other biological time series. Researchers should therefore be careful with making any inferences about other statistical analyses and time series than used in the current study. Third, all data come from a single study and is subject to method variation. To reduce such variation, we ran all analyses twice with different data from the same study. However, this does not address method variations that are the result of the study itself and future studies should replicate our findings in a different dataset. Finally, we changed the original plan for the comparison of time intervals as outlined in Appendix D1. A more tailored study design may have observed more specific results, in particular with the regard to the sensitivity to detect change in synchrony.

Future directions and conclusions

The most important lesson the current study teaches us is that researchers need to be precise in what they (aim to) investigate as defined by the parameters specified in the analyses. In the current study, dyads synchronized on a range of response windows. However, this might not always be true, especially, if the aim is to link it to specific psychological processes that might be influenced by only particular physiological processes. Future studies are therefore needed that make more refined distinctions of which components of a particular physiological signal is involved in the process of interest and how the different components interact. This will facilitate making well-informed decisions about the response windows and shed more light on the biological underpinnings of psychological processes.

Before making well-informed decisions on the parameter configurations *within* a particular method, it is important to realize what the differences are *between* methods. WCC analysis is one of many possible methods and each method has its strengths and weaknesses. While one method might be appropriate for one, it might not be for another depending on, among others, the type of data (e.g., continuous or categorical measures) and the measure of interest (e.g., strengths versus frequency of synchrony; global versus time-sensitive measure of synchrony) (Gates & Liu, 2016; Schoenherr et al., 2018). For example, we chose to treat facial muscle activity as a continuous measure. However, researchers might also be interested in investigating concrete events of, for example, smiling and its synchrony in a conversation. Here, the analysis developed by Altmann (2011) might be appropriate where time series are first categorized into intervals of synchrony and intervals of no synchrony before measures of the strength and frequency of synchrony are computed. Despite using the same data, the outcomes can be somewhat different as demonstrated by Schoenherr and her colleagues (2018). Performing an exploratory factor analysis on seven linear time series analyses and different outcome variables (among others the WCC analysis), they reported that all these methods measure the overall phenomenon of synchrony, but could be categorized into three correlated, yet distinct facets of synchrony: the strength of synchrony of the total interaction, the strength of synchrony during synchronization intervals, and the frequency of synchrony (Schoenherr et al., 2018). The WCC analysis as performed in the current study reflects the first facet. Researchers should therefore refine which facet of synchrony they are interested in and choose the appropriate methods accordingly.

The aim of the current study was to optimize the parameters for the WCC analysis from a statistical point of view. The initial idea was to provide researchers with concrete guidelines on which specific parameters would be most appropriate for the four physiological measures. However, the results show that when the aim is to detect synchrony, the parameters follow a general pattern that is not specific to the signal of interest, but rather a result of intrinsic characteristics of how the cross-correlation is calculated. That does not mean that the parameters should not be tailored to the signal of interest. Instead, theoretical considerations should be integrated with the findings observed in the current study. Here, there is no one-fits-all solution, which might not be surprising given that we aim to capture a highly complex process. The current study narrows down the range of possible parameters and we provide guidelines on how to tailor the parameters further to the interest of the researcher. Being specific and transparent about these choices will increase the comparability across studies and add more and more pieces to the puzzle of understanding the phenomenon of synchrony.

Chapter 6

General discussion

Human cooperation is an incredible phenomenon that comes in many forms including two to thousands of individuals, from a single occasion to multiple across decades, from carrying a heavy wardrobe up the stairs to international collaborations, and from infancy to adulthood. How can cooperation succeed on so many different levels? The current thesis investigates the role of nonverbal communication in successful cooperation and how such link can be most reliably tested in the lab. Specifically, in two chapters I demonstrate how face-to-face interactions can boost cooperation between strangers. Additionally, I place the tasks I use to measure cooperation into the broader context of prosocial behavior and zoom in on how to statistically capture the strength of synchrony between interaction partners. In the following, I start by summarizing the main findings from the studies presented in Chapter 2 to 5 and subsequently discuss their theoretical and methodological implications before I close up with concluding remarks.

Summary of the main findings

In **Chapter 2**, I presented a study where I aimed to investigate whether economic games and more naturalistic, interactive games measure similar behavioral tendencies, that is, prosociality. To test this, 74 participants played six different prosocial behavior tasks in a within-subject design. In dyads, participants played three variants of a social dilemma game: the original and an extended version of the Prisoner's Dilemma game (extending the response options from two to six), and a Rope-Pull game (based on the same principles as the classical Prisoner's Dilemma game, but requiring less cognitive abilities). Additionally, participants played an Egg-Hunt game (people could help another participant to collect more Eastern eggs), the Hidden-Profile game (participants needed to exchange information to make the correct decision), and a Tangram game (participants completed puzzles together). A Principle Component Analysis showed that behavior across these tasks was best captured by two components termed "social dilemma games" and "naturalistic games". The three variants of the social dilemma game loaded positively on the first component. Behavior in these games was distinct from behavior in the more naturalistic games. This finding demonstrates that the behavioral consistency observed in previous studies using economic games does not generalize to more ecologically valid games. The Egg-Hunt game loaded positively and the Hidden-Profile task loaded negatively on the second component. In other words, the more eggs a person collected for their partner during the Egg-Hunt game, the less information s/he shared during the Hidden-Profile game. One possible explanation for this finding is that people shared their information in an attempt to convince their partner of their own opinion, reflecting selfish behavior and therefore showing a negative correlation with helping behavior in the Egg Hunt game. Regarding the "social dilemma games" component, the finding that the original and extended Prisoner's Dilemma game clustered together was particularly relevant as I used these games to measure cooperation in Chapter 3 and 4, respectively. The fact that they clustered together supports the idea that they tap into similar behavioral tendencies. This is crucial because the aim of extending the classical Prisoner's Dilemma game was to preserve the same principle of the game and measure similar behavioral tendencies, while only changing the scale of the measure.

In **Chapter 3**, I investigated the beneficial effect of nonverbal communication on cooperation and how it is affected by past experience with the interaction partner. To that end, two participants ($N = 116$) played multiple rounds of the Prisoner's Dilemma game. During some rounds, participants could see each other, while during other rounds a visual cover between them prevented nonverbal communication. Additionally, dyads received either no, correct, or random (50% incorrect) feedback about each other's decisions after each round. Our results revealed that these two sources of information operated independently: face-to-face contact promoted cooperation and knowing the partner's previous decisions increased cooperation, but these two types of information would not strengthen or weaken their individual effects on cooperation. Even if the participants heard that their partner acted selfishly in the previous round, the beneficial effect from seeing each other still worked. In other words, face-to-face contact had a robust effect on cooperation, even if a person could not verify that the interaction partner reciprocated the cooperative act or if the other person had been selfish before.

In order to explain why face-to-face interaction has such positive effects on cooperative behavior, I investigated physiological synchrony as a potential underlying mechanism in **Chapter 4**. To investigate the involvement of this mechanism, I tested 152 participants in a similar set-up as the previous study with two differences: (1) throughout the experiment, participants' physiological responses were measured (i.e., heartrate and skin conductance level) and (2) the payoff structure of the Prisoner's Dilemma game changed from a 2×2 to a 6×6 structure (both versions of the game measure similar behavioral tendencies as shown in Chapter 2). Results showed that physiological synchrony emerged during social interactions and that it was related to the cooperative success of dyads. Interestingly, although physiological synchrony developed for both heartrate and skin conductance level, only the latter showed the predicted beneficial effect on cooperation. This indicates that aligning each other's responses on the sympathetic level was particularly important for how well two individuals worked together.

In the last study presented in **Chapter 5**, I dove deeper into the methodological challenges of properly quantifying interpersonal synchrony. I refined an existing analysis that was developed by Boker et al. (2002) and that I applied in Chapter 4, by tailoring its parameter settings to four physiological responses in a new dataset ($N = 68$). Specifically, I systematically investigated the effects of a range of parameters on how well the method could discriminate real dyads from people who were artificially paired into dyads but never actually interacted (i.e., surrogate dyads). I observed that the choice of parameters influenced the ability to distinguish the original dyads from the surrogate dyads and that similar patterns in parameters emerged between signals pinpointing to an intrinsic characteristic of the method. Nonetheless, the best choice of parameters differed between physiological measures as they should be tailored to the time course of the (component of the) signal of interest. Based on these considerations, I developed guidelines for each physiological measure to increase the comparability of research findings across studies.

Up until now, this dissertation has followed the order of studies starting from a board perspective and then zooming in more and more on specific aspects of the previous study. With such an approach I aimed to answer the questions of how nonverbal communication between individuals affects cooperative success and how it can be best investigated in the lab while safeguarding ecological validity. While Chapters 3 and 4 answer the first question with mainly theoretical implications, Chapters 2 and 5 address the second question which concerns predominantly methodological challenges. In the following, I will discuss the implications of the main findings in light of this distinction: theoretical and methodological implications.

Theoretical implications

The main theoretical question of the current thesis concerns how nonverbal communication affects cooperative success. In the first study (Chapter 3), I replicated previous studies showing that access to nonverbal communication is beneficial for cooperation. I extended such finding by demonstrating the robustness of the effect: face-to-face contact boosts cooperation to a similar extent if a person has past experiences with another person, or if she has no explicit knowledge. Of course, receiving information about another person still strongly influences people's willingness to cooperate. As outlined in Chapter 1, knowledge about other people's behavior facilitates the prediction of future behavior and provides a straightforward way to verify whether the prediction was accurate and whether that person can be trusted during future encounters. In a similar vein, nonverbal communication provides information about a person's intentions which facilitates the prediction of that person's next decision. Contrasting our expectations, these two sources of information operated independently on cooperation. Face-to-face contact can even "overrule" the urge to reciprocate a selfish act. In other words, our study suggests that nonverbal communication boosts cooperation to a certain degree and that that degree is constant independent of how much and what a person knows about the other individual.

The question of *how* face-to-face contact exerts its positive effects on cooperation was the focus of Chapter 4. As outlined in Chapter 1, researchers have proposed that humans have developed a refined signaling system of intentions where a variety of explicit and implicit signals facilitates nonverbal communication in social interactions (Boone & Buck, 2003). Additionally, our own emotions and their associated changes in inner states influence our decision-making (Damasio et al., 1996; Loewenstein & Lerner, 2003). Both approaches focus on how intrapersonal changes in either the observer or the observed are perceived. However, when moving from individuals' willingness to dyadic success of cooperation, I show that the signaling system incorporates an interpersonal, dynamic back-and-forth component. In fact, looking at implicit responses on the intrapersonal level was not informative of cooperative success. Instead, the study demonstrates that it is that extra layer of interpersonal communication that emerges over and above the individuals' responses during social interactions that determines how well individuals within dyads work together.

Similar to the intrapersonal level, interpersonal communication incorporates a range of different signals apparent on the explicit and implicit level. In the current thesis, I focused on two physiological responses, heartrate and skin conductance level. Other studies have shown that other types of synchrony also influence prosocial behavior such as facial mimicry, movement and vocalization synchrony (for reviews, see Mogan, Fischer, & Bulbulia, 2017; Palumbo et al., 2017; Prochazkova & Kret, 2017; Rennung & Göritz, 2016). This makes sense for two reasons: (1) given the subtle nature of physiological responses, synchrony on such level must emerge through other visible signals and (2) the more (explicit and implicit) expressions are synchronized, the better ambiguous expressions can be interpreted and the better a person can feel herself into the other person. Regarding the first reason, I observed that when people aligned their arousal responses reflected in skin conductance level changes, they were more successful to cooperate. Changes in skin conductance level are not directly visible to other people. Thus, the associated changes in arousal must be reflected in other, visible cues. This is also supported by the finding that synchrony increased with face-to-face contact compared to when a visual cover prevented nonverbal communication. In Chapter 4, I argue that pupil dilation and blushing might be potential cues as they are linked to changes in skin conductance and heartrate (Bradley et al., 2008; Dijk et al., 2011; Voncken & Bögels, 2009). In other words, people need to attend and possibly synchronize with other signals in order to reach the synchrony on the implicit, physiological level.

With respect to the second reason, our emotions and intentions are reflected in a range of explicit and implicit expressions. One expression can be interpreted in different ways depending on the context in which the expression is displayed and the combination with other expressions. For example, a smile can signal, among other states, subordination (Hecht & LaFrance, 1998), seeking of approval (Cashdan, 1998), or expressing embarrassment (Goldenthal, Johnston, & Kraut, 1981). One way to reduce the ambiguity and thereby helping to infer the meaning of a smile is to look at other signals complementing that smile. For example, expressing embarrassment is likely to be accompanied by blushing. Not only observing, but also synchronizing these different expressions can help to emotionally align with the person which subsequently affects behavior towards that person (Preston & de Waal, 2002). From this it can be argued that the richer the representation of another person's inner state through synchronization of different expressions, the stronger the emotional contagion and the more pronounced the potential effect on subsequent (prosocial) behavior. It is therefore likely that, although I only measured physiological measures, individuals synchronized on multiple levels. Following this argumentation, the answer to my research question of how nonverbal communication affects cooperative success is that we rely on a complex interpersonal signaling system incorporating different behavioral and physiological components that, when integrated, facilitates prosocial behavior towards one another.

Methodological implications

The second question the current thesis aims to answer is how we can best measure the link between nonverbal communication and cooperation in the lab from a methodological perspective. Specifically, in Chapter 2 I addressed the question of *what* we measure in light of how different prosocial behavior tasks address similar or distinct behavioral tendencies. In Chapter 5, I scrutinized the question of *how* we measure physiological synchrony. Here, I refined the method of how to optimally quantify physiological synchrony which forms the basis to investigate its causes and consequences and to compare findings between studies. In the following, I will discuss the methodological implications of these two studies with regard to the theoretical findings presented above.

In Chapter 3, I used the Prisoner's Dilemma game to measure cooperative behavior. In the follow-up study presented in Chapter 4, I changed the game from a dichotomous choice to a six-point scale, aiming to capture more fine-grained changes in cooperation. Despite this change, the two versions should still tap into the same behavioral tendencies, which was crucial as I investigated the underlying mechanisms of the effects observed in Chapter 3 in Chapter 4. Integrating the findings of both chapters with the results presented in Chapter 2, I am confident that this was indeed the case for two reasons. First, I demonstrated that behavior in these two versions were correlated suggesting that people who cooperated in one version also cooperated in the other version. Second, the cooperation rates observed in Chapter 2 were comparable to the rates seen in Chapter 3 and 4 for the original and extended version, respectively. Such consistencies support the choice of using the social dilemma games in both studies as a measure of cooperative behavior. In a next step, it is crucial to investigate whether the effects observed in Chapter 3 and 4 also generalize to more ecologically valid settings. The findings presented in Chapter 2 suggest that such generalization is challenged by methodological issues that come into play when moving away from the controlled setting of the economic games. Factors such as individual differences in skills to complete a task, ambiguity in the motivation behind behavior, and differences in the clarity of how to act prosocially are likely to influence the behavior displayed in a game. These methodological issues should not refrain researchers from studying synchrony in real life settings. In fact, our lab has successfully studied the influence of physiological synchrony in a blind-date experiment that we conducted during a festival (Prochazkova et al., 2019). However, the behavior of interest might be noisier and researchers need to take into account these differences when choosing a paradigm for their study and when comparing results between studies using different paradigms.

Moving the focus away from measuring cooperation to synchrony, one essential question is how strongly results on synchrony are influenced by the way it is quantified. Variations can originate from differences within and between methods. In Chapter 5, I focused on within-method variations and investigated how parameter settings within the Windowed Cross-Correlation analysis can cause such variation. In short, the method segments the time series of two interacting individuals into smaller, overlapping segments, also called windows, and calculates the cross-correlation between each segment. Additionally, for each segment the two time series are shifted away from each other up to a maximum lag. The size of the window and the maximum lag are two parameters that have been shown to influence the estimation of synchrony (Robinson et al., 1982; Schoenherr et al., 2019). In Chapter 5, I investigated the effect of the two parameters in the context of four physiological responses. I observed great variations between parameter configurations with a general pattern apparent in all signals: smaller window sizes were generally better in detecting synchrony. Nevertheless, there was a range of values that showed that ability, leaving the decision on which parameter to use to theoretical considerations. Regarding the maximum lag, the results revealed that this parameter was less influential than the window size, yet not trivial. The optimal maximum lag was around twice the window size. Based on these findings and theoretical considerations, I provide general recommendations on setting the window size and the maximum lag. However, I could not provide concrete optimal values for both parameters, leaving this choice to the researcher. Importantly, rather than searching for that one-fits-all solution, setting the window size to a specific value should be seen as testing a hypothesis, namely, whether people synchronize their responses that are equal to or longer than the window size chosen. In other words, different parameter choices constitute different hypotheses. Consequently, it is crucial that the researchers specify their choices both in the hypothesis and conclusion.

How can these conclusions be integrated with the results presented in Chapter 4, where I used this WCC analysis to quantify synchrony? In Chapter 4, I chose a window size (8 seconds) that falls in the range of appropriate values based on the results presented in Chapter 5. However, it is important to note that I did not apply the same surrogate data analysis in Chapter 4. Instead, I performed a down-graded version of the surrogate data analysis, where I compared the original dyads to one iteration of newly generated dyads rather than every possible dyad combination (see Chapter 4 for an explanation). As both variants are based on similar principles, I would expect similar results. Nevertheless, the surrogate data analysis allows for stronger conclusions because the level of pseudosynchrony can be estimated more reliably with a distribution of random dyads compared to one random dyad. Albeit less sensitive, the down-graded analysis performed in Chapter 4 was still sensitive enough to distinguish between the original and random dyads with the parameters chosen.

The choice of two other parameters might have undermined the effects presented in Chapter 4, in particular the maximum lag and the window increment. While the window size was 8 seconds, the maximum lag was 4 seconds. In Chapter 5, I recommend using a maximum lag of at least the size of the window. Therefore, the maximum lag used in Chapter 4 is half the recommended size. Additionally, I recommend using a window increment that is 1–5% of the window size. Therefore, for the analysis performed in Chapter 4, an increment of 80 – 400ms would have been preferred over the 2 seconds increment I used. Although the parameters chosen in Chapter 4 are not incorrect, the relatively small maximum lag and the relatively large increment result in a less sensitive analysis. With respect to the maximum lag, responses that lied further apart from each other than 4 seconds were not detected, potentially missing some moments of synchrony. This is especially the case for skin conductance level because it is a slow signal and therefore responses to one another might have lied further apart than 4 seconds. Still, there were sufficient responses that occurred within the 4 second range because the level of synchrony was higher in the original compared to the surrogate dyads. However, our results most likely showed conservative levels of synchrony, assuming that synchronized responses that lie further apart are not qualitatively different. Similarly, the relatively large window increment causes less overlap between two window segments. Consequently, the resolution of how the level of synchrony changes over time was lower, potentially missing subtle, yet crucial changes. Importantly, I would like to stress that all of this is not to say that the chosen parameters were incorrect. Instead, I would like to note that I might have missed subtle changes and consequently computed conservative estimates of synchrony. Assuming that these subtle changes would have provided only more rather than qualitatively different information, I might have obtained stronger effects in Chapter 4, if I had used the parameter recommendations of Chapter 5, but not completely different results.

In Chapter 4, I was mainly interested in the change in synchrony between two conditions and its link with cooperative behavior. The ability to detect such change in synchrony was also the aim of the study presented in Chapter 5. Unfortunately, the results lacked clear patterns across parameter configurations and showed inconsistencies between the primary and replication analysis. Such inconsistencies could have been the result of an unsuccessful manipulation between the two conditions and little differences between parameters, or the WCC analysis method that was not sensitive to detect changes that were in fact there. The results presented in Chapter 4 support the former explanation because I observed clear differences in conditions with parameters included in the range of parameters investigated in Chapter 5. Thus, in Chapter 4 the method was sensitive to detect change in synchrony. Further in line with the argument that the manipulation in Chapter 5 might not have been sensitive enough is the fact that the manipulation in Chapter 4 might have been indeed stronger as people could either see each other, allowing for nonverbal communication, or were prevented from nonverbal communication by means of a visual cover between participants. On the other hand, participants could always see each other in the study used in Chapter 5, either engaging in storytelling or looking at each other in silence. Thus, only changing the way participants interacted instead of manipulating *whether* they could interact or not might explain the small differences between conditions across parameter configurations. However, this interpretation is speculative and further research is needed on the sensitivity of parameter configurations to detect changes in synchrony.

Chapter 5 shows great variation in the strength of synchrony estimated by the same method with different parameters. Given such deviations *within* one statistical analysis method (Windowed Cross-Correlation analysis), it is likely that the differences are even more pronounced *between* statistical analysis methods. As a consequence, the comparability between studies that use different methods is likely to be low. The few studies that have looked into physiological synchrony and its link with cooperation indeed used different methods and show equivocal findings with demonstrating either increased heartrate synchrony, or elevated skin conductance level synchrony, or no link at all in a cooperative compared to a competitive context (Chanel et al., 2012; Järvelä et al., 2014; Mitkidis et al., 2015; Mønster et al., 2016; Vanutelli et al., 2017). For example, while some researchers applied a (multivariate) recurrence quantification analysis (Mitkidis et al., 2015; Mønster et al., 2016), others used slightly varying forms of simple cross-correlations and additionally calculated a weighted coherence measure of the frequency domain (Chanel et al., 2012; Järvelä et al., 2014; Vanutelli et al., 2017). These methods address different questions, have different assumptions, and operationalize synchrony in different ways. I do not mean to say that one method is better than the other, but it is important to realize that they measure different aspects of synchrony which might explain the equivocal findings in these studies.

Schoenherr et al. (2018) compared seven linear time series analysis methods (TSAMs) with different outcome scores and observed that they could be divided into three correlated, yet distinct facets of synchrony: the strength of synchrony of the total interaction, the strength of synchrony during synchronization intervals, and the frequency of synchrony. The WCC analysis as applied in the current thesis measured the first component. The reason for choosing a measure for the total interaction rather than identifying intervals of synchrony first is that I used continuous measures without clear moments of activation and deactivation. The strength of synchrony will certainly vary over time, however, not to the extent that it is on or off as can be the case, for example, in motor movements. A head or hand can move or not; a heart does not stop beating in between. For the facial expression measure, determining the synchrony intervals could have been an option. However, I wanted to be consistent across measures and for the other three signals taken under the loop in Chapter 5, it seemed most appropriate to consider the whole interaction. One could, of course, still apply a certain threshold to classify synchrony intervals as performed in the peak-picking algorithm developed by Altmann (2011). This option has been suggested to be particularly interesting for linking moments of high synchrony to specific characteristics of a conversation (Schoenherr et al., 2018). However, it is important to realize that investigating how *strong* people synchronize is a different question than how *often* they do so, and that the outcomes of these analyses are likely to be different. This comparison considers two analyses that could be considered cousins given the partial overlap in procedures (Windowed Cross-Correlation and Windowed Cross-Regression analysis; Altmann, 2011; Boker et al., 2002). Other methods concentrate on the association between participants' responses in nonlinear patterns or in the frequency domain, addressing yet other questions. It is beyond the scope of this thesis to provide an overview of these different methods and I would like to refer the interested readers to other literature (Gates & Liu, 2016; Lee-Helm, Miller, Kahle, Troxel, & Hastings, 2018; Schoenherr et al., 2018; Thorson, West, & Mendes, 2018). The important lesson here is that researchers should carefully consider different methods and be aware of what exact research question they answer with a given analysis. Once they have decided on the method, they should carefully choose its appropriate settings. Chapter 5 takes a step into this direction by providing recommendations on how to apply the WCC analysis to multiple physiological measures.

In summary, in this section I discussed two methodological implications when studying the link between nonverbal communication and cooperative behavior. First, I looked into how the finding that more synchronized dyads are more successful in cooperation could be generalized to more naturalistic games. Integrating the findings presented in Chapter 2, I encourage researchers to use more ecologically valid games and investigate whether our findings could be generalized to these situations. At the same time, I pinpoint to the methodological challenges encountered when moving away from the controlled economic games that should be considered when choosing a paradigm for a study and when comparing findings between studies using different games. Second, I discussed the implications of how synchrony is quantified with different analyses and different settings within an analysis. Researchers are faced with a great amount of (correct) choices emphasizing the need to clearly specify their choices in both their hypotheses and conclusions. I hope that the studies presented in Chapter 2 and 5 will guide researchers in making well-informed decisions which will increase the comparability across studies and shed more light on the link between nonverbal communication and interpersonal processes.

Limitations and (new) open questions

As already highlighted in the "theoretical implications" section, synchrony most likely happens on a wide range of behavioral and physiological levels. Here, I focused on two physiological measures, looking at only one piece of the puzzle. Future research is needed where multiple, both explicit and implicit measures, are measured simultaneously to address questions such as "What are the channels through which physiological synchrony emerges?", "How are different signals integrated into making a decision to cooperate or not?", "Are some signals more synchronized and more important drivers for making a cooperative decision than other signals?", "Is the number of synchronized expressions crucial for how strong their (joint) effect is on cooperation?", "Are the effects of different expressions and their integration similar across other (pro-) social behaviors?". Conducting studies where multiple signals are measured simultaneously could provide valuable insight into these questions.

The current thesis concentrates on the link between synchrony and cooperation, and discusses the implication for other prosocial behaviors. However, prosocial behavior is only one example that has been linked to synchrony. Other interpersonal processes such as sexual attraction, marital satisfaction and therapeutic outcomes have also been shown to be affected by synchrony (Levenson & Gottman, 1983; Prochazkova, Sjak-Shie, Behrens, Lindh, & Kret, 2019; Ramseyer & Tschacher, 2011). Is the link between synchrony and these different interpersonal processes caused by similar underlying processes? What other effects could synchrony have that subsequently affect the way individuals behave towards one another? Future studies are needed where the link between synchrony and social behavior is investigated further in terms of the underlying mechanisms. Given that it is unlikely that each such link is tight to one specific process, including multiple measures in one study can help us disentangle the function of synchrony on different social behaviors.

Another crucial question that remains unanswered is whether synchrony is a cause or consequence of cooperation. In other words, does the emergence of synchrony between two individuals affect how well they subsequently work together or is the strength of synchrony a reflection of how well they have cooperated? In the literature, this question is reflected in two lines of research, either manipulating synchrony or the prosocial setting. The former has concentrated on motor and vocalization synchrony asking people to dance, tap, or sing together and investigate how prosocial behavior changes between synchronized and non-synchronized conditions (for two meta-analyses, see Mogan et al., 2017; Rennung & Göritz, 2016). Another related line of research focuses on how blocking facial mimicry impairs emotion processing, for example, in response to Botox treatment and in clinical populations such as the Möbius syndrome (Bogart & Matsumoto, 2010; Neal & Chartrand, 2011). Although not directly addressing social behavior, it sheds light on how social interactions are affected by the lack of synchrony which can subsequently affect behavior. In the context of physiological synchrony, manipulating the level of synchrony in, for example, heartrate is less straightforward, which is why research has focused on manipulating the cooperative setting and investigating its effect on synchrony. Given that manipulating both variables affect the other suggests that the relationship is bi-directional. In line with the Perception-Action Model (Preston & de Waal, 2002), in my dissertation I adhere to the perspective of studying synchrony as a potential underlying mechanism for why face-to-face contact boosts cooperation. Such directional effect is reflected in the study in Chapter 4 where people first look at each other, allowing for nonverbal exchange of information, and subsequently make a decision to cooperate. Thus, synchrony precedes the decision to cooperate. However, given the repeated nature of the design with participants playing multiple rounds in succession, it is possible that the reported effects are influenced by carry-over effects between rounds mirroring reflections rather than antecedents of cooperation. Future studies should elucidate on the question of the causal link between synchrony and cooperation.

Finally, I want to emphasize the importance of conducting real-life interaction studies. Cooperation is a social concept and should therefore ideally be treated as such. This entails investigating cooperation in actual interactions, moving away from one-person computerized paradigms. Although these paradigms provide researchers with great experimental control, they undermine the interpersonal processes observed in the current thesis. I took a step in that direction by letting two participants interact during the study. However, the setting was still controlled and performed in the lab, compromising the ecological validity of the findings. Therefore, future studies should investigate whether the observed effects between physiological synchrony and cooperation are also visible outside the lab and pass the test of practical relevance (for instance, see Prochazkova et al., 2019 for a study conducted at a festival where physiological synchrony predicted blind-date success).

Conclusions

Large-scaled cooperation has been suggested as one of the driving forces of human's superiority in the evolutionary hierarchy. Its success depends on individuals working together and thereby relies on how these individuals connect on a subtle, unconscious level of nonverbal communication. Despite the technical advances that globally connect human society on a hereto unknown scale, technology cannot replace the deep-wired, evolutionary drives to communicate and bond with other individuals on the biological level for which face-to-face interactions are so essential. The current thesis sheds light on what that nonverbal communication entails revealing a new layer of interpersonal back-and-forth communication that is more than the sum of the responses of the interacting individuals. Through face-to-face contact, people pick up subtle changes of arousal in their interaction partner and adjust their own arousal levels accordingly. This connection of two bodies, emerging outside our control and consciousness, influences how well we cooperate with each other. Alongside these great theoretical implications, I have embedded these findings in a methodological cushion. First, the finding that physiological synchrony is associated with cooperation should not be blindly generalized to more naturalistic paradigms of prosocial behavior without further investigation. Therefore, researchers need to know *what* they measure. Finally, zooming in on how to statistically capture the strength of synchrony between individuals, I emphasize that researchers need to know *how* they measure it.

Appendices

Appendix A

Supplementary material for Chapter 2

Appendix A1

Table A.S1

*Descriptive statistics of the participants of the current study. This information was asked in an online questionnaire at the end of the experiment (*N = 73*)*

Descriptive	% (count)
Gender:	
Female	.69 (51)
Nationality:	
Dutch	.75 (55)
Non-Dutch	.25 (18)
Highest completed education:	
High school	.52 (38)
Applied University	.12 (9)
University	.36 (26)
Number of siblings:	
No siblings	.14 (10)
1 sibling	.44 (32)
2 siblings	.26 (19)
3 or more siblings	.16 (12)

Appendix A2

Descriptive statistics of the Liking and Knowing scale ratings

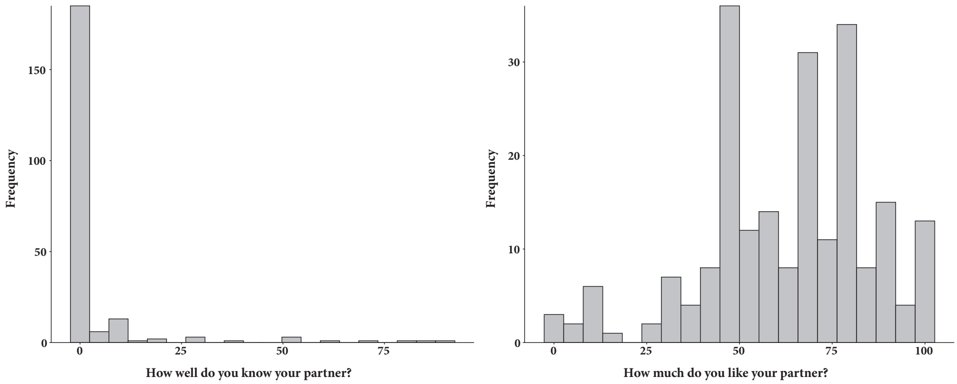


Figure A.S1. Distribution of ratings for the questions “How well do you know your partner?” (left) and “How much do you like your partner?” (right) on a scale from 0 to 100.

Table A.S2

Descriptive statistics of how well participants know (Knowing) and how much they liked their partner (Liking). Each participant rated three different partners

Descriptive	Knowing	Liking
Median	1	70
MAD	1	16
Maximum	90	100
Minimum	0	0
N	219	219
Missing	3	3

Note. MAD = median absolute deviation; *N* = sample size.

A

Table A.S3

Correlation between liking the partner and prosocial behavior

Game	Spearman's rho	95%-*CI*
Prisoner's Dilemma	.10	-.14–.33
Extended PD	.28	.03–.50
Rope Pull	.09	-.15–.32
Tangram	-.04	-.27–.20
Hidden Profile	.09	-.15–.33
Egg Hunt	.14	-.13–.40

Note. CI = confidence interval.

In Table A.S3, the Spearman's rho correlations between a person's rating of the partner and her/his prosocial behavior towards that person are shown. None of the outcomes of the games was reasonably correlated with how much participants liked each other.

Appendix A3

Alternative Prosocial Behavior Measures for the Tangram Game

In the following we provide descriptive statistics on possible alternative prosocial measures for the Tangram game on a dyadic level.

Do people perform better together than alone?

We here present three different measures to answer this question: (1) *max*, the number of completed puzzles from the participant with the highest number in the individual condition; (2) *sum*, the sum of the completed puzzles of both participants in the individual condition; (3) *mean*, the mean of the completed puzzles of both participants in the individual condition. For all measures, we calculated the difference score by subtracting the number of completed puzzles in the cooperative condition minus the corresponding measure (max, sum, or mean). A positive value indicates that people performed better together than alone. The descriptive statistics of the difference scores are presented in Table A.S4 and the distribution of the difference scores are displayed in Figure A.S2–4.

Table A.S4

Descriptive statistics of the difference score between the joint performance in the cooperative condition minus the maximum, sum, and mean performance of the two participants in the individual condition

Descriptive	Max	Sum	Mean
Median	-1.0	-3	0
MAD	1.0	1.0	1.0
Maximum	3.0	1	4
Minimum	-5.0	-6	-3
N	36	36	36
Missing	0	0	0

Note. MAD = median absolute deviation; *N* = sample size.

A

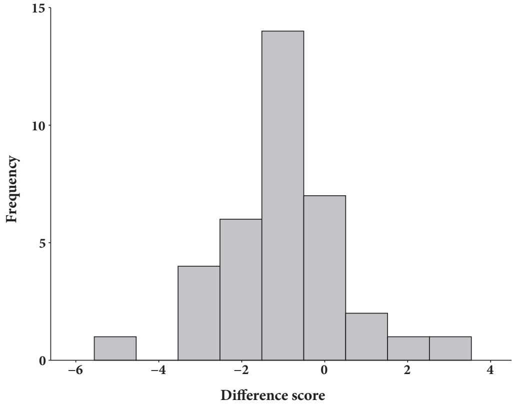


Figure A.S2. Distribution of the difference score between the joint performance in the cooperative condition minus the maximum performance of the two participants in the individual condition.

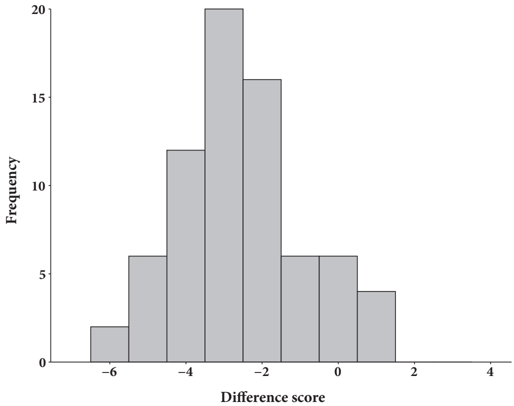


Figure A.S3. Distribution of the difference score between the joint performance in the cooperative condition minus the sum of performance of the two participants in the individual condition.

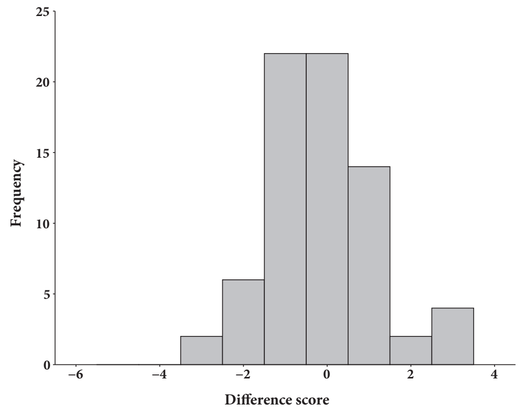


Figure A.S4. Distribution of the difference score between the joint performance in the cooperative condition minus the mean of performance of the two participants in the individual condition.

Does the difference in skills between participants influence the joint performance?

The Spearman's rho is .27 with *95% CI* (.04, .48) suggesting that there is a weak relationship between the mismatch in people's ability to perform the Tangram game (difference in completed puzzles between participants in the individual condition) and their performance together in the cooperative condition.

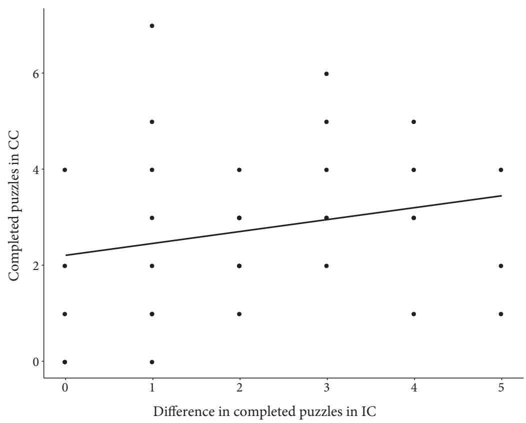


Figure A.S5. Scatterplot of the relation between the joint performance in the cooperative condition and the difference in completed puzzles between participants in the individual condition.

Appendix A4

Comparison of cooperation (success) rates between social dilemma game variants

In a follow-up analysis, we compared the cooperation rates between the Prisoner's Dilemma game, its extended version, and the Rope Pull game. The matched-pairs rank biserial correlation r is reported as a measure of effect size (Kerby, 2014). The largest difference was observed between the Prisoner's Dilemma and Rope Pull game ($Z = 4.82$, $r = .75$, $p < .001$) with higher cooperation rates in the latter. Also compared to the extended Prisoner's Dilemma game, the cooperation rate was higher in the Rope Pull game ($Z = 4.10$, $r = .62$, $p < .001$). Finally, participants cooperated more in the extended compared to the original Prisoner's Dilemma game ($Z = 2.49$, $r = .39$, $p = .013$).

Furthermore, we investigated not only the cooperation rate on the individual, but also on the dyadic level to see whether the willingness to cooperate translated into successful cooperation. Therefore, we calculated the proportion of how many times a dyad successfully cooperated from the ten trials in the three games. In the Prisoner's Dilemma and Rope Pull games, mutual cooperation occurred if both participants cooperated. In the extended Prisoner's Dilemma game, we looked at the mean of joint points participants received ranging from four (both fully defect) to six points (both fully cooperate). To make measures comparable, we transformed the joint points into proportions. The descriptive statistics are shown in Table A.S5 and visualized in Figure A.S6. Interestingly, the median of the Prisoner's Dilemma game substantially dropped compared to the cooperation rate at the individual level (from .60 to .30). In the other two games, mutual cooperation rates also decreased compared to cooperation rates on the individual level, but to a lesser extent (extended Prisoner's Dilemma: .68 versus .60; Rope Pull: .90 versus .80). Comparing the three games, participants succeeded least often in the Prisoner's Dilemma game and most often in the Rope Pull game (extended PD versus PD: $Z = 5.19$, $r = 0.52$, $p < .001$; RP versus PD: $Z = 5.61$, $r = 0.54$, $p < .001$; RP versus extended PD: $Z = 1.69$, $r = 0.21$, $p = .091$). This was also evident when looking at the proportion of trials participants cooperated successfully: in the Pull Rope game, almost half of the dyads mutually cooperated in all trials (.42), this proportion dropped to only .05 (= two dyads) in the Prisoner's Dilemma games. On the other hand, in the original Prisoner's Dilemma game, a quarter of the dyads (.25) did not successfully cooperate in any of the trials. In the extended Prisoner's Dilemma game, the success of mutual cooperation was evenly distributed throughout the spectrum.

Table A.S5

Descriptive statistics of the proportion of mutual cooperation for the games of the "social dilemma games" component

Game	Mean	Median	SD	MAD	Maximum	Minimum	Range
Prisoner's Dilemma	.35	.30	.31	.25	1.00	.00	1.00
Extended Prisoner's Dilemma	.63	.60	.23	.18	1.00	.17	.83
Rope Pull	.69	.80	.34	.20	1.00	.00	1.00

Note. SD = standard deviation; MAD = median absolute deviation.

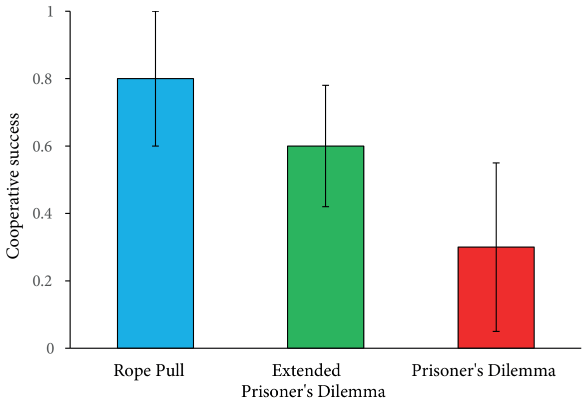


Figure A.S6. Median (± Median Absolute Deviation) of the cooperative success in the three variants of the social dilemma game.

A

In sum, on both the individual (willingness) and dyadic (success) level, cooperation was highest for the Rope Pull game and lowest for the classical Prisoner's Dilemma game. Intriguingly, in the extended Prisoner's Dilemma game, cooperation was higher compared to the original version, which is consistent with our previous studies (classical Prisoner's Dilemma game: .60 in current study and .58 in Behrens and Kret (2019); extended Prisoner's Dilemma game: .67 in the current study and .71 in Behrens et al. (2019)). Similarly, more choice options also yielded more successful cooperation, again, replicating our previous studies (classical Prisoner's Dilemma: .30 in the current study and .35 in Behrens and Kret (2019); extended Prisoner's Dilemma: .60 in the current study and .70 in Behrens et al. (2019)). The difference in cooperative success was mainly driven by a substantial proportion of dyads that always failed to cooperate (one-sided cooperation and mutual defection) when playing the classical Prisoner's Dilemma (where participants only had a dichotomous choice). Although a small majority was willing to cooperate (.60), dyads barely succeeded in *mutual* cooperation (.30). This was considerably less the case in the extended version (.67 versus .60). Our findings indicate that if people are given the option to indicate *how much* instead of *whether* they would like to work together, they are more inclined to and more successful at doing so. Thus, cooperation can be boosted by giving people multiple choices.

The results also demonstrate that the willingness and success of cooperation were considerably higher when participants played the Rope Pull game compared to both Prisoner's Dilemma games. One element that is incorporated in the former, but not the latter games is that participants received continuous feedback. Participants could adjust their behavior during a trial in response to the direct feedback of the rope. Research has indeed shown that making information about an interaction partner's decisions available to a participant facilitates cooperation (Behrens & Kret, 2019; Bixenstine & Wilson, 1963; Jorgenson & Papciak, 1981; Monterosso et al., 2002; Tedeschi et al., 1968). Another potential factor contributing to the discrepancy between the Rope Pull task and the Prisoner's Dilemma games is related to the payoff structures. In the Rope Pull task, the outcome of mutual defection and one-sided cooperation was the same (no reward), whereas in the Prisoner's Dilemma games, mutual defection led to a higher outcome (two points) than one-sided cooperation (one point). Therefore, if a participant predicts that the other person will defect, the preferred option in the latter two games is to defect as well, which is likely to elicit mutual defection in the subsequent rounds. However, if a participant predicts that the other person will defect in the Rope Pull game, she will receive no reward independent of whether she will defect or cooperate herself. Consequently, cooperation is wise because it might trigger the other to reciprocate in the next round. This shift towards mutual cooperation due to the payoff structure might therefore have inflated the cooperation rate in the Rope Pull game. We therefore argue that the greater willingness and success of cooperation in the Rope Pull game compared to the Prisoner's Dilemma games can be explained by: (1) receiving moment-to-moment feedback about the partner's intentions through pulling the rope, and (2) the payoff structure that gives no benefits to mutual defection over one-sided cooperation.

Appendix B

Supplementary material for Chapter 3

Appendix B1

Exploratory analysis of personality traits on cooperative behavior

This section includes exploratory analyses on how different personality traits (a) influence a person's willingness to cooperate and (b) modulate the effect of face-to-face contact on a person's willingness to cooperate.

There was great variation between dyads and players in how individuals were influenced by the face-to-face manipulation, as suggested by the large variances of the random effects of Dyad and Dyad * Player in the main analyses (see Result section of Chapter 3). In an attempt to explain these differences, we investigated how personality traits influenced the experimental manipulations on participants' willingness to cooperate. Participants completed questionnaires about their empathy level (IRI), social anxiety (LSAS), emotion recognition ability (Reading the Mind in the Eyes task), and social value orientation (SVO). For an overview of the descriptive statistics of these questionnaires in our sample and the correlations between them, see Tables B.S1-S2. For the IRI, we looked at the total score and the subscales perspective taking (PT), empathic concern (EC), fantasy scale (FS), and personal distress (PD) separately. Also, the two subscales anxiety and avoidance of the LSAS were analyzed individually. For both questionnaires, the mean was calculated per subscale per person. The SVO classifies individuals into four categories: prosocial, individualistic, competitive and no classification (Van Lange, 2000). The sample size for the latter three categories was too small to constitute a group, which is why we decided to combine them into the category "non-prosocial" (prosocial $n = 73$, non-prosocial $n = 33$ [consisting of 2 competitive, 15 individualistic, and 16 unclassified participants]). The performance for the Reading the Mind in the Eyes task was calculated based on the mean accuracy level (0 = incorrect, 1 = correct).

We concentrated the analysis on participants' own decisions rather than on the joint outcome as we expected individual characteristics to mainly influence individual decisions. Furthermore, we laid our main focus on the face-to-face manipulation and not on the feedback conditions as we were mainly interested in how personality traits would influence interpersonal communication rather than how people differ in their use of explicit, objective feedback.

The effect of nonverbal communication was moderated by the characteristics of the participants (Table B.S3). In particular, we observed a significant interaction between Face condition and SVO, IRI (total score and subscales PT and EC), Reading the Mind in the Eyes task, and LSAS anxiety scale (p's $\leq .002$). To disentangle these interaction effects and for the ease of interpretation, we median-splitted participants on the IRI, Reading the Mind in the Eyes task and LSAS anxiety scales. For these questionnaires, there was a significant difference between Face conditions for participants scoring low on each of the scales, but not for individuals having high scores (p's $\leq .04$; Figure B.S1 a-c). By visual inspection, it can be seen that participants who were less empathic and who had more difficulties to read another person's mind, generally cooperated less than people scoring high on these measures. On top of that, they were more strongly influenced by the experimental manipulation and were even less willing to cooperate in the face-blocked than the face-to-face condition. For the anxiety scale of the LSAS, people who scored below the median were again more influenced by the experimental conditions. In this case, low socially anxious people were more willing to cooperate when facing the other person compared to when not and high socially anxious people were generally less cooperative, regardless of the face condition. For the SVO, prosocial individuals cooperated more in the face-to-face compared to face-blocked condition, but non-prosocials were unaffected by the face manipulation (Figure B.S1d).

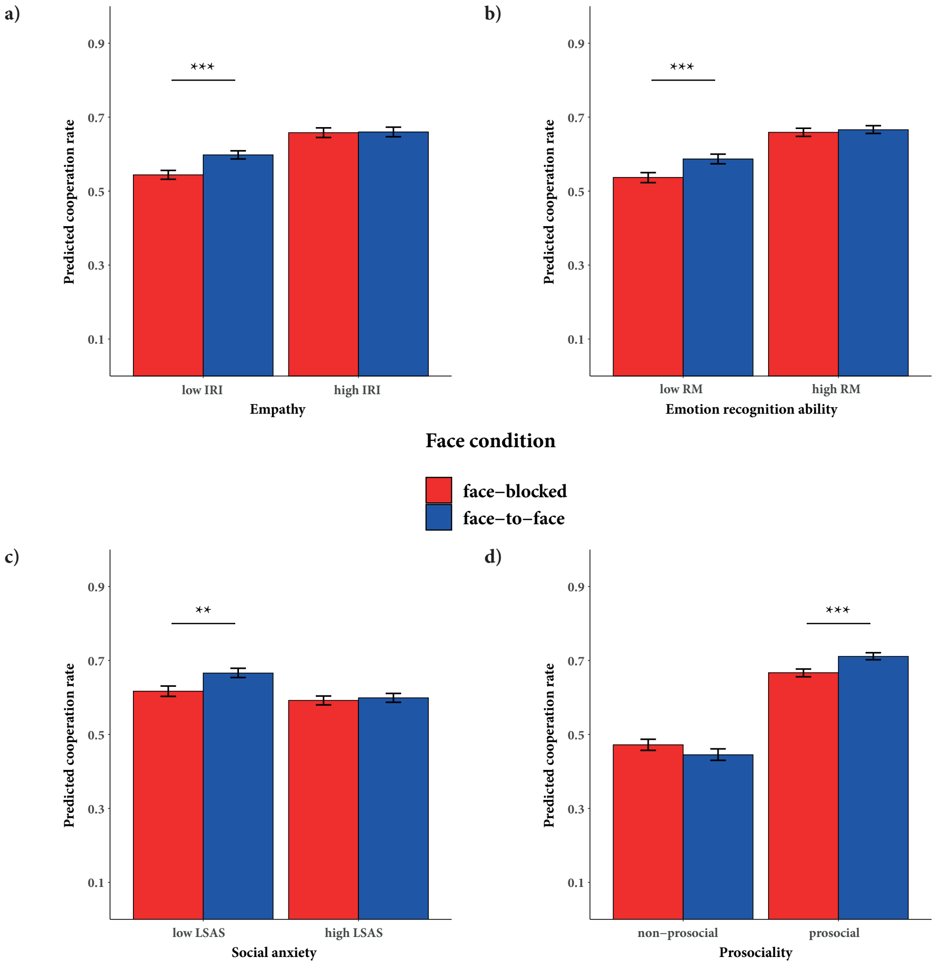


Figure B.S1. Predicted mean cooperation rate per Face condition moderated by (a) empathy (IRI total score), (b) emotion recognition abilities (RM = Reading the Mind in the Eyes task), (c) social anxiety (LSAS anxiety subscale), and (d) prosociality (SVO); ** $p < .01$, *** $p < .001$.

B

Table B.S1

Descriptive statistics of the personal characteristic questionnaires

Questionnaire	Mean	SD	Max	Min	$N_{Missing}$
RM	0.70	0.09	0.92	0.47	4
LSAS anxiety *	17	10	49	1	13
LSAS avoidance *	11	9	45	0	13
IRI Total *	101	15	134	66	13
IRI PT *	28	5	40	13	13
IRI EC *	27	5	39	11	13
IRI FS *	26	6	39	9	13
IRI PD *	20	6	34	6	13
Prosocial (SVO)	73 (68.9%)				

Note. * based on the summed score per subject.

Table B.S2

Pearson's correlation matrix with the RM, LSAS and IRI questionnaires

	RM	LSAS anxiety	LSAS avoidance	IRI Total	IRI PT	IRI EC	IRI FS
RM							
LSAS anxiety	0.160						
LSAS avoidance	0.12	0.70 ***					
IRI Total	0.071	0.308 **	0.17				
IRI PT	-0.108	0.101	0.071	0.510 ***			
IRI EC	0.038	0.213 *	0.10	0.731 ***	0.279 **		
IRI FS	0.068	0.071	-0.06	0.683 ***	0.099	0.289 **	
IRI PD	0.160	0.422 ***	0.33 **	0.710 ***	0.087	0.423 ***	0.310 **

Note. * $p < 0.05$; ** $p < 0.005$; *** $p < 0.001$; RM = Reading the Mind in the Eyes task; LSAS = Liebowitz Social Anxiety Scale; IRI = Interpersonal Reactivity Index; PT = Perspective Taking; EC = Empathic Concern; FS = Fantasy Scale; PD = Personal Distress.

Table B.S3

Descriptive statistics of model parameters regressing the interaction between each personality trait and the Face condition against the willingness to cooperate (defect = 0, cooperate = 1)

Effects	Test statistics	*p*-value
SVO * Face	$F(1, 9335) = 18.85$	**< .001**
Split by SVO		
Face (prosocials)	$B = .34$, $SE = .07$, CI (.21, .48)	**< .001**
Face (non-prosocials)		.086
IRI total * Face	F (1, 9088) = 11.22	**.001**
median-split IRI total		
Face (high IRI)		.345
Face (low IRI)	$B = .38$, $SE = .07$, CI (.24, .53)	**< .001**
IRI PT * Face	F (1, 9088) = 21.66	**< .001**
Median-split IRI PT		
Face (high IRI PT)		.465
Face (low IRI PT)	$B = .49$, $SE = .09$, CI (.33, .66)	**< .001**
IRI EC * Face	F (1, 9088) = 24.87	**< .001**
Median-split IRI EC		
Face (high IRI EC)		.052
Face (low IRI EC)	$B = .49$, $SE = .08$, CI (.34, .64)	**< .001**
IRI FS * Face		.213
IRI FS main effect		.986
IRI PD * Face		.040
IRI PD main effect		.507
RM * Face	F (1, 9870) = 10.22	**.001**
Median-split RM^{+}		
Face (high RM)		.839
Face (low RM)	$B = .37$, $SE = .08$, CI (.20, .53)	**< .001**
LSAS anxiety * Face	F (1, 9088) = 12.26	**< .001**
Median-split LSAS		
Face (high LSAS anxiety)		.080
Face (low LSAS anxiety)	$B = .26$, $SE = .09$, CI (.09, .42)	**.002**
LSAS avoidance * Face		.311
LSAS avoidance main effect		.486

Note. P-values below the significance level of .005 are indicated in bold. SVO = Social Value Orientation; IRI = Interpersonal Reactivity Index; PT = Perspective Taking; EC = Empathic Concern; FS = Fantasy Scale; PD = Personal Distress; RM = Reading the Mind in the Eyes task; LSAS = Liebowitz Social Anxiety Scale. $^{+}$random effect Dyad excluded because of the lack of enough variance.

Appendix B2

Exploratory analysis on the accuracy of predicting the partner's cooperative decision

In the current study, participants gave two responses per trial: whether they wanted to choose option A or B (corresponding to cooperating and defecting, respectively) and what they thought their partner chose. Based on this, we investigated whether individuals could read each other's intentions based on nonverbal cues only. To that extent, we conducted a one-sample t-test in the face-to-face / no feedback condition to compare the mean accuracy level to the level of chance. The results revealed that participants in this condition were not able to predict their partner's decisions ($M = .51$, $SD = .31$, $t(39) = .14$, $p = .890$).

Appendix B3

Exploratory analysis on the order effect of the Face condition on cooperative behavior

The order of the Face condition was counterbalanced between dyads. In an attempt to assure that the order did not affect the manipulation, we tested for the interaction between Face condition and the Face order. Surprisingly, this interaction was significant, $F(1, 10223) = 24.20$, $p < .001$ (see Figure B.S2). Specifically, the increase in the willingness to cooperate in the face-to-face compared to face-blocked condition was bound to those dyads who started the experiment in the face-to-face condition, $B = .42$, $SE = .07$, CI (.27, .56), $OR = 1.52$, $p < .001$, but not when they began in the face-blocked condition ($p = .180$). Similarly, the order of the Face condition influenced the successfulness of the joint outcome, $F(2, 5157) = 9.42$, $p < .001$. For all analyses of the study including the Face condition, we performed the analyses with and without including the interaction effect between Face condition and the Face order. None of the findings were influenced by it, so we only report the analyses without the interaction effect.

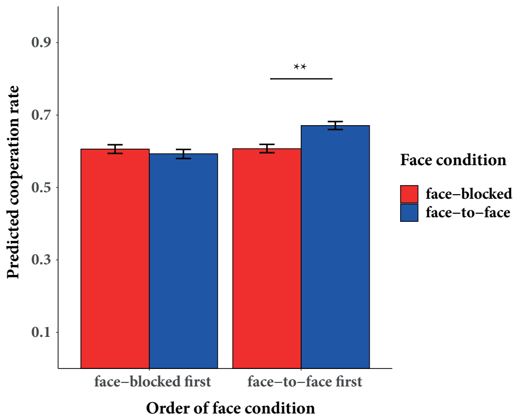


Figure B.S2. Predicted mean cooperation rate (± 2 SE) in the face-blocked and face-to-face condition moderated by the order of the condition (** $p < 0.005$).

B

Appendix B4

Descriptive statistics of participants' experiences during the study

Here, we present descriptive statistics of the experiences of the participants during the experiment. They filled out a visual analogue scale (VAS) after the first practice trails, after the first completed session, after the second practice trials and after the second session (the end of the experiment). Among others, they indicated how motivated they were, how much difficulty participants had to keep their attention to the task, how much they felt connected to their interaction partner and how anxious they felt. All questions were answered by setting a marker on a 10 cm long line ranging from "not at all" on the left to "very much" on the right. Additionally, participants completed the Positive And Negative Affect Schedule (PANAS) questionnaire before they started the game. Finally, participants filled out the Desire for Future Interaction scale (DFI) to indicate how much participants would like to meet their interaction partner again in different situations on a 5-point Likert scale. In Table B.S4, we present the mean and standard deviation of the VAS for each time point, the PANAS and the DFI.

Table B.S4

Descriptive statistics (mean and standard deviation) of the PANAS (Positive And Negative Affect Schedule), the VAS (Visual Analogue Scale), and the DFI (Desire of Future Interaction scale) participants completed before, during, and after the experiment, respectively

Question	After 1st practice trials	After 1st session	After 2nd practice trials	After 2nd session
How…do you feel at this moment?				
tense	2.5 (2.1)	1.8 (2.4)	1.9 (2.2)	1.4 (2.1)
awkward	3.0 (2.5)	2.2 (2.4)	2.8 (2.7)	1.5 (2.0)
shy	1.8 (2.0)	1.4 (1.9)	2.0 (3.4)	1.1 (1.7)
anxious	0.8 (1.1)	0.6 (1.0)	0.6 (1.1)	0.3 (0.6)
observed	4.5 (2.8)	3.8 (3.0)	3.7 (3.4)	2.8 (3.1)
Do you feel like the other sees right through you?	2.5 (2.4)	2.3 (2.4)	2.0 (2.5)	2.2 (2.7)
Do you feel connected to the other person?	7.4 (2.2)	7.0 (2.0)	7.5 (2.5)	7.3 (2.5)
How motivated are you to complete this task?	7.0 (1.9)	5.8 (2.8)	5.9 (2.6)	5.6 (2.7)
How difficult is it for you to keep your attention directed to the task?	2.2 (2.4)	5.0 (3.1)	4.5 (3.1)	5.7 (3.1)
Desire for Future Interaction (DFI)				3.21 (.62)
Positive And Negative Affect Schedule (PANAS)				
Positive subscale	3.22 (.51)			
Negative subscale	1.30 (.31)			

Note. The questions were answered by setting a mark on a 10 cm line, therefore the scale ranges from 0 to 10. The two PANAS subscales and the DFI were rated on a 5-point Likert scale before and after the experiment, respectively.

Appendix C

Supplementary material for Chapter 4

Appendix C1

Sensitivity Analysis

The sensitivity analysis has been proposed to be a valid post-hoc analysis in case an a priori power analysis has not been conducted before the study (Davis et al., 2018). In contrast to the traditional power analysis, where the relationship between power and sample size given a specified effect size is computed, the sensitivity analysis investigates the relationship between power and effect size given a particular sample size. The idea is to run simulation-based power analyses and detect the minimum true effect size that a study is sensitive enough to detect given a certain level of power (mostly, 80%) and a specific sample size.

The simulation-based sensitivity analysis includes the following steps that are repeated 1000 times: (i) simulate new data for the response variables based on the specified model (in our case, the full model shown in Table S2); (ii) refit the model to the new data; (iii) perform a statistical test on the effect of interest (in our case, the interaction effect between skin conductance level synchrony and Face condition). The assumption is that the effect of interest reflects the true population effect size, so every positive test is a true positive and every negative test is a false negative (i.e., a Type II error). Based on these results, the power can be directly calculated from the number of successes and failures (Green & Macleod, 2016). This power analysis is not only performed for the observed effect (in our case, the estimated interaction effect between skin conductance level synchrony and Face condition [.86], see Table S2), but also for a range of other effect sizes. Notice that the effect size is based on the scaled estimate of the model rather than a standardized effect size. For each effect size, the power to detect that effect (assuming that it is the true population effect size) is calculated resulting in the curve shown in Figure S1. The dashed line indicates the 80% power criterion and its associated true effect size (.70) that we can detect given our sample size. In other words, with our design, we would find a significant p-value in 80% of the cases if the true effect size was .70. The observed effect size of .86 is associated with a power of 89%, again assuming that the observed effect size reflects the true population effect size.

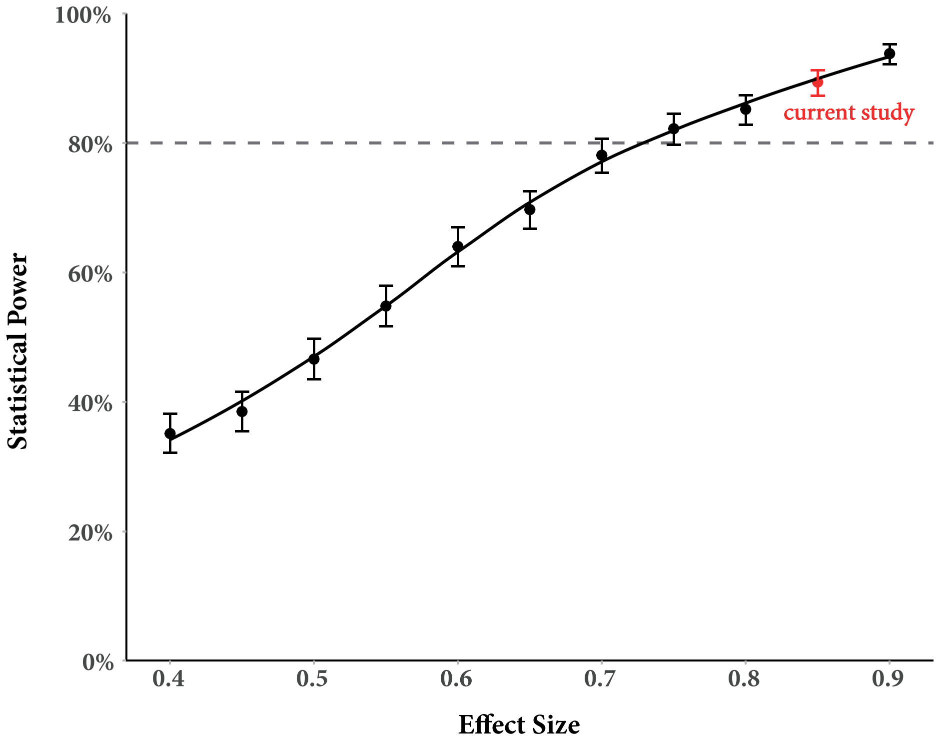


Figure C.S1. Simulation-based sensitivity analysis with statistical power as a function of different effect sizes. The observed effect size and associated power of the current study is marked in red. The dashed line marks an 80%-power threshold. The error bars reflect 95% confidence intervals. Notice that the effect sizes are based on the raw scale of the model and should not be interpreted following rule of thumb guidelines regarding the strength of the effect (e.g., Cohen's d).

C

Appendix C2

Information about the self-reported questionnaires

Table C.S1

Descriptive statistics of the self-reported questionnaires

Questionnaire	Mean	SD	Range	Theoretical range	$N_{missing}$
RM	26.05	3.77	12–34	0–37	1
LSAS	34.34	20.17	0–100	0–144	10
IRI	124.10	16.03	78–165	28–196	9
PANAS POS	30.30	6.99	11–46	10–50	1
PANAS NEG	13.44	3.42	9–25	10–50	1
DFI	3.17	.64	1.25–5	1–5	1
SVO (prosocial)	66.4%				9

Note. RM = Reading the Mind in the Eyes game; LSAS = Liebowitz Social Anxiety Scale; IRI = Interpersonal Reactivity Index; PANAS = Positive and Negative Affect Schedule; DFI = Desire for Future Interaction Scale.

Appendix C3

Quantification of physiological synchrony

Two methods that take non-stationarity into account are lagged windowed cross-correlation (Boker et al., 2002) and recurrence quantification analysis (Gates & Liu, 2016). The latter method is frequently used which has the advantage of having very few assumptions. However, the disadvantage is that it determines synchrony on a binary scale of moments being classified as either synchronized or not. The former method, albeit constraint by more assumptions, has the advantage of differentiating the degree of synchronization by quantifying it on a continuous (correlation) scale. Additionally, we feel that windowed cross-correlation is more intuitive to interpret. Consequently, we decided to apply this method which provides measures of the strength of synchrony and its variability.

The objective of the lagged windows-cross correlations analysis (Boker et al., 2002) is to calculate the strength of association between two time series while taking into account the non-stationarity of the signals and the lag between responses, that is, to consider the dynamics of a dyadic interaction. Specifically, the time series are segmented into smaller intervals, calculating the cross-correlation for each segment. This allows the means and variances to differ between segments accounting for non-stationarity. This is important as the level of synchrony may change during the experiment, sometimes having moments of strong synchronization while during other times responding less strongly to one another. Additionally, as the strength of association between two time points may differ depending on how far apart they are from each other, the segments are moved along the time series by an increment such that two adjacent segments overlap. Hence, segmenting the time series into smaller intervals and partially overlapping these intervals while moving along the time series provides a better estimate of the local strength of association between the physiological signals of two participants.

Besides the dynamics in the strength of synchronization during the course of the experiment, participants differ in how fast one might respond to a certain event or the other person. In other words, participants might not always be perfectly "in sync" whereby one participant might sometimes respond to the other person or vice versa introducing a delay between the responses of two individuals. To account for this, for each segment, the signals of the two participants are lagged in relation to one another. Specifically, the signal of participant 1 is kept constant while the signal of participant 2 is shifted more and more by a specified lag increment until a maximum lag is reached. Next, the same procedure is performed the other way around with participant 2 being kept constant. The maximum lag determines what is still considered synchrony. For example, if the maximum lag is four seconds, responses from two participants that are four seconds apart from each other are still considered synchronized. On the other hand, if one participant reacts to a certain event and the other participant shows a response 5 seconds later, it is not considered a response to the same event anymore and therefore does not count as synchrony. Based on this approach, there are four parameters that need to be determined: (1) the length of each segment, referred to the window size w_{max}; (2) the increment with which the segments are moved along the time series, the window increment w_{inc}; (3) the maximum with which two segments can be lagged

C

from one another, the maximum lag τ_{max}; and (4) the increment with which two segments are lagged from each other, the lag increment τ_{inc}. We determined the parameters following an extensive process by comparing previous studies using similar statistical methods, by looking at what is physiologically plausible given the time course of the physiological signals and by employing a data-driven bottom-up approach where we investigated how changing the parameters affected the outcomes using a different dataset. As expected, the absolute values of the synchrony measures varied depending on the parameters, but as supported by (McAssey et al., 2013), the relative results were not affected (e.g. a dyad manifesting relatively high synchrony showed such tendency for the different parameters). Based on these three factors, we set the parameters as follows: the window size was 8 seconds (160 samples), the window increment was 2 seconds (40 samples), the maximum lag was 4 seconds (80 samples) and the lag increment was 100ms (2 samples).

Calculating the cross correlations of each lag for each window segment generates a result matrix with each row representing one window segment and each column indicating a lag. The middle column represents the cross-correlation with a lag of zero, while the first and last column contain the cross-correlations for the maximum lag of participant 1 and 2. Hence, the number of columns in the result matrix is $(2^* \tau_{max} / \tau_{inc}) + 1$. The number of rows is given by $(N - w_{max} - \tau_{max})/w_{inc}$, with N being the number of observations in the whole time series.

Based on this result matrix, a so-called peak picking algorithm is applied. For each segment (i.e., each row in the matrix), the maximum cross-correlation across the lags is detected closest to the zero-lag (i.e., across all columns in a given row). If that maximum correlation is preceded and followed by smaller correlations, it is marked as a peak. For example, if participant 2 synchronizes with participant 1 with a lag of one second, the cross-correlations will become higher the closer the segments from the two participants are shifted towards the point where they are one second apart from each other. When the two signals are lagged by exactly one second the cross-correlation is highest (the peak). If the signals are lagged further away from each other, the cross-correlation decreases again. If, however, a peak cannot be detected, the algorithm assigns a missing value for that segment. This might be the case, for example, if people do not respond to an event or to each other (e.g., both participants wait and do nothing). The peak picking algorithm outputs a matrix with two columns, containing the value of the maximum cross-correlation (the peak) and the corresponding lag at which the peak cross-correlation is detected. The output has the same number of rows as the result matrix as it searches for a peak cross-correlation for each window segment.

Both the windowed cross-correlations and the peak picking algorithm are conducted four times per dyad, once for the heart rate responses and once for the skin conductance level responses for the face-to-face session and for the face-blocked condition resulting in N_{dyads} * 4 result and peak picking matrices. Finally, the mean of the peak cross-correlations of all window segments (i.e., all rows of the peak picking matrix) is calculated for both physiological measures per Face condition per dyad as the measure of synchrony and is grand-mean centered for the analysis predicting cooperative success.

Appendix C4

Model summary – main analysis

Table C.S2

Model summary of the multilevel linear regression analysis predicting cooperative success based on the level of synchrony in heart rate (HR) and skin conductance level (SCL) and their interaction with Face condition (face-blocked = 0; face-to-face = 1). Feedback condition (feedback no = 0; yes = 1) was included as a control variable and Dyad as a random intercept effect.

	Cooperative success				
Predictors	**Estimates**	**CI**	**t-value**	**df**	**p**
Intercept	5.07	4.85 – 5.29	46.65	49.31	**<0.001**
Feedback condition	0.20	-0.10 – 0.50	1.33	48.51	0.188
Face condition	0.10	0.06 – 0.13	5.47	2890.15	**<0.001**
HR synchrony	0.02	-0.64 – 0.67	0.05	2668.12	0.962
SCL synchrony	-0.01	-0.52 – 0.50	-0.04	2884.94	0.968
HR synchrony * Face condition	0.22	-0.28 – 0.72	0.86	2861.92	0.389
SCL synchrony * Face condition	0.86	0.34 – 1.38	3.24	2882.33	**0.001**
Random Effects					
σ^2	0.18				
$\tau_{00\ Dyad}$	0.28				
ICC	0.61				
N_{Dyad}	50				
Observations	2905				
Marginal R^2 / Conditional R^2	0.033 / 0.619				

Note. SCL = Skin Conductance Level; PPN = participant; CI = 95% confidence interval; σ^2 = residuals; $\tau_{00\ Dyad}$ = random intercept effect for Dyad; ICC = intraclass correlation.

C

Appendix C5

Control analysis – does arousal predict cooperative success?

In the current study we observed that physiological synchrony could predict cooperative success. One possible confound is that it is not the synchrony between two participants, but the co-occurrence of the arousal responses of the two individuals that drive these findings. For example, skin conductance levels might rise if a participant decides to cooperate due to the increased risk of being exploited. Similarly, if the other participant decides to cooperate as well, the same physiological reaction could be expected. Consequently, the responses of the two participants would highly correlate reflecting the individuals' decisions rather than an interpersonal process. To test this, we conducted a control analysis where cooperative success (the joint points won per trial) was regressed against the participants' skin conductance level and their interaction with Face condition (face-blocked = 0; face-to-face = 1). For the skin conductance level, we first standardized the responses per participant and then computed the mean skin conductance level per trial. Consistent with the model of the main analysis, we included the Feedback condition (feedback no = 0; yes = 1) as a control variable and Dyad as a random intercept effect. The model summary is displayed in Table C.S3 which shows that cooperative success could not be predicted by the arousal responses of the two individuals.

Table C.S3

Model summary of the control analysis (multilevel linear regression analysis) with participants' own skin conductance level (SCL PPN) and the interaction with Face condition (face-blocked = 0; face-to-face = 1) predicting cooperative success. Feedback condition (feedback no = 0; yes = 1) was added as a control variable and Dyad was included as a random intercept effect

	Cooperative success				
Predictors	**Estimates**	**CI**	**t-value**	**df**	**p**
Intercept	5.15	4.97 – 5.34	54.41	61.98	**$<$ 0.001**
Feedback condition	0.12	-0.14 – 0.38	0.91	60.95	0.362
Face condition	0.10	0.07 – 0.13	7.00	3566.66	**$<$ 0.001**
SCL PPN1	0.00	-0.02 – 0.03	0.13	3591.01	0.895
SCL PPN2	-0.02	-0.04 – 0.00	-1.64	3585.37	0.100
SCL PPN1 * Face condition	0.02	-0.02 – 0.06	1.12	3603.30	0.262
SCL PPN2 * Face condition	0.03	-0.01 – 0.06	1.39	3597.28	0.164
Random Effects					
σ^2	0.18				
$\tau_{00\ Dyad}$	0.27				
ICC	0.61				
N_{Dyad}	63				
Observations	3634				
Marginal R^2 / Conditional R^2	0.016 / 0.614				

Note. SCL = Skin Conductance Level; PPN = participant; CI = 95% confidence interval; σ^2 = residuals; $\tau_{00\ Dyad}$ = random intercept effect for Dyad; ICC = intraclass correlation.

C

Appendix C6

Control analysis – is the level of synchrony an artifact of the experimental set-up?

Because the heart rate and skin conductance level will always show a certain level of synchrony between participants due to the nature of the signals and the experimental set-up (Moulder et al., 2018), we conducted a control analysis to show that synchrony was elevated due to the interaction itself. Specifically, we compared the original dyads with newly generated dyads (Player 1 from Dyad$_i$ and Player 2 from Dyad$_{i+1}$). Because the trial length varied (there was no time restriction for making a decision), each trial was cut to the shorter trial of the newly generated dyad. Subsequently, the correlation between the responses of the two individuals was calculated per trial per dyad for heart rate and skin conductance level. Finally, we ran an independent t-test on the Fisher-Z-transformed correlation values between the original and the newly generated dyads. As a measure of effect size, we report Cohen's *d*. The results revealed that for both heart rate and skin conductance level synchrony, the level of synchrony was significantly higher in the original dyads compared to the newly generated dyads (HR: $t(3622.7) = 8.06$, $p < .001$, $d = .27$; SCL: $t(3015.5) = 4.38$, $p < .001$, $d = .15$). This indicates that the level of synchrony was due to the interaction rather than the experimental set-up of the study.

Appendix C7

Behavioral results

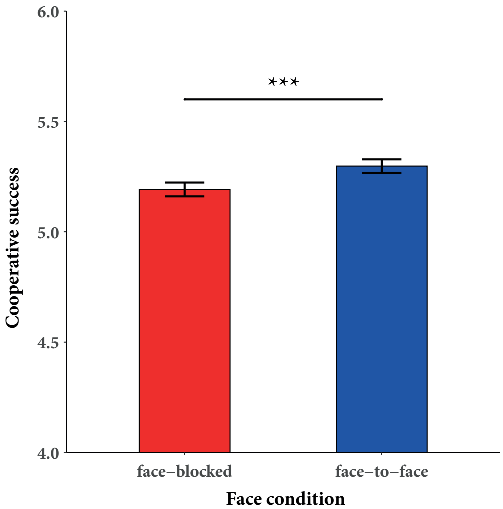


Figure C.S2. The cooperative success rate for the face-blocked and face-to-face conditions with error bars representing 95%-confidence intervals. * $p < .05$; ** $p < .01$; *** $p < .001$.

Appendix D

Supplementary material for Chapter 5

Appendix D1

Choice of comparisons

The initial plan was to make two comparisons for the "change in synchrony" criterion: (1) compare the first baseline measure with the breathing exercise interval, and (2) compare the positive and neutral stories. Regarding the first comparison, the breathing exercise was meant to manipulate synchrony as people were explicitly instructed to breathe synchronously. Although this manipulation worked for the heartrate measure, there were no differences between baseline and breathing intervals evident in the other three signals across parameter configurations. Similarly, based on a previous study showing more synchrony during emotional periods during storytelling, we expected differences in synchrony between the positive and neutral stories in the current study (Kang & Wheatley, 2017). However, across parameter configurations the distance in means between the two conditions were negligible for all four physiological measures, despite significant differences in ratings for both the valence and the intensity ($M_{pos} = 7.53$, $M_{neu} = 5.65$, $t(135) = 15.49$, $p < .001$; $M_{pos} = 5.32$, $M_{neu} = 3.13$, $t(135) = 10.31$, $p < .001$, respectively). We could have used the comparison between the baseline measure and the breathing exercise for the heartrate measure, however, we wanted to use the same conditions across signals to be consistent between measures. Additionally, we wanted to prevent losing collected data and use comparisons that would be comparable for the primary and replication analysis. We therefore decided to include two intervals per condition and compare the two baseline measures with two storytelling intervals. It is important to note that the aim of our study was not to find differences between the conditions to support a theoretical research hypothesis. Instead, we wanted to perform comparisons between conditions where the difference was as large as possible and where variance between parameter configurations could potentially show. Using data without any effects observed across parameter configurations would raise the question of whether the results were due to a lack of actual differences or an insensitive method. Unfortunately, we still faced exactly that dilemma in the results. Nevertheless, we think that the chosen comparisons had the most potential to show effects and possible differences in parameter configurations.

Appendix D2

Descriptive statistics of the questionnaires and story ratings

Table D.S1

Descriptive statistics of the Positive And Negative Affect Schedule (PANAS), the story ratings, the Interpersonal Reactivity Index (IRI), and the Five Facet Mindfulness Questionnaire (FFMQ)

Questionnaire	Mean	SD	Minimum	Maximum
PANAS positive scale:				
Baseline	2.69	.70	1.20	4.60
Positive stories	2.75	.74	1.00	4.40
Neutral stories	2.38	.76	1.10	4.10
PANAS negative scale:				
Baseline	1.56	.44	1.00	2.70
Positive stories	1.29	.38	1.00	2.50
Neutral stories	1.22	.29	1.00	2.40
Story rating (valence):				
Positive stories	7.54	.91	5	9
Neutral stories	5.74	.97	3	8
Story rating (intensity):				
Positive stories	5.46	1.90	1	9
Neutral stories	3.23	1.74	1	7
Questionnaires:				
IRI	3.16	.25	2.47	3.82
FFMQ	2.99	.25	2.25	3.58

Note. SD = standard deviation.

Appendix D3

Replication analysis

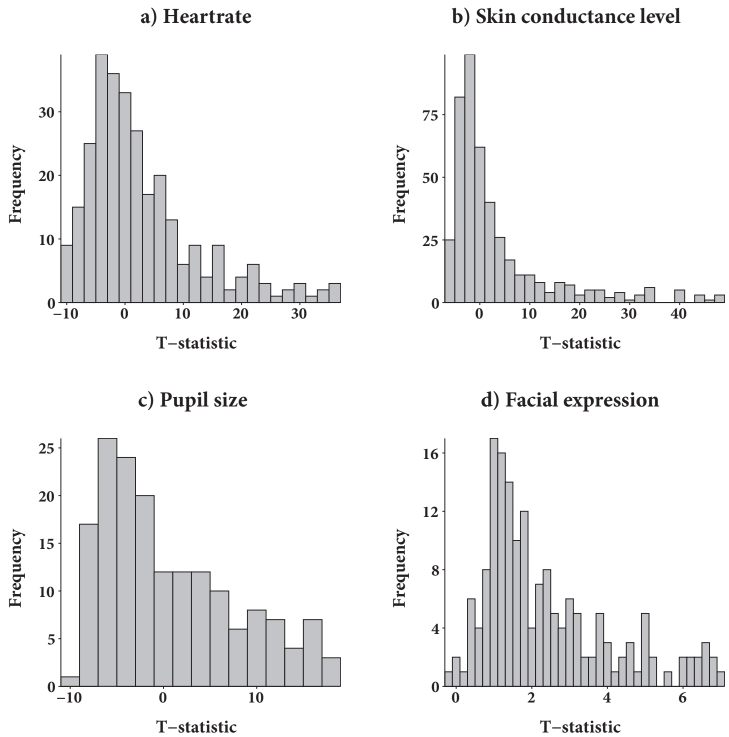


Figure D.S1. Distribution of t-statistics of the comparison between the original and surrogate dyads for each physiological measure (replication analysis). A positive value indicates higher synchrony levels in the original compared to the surrogate dyads. Each data point represents one parameter configuration. For the analyses, data from the second baseline measure were used.

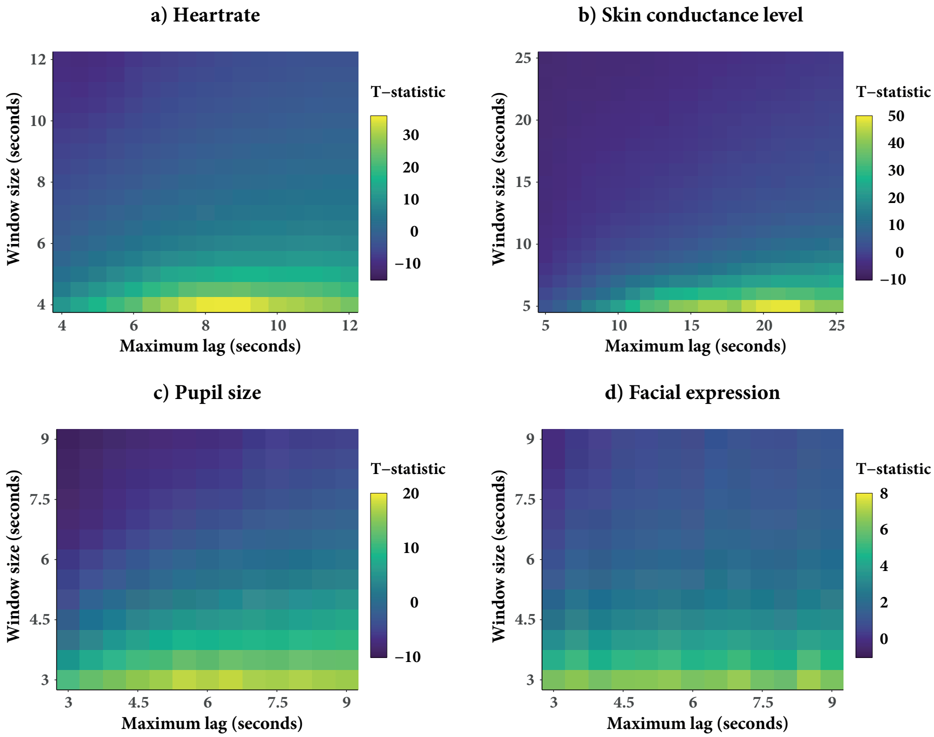


Figure D.S2. Distribution of the t-statistics of the comparison between the originate and surrogate dyads for all parameter configurations and each physiological measure (replication analysis). The color coding runs from the lowest (blue) to the highest (yellow) t-statistic. A positive t-statistic indicates that the original dyads showed higher synchrony levels than the surrogate dyads. The more yellow, the better the discrimination between the original and surrogate dyads. Data from the second baseline measure were used. Notice that the scaling of the axes and the color coding are adjusted to each physiological measure to increase comparability between parameters.

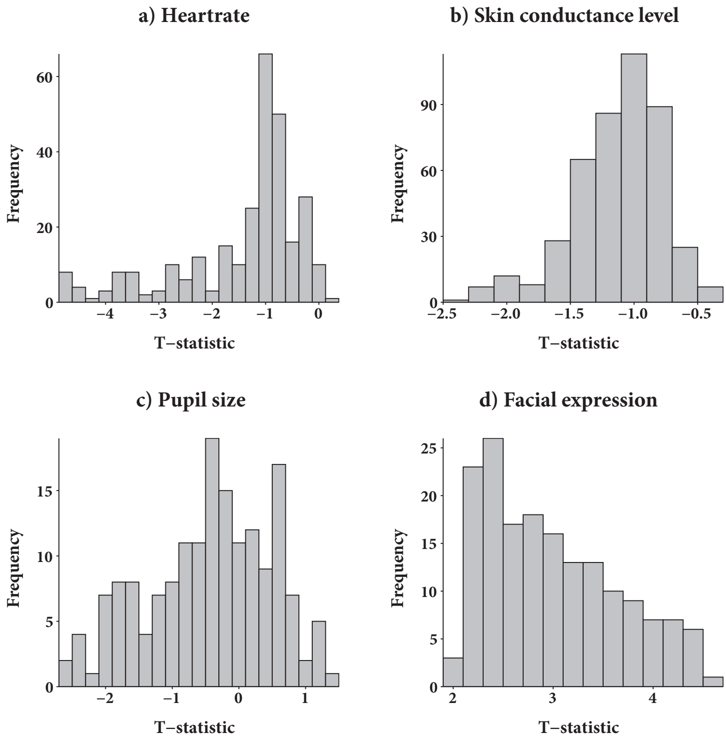


Figure D.S3. Distribution of t-statistics of the comparison between storytelling and baseline for each physiological measure (replication analysis). A positive value indicates higher synchrony levels during storytelling compared to baseline. Each data point represents one parameter configuration. Analysis was based on data from the second and fourth stories and both baseline measures.

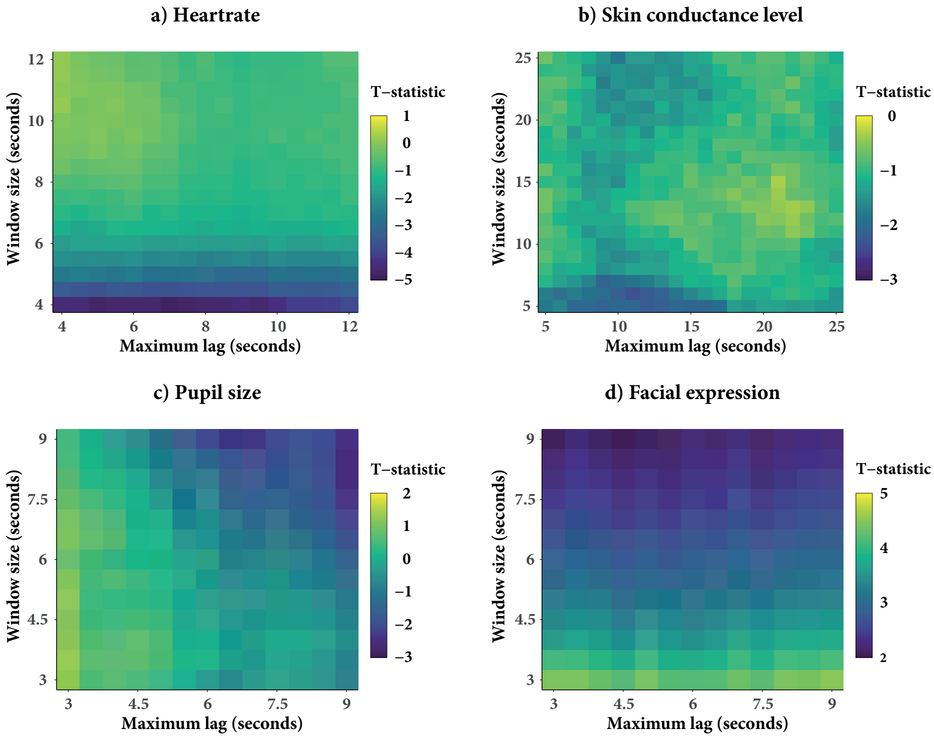


Figure D.S4. Distribution of t-statistics of the comparison between storytelling and baseline of all parameter configurations for each physiological measure (replication analysis). The color coding runs from the lowest (blue) to the highest (yellow) t-statistic. A positive t-statistic indicates that the level of synchrony was higher during storytelling than during baseline. Analysis was based on data from both baseline measures and the second and fourth stories. Notice that the scaling of the axes and the color coding are adjusted to each physiological measure to increase comparability between parameters. Also, the highest t-statistic was not always the highest absolute value with the latter value being discussed in the result section. However, the general idea of greater (absolute) t-statistics indicating better discrimination between the two conditions remains.

References

Achaibou, A., Pourtois, G., Schwartz, S., & Vuilleumier, P. (2008). Simultaneous recording of EEG and facial muscle reactions during spontaneous emotional mimicry. *Neuropsychologia*, *46*(4), 1104–1113. https://doi.org/10.1016/j.neuropsychologia.2007.10.019

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal Processing of Social Information from Faces in Autism. *Journal of Cognitive Neuroscience*, *13*(2), 232–240. https://doi.org/10.1162/089892901564289

Adolphs, R., & Tusche, A. (2017). From Faces to Prosocial Behavior: Cues, Tools, and Mechanisms. *Current Directions in Psychological Science*, *26*(3), 282–287. https://doi.org/10.1177/0963721417694656

Altmann, U. (2011). Investigation of movement synchrony using windowed cross-lagged regression. In *Analysis of verbal and nonverbal communication* (pp. 335–345). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978–3–642–25775–9

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using Self-Report Assessment Methods to Explore Facets of Mindfulness. *Assessment*, *13*(1), 27–45. https://doi.org/10.1177/1073191105283504

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, *54*(1), 39–57. https://doi.org/10.1177/0022002709352443

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex Differences in Cooperation: A Meta-Analytic Review of Social Dilemmas. *Psychological Bulletin*, *137*(6), 881–909. https://doi.org/10.1037/a0025354

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes and Intergroup Relations*, *12*(4), 533–547. https://doi.org/10.1177/1368430209105040

Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, *139*(5), 1090–1112. https://doi.org/10.1037/a0030939

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal Child Psychol. Psychiat.*, *42*(2), 241–251. https://doi.org/https://doi.org/10.1017/S0021963001006643

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290. https://doi.org/10.1177/0963721411422522

Batson, C. D., & Powell, A. A. (2003). Altruism and prosocial behavior. In *Handbook of Psychology* (5th ed., pp. 463–484). John Wiley & Sons, Inc. https://doi.org/10.1016/0005–7967(65)90022–7

Beard, C., Rodriguez, B., Moitra, E., Sibrava, N., Bjornsson, A., Weisberg, R., & Keller, M. (2011). Psychometric Properties of the Liebowitz Social Anxiety Scale (LSAS) in a Longitudinal Study of African Americans with Anxiety Disorders. *Journal of Anxiety Disorders*, *25*(5), 722–726. https://doi.org/10.1016/j.janxdis.2011.03.009

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In *Handbook of Psychophysiology* (pp. 142–162). Press, Cambridge University.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295. https://doi.org/10.1126/science.275.5304.1293

Behrens, F., & Kret, M. E. (2019). The Interplay Between Face-To-Face Contact and Feedback on Cooperation During Real-Life Interactions. *Journal of Nonverbal Behavior*, *43*(4), 513–528. https://doi.org/10.1007/s10919–019–00314–1

Behrens, F., & Kret, M. E. (2020). Under the umbrella of prosocial behavior – a critical comparison of paradigms. https://doi.org/https://doi.org/10.31234/osf.io/9uebc

Behrens, F., Moulder, R. G., Boker, S. M., & Kret, M. E. (2020). Quantifying physiological synchrony through Windowed Cross-Correlation Analysis: Statistical and theoretical considerations. https://doi.org/10.1101/2020.08.27.269746

Behrens, F., Snijdewint, J. A., Moulder, R. G., Prochazkova, E., Sjak-Shie, E. E., Boker, S. M., & Kret, M. E. (2019). Physiological synchrony promotes cooperative success in real-life interactions. https://doi.org/http://dx.doi.org/10.1101/792416

Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field?-Evidence from donations. *Experimental Economics*, *11*(3), 268–281. https://doi.org/10.1007/s10683–007–9192-y

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review*, *98*(4), 459–487. https://doi.org/10.1037/0033–295X.98.4.459

Bicchieri, C., & Lev-on, A. (2007). Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. *Politics, Philosophy & Economics*, *6*(2), 139–168. https://doi.org/10.1177/1470594X07077267

Biel, A., & Thøgersen, J. (2007). Activation of social norms in social dilemmas: A review of the evidence and reflections on the implications for environmental behaviour. *Journal of Economic Psychology*, *28*(1), 93–112. https://doi.org/10.1016/j.joep.2006.03.003

Bixenstine, V. E., & Wilson, K. V. (1963). Effects of level of cooperative choice by the other player on choices in a prisoner's dilemma game. Part II. *Journal of Abnormal and Social Psychology*, *67*(2), 139–147. https://doi.org/10.1037/h0044242

Blanco, M., Engelmann, D., & Normann, H.-T. (2010). A within-subject analysis of other-regarding preferences. In *DICE Discussion Paper* (Vol. 6). Düsseldorf Institute for Competition Economics (DICE). Retrieved from http://hdl.handle.net/10419/41419

Bó, P. (2005). Cooperation under the Shadow of the Future : Experimental Evidence from Infinitely Repeated Games. *The American Economic Review*, *95*(5), 1591–1604.

Böckler, A., Tusche, A., & Singer, T. (2016). The Structure of Human Prosociality: Differentiating Altruistically Motivated, Norm Motivated, Strategically Motivated, and Self-Reported Prosocial Behavior. *Social Psychological and Personality Science*, *7*(6), 530–541. https://doi.org/10.1177/1948550616639650

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, *47*(3), 453–480. https://doi.org/10.1348/014466607X244970

Bogart, K. R., & Matsumoto, D. (2010). Facial mimicry is not necessary to recognize emotion: Facial expression recognition by people with Moebius syndrome. *Social Neuroscience*, *5*(2), 241–251. https://doi.org/10.1080/17470910903395692

Bohnet, I., & Frey, B. S. (1999). The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior & Organization*, *38*(1), 43–57. https://doi.org/10.1016/S0167–2681(98)00121–8

Boker, S. M., & Rotondo, J. L. (2002). Symmetry building and symmetry breaking in synchronized movement. *Mirror Neurons and the Evolution of Brain and Language*, *42*, 163–171. https://doi.org/10.1075/aicr.42.14bok

Boker, S. M., Xu, M., Rotondo, J. L., & King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods*, *7*(3), 338–355. https://doi.org/10.1037/1082–989X.7.3.338

Bonnefon, J. F., Hopfensitz, A., & De Neys, W. (2017). Can We Detect Cooperators by Looking at Their Face? *Current Directions in Psychological Science*, *26*(3), 276–281. https://doi.org/10.1177/0963721417693352

Boone, C., Declerck, C., & Kiyonari, T. (2010). Inducing cooperative behavior among proselfs versus prosocials: The moderating role of incentives and trust. *Journal of Conflict Resolution*, *54*, 799–824. https://doi.org/10.1177/0022002710372329

Boone, R. T., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, *27*(3), 163–182. https://doi.org/https://doi.org/10.1023/A:1025341931128

Bowles, S., & Gintis, H. (2013). *A cooperative species: Human reciprocity and its evolution*. Princeton University Press. https://doi.org/10.1515/9781400838837

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and Motivation I: Defensive and Appetitive Reactions in Picture Processing. *Emotion*, *1*(3), 276–298. https://doi.org/10.1037/1528–3542.1.3.276

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607. https://doi.org/10.1111/j.1469–8986.2008.00654.x

Brosig, J., Weimann, J., & Ockenfels, A. (2003). The effect of Communication Media on Cooperation. *German Economic Review*, *4*(2), 217–241. https://doi.org/10.1111/1468–0475.00080

Buck, R., & VanLear, A. C. (2002). Verbal and nonverbal communication: Distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior. *Journal of Communication*, *52*(3), 522–541. https://doi.org/10.1093/joc/52.3.522

Cappella, J. (1996). Dynamic coordination of vocal and kinesic behavior in dyadic interaction: Methods, problems, and interpersonal outcomes. In *Dynamic patterns in communication processes* (pp. 353–386). Thousand Oaks, CA, US: Sage.

Cashdan, E. (1998). Smiles, speech, and body posture: How women and men display sociometric status and power. *Journal of Nonverbal Behavior*, *22*(4), 209–228. https://doi.org/10.1023/A:1022967721884

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. https://doi.org/10.1016/j.intcom.2012.04.012

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910. https://doi.org/10.1037/0022–3514.76.6.893

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/https://doi.org/10.1037/0033–2909.112.1.155

Coyne, J. C. (1976). Depression and response of others. *Journal of Abnormal Psychology*, *85*(2), 186–193. https://doi.org/https://doi.org/10.1037/0021–843X.85.2.186

Critchley, H. D., & Harrison, N. A. (2013). Visceral Influences on Brain and Behavior. *Neuron*, *77*(4), 624–638. https://doi.org/10.1016/j.neuron.2013.02.008

Critchley, H. D., Rotshtein, P., Nagai, Y., O'Doherty, J., Mathias, C. J., & Dolan, R. J. (2005). Activity in the human brain predicting differential heart rate responses to emotional facial expressions. *NeuroImage*, *24*(3), 751–762. https://doi.org/10.1016/j.neuroimage.2004.10.013

Crone, E. A., Somsen, R. J. M., Van Beek, B., & Van Der Molen, M. W. (2004). Heart rate and skin conductance analysis of antecendents and consequences of decision making. *Psychophysiology*, *41*(4), 531–540. https://doi.org/10.1111/j.1469–8986.2004.00197.x

Damasio, A. R., Everitt, B. J., & Bishop, D. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions: Biological Sciences*, *351*(1346), 1413–1420. https://doi.org/https://doi.org/10.1098/rstb.1996.0125

Danyluck, C., & Page-Gould, E. (2019). Social and Physiological Context can Affect the Meaning of Physiological Synchrony. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598–019–44667–5

Davis, M. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*. https://doi.org/https://doi.org/10.1037/0022–3514.44.1.113

Davis, W. E., Giner-Sorolla, R., Lindsay, D. S., Lougheed, J. P., Makel, M. C., Meier, M. E., … Zelenski, J. M. (2018). Peer-Review Guidelines Promoting Replicability and Transparency in Psychological Science. *Advances in Methods and Practices in Psychological Science*, *1*(4), 556–573. https://doi.org/10.1177/2515245918806489

Dawes, R. M. (1980). Social Dilemmas. *Annual Review of Psychology*, *31*, 169–193. https://doi.org/10.1146/annurev.ps.31.020180.001125

Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In *Handbook of Psychophysiology* (pp. 200–223). Cambridge University Press.

De Dreu, C. K. W., Nijstad, B. A., & Van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review*, *12*(1), 22–49. https://doi.org/10.1177/1088868307304092

de Klerk, C. C. J. M., Hamilton, A. F. d. C., & Southgate, V. (2018). Eye contact modulates facial mimicry in 4-month-old infants: An EMG and fNIRS study. *Cortex*, *106*, 93–103. https://doi.org/10.1016/j.cortex.2018.05.002

De Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, *106*(1), 73–88. https://doi.org/10.1037/a0034251

De Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498–509. https://doi.org/10.1038/nrn.2017.72

Dijk, C., Koenig, B., Ketelaar, T., & de Jong, P. J. (2011). Saved by the blush: Being trusted despite defecting. *Emotion*, *11*(2), 313–319. https://doi.org/10.1037/a0022774

Doesum, N. J. V., Van Lange, D. A. W., & Van Lange, P. A. M. (2013). Social mindfulness: Skill and will to navigate the social world. *Journal of Personality and Social Psychology*, *105*(1), 86–103. https://doi.org/10.1037/a0032540

Drimalla, H., Landwehr, N., Hess, U., & Dziobek, I. (2019). From face to face: the contribution of facial mimicry to cognitive and emotional empathy. *Cognition and Emotion*, *33*(8), 1672–1686. https://doi.org/10.1080/02699931.2019.1596068

Drolet, A. L., & Morris, M. W. (2000). Rapport in conflict resolution: Accounting for how nonverbal exchange fosters cooperation on mutually beneficial settlements to mixed-motive conflicts. *Journal of Experimental Social Psychology*, *36*(1), 26–50. https://doi.org/https://doi.org/10.1006/jesp.1999.1395

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology*, *8*(2), 100–110. https://doi.org/10.1037/1089–2680.8.2.100

Eisenberg, N., & Miller, P. A. (1987). The Relation of Empathy to Prosocial and Related Behaviors. *Psychological Bulletin*, *101*(1), 91–119. https://doi.org/10.1037/0033–2909.101.1.91

Ekman, P. (1997). Should we call it expression or communication? *Innovation*, *10*, 333–344. https://doi.org/10.1080/13511610.1997.9968538

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037//0033–2909.128.2.203

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E. J. M., & Parizel, P. M. (2011). Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, a fMRI study. *Journal of Neuroscience, Psychology, and Economics*, *4*(1), 11–24. https://doi.org/10.1037/a0020151

Epstein, Z., Peysakhovich, A., & Rand, D. G. (2016). The good, the bad, and the unflinchingly selfish: Cooperative decision-making can be predicted with high accuracy when using only three behavioral types. *EC 2016–Proceedings of the 2016 ACM Conference on Economics and Computation*, 547–559. https://doi.org/10.1145/2940716.2940761

Fawcett, C., Arslan, M., Falck-Ytter, T., Roeyers, H., & Gredebäck, G. (2017). Human eyes with dilated pupils induce pupillary contagion in infants. *Scientific Reports*, *7*, 1–7. https://doi.org/10.1038/s41598–017–08223–3

Fawcett, C., Wesevich, V., & Gredebäck, G. (2016). Pupillary Contagion in Infancy: Evidence for Spontaneous Transfer of Arousal. *Psychological Science*, *27*(7), 997–1003. https://doi.org/10.1177/0956797616643924

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., & Louzoun, Y. (2011). Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development*, *34*(4), 569–577. https://doi.org/10.1016/j.infbeh.2011.06.008

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral Markers and Recognizability of the Smile of Enjoyment. *Journal of Personality and Social Psychology*, *64*(1), 83–93. https://doi.org/10.1037/0022–3514.64.1.83

Frank, R. H. (1988). *Passion within reason: the strategic role of the emotions*. WW Norton & Co.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993a). Does Studying Economics Inhibit Cooperation ? *The Journal of Economic Perspectives*, *7*(2), 159–171. https://doi.org/10.1257/jep.7.2.159

Frank, R. H., Gilovich, T., & Regan, D. T. (1993b). The evolution of one-shot cooperation: An experiment. *Ethology and Sociobiology*, *14*(4), 247–256. https://doi.org/10.1016/0162–3095(93)90020-I

Franzen, A., & Pointner, S. (2013). The external validity of giving in the dictator game: A field experiment using the misdirected letter technique. *Experimental Economics*, *16*(2), 155–169. https://doi.org/10.1007/s10683–012–9337–5

Frohlich, N., & Oppenheimer, J. (1998). Some consequences of e-mail vs. face-to-face communication in experiment. *Journal of Economic Behavior & Organization*, *35*, 389–403. https://doi.org/https://doi.org/10.1016/S0167–2681(98)00044–4

Galazka, M. A., Åsberg Johnels, J., Zürcher, N. R., Hippolyte, L., Lemonnier, E., Billstedt, E., … Hadjikhani, N. (2019). Pupillary Contagion in Autism. *Psychological Science*, *30*(2), 309–315. https://doi.org/10.1177/0956797618809382

Gates, K. M., & Liu, S. (2016). Methods for Quantifying Patterns of Dynamic Interactions in Dyads. *Assessment*, *23*(4), 459–471. https://doi.org/10.1177/1073191116641508

Gibbs, R. W., & Van Orden, G. C. (2003). Are emotional expressions intentional? A self-organizational approach. *Consciousness & Emotion*, *4*, 1–16. https://doi.org/10.1075/ce.4.1.02gib

Goldenthal, P., Johnston, R. E., & Kraut, R. E. (1981). Smiling, appeasement, and the silent bared-teeth display. *Ethology and Sociobiology*, *2*(3), 127–133. https://doi.org/10.1016/0162–3095(81)90025-X

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041–210X.12504

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*, 1634–1640. https://doi.org/10.1126/science.1089506

Hatfield, E., Cacioppo, J., & Rapson, R. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–99. https://doi.org/10.1111/1467–8721.ep10770953

Hecht, M., & LaFrance, M. (1998). License or obligation to smile-the effect of power and sex on amount and type of smiling. *Personality and Social Psychology Bulletin*, *24*(12), 1332–1342. https://doi.org/https://doi.org/10.1177/01461672982412007

Hess, U., & Bourgeois, P. (2010). You smile-I smile: Emotion expression in social interaction. *Biological Psychology*, *84*(3), 514–520. https://doi.org/10.1016/j.biopsycho.2009.11.001

Jahng, J., Kralik, J. D., Hwang, D. U., & Jeong, J. (2017). Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the Prisoner's Dilemma Game. *NeuroImage*, *157*(June), 263–274. https://doi.org/10.1016/j.neuroimage.2017.06.024

Järvelä, S., Kivikangas, J. M., Kätsyri, J., & Ravaja, N. (2014). Physiological Linkage of Dyadic Gaming Experience. *Simulation and Gaming*, *45*(1), 24–40. https://doi.org/10.1177/1046878113513080

JASP-Team. (2019). JASP (Version 0.10).

Jorgenson, D. O., & Papciak, A. S. (1981). The effects of communication, resource feedback, and identifiability on behavior in a simulated commons. *Journal of Experimental Social Psychology*, *17*(4), 373–385. https://doi.org/10.1016/0022–1031(81)90044–5

Kang, O., & Wheatley, T. (2015). Pupil dilation patterns reflect the contents of consciousness. *Consciousness and Cognition*, *35*(May 2015), 128–135. https://doi.org/10.1016/j.concog.2015.05.001

Kang, O., & Wheatley, T. (2017). Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *Journal of Experimental Psychology: General*, *146*(4), 569–576. https://doi.org/10.1037/xge0000271

Kaufmann, H. (1967). Similarity and cooperation received as determinants of cooperation rendered. *Psychonomic Science*, *9*(2), 73–74. https://doi.org/https://doi.org/10.3758/BF03330765

Kerby, D. S. (2014). The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. *Comprehensive Psychology*, *3*(1), 1–9. https://doi.org/10.2466/11.it.3.1

Kiesler, S., Sproull, L., & Waters, K. (1996). A prisoner's dilemma experiment on cooperation with people and human-like computers. *Journal of Personality and Social Psychology*, *70*(1), 47–65. https://doi.org/10.1037/0022–3514.70.1.47

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1), 78–100. https://doi.org/10.1037//0033–2909.100.1.78

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Frontiers in Psychology*, *6*(May), 1–10. https://doi.org/10.3389/fpsyg.2015.00711

Kret, M. E., & De Dreu, C. K. W. (2017). Pupil-mimicry conditions trust in partners: Moderation by oxytocin and group membership. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1850), 1–10. https://doi.org/10.1098/rspb.2016.2554

Kret, M. E., Fischer, A. H., & De Dreu, C. K. W. (2015). Pupil Mimicry Correlates With Trust in In-Group Partners With Dilating Pupils. *Psychological Science*, *26*(9), 1401–1410. https://doi.org/10.1177/0956797615588306

Kret, M. E., Stekelenburg, J. J., De Gelder, B., & Roelofs, K. (2017). From face to hand: Attentional bias towards expressive hands in social anxiety. *Biological Psychology*, *122*, 42–50. https://doi.org/10.1016/j.biopsycho.2015.11.016

Kret, M. E., Tomonaga, M., & Matsuzawa, T. (2014). Chimpanzees and Humans Mimic Pupil-Size of Conspecifics. *PLoS ONE*, *9*(8), e104886. https://doi.org/10.1371/journal.pone.0104886

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial Dynamics as Indicators of Trustworthiness and Cooperative Behavior. *Emotion*, *7*(4), 730–735. https://doi.org/10.1037/1528–3542.7.4.730

Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, *14*(4), 334–339. https://doi.org/10.1111/1467–9280.14481

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273. https://doi.org/10.1111/j.1469–8986.1993.tb03352.x

Lee-Helm, J., Miller, J. G., Kahle, S., Troxel, N. R., & Hastings, P. D. (2018). On Measuring and Modeling Physiological Synchrony in Dyads. *Multivariate Behavioral Research*, *0*, 1–23. https://doi.org/10.1080/00273171.2018.1459292

Leiberg, S., Klimecki, O., & Singer, T. (2011). Short-term compassion training increases prosocial behavior in a newly developed prosocial game. *PLoS ONE*, *6*(3). https://doi.org/10.1371/journal.pone.0017798

Lenth, R. V. (2007). *Post Hoc Power : Tables and Commentary*. Department of Statistics and Actuarial Science, University of Iowa.

Lev-on, A., Chavez, A., & Bicchieri, C. (2010). Group and Dyadic Communication in Trust Games. *Rationality and Society*, *22*(1), 37–54. https://doi.org/10.1177/1043463109337100

Levenson, R. W., & Gottman, J. M. (1983). Marital Interaction- PhysiologicalLinkage and Affective Exchange. *Journal of Personality and Social Psychology*, *45*(3), 587–597.

Lewkowicz, D., Quesque, F., Coello, Y., & Delevoye-Turrell, Y. N. (2015). Individual differences in reading social intentions from motor deviants. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01175

Loewenstein, G., & Lerner, J. (2003). The role of affect in decision making. In *Handbook of Affective Sciences* (pp. 619–642). Oxford University Press. https://doi.org/10.1016/B978–0–444–62604–2.00003–4

Lorah, J. (2018). Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large-Scale Assessments in Education*, *6*(8), 1–11. https://doi.org/10.1186/s40536–018–0061–2

Mancini, G., Ferrari, P. F., & Palagi, E. (2013). Rapid facial mimicry in geladas. *Scientific Reports*, *3*, 1–6. https://doi.org/10.1038/srep01527

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

McAssey, M. P., Helm, J., Hsieh, F., Sbarra, D. A., & Ferrer, E. (2013). Methodological advances for detecting physiological synchrony during dyadic interactions. *Methodology*, *9*(2), 41–53. https://doi.org/10.1027/1614–2241/a000053

McClung, J. S., Placì, S., Bangerter, A., Clément, F., & Bshary, R. (2017). The language of cooperation: Shared intentionality drives variation in helping as a function of group membership. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1863). https://doi.org/10.1098/rspb.2017.1682

Mitkidis, 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 and Behavior*, *149*, 101–106. https://doi.org/10.1016/j.physbeh.2015.05.033

Mogan, R., Fischer, R., & Bulbulia, J. A. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *Journal of Experimental Social Psychology*, *72*(September 2016), 13–20. https://doi.org/10.1016/j.jesp.2017.03.009

Mønster, D., Håkonsson, D. D., Eskildsen, J. K., & Wallot, S. (2016). Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology and Behavior*, *156*, 24–34. https://doi.org/10.1016/j.physbeh.2016.01.004

Monterosso, J., Ainslie, G., Pamela Toppi Mullen, P. A. C., & Gault, B. (2002). The fragility of cooperation: A false feedback study of a sequential iterated prisoner's dilemma. *Journal of Economic Psychology*, *23*(4), 437–448. https://doi.org/10.1016/S0167–4870(02)00095–8

Moulder, R. G., Boker, S. M., Ramseyer, F., & Tschacher, W. (2018). Determining synchrony between behavioral time series: An application of surrogate data generation for establishing falsifiable null-hypotheses. *Psychological Methods*, *23*(4), 757–773. https://doi.org/10.1037/met0000172

Myllyneva, A., & Hietanen, J. K. (2015). There is more to eye contact than meets the eye. *Cognition*, *134*, 100–109. https://doi.org/10.1016/j.cognition.2014.09.011

Neal, D. T., & Chartrand, T. L. (2011). Embodied emotion perception: Amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673–678. https://doi.org/10.1177/1948550611406138

Nevicka, B., Ten Velden, F. S., de Hoogh, A. H. B., & van Vianen, A. E. M. (2011). Reality at odds with perceptions: Narcissistic leaders and group performance. *Psychological Science*, *22*(10), 1259–1264. https://doi.org/10.1177/0956797611417259

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2009). Slow echo: facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. *Developmental Science*, *12*(4), 510–520. https://doi.org/10.1111/j.1467–7687.2008.00796.x

Oliva, M., & Anikin, A. (2018). Pupil dilation reflects the time course of emotion recognition in human vocalizations. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598–018–23265-x

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research and Evaluation*, *9*(11). https://doi.org/https://doi.org/10.7275/ktzq-jq66

Palagi, E., Leone, A., Mancini, G., & Ferrari, P. F. (2009). Contagious yawning in gelada baboons as a possible expression of empathy. *Proceedings of the National Academy of Sciences*, *106*(46), 19262–19267. https://doi.org/10.1073/pnas.0910891106

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2017). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review*, *21*(2), 99–141. https://doi.org/10.1177/1088868316628405

Parks, C. D., & Vu, A. D. (1994). Social dilemma behavior of individuals from highly individualist and collectivist cultures. *Journal of Conflict Resolution*, *38*(4), 708–718. https://doi.org/10.1177/0022002794038004006

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial Behavior: Multilevel Perspectives. *Annual Review of Psychology*, *56*(1), 365–392. https://doi.org/10.1146/annurev.psych.56.091103.070141

Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a "cooperative phenotype" that is domain general and temporally stable. *Nature Communications*, *5*, 1–8. https://doi.org/10.1038/ncomms5939

Pierce, T. (2009). Social anxiety and technology: Face-to-face communication versus technological communication among teens. *Computers in Human Behavior*, *25*(6), 1367–1372. https://doi.org/10.1016/j.chb.2009.06.003

Pillutla, M. M., & Chen, X.-P. (1999). Social Norms and Cooperation in Social Dilemmas: The Effects of Context and Feedback. *Organizational Behavior and Human Decision Processes*, *78*(2), 81–103. https://doi.org/10.1006/obhd.1999.2825

Plate, J. D. J., Borggreve, A. S., van Hillegersberg, R., & Peelen, L. M. (2019). Post Hoc Power Calculation: Observing the Expected. *Annals of Surgery*, *269*(1), 11. https://doi.org/10.1097/SLA.0000000000002910

Poncela-Casasnovas, J., Gutiérrez-Roig, M., Gracia-Lázaro, C., Vicens, J., Gómez-Gardeñes, J., Perelló, J., … Sánchez, A. (2016). Humans display a reduced set of consistent behavioral phenotypes in dyadic games. *Science Advances*, *2*(8), 1–9. https://doi.org/10.1126/sciadv.1600451

Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., Aljama-Corrales, T., Charleston-Villalobos, S., & Chon, K. H. (2016). Power Spectral Density Analysis of Electrodermal Activity for Sympathetic Function Assessment. *Annals of Biomedical Engineering*, *44*(10), 3124–3135. https://doi.org/10.1007/s10439–016–1606–6

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–72. https://doi.org/10.1017/S0140525X02000018

Prochazkova, E., & Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience and Biobehavioral Reviews*, *80*(April), 99–114. https://doi.org/10.1016/j.neubiorev.2017.05.013

Prochazkova, E., Prochazkova, L., Giffin, M. R., Scholte, H. S., De Dreu, C. K. W., & Kret, M. E. (2018). Pupil mimicry promotes trust through the theory-of-mind network. *Proceedings of the National Academy of Sciences*, *115*(31), E7265–E7274. https://doi.org/10.1073/pnas.1803916115

Prochazkova, E., Sjak-Shie, E. E., Behrens, F., Lindh, D., & Kret, M. E. (2019). The choreography of human attraction : physiological synchrony in a blind date setting, 1–33. https://doi.org/http://dx.doi.org/10.1101/748707

Quesque, F., Behrens, F., & Kret, M. E. (2019). Pupils say more than a thousand words: Pupil size reflects how observed actions are interpreted. *Cognition*, *190*(April), 93–98. https://doi.org/10.1016/j.cognition.2019.04.016

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, *79*(3), 284–295. https://doi.org/10.1037/a0023419

Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game. *Evolution and Human Behavior*, *33*(3), 200–209. https://doi.org/10.1016/j.evolhumbehav.2011.09.003

Rennung, M., & Göritz, A. S. (2016). Prosocial consequences of interpersonal synchrony: A Meta-Analysis. *Zeitschrift Für Psychologie / Journal of Psychology*, *224*(3), 168–189. https://doi.org/10.1027/2151–2604/a000252

Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., … Pagnoni, G. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia*, *46*(5), 1256–1266. https://doi.org/10.1016/j.neuropsychologia.2007.11.033

Robinson, J. W., Herman, A., & Kaplan, B. J. (1982). Autonomic responses correlate with counselor-client empathy. *Journal of Counseling Psychology*, *29*(2), 195–198. https://doi.org/10.1037/0022–0167.29.2.195

Saleem, M., Anderson, C. A., & Barlett, C. P. (2015). Assessing Helping and Hurting Behaviors Through the Tangram Help/Hurt Task. *Personality and Social Psychology Bulletin*, *41*(10), 1345–1362. https://doi.org/10.1177/0146167215594348

Sally, D. (1995). Conversation and cooperation in social dilemmas–a meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*(1), 58–92. https://doi.org/10.1177/0049124187016001006

Sánchez-Amaro, A., Duguid, S., Call, J., & Tomasello, M. (2019). Chimpanzees and children avoid mutual defection in a social dilemma. *Evolution and Human Behavior*, *40*(1), 46–54. https://doi.org/10.1016/j.evolhumbehav.2018.07.004

Scharlemann, P. W., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, *22*, 617–640. https://doi.org/10.1016/S0167–4870(01)00059–9

Schoeneberger, J. A. (2016). The Impact of Sample Size and Other Factors When Estimating Multilevel Logistic Models. *Journal of Experimental Education*, *84*(2), 373–397. https://doi.org/10.1080/00220973.2015.1027805

Schoenherr, D., Paulick, J., Strauss, B. M., Deisenhofer, A. K., Schwartz, B., Rubel, J. A., … Altmann, U. (2019). Identification of movement synchrony: Validation of windowed cross-lagged correlation and -regression with peak-picking algorithm. *PLoS ONE*, *14*(2), 1–24. https://doi.org/10.1371/journal.pone.0211494

Schoenherr, D., Paulick, J., Worrack, S., Strauss, B. M., Rubel, J. A., Schwartz, B., … Altmann, U. (2018). Quantification of nonverbal synchrony using linear time series analysis methods: Lack of convergent validity and evidence for facets of synchrony. *Behavior Research Methods*, *51*(1), 361–383. https://doi.org/10.3758/s13428–018–1139-z

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Scott-Phillips, T. C., Blythe, R. A., Gardner, A., & West, S. A. (2012). How do communication systems emerge? *Proceedings of the Royal Society B: Biological Sciences*, *279*(1735), 1943–1949. https://doi.org/10.1098/rspb.2011.2181

Sheather, S. (2009). *A modern approach to regression with R*. Springer Science & Business Media. https://doi.org/10.1007/978–0–387–09608–7

Shih, E. W., Quiñones-Camacho, L. E., Karan, A., & Davis, E. L. (2019). Physiological contagion in parent–child dyads during an emotional challenge. *Social Development*, (December 2018), 1–17. https://doi.org/10.1111/sode.12359

Sjak-Shie, E. E. (2017). PhysioData Toolbox 0.3.5.

Smith, D., Schlaepfer, P., Major, K., Dyble, M., Page, A. E., Thompson, J., … Migliano, A. B. (2017). Cooperation and the evolution of hunter-gatherer storytelling. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467–017–02036–8

Sparks, A., Burleigh, T., & Barclay, P. (2016). We can see inside: Accurate prediction of Prisoner's Dilemma decisions in announced games following a face-to-face interaction. *Evolution and Human Behavior*, *37*(3), 210–216. https://doi.org/10.1016/j.evolhumbehav.2015.11.003

Sprecher, S. (2014). Initial interactions online-text, online-audio, online-video, or face-to-face: Effects of modality on liking, closeness, and other interpersonal outcomes. *Computers in Human Behavior*, *31*(1), 190–197. https://doi.org/10.1016/j.chb.2013.10.029

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, *48*(6), 1467–1478. https://doi.org/10.4324/9780203647585

Stoop, J. (2014). From the lab to the field: Envelopes, dictators and manners. *Experimental Economics*, *17*(2), 304–313. https://doi.org/10.1007/s10683–013–9368–6

Sylwester, K., Lyons, M., Buchanan, C., Nettle, D., & Roberts, G. (2012). The role of Theory of Mind in assessing cooperative intentions. *Personality and Individual Differences*, *52*(2), 113–117. https://doi.org/10.1016/j.paid.2011.09.005

Tarr, B., Launay, J., & Dunbar, R. I. M. (2016a). Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and Human Behavior*, *37*(5), 343–349. https://doi.org/10.1016/j.evolhumbehav.2016.02.004

Tarr, B., Launay, J., & Dunbar, R. I. M. (2016b). Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and Human Behavior*, *37*(5), 343–349. https://doi.org/10.1016/j.evolhumbehav.2016.02.004

Tedeschi, J., Lesnick, S., & Gahagan, J. (1968). Feedback and "washout" effects in the Prisoner's Dilemma game. *Journal of Personality and Social Psychology*, *10*(1), 31–34. https://doi.org/10.1037/h0026312

Thorson, K. R., West, T. V., & Mendes, W. B. (2017). Measuring Physiological Influence in Dyads: A Guide to Designing, Implementing, and Analyzing Dyadic Physiological Studies. *Psychological Methods*. https://doi.org/10.1037/met0000166

Thorson, K. R., West, T. V., & Mendes, W. B. (2018). Measuring physiological influence in dyads: A guide to designing, implementing, and analyzing dyadic physiological studies. *Psychological Methods*, *23*(4), 595–616. https://doi.org/10.1037/met0000166

Toma, C., & Butera, F. (2009). Hidden profiles and concealed information: Strategic information sharing and use in group decision making. *Personality and Social Psychology Bulletin*, *35*(6), 793–806. https://doi.org/10.1177/0146167209333176

Van Kleef, G. A. (2010). The Emerging View of Emotion as Social Information. *Social and Personality Psychology Compass*, *4*(5), 331–343. https://doi.org/10.1111/j.1751–9004.2010.00262.x

van Lange, P. A. M. (1999). The Pursuit of Joint Outomes and Equality of Outcomes: An Integrative Model of Social Value Orientation. *Journal of Personality and Social Psychology*, *77*(2), 337–349. https://doi.org/https://doi.org/10.1037/0022–3514.77.2.337

Van Lange, P. A. M. (2000). Beyond Self-interest: A Set of Propositions Relevant to Interpersonal Orientations. *European Review of Social Psychology*, *11*(1), 297–331. https://doi.org/10.1080/14792772043000068

Van Lange, P. A. M., Bekkers, R., Schuyt, T. N. M., & Van Vugt, M. (2007). From games to giving: Social value orientation predicts donations to noble causes. *Basic and Applied Social Psychology*, *29*(4), 375–384. https://doi.org/10.1080/01973530701665223

Van Lange, P. A. M., De Bruin, E. M. N., Otten, W., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, *73*(4), 733–746. https://doi.org/10.1037/0022–3514.73.4.733

Vanutelli, M. E., Gatti, L., Angioletti, L., & Balconi, M. (2017). Affective Synchrony and Autonomic Coupling during Cooperation: A Hyperscanning Study. *BioMed Research International*, 1–10. https://doi.org/10.1155/2017/3104564

Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: the sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, *28*(4), 260–271. https://doi.org/10.1016/j.evolhumbehav.2007.04.006

Vink, R., Hasselman, F., Cillessen, A. H. N., Wijnants, M. L., & Bosman, A. M. T. (2018). Does competence determine who leads in a dyadic cooperative task? A study of children with and without a neurodevelopmental disorder. *Complexity*, *2018*. https://doi.org/10.1155/2018/5379531

Volk, S., Thöni, C., & Ruigrok, W. (2012). Temporal stability and psychological foundations of cooperation preferences. *Journal of Economic Behavior and Organization*, *81*(2), 664–676. https://doi.org/10.1016/j.jebo.2011.10.006

Voncken, M. J., & Bögels, S. M. (2009). Physiological blushing in social anxiety disorder patients with and without blushing complaints: Two subtypes? *Biological Psychology*, *81*(2), 86–94. https://doi.org/10.1016/j.biopsycho.2009.02.004

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022–3514.54.6.1063

Wehebrink, K. S., Koelkebeck, K., Piest, S., De Dreu, C. K. W., & Kret, M. E. (2018). Pupil mimicry and trust–Implication for depression. *Journal of Psychiatric Research*, *97*, 70–76. https://doi.org/https://doi.org/10.1016/j.jpsychires.2017.11.007

Werner, N. S., Duschek, S., & Schandry, R. (2009). Relationships between affective states and decision-making. *International Journal of Psychophysiology*, *74*(3), 259–265. https://doi.org/10.1016/j.ijpsycho.2009.09.010

Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. *Communication Monographs*, *71*(3), 286–310. https://doi.org/10.1080/0363452042000299894

Wood, A., Rychlowska, M., Korb, S., & Niedenthal, P. (2016). Fashioning the Face: Sensorimotor Simulation Contributes to Facial Expression Recognition. *Trends in Cognitive Sciences*, *20*(3), 227–240. https://doi.org/10.1016/j.tics.2015.12.010

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., … Simunovic, D. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes*, *120*(2), 260–271. https://doi.org/10.1016/j.obhdp.2012.06.002

Summary

Cooperation is of great importance in our society and an essential ingredient for the success of humanity. While the news is often filled with horrific misdeeds committed by humans, there are just as many examples of unimaginable acts of cooperation and goodness. People are willing to donate money to people they will never meet, research projects are set up with researchers living on the other side of the world, and dozens of countries are currently fighting together to get the corona virus under control. In this thesis I investigate the question of how nonverbal communication influences how well people work together and how this can best be tested in the lab. The four empirical studies presented build upon each other by holding a magnifying glass over one aspect of the previous study.

The first study discussed in **Chapter 2** describes a methodological paper where I compare tasks (or "games") that measure prosocial behavior in different ways. These tasks have all been used to measure cooperative behavior in the lab. Three of them are so-called social dilemma tasks, where a dilemma is created between the interest of an individual and that of the group by means of simple rules to distribute resources between people. The other tasks are closer to life outside the lab, where participants collect Easter eggs, discuss which candidate is the best fit for a job, and solve puzzles together. By comparing the two types of games, I was able to investigate whether the games measure the same behavior and are therefore interchangeable between studies. The results show that this is not always the case. People who cooperated with others in the social dilemma tasks did not show more prosocial behavior in the more naturalistic games. This difference is best explained by differences in how good people were in a game (e.g., how well they can solve puzzles) and how clear it was *whether* and *how* people could work together. In other words, just like in real life, it was not always a question of wanting to act prosocially, but also of being able to do so. Importantly, I was able to demonstrate that two versions of the social dilemma tasks do measure the same behavior, because I use these tasks in the following chapters to measure cooperative behavior. The only difference between the tasks was that in one version people could choose between working together or not and in the other version they could choose from six options that represented a kind of "scale of wanting to work together".

In **Chapter 3**, I zoom in on cooperative behavior and look at what makes people succeed in working together. One important ingredient is that people not only make verbal or written agreements with each other, but look each other in the eye when they agree to cooperate. It is not without reason that people fly around the world to see each other during negotiations instead of just calling each other. Research supports the efficacy of this behavior and shows that people do work better together when facing each other than when they call or send emails. In the study described in Chapter 3, I go a step further and investigate how this positive effect is influenced by what people know about each other. Past experiences make it easier to predict what a person will do in the future. Likewise, it helps to look at a person in the eyes to estimate whether or not the person can be trusted. I was interested in how these two sources of information are integrated into the decision to work with someone. The results show that both sources had a positive effect on the willingness and success of cooperation. Interestingly, the effects did not influence each

other. People cooperated more when they saw each other regardless of how much they knew about the other person and whether they could find out whether their willingness to cooperate was reciprocated or not. The "boost" in cooperation after seeing each other worked even when they were told that the other person had been selfish before. In other words, the study shows that the positive effect of looking at each other on cooperation is quite robust.

The next question is then of course: What exactly is it in the face that makes people work better together? People have developed a so-called signaling system where nonverbal signals such as body language and facial expressions communicate to the people around us what we think and feel. In addition to the visible signals, there are also many changes within a person that influence how we perceive others and what decisions we make during an interaction with that person. When a man looks at the woman he is in love with, not only does a big smile appear on his face, but his hands start to sweat and his heart starts to beat wildly. Such changes, albeit less extreme, also occur when we make decisions about whether we trust others and consequently want to work with them or not, especially if these decisions have far-reaching consequences.

The visible and invisible changes associated with a decision whether or not to work with someone have been largely investigated through computer tasks. For example, photos of fictional interaction partners are manipulated to investigate the influence of certain signals (e.g., a person who smiles or not). Another method is to look at the nonverbal (physical) reactions of participants while they look at the photos and make decisions. This controlled way of investigating how we express our feelings and intentions and perceive them in others has given us many insights. However, cooperation by definition takes place between at least two people. To understand how people work together successfully, it is therefore necessary to study actual interactions rather than one-person computer tasks. Therefore, the four empirical articles I present in the thesis are based on studies where two people interact with each other, sometimes by playing games to measure their prosocial behavior (Chapters 2 to 4) and sometimes by telling stories (Chapter 5).

Bringing two participants together gives a new perspective to look at nonverbal communication because there is an interaction between the signals from the two people. A person can respond directly to the nonverbal signals from the other person and vice versa. In fact, research shows that people mirror the signals from each other. Such mirroring, also called mimicry or synchrony, takes place at different levels such that people engaged in a social interaction show similar patterns in their behavior (for example, in facial expressions), physiological responses (for example, changes in heart rate), and neural activity. The mirroring ensures that people are able to put themselves in the shoes of another person, to feel their emotions and to adjust their behavior accordingly, for example, by showing empathy or helping.

In the study discussed in **Chapter 4**, I investigated the influence of synchronizing physiological responses on cooperative behavior. Are people who synchronize more successful in working together? The study shows the answer to this question is: yes. Dyads that showed a similar arousal level during the experiment were better at cooperation. This effect was amplified when people looked at each other, that is, when they could exchange nonverbal signals. Arousal level was measured by looking at the skin conduction level of their fingertips. The more someone gets excited, the more the person sweats and the higher the skin conductance level. The fact that people synchronized their arousal level more when they looked at each other compared to when there was a visual cover between them shows that people can pick up subtle changes in their physiological responses through changes in their faces and adjust their own responses to them. Subsequently, even if we are not aware of them, these small changes can affect the way we interact with other people.

The question which then interested me the most was how best to express the synchronization of physiological responses between two people in numbers. The study discussed in **Chapter 5** addresses this question. Ideally, you want to have a measure of how well people synchronize with each other, taking into account the dynamics of a natural conversation. There are two aspects that play a role in this. First, there are changes in the degree of synchrony over time because there are always times when people mirror each other very well and other times when things go less smoothly. Secondly, delays in reactions between two people arise because people do not perfectly synchronize the same reactions at the exact same time. Windowed Cross-Correlation is a statistical analysis that takes both aspects into account. The advantage is that the analysis can be tailored to the signal you are interested in by adjusting certain parameters. For example, changes in skin conductance level are quite slow, so how well people synchronize at this level also changes slowly. On the other hand, if you are interested in mirrored facial expressions, the changes will occur faster because the facial expressions themselves change faster. These differences in the speed of signals can be included in the analysis. However, this advantage is at the same time a drawback because there have been no guidelines on how to choose the parameters. As a result, the parameters diverge considerably between studies while this can have a major impact on the estimated degree of synchronization. In Chapter 5, I present a study setting up these guidelines for four different physiological measurements: heartrate, skin conductance level, pupil size, and facial expressions. Using two criteria, I compare a range of options for the parameters for each measurement and see which are the most suitable. The results show that there is not one optimal parameter setting, but that multiple parameters are appropriate from a statistical point of view. By integrating these findings with theoretical considerations, I develop guidelines for choosing the right parameters.

In summary, the current dissertation shows that successful cooperation is more than the sum of the contributions of two individuals. Our behavior is influenced by how our bodies respond to each other, a process that happens automatically and unconsciously. Whether these results will hold up outside the lab is a question for further research. However, this thesis shows that methodological challenges arise when researchers leave the safe path of the controlled, somewhat artificial setting of the lab. These challenges are not insurmountable, but must be taken into account when researchers want to set up follow-up studies and compare findings where different tasks were used. This dissertation also shows that the appropriate statistical analysis and guidelines for the correct application of analyses can help to make results from different studies more comparable with each other. This brings us a step closer to better understanding complex processes such as nonverbal communication, and its influence on behavior such as cooperation.

Samenvatting

Coöperatie is van groot belang in onze maatschappij en vormt een essentieel ingrediënt voor het succes van de mensheid. Ook al is het nieuws vaak gevuld met afschuwelijke wandaden uitgevoerd door mensen, er zijn net zo veel voorbeelden te vinden van onvoorstelbare daden van samenwerking en goedheid. Mensen doneren geld aan mensen die ze nooit zullen ontmoeten, onderzoeksprojecten worden opgezet met onderzoekers die aan de andere kant van de wereld leven, en tientallen landen strijden samen om het corona-virus onder controle te krijgen. In dit proefschrift onderzoek ik de vraag hoe nonverbale communicatie invloed heeft op hoe goed mensen met elkaar samenwerken en hoe dit het beste getest kan worden in het lab. De vier gepresenteerde empirische artikelen bouwen op elkaar voort door met een vergrootglas op één aspect van het hoofdstuk daarvoor in te zoomen.

De eerste studie die wordt besproken in **hoofdstuk 2** beschrijft een methodologisch paper waar ik taken (of "spellen") vergelijk die op verschillende manieren prosociaal gedrag meten. Deze taken zijn allen vaker gebruikt om coöperatief gedrag in het lab te meten. Drie van de taken zijn zogenaamde sociale dilemma taken die door eenvoudige regels om bronnen te verdelen tussen proefpersonen een dilemma creëren waar mensen moeten kiezen tussen het belang van zichzelf en dat van de groep. De andere taken staan wat dichter bij het leven buiten het lab, waar proefpersonen paaseitjes verzamelen, een discussie voeren over welke kandidaat het beste bij een baan past, en gezamenlijk puzzels oplossen. Door de twee typen spellen te vergelijken kon ik onderzoeken of de spellen hetzelfde gedrag meten en dus uitwisselbaar zijn tussen studies. De resultaten laten zien dat dat niet altijd het geval is. Mensen die veel met anderen samenwerkten in de sociale dilemma taken lieten niet méér prosociaal gedrag zien in de natuurlijkere spellen. De verschillen zijn het beste te verklaren door verschillen in hoe goed mensen waren in een spel (bv. hoe goed ze puzzels kunnen oplossen) of hoe duidelijk het was *of* en *hoe* mensen samen konden werken. Dus, net zoals in het echte leven was het niet altijd een kwestie van willen, maar ook van kunnen. Belangrijk was dat ik kon aantonen dat twee versies van de sociale dilemma taken wél hetzelfde gedrag meten omdat ik deze taken in de volgende hoofdstukken gebruik om coöperatief gedrag te meten. Het enige verschil tussen de taken was dat in de ene versie mensen konden kiezen tussen samenwerken of niet en in de andere versie konden kiezen uit zes opties die een soort "mate van willen samenwerken" representeerden.

In **hoofdstuk 3** zoom ik in op coöperatief gedrag en kijk naar wát mensen goed samen laat werken. Eén ingrediënt wat van groot belang is, is dat mensen niet alleen verbale of schriftelijke afspraken met elkaar maken, maar elkaar in de ogen kunnen kijken als ze een deal maken om samen te werken. Niet voor niets vliegen mensen de wereld rond om elkaar te zien tijdens onderhandelingen terwijl ze ook met elkaar zouden kunnen bellen. Onderzoek geeft deze mensen gelijk en laat zien dat mensen inderdaad beter samenwerken als ze tegenover elkaar staan dan dat ze bellen of emails sturen. In de studie beschreven in hoofdstuk 3 ga ik een stap verder en onderzoek hoe dit positieve effect beïnvloed wordt door wat mensen over elkaar weten. Ervaringen uit het verleden maakt het voorspellen van wat een persoon in de toekomst gaat doen makkelijker. Evenzo helpt het om een persoon in de ogen te kunnen kijken om in te schatten of de persoon wel of niet te vertrouwen is. Ik was geïnteresseerd hoe deze twee bronnen van informatie geïntegreerd zouden worden in de beslissing om met iemand samen te werken. Uit de resultaten bleek dat beide bronnen een positief effect hadden op de bereidheid tot en het succes van samenwerken. Interessant genoeg beïnvloedden de effecten elkaar niet. Mensen coöpereerden meer als ze elkaar zagen onafhankelijk van hoe veel ze over de andere persoon wisten en of ze konden achterhalen of hun bereidheid om samen te werken wederzijds was of niet. De "boost" in coöperatie door elkaar te kunnen zien werkte zelfs als ze te horen kregen dat de andere persoon op een eerder moment egoïstisch was geweest. Met andere woorden, deze studie laat zien dat de werking van elkaar aankijken op hoe goed mensen samenwerken vrij robuust is.

De volgende vraag is dan natuurlijk: Wat is het precies in het gezicht dat mensen beter laat samenwerken? Mensen hebben een zogenaamd signaling-system ontwikkeld waarbij nonverbale signalen zoals lichaamstaal en gezichtsexpressies aan onze medemensen communiceren wat we denken en voelen. Naast de zichtbare signalen zijn er ook veel veranderingen binnen een persoon die beïnvloeden hoe we de ander ervaren en welke beslissingen we tijdens een interactie met die persoon nemen. Als een jongen verliefd naar een meisje kijkt, verschijnt niet alleen een brede glimlach op zijn gezicht, maar zijn handen beginnen te zweten en zijn hart begint wild te kloppen. Dit soort veranderingen vinden ook plaats als we beslissingen maken over of we iemand vertrouwen en vervolgens met degene samenwerken of niet, zeker als deze beslissingen verregaande consequenties hebben.

De zichtbare en onzichtbare veranderingen die gepaard gaan met een beslissing om wel of niet met iemand samen te werken zijn grotendeels onderzocht aan de hand van computer taken. Bijvoorbeeld worden foto's van fictieve interactie partners gemanipuleerd om de invloed van bepaalde signalen te onderzoeken. Ook wordt gekeken naar de nonverbale (lichamelijke) reacties van proefpersonen terwijl ze naar de foto's kijken en een beslissing maken. Deze gecontroleerde manier van onderzoek doen naar hoe we onze gevoelens en intenties uitdrukken en bij anderen ervaren heeft ons veel inzichten gegeven. Echter, coöperatie vindt per definitie tussen minstens twee personen plaats. Om te begrijpen hoe mensen succesvol samenwerken is het dus noodzakelijk om onderzoek te doen naar daadwerkelijke interacties in plaats van naar een-persoons-computer taken. Daarom zijn de vier empirische artikelen die ik in de proefschrift presenteer gebaseerd op studies waar altijd twee mensen met elkaar interacteren, soms door spellen te spelen om hun prosociaal gedrag te meten (hoofdstukken 2–4) en soms door elkaar verhalen te vertellen (hoofdstuk 5).

Het samenbrengen van twee proefpersonen geeft een nieuw perspectief om naar nonverbale communicatie te kijken omdat er een wisselwerking ontstaat tussen de signalen van de twee personen. Een persoon kan direct reageren op de nonverbale signalen van de ander persoon en omgekeerd. Sterker nog, uit onderzoek blijkt dat mensen de signalen van elkaar spiegelen. Het spiegelen, ook mimicry of synchrony genoemd, van signalen vindt plaats op verschillende niveaus zodat mensen die in een sociale interactie zijn verwikkeld vergelijkbare patronen tonen in hun gedrag (bijvoorbeeld in gezichtsexpressies), fysiologische reacties (bijvoorbeeld veranderingen in hartslag), en neurale activiteit. Het spiegelen leidt ertoe dat mensen zich kunnen verplaatsen in een ander persoon, hun emoties kunnen voelen en hun gedrag hier vervolgens op aanpassen door bijvoorbeeld empathie te tonen of te helpen.

In de studie die in **hoofdstuk 4** wordt besproken heb ik gekeken naar de invloed van het synchroniseren van fysiologische reacties op coöperatief gedrag. Zijn mensen die elkaar meer synchroniseren succesvoller in samenwerken? Uit deze studie blijkt het antwoord op deze vraag: ja. Koppels die een vergelijkbaar opwindingsniveau lieten zien tijdens het experiment, waren beter in samenwerken. Dit effect werd versterkt als mensen elkaar aankeken, dus als ze nonverbale signalen konden uitwisselen. Het opwindingsniveau werd gemeten door naar het huidgeleidingsniveau van hun vingertoppen te kijken. Hoe meer iemand zich opwindt, hoe meer degene zweet en hoe hoger het huidgeleidingsniveau ligt. Het feit dat mensen elkaars opwindingsnivau meer spiegelden als ze elkaar konden aankijken in vergelijking met als er een schot tussen hen zat, laat zien dat mensen subtiele veranderingen in hun fysiologische reacties door veranderingen in hun gezicht kunnen oppikken en hun eigen reacties erop aanpassen. Vervolgens kunnen deze kleine veranderingen, ook al zijn we ons er niet van bewust, de manier van hoe we met andere mensen omgaan beïnvloeden.

De vraag die mij vervolgens vooral interesseerde was hoe je het beste de synchronisatie van fysiologische reacties tussen twee personen kunt uitdrukken in getallen. De studie die in **hoofdstuk 5** besproken wordt richt zich op deze vraag. Het liefst wil je een maat hebben van hoe goed mensen elkaar spiegelen en daarbij de dynamiek van een natuurlijke conversatie meenemen. Er zijn twee aspecten die hierbij een rol spelen. Ten eerste treden er veranderingen op in de mate van synchronisatie over tijd omdat er altijd momenten zijn waar mensen elkaar erg goed spiegelen en andere momenten waar het minder goed gaat. Ten tweede ontstaan vertragingen in reacties tussen twee personen doordat mensen niet perfect synchroon dezelfde reacties vertonen. Windowed Cross-Correlation is een statistische analyse die beide aspecten meeneemt. Het voordeel is dat de analyse kan worden toegespitst op het signaal waar je geïnteresseerd in bent door bepaalde parameters aan te passen. Bijvoorbeeld het huidgeleidingsniveau is vrij traag, dus hoe goed mensen synchroniseren op dit niveau verandert ook langzaam. Aan de andere kant, als je geïnteresseerd bent in het spiegelen van gezichtsuitdrukkingen zullen de veranderingen sneller optreden omdat de gezichtsuitdrukkingen zelf sneller veranderen. Deze verschillen in de snelheid van signalen kunnen worden meegenomen in de analyse. Dit voordeel is echter tegelijkertijd een nadeel omdat er geen richtlijnen waren van hoe de parameters gekozen moesten worden. Als gevolg hiervan lopen de parameters behoorlijk uit elkaar tussen studies terwijl dit grote invloed kan hebben op de mate van synchronisatie zoals deze gemeten wordt. In hoofdstuk 5 presenteer ik een studie die deze richtlijnen opzet voor vier verschillende fysiologische metingen: hartslag, huidgeleidingsniveau, pupilgrootte en gezichtsexpressies. Aan de hand van twee criteria vergelijk ik voor elke meting een reeks mogelijkheden voor de parameters en kijk welke het meest geschikt zijn. Uit de resultaten blijkt dat er geen optimale parameter setting gekozen kan worden, maar dat meerdere parameters goed zijn vanuit een statistisch oogpunt. Door deze bevindingen te integreren met theoretische overwegingen ontwikkelen we richtlijnen om de juiste parameters te kiezen.

Samengevat toont de huidige dissertatie aan dat een succesvolle samenwerking meer is dan de som van de bijdragen van twee individuen. Ons gedrag wordt beïnvloed door hoe onze lichamen op elkaar reageren, een proces dat automatisch en onbewust plaatsvindt. Of deze resultaten ook standhouden buiten het lab is een vraagstuk voor verder onderzoek. Dit proefschrift laat echter zien dat er methodologische uitdagingen optreden als onderzoekers het veilige pad van de gecontroleerde, ietwat artificiële setting van het lab verlaten. Deze uitdadingen zijn niet onoverkomelijk, maar moeten wel meegenomen worden als onderzoekers vervolgstudies opzetten en bevindingen willen vergelijken waar verschillende taken gebruikt worden. Ook laat deze dissertatie zien dat de juiste statistische analyse en het opzetten van richtlijnen van het correct toepassen van analyses kunnen helpen de vergelijking van resultaten tussen studies te vergroten. Dit brengt ons een stap dichterbij om complexe processen zoals nonverbale communicatie en de invloed op gedrag zoals coöperatie beter te begrijpen.

Curriculum Vitae

Friederike Behrens was born in Bad Oldesloe, Germany (1989) and graduated from "Oedeme Gymnasium" high school in Lüneburg in 2009. She completed her Bachelor of Science in Psychology (2009–2012), Master of Science in Psychology (2012–2015), and Research Master of Behavioural and Social Sciences (2013–2015) at Rijksuniversiteit Groningen, the latter with distinction. For the first master thesis, Friederike investigated the role and temporal profile of the early visual cortex in perception with a link to blindsight by means of transcranial magnetic stimulation (TMS). This project was supervised by dr. J. Jolij and dr. J.H.C Heutink. For the research project of her second master thesis, she had the opportunity to visit the lab of dr. W.J. Matthews at University of Cambridge to conduct a MEG/EEG study, where she investigated the effect of repetition and expectation on time perception. This project was co-supervised by prof. dr. D.H. van Rijn at Rijksuniversiteit Groningen.

During her PhD research at the unit Cognitive Psychology of the Institute of Psychology (Leiden University, 2015 - 2020), Friederike completed four studies to collect the data presented in this dissertation. During this time, she learned the necessary techniques to collect, pre-process, and analyze physiological data. In her third year, she visited the lab of prof. dr. S.M. Boker at University of Virginia to deepen her knowledge in statistical methods such as the Windowed Cross-Correlation analysis. Friederike advanced this analysis together with one of the lab members, dr. R.G. Moulder, to tailor it to the context of physiological synchrony. Besides conducting research, she also enjoyed being involved in extracurricular activities such as representing the PhD students of the faculty as a member of the PhD Platform board and organizing PhD Career events.

After her PhD, Friederike decided to explore the career possibilities outside the university and started a new job as a data scientist in Groningen, where she can apply her statistical knowledge gained during her PhD to varies projects outside academia.

List of publications

Behrens, F., & Kret, M. E. (2019). The interplay between face-to-face contact and feedback on cooperation during real-life interactions. *Journal of Nonverbal Behavior*, *43*, 513–528.

Huntjens, R. J., Wessel, I., Ostafin, B. D., Boelen, P. A., **Behrens, F.**, & van Minnen, A. (2016). Trauma-related self-defining memories and future goals in Dissociative Identity Disorder. *Behaviour Research and Therapy*, *87*, 216–224.

Massen, J. J., **Behrens, F.**, Martin, J. S., Stocker, M., & Brosnan, S. F. (2019). A comparative approach to affect and cooperation. *Neuroscience & Biobehavioral Reviews*, *107*, 370–387.

Quesque, F., **Behrens, F.**, & Kret, M. E. (2019). Pupils say more than a thousand words: Pupil size reflects how observed actions are interpreted. *Cognition*, *190*, 93–98.

Sense, F., **Behrens, F.**, Meijer, R. R., & van Rijn, H. (2015). Stability of Individual Parameters in a Model of Optimal Fact Learning. In *Proceedings of the 13th International Conference on Cognitive Modeling*, 136–141.

Sense, F., **Behrens, F.**, Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in cognitive science*, *8*, 305–321.

Manuscripts in preparation

Behrens, F., & Kret, M.E. (2020). Under the umbrella of prosocial behavior – a critical comparison of paradigms. *PsyArXiv*.

Behrens, F., Moulder, R. G., Boker, S. M., & Kret, M. E. (2020). Quantifying Physiological Synchrony through Windowed Cross-Correlation Analysis: Statistical and Theoretical Considerations. *bioRxiv*.

Behrens, F., Snijdewint, J. A., Moulder, R. G., Prochazkova, E., Sjak-Shie, E. E., Boker, S. M., & Kret, M. E. (in press). Physiological synchrony is associated with cooperative success in real-life interactions. *Scientific Reports*.

Prochazkova, E., Sjak-Shie, E. E., **Behrens, F.**, Lindh, D., & Kret, M. E. (submitted). Physiological synchrony predicts attraction in a blind date setting.