SOCIAL ATTENTION IN THE REAL WORLD

THEORETICAL AND EMPIRICAL LIMITS

Gijs A. Holleman

SOCIAL ATTENTION IN THE REAL WORLD: THEORETICAL AND EMPIRICAL LIMITS

Gijs A. Holleman

ISBN: 978-94-6416-728-3
Cover design: Frederique Scholtes
Chapter illustrations: Roos Holleman
Lay-out: Publiiss | www.publiiss.nl
Print: Ridderprint | www.ridderprint.nl

© Copyright 2021: Gijs A. Holleman

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording, or otherwise, without the prior written permission of the author.

SOCIAL ATTENTION IN THE REAL WORLD: THEORETICAL AND EMPIRICAL LIMITS

Sociale aandacht in de echte wereld

Theoretische en empirische grenzen
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 3 december 2021 des middags te 2.15 uur

door

Gijs A. Holleman

geboren op 22 mei 1991 te Tilburg

Promotores

Prof. dr. C. Kemner

Prof dr. M. Deković

Copromotores

Dr. R. S. Hessels

Dr. I. T. C. Hooge

Table of Contents

Chapter 1:	General introduction	7
Chapter 2:	The “real-world” approach and its problems	17
Chapter 3:	The reality of “real-life” neuroscience	41
Chapter 4:	Implying social interaction	51
Chapter 5:	Gaze and speech in parent-child interactions	89
Chapter 6:	Gaze behavior and turn-taking in parent-child conversations	125
Chapter 7:	General discussion	153
Chapter 8:	Nederlandse samenvatting	167
Chapter 9:	Dankwoord	177
Chapter 10:	List of publications	183
Chapter 11:	Curriculum Vitae	187
Chapter 12:	Bibliography	191

Chapter 1

General introduction

BACKGROUND AND AIM OF THIS DISSERTATION

This thesis is about social interaction. Social interactions are ubiquitous in nature, from the division of labor in ant colonies (Möglich & Hölldobler, 1974), to parental caring behavior in penguins (Ancel, Beaulieu, & Gilbert, 2013), and conflict resolution among primates (De Waal, 1989). What constitutes a social interaction? A social interaction can be defined as an encounter between two or more individuals, whether within or between species, characterized by coordinated and reciprocal patterns of behavior. In this thesis, I am interested in human social interaction. Higher primates, and especially human beings, exhibit a remarkable capacity for social interaction, as is evident from how they communicate, cooperate, compete, and dispute with each other. A primary mode of human social interaction is face-to-face interaction. Face-to-face interactions consist of a variety of verbal and nonverbal behaviors, including speech acts, vocalizations, eye contact, gazing, facial displays, gestures, and other bodily movements which together facilitate and regulate interpersonal communication (Duncan & Fiske, 2015; Goffman, 1967a; Kendon, 1990; Levinson & Holler, 2014).

One research field in which scientists study aspects of human social interaction is colloquially known as the field of ‘social attention’ (Langton, Watt, & Bruce, 2000; Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012). A central idea in this field of study is that gaze behavior - where someone looks and when - is not only important to how humans perceive and acquire information about the world, but can also signal valuable information to others (Argyle & Cook, 1976; Kendon, 1967; Risko, Richardson, & Kingstone, 2016). The human eyes - their physical appearance and movements, as well as the features of the face to which they belong - are important sources of information when perceived by another person. This ‘dual function’ of gaze to signal and perceive simultaneously has important implications for social cognition, which refers to the psychological capacity of understanding the thoughts, actions, and goals of others. For example, someone’s gaze direction can be used by another person to infer where that someone is looking, and observing someone’s eyes and face can provide insights about that person’s cognitive or emotional state (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Stephenson, Edwards, & Bayliss, 2021)¹. Over the past decades, researchers in the

¹ In everyday language, there are many verbs that imply the information-acquisition (perceptual) function of the eyes, such as *gaze*, *watch*, *stare*, *peer*, *glance*, *scan* and *peek*, while other verbs imply the information-signalling function of the eyes. For example, to *glower*, *glare*, or *scowl* implies looking at another person with anger or discontent. Eyes can express a sullen or gloomy look (resentful, grumpy, joyless), a doe eyed look (naïve, guileless), eyes can look heavy (sad or tired), or twinkle with excitement and joy.

field of social attention have heavily relied on a set of laboratory procedures, or so-called ‘experimental paradigms’², to investigate aspects of human face-to-face interaction, e.g., face perception, joint attention, and facial emotion recognition. For illustrative purposes, two examples of popular experimental paradigms used in the field of social attention are described below.

Example 1: The gaze-cueing paradigm

The gaze-cueing paradigm has been used extensively to study aspects of joint attention. Joint attention occurs when two or more individuals are directing attention to the same event or object in the environment, which can be induced by nonverbal communicative signals such as gaze and gestures. Figure 1 depicts a single trial from a gaze-cueing experiment. A participant is seated in front a computer screen. First, a first a fixation cross appears centrally on the screen. After approximately 500 milliseconds, a face is presented with its gaze directed straight ahead, and then with its gaze directed towards the left or the right side of the screen. Next, a target object (e.g., a letter) appears on the left or the right side of the screen. After the target appears, the participant must respond as fast as possible by pressing a key press to indicate at what side of the screen the target appeared. Results from gaze-cueing experiments typically show that reaction times of participants to determine the target location is shorter when the target appears in the gazed-at location (Friesen & Kingstone, 1998). This is usually regarded as evidence that someone’s gaze direction is a powerful cue which can trigger shifts of attention in the observer (for a review, see Frischen, Bayliss, & Tipper, 2007).

² The notion of experimental paradigm, as I use it here, is based on (but is certainly not the same as) how the historian and philosopher of science Thomas Kuhn describes his concept of *scientific paradigm*, which refers to how researchers within a scientific discipline are for a period of time “...committed to the same rules and standards for scientific practice...” (Kuhn, 2012, p. 11), until a revolutionary period (a paradigm shift) occurs and the consensus about the current rules, standards, questions, results, and theories within that scientific community undergoes changes.

Figure 1. A single trial from a gaze-cueing experiment. Adapted from Driver et al. (1999) and Zhang et al. (2020). The trial begins with a fixation cross. After the fixation cross, a face with direct gaze is presented for 500 milliseconds in the center of the screen. Next, the face directs its gaze to either the left or the right side of the screen. This is the ‘gaze cueing’ element. After this, a target appears at either the left side or the right side of the face. If the target appears at the same side as the face directs its gaze, this is called an ‘valid’ or ‘congruent’ trial. If the target appears at the opposite side of the face’s gaze direction, this is called an ‘invalid’ or ‘incongruent’ trial. In the final display, it is visible that the target appeared at the same side as the gaze cue, hence a ‘valid’ trial. The task of the participant is to respond as fast as possible by means of a key press to indicate where (i.e., left or right) the target appeared.

Example 2: The facial emotion-recognition paradigm

Similarly, researchers have used a set of experimental procedures to study how humans perceive and process emotional expressions from human eyes and faces (Ekman, 1976; Ekman, Friesen, & Ellsworth, 2013; Mavratzakis, Herbert, & Walla, 2016; Terburg, Hooiveld, Aarts, Kenemans, & van Honk, 2011). In typical facial emotion-recognition experiments, participants are asked to either look at, judge, match, recognize, or describe the emotional expressions of face images which are presented to them on a computer screen. This experiment usually goes as follows: after being presented with a fixation cross

(see previous example), a face with an emotional expression is displayed in the center of the screen. In some experiments, participants need to indicate whether the emotional expression matches with a given written emotion: e.g., happy or sad. In other versions, the participants may need to indicate the emotional expressions from a list or describe what emotional expression they perceive. For a study in which researchers used different versions, see Mavratzakis, Herbert & Walla (2016). Now try it for yourself. What facial emotional expressions are depicted in Figure 2, panels A, B, C, D, and E?

Figure 2. Face images used in emotion-recognition experiments. What emotions are depicted?
Answers: A. Sad B. Happy C. Surprise D. Anger E. Boredom, or the author's facial expression after > 4 alcoholic beverages.

Based on such experimental paradigms, in which human subjects are presented with images or videos of human faces on a computer screen in a controlled laboratory setting, researchers have drawn conclusions about how human faces and eyes attract and retain visual attention (Bindemann, Burton, Hooge, Jenkins, & De Haan, 2005; Birmingham, Bischof, & Kingstone, 2009; Senju, Hasegawa, & Tojo, 2005), what information is used to recognize emotional and mental states from eyes and faces (Baron-Cohen et al., 2001; Ekman et al., 2013; Leppänen & Nelson, 2009), what features of the face are looked at when learning or recognizing faces (Arizpe, Walsh, Yovel, & Baker, 2017; Henderson, Williams, & Falk, 2005; Peterson & Eckstein, 2013), how gaze direction is inferred from eye and head orientation (Langton, 2000; Langton et al., 2000; Stephenson et al., 2021), how someone's gaze direction triggers reflexive shifts of attention (Driver et al., 1999;

Friesen & Kingstone, 1998; Frischen et al., 2007), what areas of the face are looked at when listening to someone speak (Buchan, Paré, & Munhall, 2007; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998; Vő, Smith, Mital, & Henderson, 2012), and what areas in the human brain are involved when processing visual displays of gaze direction and facial expressions (George, Driver, & Dolan, 2001; Senju & Johnson, 2009a, 2009b; Sprengelmeyer, Rausch, Eysel, & Przuntek, 1998).

However, while researchers have extensively used such experimental laboratory paradigms in past decades to investigate various aspects human social interactions, such as face perception, joint attention, and facial emotion recognition (see e.g., Ekman et al., 2013; Frischen et al., 2007; Senju & Johnson, 2009a, 2009b; Stephenson et al., 2021), other researchers have begun to raise concerns about the validity and generalizability of these experimental paradigms, and have questioned whether the conclusions and theories derived from these paradigms accurately describe and predict how these aspects of human social interaction function in the ‘real world’ (see e.g., Kompatsiari, Ciardo, Tikhonoff, Metta, & Wykowska, 2018; Osborne-Crowley, 2020; Risko et al., 2012; Risko et al., 2016; Schilbach et al., 2013; Shamay-Tsoory & Mendelsohn, 2019). For example, Kompatsiari et al. (2018, p. 1) wrote that:

“the majority of studies examine JA [joint attention] through offline protocols with social stimuli presented statically on a screen. However, screen-based offline paradigms lack ecological validity, and they might fail to capture true social cognitive mechanisms evoked in natural social interactions.”

Likewise, in a recent review by Shamay-Tsoory and Mendelsohn (2019, p. 4) it was argued that:

“Conclusions drawn from controlled experimental designs with a limited number of variables may not be valid in real-life behavior... an experiment that tests emotion recognition in facial expression in a computerized task with an artificial, decontextualized, still picture may not be applicable to emotion recognition in naturalistic settings.”

The statements by Kompatsiari et al. (2016) and Shamay-Tsoory & Mendelsohn (2019) clearly indicate that they have concerns about the validity, applicability, and generalizability of results and conclusions based on experimental research with some experimental laboratory paradigms. I wondered to what extent their concerns are justified, given that such experimental paradigms have been used extensively by researchers in the field of social attention over the past decades. Specifically, if it is true that some (if not most)

experimental paradigms are not adequately suited to capture relevant aspects of how people interact with each other, then how should researchers in the field of social attention proceed to investigate the relevant aspects of human social interactions? What are the possibilities to study cognitive and behavioral aspects of human social interaction in a way that does allow one to draw valid conclusions? In this thesis, I aim to answer these questions. As we shall see, discussions about the validity and generalizability of experimental research have a long pedigree in psychological science, and the nature of this discussion raises theoretical, methodological, technical, and even meta-scientific questions about the ways by which psychologists and cognitive scientists can study aspects of human cognition and behavior. But before we will dive deeper into these issues, I shall first briefly describe my personal motivation behind the topics discussed in the current thesis.

MOTIVATION

During my master's program *Applied Cognitive Psychology* at Utrecht University in 2015 – 2016, I first became interested in the field of social attention. For my master's thesis, I had the opportunity to carry out an experiment together with Roy Hessels, Ignace Hooge, Tim Cornelissen, and Chantal Kemner. We used a dual eye-tracking setup to study how people look at each other in video-mediated face-to-face interactions (Hessels, Holleman, Cornelissen, Hooge, & Kemner, 2018). During this project, I learned from firsthand experience that designing and conducting research to study aspects of gaze behavior in human social interaction is not an easy task. Social attention research is still very much a field in development. Within the field there has been, and continues to be, a lot of discussion about what experimental methodology is adequately suited to investigate the cognitive, behavioral, and neurobiological processes that underlie human social interaction (for reviews and discussions, see e.g., Osborne-Crowley, 2020; Pfeiffer, Vogeley, & Schilbach, 2013; Risko et al., 2012; Risko et al., 2016; Schilbach, 2015; Schilbach et al., 2013; Shamay-Tsoory & Mendelsohn, 2019). Thus, when I started as a PhD-candidate in 2017 with the goal to develop new experimental paradigms to study the role of gaze behavior in face-to-face interaction, I was already faced with two problems. First, if one decides to conduct research in the field of social attention, how does one know whether one's experimental methodology is appropriate to study relevant aspects of gaze behavior and its role in human social interactions? Second, given the different views and opinions regarding the validity, applicability, and generalizability of experimental laboratory paradigms, how can one effectively deal with potential criticism on one's methodological decisions?

Being interested in the history and philosophy of psychological science, I soon came to understand that this discussion – whether some experimental laboratory paradigms used by researchers may or may not be adequate to study aspects of human cognition and behavior in the ‘real-world’– is by no means unique to the field of social attention. In fact, this so-called ‘real-world or the lab’-dilemma (Banaji & Crowder, 1989; Hammond & Stewart, 2001; Kingstone, Smilek, & Eastwood, 2008; Neisser, 1976; Shamay-Tsoory & Mendelsohn, 2019) has been a recurring theme in the history of psychological science. Therefore, I ask what researchers in the field of social attention may learn from the arguments and discussions about experimental methodology in other research fields of psychology. How have researchers in psychological science argued about these questions in the past, and how can knowledge of this history guide the directions of future research? Since my goal was to develop new eye-tracking paradigms to study gaze behavior in face-to-face interactions, I came to the conclusion that better understanding of this discussion would also benefit my own research and would inform other researchers in the field of social attention.

OUTLINE OF DISSERTATION

This dissertation consists of two parts. In **Part 1**, I examine some of the conceptual and methodological issues underlying the ‘real-world or the laboratory’-dilemma in psychological science. In **Chapter 2**, I focus on the much-debated concept of *ecological validity*, which has dominated the ways in which researchers have addressed the correspondence between human behavior and cognition as studied in the ‘laboratory’ and human behavior and cognition as it supposedly occurs in the ‘real world’. I argue that this concept has caused considerable confusion about how researchers think and make statements about the validity and generalizability of research findings. I show how and where this discussion has the tendency to go astray, and I offer an alternative to move beyond the so-called ‘real-world or the lab’-dilemma which has permeated throughout the history of psychology. In **Chapter 3**, I apply my analysis of the concept of ecological validity to a recent discussion in the field of social attention and social cognition. To do so, I contrast the contemporary usage of the term ecological validity with its original formulations by the Hungarian-American psychologist Egon Brunswik (Brunswik, 1955, 1956b; Hammond & Stewart, 2001).

In **Part 2** of this dissertation, three empirical studies on different aspects of gaze behavior in face-to-face interactions are presented. These studies are informed by the theoretical

discussions of Part 1. In short, the ‘real-world or ‘the lab’-dichotomy proves to be an unconstructive model to address the fundamental problem of generalizability (as discussed in Chapter 2 and 3). Therefore, to investigate the role of gaze behavior in face-to-face interactions, one must be specific about the particular social and behavioral contexts in which one is interested. To do so, I formulate several sub questions to study specific aspects of human gaze behavior and its role in face-to-face interaction. In **Chapter 4**, I present a study in which we used a controlled eye-tracking experiment in the context of a public festival setting. The goal of this study was to uncover whether social interaction can be implied – that is, whether people could be led to believe they are interacting with another person – even though they were not. We used a ‘Turing-test’-approach to study what factors are important for people to experience social interaction, and whether implied social interaction influences gaze behavior to faces. In **Chapter 5**, I present a study in which we used a dual eye-tracking setup to study gaze behavior in parent-child conversations. In this study, I focused on the role of gaze behavior in relation to speech behavior and conversation topic. Parent-child interactions are an important context of social interaction and are characterized by particular interpersonal dynamics (e.g., conflict and cooperation, parental authority and child compliance). In this study, I used a semi-structured conversation paradigm to engage parents and children in two brief conversations about common family issues and investigated the role of gaze behavior during these conversations. I provided detailed descriptions of how gaze and speech behavior during face-to-face conversations are modulated by the interpersonal context of the interaction, in terms of conflict and cooperation, and show how participant-roles (speaker – listener, parent – child) are associated with where interactants looked at each other’s faces. In **Chapter 6**, I present an in-depth analysis of gaze and turn-taking behavior during parent-child conversations. Using a selection of the data presented in Chapter 5, in this chapter I focus on where parents and children looked at each other’s faces during turn transitions; when speakers and listeners switch roles, and whether gaze to facial features is dependent on turn transition direction and topic of conversation. In **Chapter 7**, I discuss the general conclusions and implications of all the chapters presented in this dissertation. This final chapter also includes a brief evaluation of the experimental paradigms used for the empirical studies presented in this thesis, and a brief commentary on how other researchers in the field have reacted to some of the issues discussed in this thesis. I conclude by discussing some future directions for the study of gaze in human interaction within the field of social attention.

Chapter 2

The “real world” approach and its problems

Published as:

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020).
The ‘real-world approach’ and its problems: A critique of the term ecological validity.
Frontiers in Psychology, 11, 721.

Author contributions:

GH drafted the manuscript. GH, IH, CK, and RH edited and finalized the manuscript.

ABSTRACT

A popular goal in psychological science is to understand human cognition and behavior in the ‘real world’. In contrast, researchers have typically conducted their research in experimental research settings, a.k.a. the ‘psychologist’s laboratory’. Critics have often questioned whether psychology’s laboratory experiments permit generalizable results. This is known as the ‘real-world or the lab’-dilemma. To bridge the gap between lab and life, many researchers have called for experiments with more ‘ecological validity’ to ensure that experiments more closely resemble and generalize to the ‘real world’. However, researchers seldom explain what they mean with this term, nor how more ecological validity should be achieved. In our opinion, the popular concept of ecological validity is ill-formed, lacks specificity, and falls short of addressing the problem of generalizability. To move beyond the ‘real world or the lab’-dilemma, we believe that researchers in psychological science should always specify the particular context of cognitive and behavioral functioning in which they are interested, instead of advocating that experiments should be more ‘ecologically valid’ in order to generalize to the ‘real world’. We believe this will be a more constructive way to uncover the context-specific and context-generic principles of cognition and behavior.

INTRODUCTION

A popular goal in psychological science is to understand human cognition and behavior in the ‘real world’. In contrast, researchers have traditionally conducted experiments in specialized research settings, a.k.a. the ‘psychologist’s laboratory’ (Danziger, 1994; Hatfield, 2002). Over the course of psychology’s history, critics have often questioned whether psychology’s lab-based experiments permit the generalization of results beyond the laboratory settings within which these results are typically obtained. In response, many researchers have advocated for more ‘ecologically valid’ experiments, as opposed to the so-called ‘conventional’ laboratory methods (Aanstoos, 1991; Kingstone et al., 2008; Neisser, 1976; Osborne-Crowley, 2020; Shamay-Tsoory & Mendelsohn, 2019). In recent years, several technological advances (e.g., virtual reality, wearable eye trackers, mobile electroencephalography [EEG] devices, functional near-infrared spectroscopy [fNIRS], biosensors, etc.) have further galvanized researchers to emphasize the importance of studying human cognition and behavior in the ‘real world’, as new technologies will aid researchers in overcoming some of the inherent limitations of laboratory experiments (Schilbach, 2015; Shamay-Tsoory & Mendelsohn, 2019; Sonkusare, Breakspear, & Guo, 2019).

In this article, we will argue that the general aspiration of researchers to understand human cognition and behavior in the ‘real world’ by conducting experiments that are more ‘ecologically valid’ (henceforth referred to as the ‘real-world approach’) is not without its problems. Most notably, we will argue that the popular term ‘ecological validity’, which is widely used nowadays by researchers to discuss whether experimental research resembles and generalizes to the ‘real world’, is shrouded in both conceptual and methodological confusion. As we ourselves are interested in cognitive and behavioral functioning in the context of people’s everyday experience, and conduct experiments across various ‘laboratory’ and ‘real world’ environments, we have seen how the uncritical use of the term ‘ecological validity’ can lead to rather misleading and counterproductive discussions. This not only holds for how this concept is used in many scholarly articles and textbooks, but also in presentations and discussions of experimental research at conferences, during the review process, and when talking with students about experimental design and the analysis of evidence.

Although the usage of the term ecological validity has previously been criticized by several scholars (cf. Araujo, Davids, & Passos, 2007; Dunlosky, Bottiroli, & Hartwig, 2009; Hammond, 1998; Schmuckler, 2001), we think that these critiques have largely

been overlooked. Therefore, it will be necessary to cover some of the same ground. The contribution of this article is threefold. First, we extend the critique of the term ecological validity and apply it to the field of social attention. Second, we scrutinize some of the assumptions that guide the contemporary framework of ecological validity, specifically those regarding artificiality–naturalness and simplicity–complexity. Finally, our article is meant to educate a new generation of students and researchers on the historical roots and conceptual issues of the term ecological validity. This article consists of four parts. First, we will provide a brief history of the so-called ‘real-world or the lab’-dilemma and discuss several definitions and interpretations of the term ecological validity. Second, we will go into the historical roots of the concept of ecological validity and describe how the original meaning of this concept has transformed significantly. Third, we will scrutinize the prevailing assumptions that seem to guide how researchers are currently using the term ecological validity. Finally, we will apply our conceptual analysis to a specific field of study, namely the field of social attention. In recent years, this field has been particularly concerned with issues of ecological validity and generalizability. Therefore, the field of social attention offers an exemplary case to explain how the uncritical use of the terms ‘ecological validity’ and the ‘real world’ may lead to misleading and counterproductive conclusions.

A BRIEF HISTORY OF THE ‘REAL-WORLD OR THE LAB’-DILEMMA

The popular story of psychology (or the broader ‘cognitive sciences’) has it that *“psychology became a science by rising from the ‘armchair’ of speculation and uncontrolled observation, and entering the laboratory to undertake controlled observation and measurement”* (Hatfield, 2002, p. 208). The ‘psychologist’s laboratory’, a special room furnished with all kinds of lab paraphernalia and sophisticated equipment, has been regarded as the celebrated vehicle of psychology’s journey into sciencehood (Danziger, 1994; C. J. Goodwin, 2015). However, despite psychologists’ long tradition of laboratory experimentation (for a history and discussion, see Gillis & Schneider, 1966), there also have been many critical voices saying that psychology’s laboratory experiments are too limited in scope to study how people function in daily life. For example, Brunswik (1943, p. 262) once wrote that experimental psychology was limited to *“narrow-spanning problems of artificially isolated proximal or peripheral technicalities of mediation which are not representative of the larger patterns of life”*. Barker (1968, p. 3) wrote that *“it is impossible to create in the laboratory*

the frequency, duration, scope and magnitude of some important human conditions”. Neisser (1976, p. 34) wrote that “*contemporary studies of cognitive processes usually use stimulus material that is abstract, discontinuous, and only marginally real*”. Bronfenbrenner (1977, p. 513) wrote that “*many of these experiments involve situations that are unfamiliar, artificial, and short-lived and that call for unusual behaviors that are difficult to generalize to other settings*”. Kingstone et al. (2008, p. 355) declared that “*the research performed in labs, and the findings they generate, are in principle and in practice unlikely to be of relevance to the more complex situations that people experience in everyday life*”, and Shamay-Tsoory and Mendelsohn (2019, p. 1) stated that “*conventional experimental psychological approaches have mainly focused on investigating behavior of individuals as isolated agents situated in artificial, sensory, and socially deprived environments, limiting our understanding of naturalistic cognitive, emotional, and social phenomena*”.

According to these scholars, psychological science is faced with a gloomy predicament: findings and results based on highly controlled and systematically designed laboratory experiments may not be a great discovery but only a “*mere laboratory curiosity*” (Gibson, 1970, pp. 426-427). As Anderson, Lindsay, and Bushman (1999, p. 3) put it: “*A common truism has been that... laboratory studies are good at telling whether or not some manipulation of an independent variable causes changes in the dependent variable, but many scholars assume that these results do not generalize to the “real world”*”. The general concern is that, due to the ‘artificiality’ and ‘simplicity’ of the laboratory, some (if not many) lab-based experiments do not adequately represent the ‘naturalness’ and ‘complexity’ of psychological phenomena in everyday life (see Figure 1). This problem has become familiar to psychologists as the ‘real world or the lab’-dilemma (Hammond & Stewart, 2001). At the heart of psychology’s ‘real world or the lab’-dilemma lies a pernicious methodological choice: “*Should it [psychological science] pursue the goal of generality by demanding that research be generalizable to “real life” (aka the “real world”), or should it pursue generalizability by holding onto its traditional laboratory research paradigm?*” (Hammond & Stewart, 2001, p. 7).

Figure 1. Examples of historical and contemporary laboratory rooms and field experiments. **(A)** A laboratory room from the early 20th century. A participant is seated in front a ‘disc tachistoscope’, an apparatus to display visual images (adapted from Hilton, 1920). **(B)** A picture of a field experiment by J.J. Gibson. Observers had to judge the size of an object in the distance (adapted from Gibson, 1950). **(C)** A 21st century eye tracking laboratory. A participant is seated in front of a *SMI Hi-Speed* tower-mounted eye tracker (based on Valtakari, Hooge, Benjamins, & Keizer, 2019). **(D)** A wearable eye-tracker (barely visible) is used to measure gaze behavior while participants walked through corridors with human crowds (Hessels, van Doorn, Benjamins, Holleman, & Hooge, 2020).

Although psychological science is comprised of many specialized research areas, the goal to understand human cognition and behavior in the ‘real world’ has become a critically acclaimed goal for psychologists and cognitive scientists of all stripes. Indeed, examples of the ‘real-world or the lab’-dilemma can be found not only in various ‘applied’ fields of psychology, such as ergonomics (Hoc, 2001), clinical (neuro)psychology (Parsons, 2015; Wilson, 1993), educational psychology (Dunlosky et al., 2009), sport psychology (Davids, 1988), marketing and consumer psychology (Smith, Feinberg, & Burns, 1998), and the psychology of driving (S. D. Rogers, Kadar, & Costall, 2005), but also in the so-called ‘basic’ fields of psychological science, such as the study of perception (Brunswik, 1956b; Gibson, 2014), attention (Peelen & Kastner, 2014; Simons & Levin, 1998), memory (Banaji & Crowder, 1989; Cohen & Conway, 2007; Neisser, 1991), social

cognition (Osborne-Crowley, 2020; Schilbach, 2015; Schilbach et al., 2013; Shamay-Tsoory & Mendelsohn, 2019), judgement-and-decision making (Koehler, 1996), and child development (Adolph, 2019; Lewkowicz, 2001; Schmuckler, 2001).

THE REAL-WORLD APPROACH: A CALL FOR ECOLOGICAL VALIDITY

In the past decades, researchers have often discussed how they may overcome some of the limitations of laboratory-based experiments. Perhaps the largest common denominator of what we call the ‘real-world approach’ is a strong emphasis on ‘ecological validity’. Over the past decades, the term ecological validity has made its appearance whenever researchers became concerned with the potential limitations of laboratory experiments (see e.g., Aanstoos, 1991; Banaji & Crowder, 1989; Caruana, McArthur, Woolgar, & Brock, 2017; Jenkins, 1974; Koehler, 1996; Neisser, 1976; Osborne-Crowley, 2020; Risko et al., 2012; Schilbach, 2015; Shamay-Tsoory & Mendelsohn, 2019; Smilek, Birmingham, Cameron, Bischof, & Kingstone, 2006). As Neisser (1976, p. 33) famously put it:

“The concept of ecological validity has become familiar to psychologists. It reminds them that the artificial situation created for an experiment may differ from the everyday world in crucial ways. When this is so, the results may be irrelevant to the phenomena that one would really like to explain”.

The main problem, according to Neisser and many others, is that experiments in psychological science are generally “*lacking in ecological validity*” (Neisser, 1976, p. 7; Shamay-Tsoory & Mendelsohn, 2019; Smilek et al., 2006; Sonkusare et al., 2019). Aanstoos (1991, p. 77) even referred to this problem as the “*ecological validity crisis*”. To counter this problem, many researchers have called for studies with ‘more’ or ‘greater’ ecological validity. For example, Koehler (1996, p. 1) advocated for a “*more ecologically valid research program*”, Schilbach (2015, p. 130) argued for “*the inclusion of more ecologically valid conditions*”, and Smilek et al. (2006, p. 104) suggested that “*in order for results to generalize to real-world scenarios we need to use tasks with greater ecological validity*”. Clearly, ecological validity is regarded as an important feature of experimental research by researchers who pursue the ‘real-world approach’. However, in our opinion, and we are not alone in this regard (see also Araujo et al., 2007; Dunlosky et al., 2009; Hammond, 1998), this notion of ecological validity has caused considerable confusion. To foreshadow some of our criticism of ecological validity, we will show that this concept

has largely been detached from its original parentage (cf. Brunswik, 1949), and is now host to different interpretations guided by questionable assumptions (for a history, see Hammond, 1998). Worst of all, the concept is often wielded as a blunt weapon to criticize and dismiss experiments, even though researchers seldom make explicit what definition of ecological validity they use or by which set of criteria they have evaluated a study's ecological validity (as previously pointed out by Dunlosky et al., 2009; Hammond, 1998; Schmuckler, 2001).

The big umbrella of ecological validity

In past decades, the concept of ecological validity has been related to various facets of psychological research, for example, the ecological validity of stimuli (Jack & Schyns, 2017; Neisser, 1976; Risko et al., 2012), the ecological validity of tasks (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017; Smilek et al., 2006), the ecological validity of conditions (Blanco-Elorrieta & Pykkänen, 2018; Schilbach, 2015), the ecological validity of research settings (Bronfenbrenner, 1977; Schmuckler, 2001), the ecological validity of results (Eaton & Clore, 1975; Greenwald, 1976; Silverstein & Stang, 1976), the ecological validity of theories (Neisser, 1976), the ecological validity of research designs (S. D. Rogers et al., 2005), the ecological validity of methods (Banaji & Crowder, 1989), the ecological validity of phenomena (Johnston, Molyneux, & Young, 2014), the ecological validity of data (Aspland & Gardner, 2003), and the ecological validity of paradigms (Macdonald & Tatler, 2013; Schilbach et al., 2013). However, despite the popular usage of this term, specific definitions and requirements of ecological validity are not always clear.

A closer look at the literature suggests that different definitions and interpretations are used by researchers. Let's consider some examples of the literature where researchers have been more explicit in their definitions of ecological validity. For example, Ashcraft and Radvansky (2009, p. 511) defined ecological validity as: *"The hotly debated principle that research must resemble the situations and task demands that are characteristic of the real world rather than rely on artificial laboratory settings and tasks so that results will generalize to the real world, that is, will have ecological validity"*. Another influential definition of ecological validity was given by Bronfenbrenner (1977), who defined ecological validity as *"the extent to which the environment experienced by the subjects in a scientific investigation has the properties it is supposed or assumed to have by the investigator"* (p. 516). In Bronfenbrenner's view, a study's ecological validity should not be predicated on the extent to which the

research context resembles or is carried out in a ‘real-life’ environment. Instead, theoretical considerations should guide one’s methodological decisions on what type of research context is most appropriate given one’s focus of inquiry. For example, if one is interested in the behavioral responses of children when they are placed in a ‘strange situation’ then a laboratory room may be adequately suited for that particular research goal. However, if one is interested in how children behave within their home environment, then a laboratory room may not be the most suitable research context. As Bronfenbrenner (1977, p. 516) remarked: *“Specifically, so far as young children are concerned, the results indicate that the strangeness of the laboratory situation tends to increase anxiety and other negative feeling states and to decrease manifestations of social competence”*.

Ecological validity has also been used interchangeably with (or regarded as a necessary component of) ‘external validity’ (Berkowitz & Donnerstein, 1982; Hoc, 2001; Mook, 1983). The concept of external validity typically refers to whether a given study result or conclusion, usually obtained under one set of conditions and with one group of participants, can also be generalized to other people, tasks, and situations (Campbell, 1957). For example, in the literature on neuropsychological assessment and rehabilitation, ecological validity has primarily been conceptualized as *“...the degree to which clinical tests of cognitive functioning predict functional impairment”* (Higginson, Arnett, & Voss, 2000, p. 185). In this field, there has been much discussion about whether the neuropsychological tests used by clinicians accurately predict cognitive and behavioral impairments in everyday life (Heinrichs, 1990; Wilson, 1993). One major concern is that the test materials are either too abstract or too general to adequately represent the kind of problems that people with cognitive and neurological impairments encounter in their daily routines, for example, while cooking or buying food at the supermarket. In response, various efforts have been made to increase the ecological validity of neuropsychological tests, for example, by developing performance measures with relevance for everyday tasks and activities (Alderman, Burgess, Knight, & Henman, 2003; Shallice & Burgess, 1991), by combining and correlating tests results with behavioral observations and self-reports (Higginson et al., 2000; Wilson, 1993), and by using *Virtual Reality* (VR) applications to create test situations in which a patient’s cognitive and functional impairments are likely to be expressed (Parsons, 2015; Parsons, Carlew, Magtoto, & Stonecipher, 2017).

The historical roots of ecological validity

As we have seen, definitions and interpretations of ecological validity may not only differ among researchers, but also across various subfields within psychology. As such, it is not always clear how the concept should be interpreted. Interestingly, the term ecological validity used to have a very precise meaning when it was first introduced to psychological science by Egon Brunswik (Brunswik, 1949, 1952, 1955, 1956b). Brunswik coined the term ‘ecological validity’ to describe the correlation between a proximal sensory cue (e.g., retinal stimulation) and a distal object-variable (e.g., object in the environment). In Brunswik’s terminology, ecological validity refers to a measure (a correlation coefficient) that describes a probabilistic relationship between the distal and proximal layers of an organism-environment system. According to Brunswik (1955): “*A correlation between ecological variables, one which is capable of standing in this manner as a probability cue for the other, may thus be labeled “ecological validity”*” (p. 199). Brunswik (1952) believed psychology to primarily be a science of organism-environment relations in which the “*organism has to cope with an environment full of uncertainties*” (p. 22). In Brunswik’s ‘lens model’ (Brunswik, 1952), the ecological validities of perceptual cues indicate the potential utility of these cues for the organism to achieve its behavioral goals. Note that Brunswik’s concept of ecological validity is very different from how the term is generally used nowadays, namely to discuss and evaluate whether some laboratory-based experiments resemble and generalize to the ‘real world’ (cf. Ashcraft & Radvansky, 2009; Neisser, 1976; Shamay-Tsoory & Mendelsohn, 2019; Smilek et al., 2006).

The erosion and distortion of Brunswik’s definition of ecological validity has been documented by several scholars (e.g., Araujo et al., 2007; Hammond, 1998; Holleman, Hooge, Kemner, & Hessels, in press). As explained by Hammond (1998), the original definition of ecological validity, as Brunswik introduced it (1949, 1952), has been conflated with Brunswik’s ‘representative design’ of experiments (Brunswik, 1955, 1956b). Representative design was Brunswik’s methodological program for psychological science to achieve generalizability of results. To achieve this, researchers should not only conduct proper sampling on the side of the subjects, by sampling subjects who are representative of a specific ‘target population’ (e.g., children, patients), but researchers should also sample stimuli, tasks, and situations which are representative of a specific ‘target ecology’. As such, an experiment may be treated as a sample of this ‘target ecology’. By virtue of sampling theory, researchers may then determine whether results can be generalized to the intended conditions. In short, representative design requires researchers to first specify the conditions toward which they intend to generalize their findings, and

then specify how those conditions are represented in the experimental arrangement (Brunswik, 1956b). For more in-depth discussions on representative design, see Dhimi, Hertwig, and Hoffrage (2004), Hogarth (2005), and Hammond and Stewart (2001).

A SYSTEMATIC APPROACH TO ECOLOGICAL VALIDITY?

The current lack of terminological precision surrounding ecological validity is, to say the least, problematic. There seems to be no agreed upon definition in the literature, nor any means of classification to determine or evaluate a study’s ecological validity. This seems to be at odds with the relative ease by which researchers routinely invoke this concept to discuss the limitations and shortcomings of laboratory experiments. All the while, researchers seldom make clear how they have determined a study’s ecological (in)validity. As Schmuckler (2001, p. 419) pointed out: *“One consequence of this problem is that concerns with ecological validity can be raised in most experimental situations”*. To overcome these problems, several scholars have emphasized the need for a more systematic approach to ecological validity (Kingstone et al., 2008; Lewkowicz, 2001; Risko et al., 2012; Schmuckler, 2001). For example, Lewkowicz (2001, p. 443) wrote that: *“What is missing is an independent, objective, and operational definition of the concept of ecological validity that makes it possible to quantify a stimulus or event as more or less ecologically valid”*. According to Schmuckler (2001), ecological validity can be evaluated on at least three dimensions: (1) *the nature of the stimuli*; (2) *the nature of task, behavior, or response*; (3) *the nature of the research context*. Researchers have primarily discussed these dimensions in terms of their artificiality–naturalness (e.g., Hoc, 2001; Risko et al., 2012; Schmuckler, 2001; Shamay-Tsoory & Mendelsohn, 2019; Sonkusare et al., 2019), and their simplicity–complexity (e.g., Kingstone et al., 2008; Lappi, 2015; Peelen & Kastner, 2014). As such, a general framework can be construed where stimuli, tasks, behaviors, and research contexts can be evaluated on a continuum of artificiality–naturalness and simplicity–complexity (see also Lappi, 2015; Osborne-Crowley, 2020; Risko et al., 2012; Shamay-Tsoory & Mendelsohn, 2019). At one extreme is the laboratory, characterized by its artificiality and simplicity. At the other extreme is the ‘real world’, characterized by its naturalness and complexity. According to this multidimensional framework, researchers may determine a study’s overall ecological validity by combining (e.g., averaging or summing) the main components of ecological validity (i.e., stimuli, tasks/behaviors, research context) in terms of their relative artificiality–naturalness and simplicity–complexity. However, while many researchers have conceptualized ecological validity alongside these dimensions, we think there are several problems to consider. Since the dimensions of this framework

are supposedly important to determine the ecological validity of experimental research, this then raises the question of how researchers can judge the artificiality–naturalness and simplicity–complexity of particular experiments. This question will be explored in the following sections.

Artificiality - naturalness

The contrast between ‘artificiality’ and ‘naturalness’ is a particularly prominent point of discussion in the ‘real world or the lab’-dilemma and when researchers talk about the ecological validity of experimental research practices (Hoc, 2001; Kingstone et al., 2008; Shamay-Tsoory & Mendelsohn, 2019). According to Hoc (2001, pp. 282-283), ‘artificial’ situations are “*those that are specifically designed for research*” and ‘natural’ situations are “*the target situations to be understood by research*”. Importantly, Hoc (2001) notes that this distinction is made from the perspective of the researcher. However, this artificiality–naturalness distinction should also be considered from the subject’s point of view. For example, according to Sonkusare et al. (2019): “*naturalistic paradigms can be heuristically defined as those that employ the rich, multimodal dynamic stimuli that represent our daily lived experience, such as film clips, TV advertisements, news items, and spoken narratives, or that embody relatively unconstrained interactions with other agents, gaming environments, or virtual realities*” (p. 700). Furthermore, researchers have long recognized that artificiality arises when the experimental methods employed by researchers interfere with the naturalness of the psychological phenomena one aims to study. Consequently, there is always an inherent trade-off between the degree of artificiality imposed by the experimental conditions and the naturalness of the phenomena under scientific investigation (Banaji & Crowder, 1989; Barker, 1968; Brunswik, 1956b; Caruana et al., 2017; Kingstone et al., 2008; Risko et al., 2012). However, as Winograd (1988) has previously remarked, it remains difficult to “*draw a line where artificiality ends and ecological validity... for real events begins*” (p. 18).

Interestingly, discussions on the naturalness–artificiality of experimental methods have a long pedigree in psychological science. By the end of the 19th century, Thorndike (1899) and Mills (1899) already argued fiercely about what methodology should be favored to study the behavior of cats. Mills dismissed Thorndike’s work because of the artificiality of the experimental methods employed by Thorndike (see Figure 2), whereas Thorndike regarded the ethological approach favored by Mills as a collection of uncritical observations and anecdotes. Mills (1899, p. 264) wrote that: “*Dr. Thorndike... has given*

the impression that I have not made experiments, or ‘crucial experiments’... I may remark that a laboratory as ordinarily understood is not well suited for making psychological experiments on animals”. Mills’ point was that, “cats placed in small enclosures...cannot be expected to act naturally. Thus, nothing from about their normal behavior can be determined from their behavior in highly artificial, abnormal surroundings” (C. J. Goodwin, 2015, p. 200). In response to Mills, Thorndike (1899, p. 414) replied: “Professor Mills does not argue in concrete terms, does not criticize concrete unfitness in the situations I devised for the animals. He simply names them unnatural”. Thorndike clearly did not accept Mills’ charge on the artificiality of his experimental arrangements to study the behavior of cats because Mills did not define what should be considered natural behavior in the first place.

Figure 2. A ‘puzzle box’ devised by Thorndike (1899, 2017) to study learning behavior of cats. A hungry cat is placed in a box which can be opened if the cat pushes a latch. A food reward (‘positive reinforcer’) will be obtained by the cat if it figures out how to escape from the box. Thorndike discovered that after several trials, the time it takes the cat to escape from the box decreases. Experiments with puzzle boxes remain popular today to study the cognitive capacities of animals, for example, see Richter, Hochner, and Kuba (2016) for a study with octopuses. Copyright statement – Image created and owned by author IH and is based on E.L. Thorndike’s book: *Animal Intelligence* (Original date of publication, 1911, Fig. 1, p. 30).

We think that this historical discussion between Thorndike and Mills is illuminating, because it characterizes the heart of the discussion on ecological validity nowadays. Namely, what exactly did Mills consider to be ‘natural’ or ‘normal’ behavior? And how did Mills determine that Thorndike’s experiments failed to capture the ‘natural’ behavior of cats? Following Thorndike’s point on the matter, we think that researchers cannot readily determine the naturality–artificiality of any given experimental arrangement, at least not without specifying what is entailed by these ascriptions. As Dunlosky et al. (2009, p. 431) previously remarked: *“A naturalistic setting guarantees nothing, especially given that “naturalistic” is never unpacked – what does it mean?”*. Indeed, our survey of the literature also shows that the historical discussion between Thorndike and Mills is by no means a discussion of the past. In fact, we regularly encounter discussions on the ‘artificiality’ and ‘naturalness’ of experimental setups, the presentation of stimuli, the behavior of participants, or the specific tasks and procedures used in experiments – not only in the literature, but also among our colleagues and reviewers. We must often ask for the specifics, because such remarks typically remain undefined by those who toss them around.

Simplicity - complexity

The contemporary framework of ecological validity also posits that the laboratory and the ‘real world’ are inversely proportional in terms of their simplicity–complexity. Many researchers have lamented that laboratory experiments have a ‘reductionistic’ tendency to simplify the complexity of the psychological phenomena under study (e.g., Kingstone et al., 2008; Neisser, 1976; Shamay-Tsoory & Mendelsohn, 2019; Sonkusare et al., 2019). For example, Sonkusare et al. (2019, p. 699) stated that *“the ecological validity of these abstract, laboratory-style experiments is debatable, as in many ways they do not resemble the complexity and dynamics of stimuli and behaviors in real-life”*. But what exactly is meant by complexity? Let’s consider some examples from the literature. In the field of social attention, researchers have often used schematic images, photographs and videos of people and social scenes as stimuli to study the cognitive, behavioral, and physiological processes of face perception, gaze following and joint attention (Frischen et al., 2007; Langton et al., 2000; Puce & Bertenthal, 2015). However, in recent years, there has been considerable debate that such stimuli are not ‘ecologically valid’ because they do not *“capture the complexity of real social situations”* (Birmingham, Ristic, & Kingstone, 2012, p. 30). While we agree that looking at a photographic image of a person’s face is different from looking at a living and breathing person, in what ways do these situations differ in

complexity? Do these scholars mean that looking at a ‘live’ person is more complex than looking at a picture of that person? Or do they mean that the former is more complex than the latter from the perspective of the researcher who wants to understand the cognitive, behavioral, and physiological processes of face perception and social attention?

Figure 3. Are Gabor patches simple or complex compared to a picture of zebras? **(A)** A Gabor patch. **(B)** A photograph of zebras. The uniquely striped patterns of the zebra makes them most familiar to humans, whereas the question why zebras have such beautiful stripes remains the topic of much discussion among biologists, see e.g., Larison et al. (2015) and Caro and Stankowich (2015).

To take another example, Gabor patches are often used as stimuli by experimental psychologists to study ‘low-level visual processing’ (see Figure 3). Experimental psychologists use Gabor patches as visual stimuli because they offer a high degree of experimental control over various stimulus parameters (e.g., spatial frequency bandwidths, orientation, contrast, size, location). Gabor patches can be described with mathematical precision (i.e.,

“Gaussian-windowed sinusoidal gratings”, Fredericksen, Bex, & Verstraten, 1997, p. 1), and their spatial properties are considered to be a good representation of the receptive field profiles in the primary visual cortex. While Gabor patches may be considered ‘simple’ to researchers who study the relation between low-level visual processing and neural activity in terms of orientation-tuning and hemodynamic response functions, they also point to the yet to be explained ‘complexity’ of the many possible relations between other cognitive processes and patterns of neural activity in the brain. On the other hand, a naïve participant (who likely has no clue about what researchers have discovered about low-level visual processing) may describe these Gabor patches as blurry, kind of stripy, zebra-like circles, and think that they are incredibly boring to look at for many trials while lying quietly in a MRI scanner.

Our point here is that simplicity–complexity is in the eye of the beholder. Who is to say what is more simple or complex? Physicists, computer scientists, information theorists, and evolutionary biologists have developed various definitions and measures of complexity (e.g., physical complexity, computational complexity, effective complexity, algorithmic complexity, statistical complexity, structural complexity, functional complexity, etc.), typically expressed in strictly mathematical terms (Adami, 2002; Edmonds, 1995; Gell-Mann, 1995). But what definitions and measures of complexity are used by psychologists and cognitive scientists? Researchers in psychological science seem to have more loosely used the term complexity, for example, to describe a wide range of biological, behavioral, cognitive, social, and cultural phenomena, which typically contain lots of *many*’s (i.e., many parts, many variables, many degrees of freedom). Researchers may refer to various phenomena as ‘complex’ because they are simply not (yet) understood, as in “*the brain is too complex for us to understand*” (Edmonds, 1995, p. 4). Yet, such intuitive notions of complexity, whether they are caused by ignorance or whether they are used to describe something’s size, number, or variety (Edmonds, 1995), are not very helpful to evaluate the simplicity–complexity of stimuli, tasks, and situations, nor do such notions provide any formula by which these components can be summed to determine the total ecological validity of a given study. According to Gell-Mann (1995, p. 16):

“As measures of something like complexity for an entity in the real world, all such quantities are to some extent context-dependent or even subjective. They depend on the coarse graining (level of detail) of the description of the entity, on the previous knowledge and understanding of the world that is assumed, on the language employed, on the coding method used for conversion from that language into a string of bits, and on the particular idealized computer chosen as a standard.”

The ‘real-world’ or the ‘lab’: A false dilemma?

We have discussed several problems with how researchers have used the term ‘ecological validity’. In short, the concept of ecological validity has transformed significantly over the past several decades since it was introduced by Egon Brunswik (1949). It has lost most of its former theoretical and methodological cohesion (for a history, see Hammond, 1998), and the definitions and requirements of ecological validity used by researchers nowadays are seldom made explicit. As such, some experiments may be regarded as ‘ecologically valid’ by one researcher while they can be casually dismissed as ‘ecologically invalid’ by others. A closer look at the literature suggests that many researchers seem to assume that everyone understands what is meant by this term, while in fact the concept of ecological validity is seldom defined. As such, the concept of ecological validity is primarily used nowadays to make hand-waving statements about whether some (lab-based) experiments resemble ‘real life’, or whether some results obtained in the laboratory may or may not generalize to the ‘real world’.

In our opinion, the contemporary framework of ecological validity eventually falls short of providing researchers with a tractable research program. Researchers seem to primarily base their judgements of ecological validity upon their own particular theoretical assumptions and considerations about the so-called artificiality–naturalness and simplicity–complexity of experimental situations, typically in the absence of a more formal set of criteria. As such, while we certainly sympathize with the ‘call for ecological validity’, insofar it has motivated researchers to be critical about the limitations of experimental methods, we also think that the uncritical use of the term ecological validity has caused a lot of confusion, and in some cases has even been counterproductive. Perhaps the most problematic consequence of using the term ecological validity as an easy substitute for the ‘real world’ was previously pointed out by Hammond (1998). He commented that:

“There is, of course, no such thing as a “real world.” It has been assigned no properties, and no definition; it is used simply because of the absence of a theory of tasks or other environments, and thus does not responsibly offer a frame of reference for the generalization”

In Hammond’s view, the aim to understand cognitive and behavioral functioning in the ‘real world’ is basically pointless if one does not first define this notion of the ‘real world’. As such, researchers have locked themselves “onto the horns of a false dilemma” (Hammond & Stewart, 2001, p. 7). Thus, in order to talk sensibly about whether some results can also be generalized to particular situations beyond the experimental conditions in which those results were obtained, researchers first need to specify the range and distributions

of the variables and conditions to which their results are supposed to be applicable. Since the notion of the ‘real world’ patently lacks specificity, this phrase inevitably hampers researchers to specify the range and boundary conditions of cognitive and behavioral functioning in any given research context, and thus precludes one from getting at the context-specific and context-generic principles of cognition and behavior (see also Kruglanski, 1975; Simons, Shoda, & Lindsay, 2017).

The nature of the environment?

Instead of trying to understand cognitive and behavioral functioning in the ‘real world’, we completely agree with Hammond (1998) that the charge of researchers is to always specify and describe the particular context of behavior in which one is interested. Ultimately, the real challenge for researchers is to develop a theory of how specific environmental contexts are related to various forms of cognitive and behavioral functioning. But what constitutes a psychologist’s theory of the environment? Researchers in psychological science are typically concerned with the nature of the organism, yet, the nature of the environment and its relation to cognitive and behavioral functioning has received considerably less attention from a theoretical point of view (Barker, 1966; Heft, 2013). Interestingly, there have been several scholars who have dedicated themselves to precisely this question, and whose theories of cognition and behavior included a clear perspective on the nature of the environment.

According to Tolman and Brunswik (1935), the nature of the environment, as it appears to the organism, is full of uncertainties. The organism perceives the environment as an array of proximal ‘cues’ and ‘signs’ (i.e., information sources), which are the ‘local representatives’ of various distal objects and events in the organism’s environment. To function more or less efficiently, the organism needs to accumulate, combine, and substitute the information it derives from the available ‘cues’ and ‘signs’, so that it can adequately adjust its means to achieve its behavioral goals (e.g., finding food or shelter). However, since the environment is inherently probabilistic and only partly predictable, the organism continually needs to adjust its assumptions about the state of the environment based on the available information sources. Another example is given by Barker (1968), whose concept of ‘behavior settings’ (see also Heft, 2001) is key in describing how the environment shapes the frequency and occurrence of human cognition and behavior. Important to behavior settings is that they are the product of the collective actions of a group of individuals. Their geographical location can be specified (e.g., the supermarket, the cinema, etc.), and they have clear temporal and physical

boundaries (e.g., opening hours, a door to enter and exit the building). Behavior settings are ‘independent’ of an individual’s subjective experience, yet what goes on inside any behavior setting is characterized by a high degree of interdependency and equivalence of actions between individuals (e.g., most people who are inside a supermarket are shopping for groceries and people in cinemas are watching movies). Another ‘classic’ example of a theory of the environment can be found in J.J. Gibson’s book *The Ecological Approach to Visual Perception* (1979/2014). According to Gibson, there exists a strong mutuality and reciprocity between the organism and its environment. He introduced the concept of ‘affordances’ to explain how the inherent ‘meaning’ of things (i.e., functional significance to the individual) can be directly perceived by an individual perceiver and how this ‘information’ shapes the possibilities for potential actions and experiences. For example, a sufficiently firm and smooth surface may be walk-on-able, run-on-able, or dance-on-able, whereas a rough surface cluttered with obstacles does not afford such actions (Heft, 2001). In short, affordances are properties of an organism-environment system. They are perceiver-relative functional qualities of an object, event or place in the environment and they are dependent on the particular features of the environment and their relationships with the functional capabilities of a particular individual (for more in-depth discussions, see e.g., Heft, 2001; Stoffregen, 2003).

In order to describe and specify the environment and its relation to cognitive and behavioral functioning, we may draw on these scholars to guide us in a more specific direction. While we do not specifically recommend any of these perspectives, we think they are illuminating because these scholars motivate us to ask questions such as: What is the specific functional context of the cognitive and behavioral processes one is interested in? What are the relevant variables and conditions in this context given one’s focus of inquiry and level of analysis? What do we know or assume to know about the range and distribution of these variables and conditions? And how can these variables and conditions be represented in experimental designs to study specific patterns of cognitive and behavioral functioning? In order to answer some these questions, several researchers have emphasized the importance of first observing how people behave in everyday situations prior to experimentation. For example, Kingstone et al. (2008) advocated for an approach called *Cognitive Ethology*, which proposes that researchers should first observe how people behave in everyday situations before moving into the laboratory. In a similar vein, Adolph (2019) proposes that researchers should start with a rich description of the behaviors they are interested in order to first identify the “essential invariants” of these behaviors (p. 187).

THE FIELD OF SOCIAL ATTENTION: AWAY FROM THE REAL-WORLD AND TOWARDS SPECIFICITY ABOUT CONTEXT

To exemplify how some of the ideas outlined above may be useful to researchers, we will apply these ideas to a research topic of our interest: social attention. The field of social attention, as briefly discussed previously, is primarily focused on how attention is influenced by socially relevant objects, events, and situations, most notably, interactions with other social agents. In recent decades, it has been argued extensively that the experimental arrangements used by researchers in this field need more ‘ecological validity’ in order to adequately study the relevant characteristics of social attention in the ‘real world’ (Caruana et al., 2017; Macdonald & Tatler, 2018; Risko et al., 2012; Risko et al., 2016; Schilbach et al., 2013; Shamay-Tsoory & Mendelsohn, 2019). In the light of these concerns, several researchers have advocated to study “*real-world social attention*” (Risko et al., 2016, p. 1) and “*real-world social interaction*” (Macdonald & Tatler, 2018, p. 1; see also Shamay-Tsoory & Mendelsohn, 2019). One example of this is given by Macdonald and Tatler (2018). In this study, Macdonald & Tatler (2018) investigated how social roles given to participants influenced their social gaze behavior during a collaborative task: baking a cake together. Participants were either not given explicit social roles, or they were given a ‘Chef’ or ‘Gatherer’ role. Macdonald & Tatler (2018) showed that, regardless of whether social roles were assigned or not, participants did not gaze at their cake-baking partners very often while carrying out the task. After comparing their results with other so-called ‘real-world interaction studies’ (e.g., Laidlaw, Foulsham, Kuhn, & Kingstone, 2011; Wu, Bischof, & Kingstone, 2013), the authors stated that: “*we are not able to generalise about the specific amount of partner gaze during any given real-world interaction*” (Macdonald & Tatler, 2018, p. 2171). We think that this statement clearly illustrates how the use of ‘real world’ and ‘real life’ labels may lead to misleading and potentially counterproductive conclusions, as it seems to imply that ‘real-world interactions’ encompass a clearly defined category of behaviors. However, as argued previously, these so-called ‘real-world interactions’ are *not* a clearly defined category of behaviors. Instead, statements about generalizability need to be considered within a more constrained and carefully defined context (cf. Brunswik, 1956b; Simons et al., 2017). This would make it more clear what researchers are talking about, instead of subsuming studies under the big umbrella of the ‘real world’. For example, if the goal is to study how the cognitive and behavioral processes of social attention are influenced by different contexts and situations, researchers need to specify social gaze behavior as a function of these different contexts and situations.

Thus, instead of studying ‘real-world’ social attention in the context of ‘real-world’ social interactions, researchers should first try to describe and understand cake-baking attention (Macdonald & Tatler, 2018), sharing-a-meal attention (Wu et al., 2013), waiting-room attention (Laidlaw et al., 2011), walking-on-campus attention (Foulsham, Walker, & Kingstone, 2011), Lego-block-building attention (Macdonald & Tatler, 2013), playing-word-games attention (Ho, Foulsham, & Kingstone, 2015), interviewee-attention (Freeth, Foulsham, & Kingstone, 2013), and garage-sale attention (Rubo & Gamer, 2018). By doing so, we may begin to understand the context-generic and context-specific aspects of attentional processes, allowing for a more sophisticated theory of social attention. These examples not only show the wide variety of behavioral tasks and contexts that are possible to study in relation to social attention, they also show that uncritical references to ‘ecological validity’ a.k.a. ‘real-worldliness’ are not very helpful to specify the relevant characteristics of particular behavioral contexts.

There are also good examples where researchers have been more explicit about the specific characteristics of social situations that they are interested in. Researchers in the field of social attention have, for example, tried to unravel the different functions of gaze behavior. One important function of gaze behavior is to acquire visual information from the world, however, within a social context, gaze may also signal important information to others which may be used to initiate and facilitate social interaction (see e.g., Gobel, Kim, & Richardson, 2015; Risko et al., 2016). In a series of experiments, researchers have systematically varied whether, and the degree to which social interaction between two people was possible, and measured how gaze was modulated as a function of the social context (Gobel et al., 2015; Gregory & Antolin, 2019; Holleman, Hessels, Kemner, & Hooge, 2020; Laidlaw et al., 2011). In other studies, researchers have been explicit about the task-demands and social contexts that elicit specific patterns of gaze behavior, for example, in the context of face-to-face interactions and conversational exchanges (Hessels, Holleman, Kingstone, Hooge, & Kemner, 2019; Ho et al., 2015). We think that, if researchers would try to be more explicit in their descriptions of task-demands and social contexts in relation to gaze, this may prove to be a solid basis for a more sophisticated theory of social attention, yet such work remains challenging (for a recent review, see Hessels, *in press*).

CONCLUSION

We have argued that the ‘real-world approach’ and its call for ecological validity has several problems. The concept of ecological validity itself is seldom defined and interpretations differ among researchers. We believe that references to ecological validity and the ‘real world’ can become superfluous if researchers would clearly specify and describe the particular context of behavior in which they are interested. This will be a more constructive way to uncover the context-specific and context-generic principles of cognition and behavior. As a final note, we hope that editors and reviewers will safeguard journals from publishing papers where terms such as ‘ecological validity’ and the ‘real world’ are used without specification.

Chapter 3

The reality of “real life” neuroscience

Published as:

Holleman, G. A., Hooge, I. T. C., Kemner, C., Hessels, R. S. (2021). The reality of “real-life” neuroscience: A commentary on Shamay-Tsoory and Mendelsohn (2019). *Perspectives on Psychological Science*, 16(2), 461–465.

Author contributions:

GH drafted the manuscript. GH, IH, CK, and RH edited and finalized the manuscript.

ABSTRACT

The main thrust of Shamay-Tsoory & Mendelsohn's ecological approach is that: "...*the use of real-life complex, dynamic, naturalistic stimuli provides a solid basis for understanding brain and behavior*" (p. 851). While we support the overall goal and objectives of Shamay-Tsoory & Mendelsohn's approach to 'real-life' neuroscience, their review refers to the terms 'ecological validity' and 'representative design' in a different manner than how they were originally introduced by Egon Brunswik. Our aim is to clarify Brunswik's original definitions and briefly explain how these concepts pertain to the larger problem of generalizability. Not just for history's sake, but because we believe that a proper understanding of these concepts is important for researchers who want to understand human behavior and the brain in the context of everyday experience, and because Brunswik's original ideas may contribute to Shamay-Tsoory's & Mendelsohn's ecological approach. Finally, we'll argue that the popular and often misused concept of 'ecological validity' is ill-formed, lacks specificity and may even undermine the development of theoretically sound and tractable research.

INTRODUCTION

In “Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research”, Shamay-Tsoory & Mendelsohn (2019, henceforth ST&M) argue that: “...*the field of cognitive neuroscience may be hampered by the limited ecological validity that characterizes the bulk of paradigms...*” (p. 841). To counter this problem, ST&M advocate that researchers should strive for experimental paradigms that possess “*high ecological validity*” (p. 851). The goal of ST&M’s ecological approach to cognitive neuroscience is to understand human behavior and the brain within the context of everyday human experience. To achieve this goal, ST&M advocate that “...*the use of real-life complex, dynamic, naturalistic stimuli provides a solid basis for understanding brain and behavior*” (p. 851). Specifically, given that humans are active, embodied, social agents who interact with each other in everyday situations, experimental paradigms used by researchers must be aimed at preserving these characteristics. Finally, ST&M argue that with the advent of new portable measuring devices (e.g., mobile electroencephalography [EEG], functional near-infrared spectroscopy [fNIRS], and wearable eye trackers), the investigation of brain and behavior in ‘real life’ has become a methodologically feasible goal for cognitive neuroscientists. While we support the goals and objectives of ST&M’s ecological approach, the aim of this commentary is to clarify two important concepts of psychological theory, namely: *ecological validity* and *representative design*. In ST&M’s review, the authors acknowledge Egon Brunswik as the originator of these two concepts (p. 843). However, ST&M do not provide a definition of these concepts (neither Brunswik’s definitions, nor their own), and the ways in which ST&M make use of these concepts is different from how Brunswik originally introduced them. We think it’s important to clarify Brunswik’s original definitions. This is not just important for history’s sake, but primarily because Brunswik’s ideas may contribute to ST&M’s ecological approach to study brain and behavior in everyday life. In what follows, we will explain how Brunswik originally defined the concept of ecological validity and how it is related to his methodological program of representative design. Finally, we’ll discuss how ST&M’s ecological approach may benefit from Brunswik’s perspective on the problem of generalizability.

BRUNSWIK'S FORGOTTEN LEGACY: ECOLOGICAL VALIDITY AND REPRESENTATIVE DESIGN

Brunswik (1952, 1955, 1956b) defined ecological validity as the potential utility of a proximal cue (i.e., stimulus information) available to an organism to infer the state of a distal variable (e.g., object in the environment). To illustrate Brunswik's definition of ecological validity, we'll take a situation where a human observer judges the age of another person. Consider Brunswik's lens model (Brunswik, 1955, 1956b) depicted in Figure 1. The distal variable in our example is the person-object's age and the observer's judgement is the central response. The degree to which the actual age of the person matches the observer's judgement is what Brunswik called 'functional validity', or 'achievement'. A variety of proximal cues (i.e., information sources) may be available and utilized by the observer to infer someone's age. The correlation between the distal variable (i.e., age) and a proximal cue is called 'ecological validity', and the correlation between cues and the observer's judgement (i.e., central response) is called 'cue utilization'.

Figure 2. The lens model. Redrawn from Brunswik (1955; 1956b).

In Brunswik’s lens model, ecological validity and cue utilization can but do not necessarily have to correspond to each other. If, say, an observer infers someone’s age from a person’s physical appearance, the observer may use cues such as hair color, skin condition, posture, and clothing. Some of these cues may have ‘limited’ or ‘low’ ecological validity but can still be useful in perception. According to Brunswik (1956b), the observer may use “*relatively superficial and stereotyped cues of limited ecological validity, preferably a multitude of them*” (p. 89). In other words, the observer may use a variety of cues, some of which are more reliable than others, and these cues may be used interchangeably. In contrast to how Brunswik defined ecological validity, ST&M’s usage of ecological validity is quite different. While ST&M do not explicitly define ecological validity, ST&M seem to employ the phrase ‘ecological validity’ to evaluate and compare lab-based research with ‘real life’. This is evident from several phrases in ST&M’s review. For example, ST&M (2019) write: “...*the field of cognitive neuroscience may be hampered by the **limited ecological validity** that characterizes the bulk of paradigms and settings in the field...*” (p. 841, emphasis ours), and when they write that “*real-life situations provide a natural context and allow dynamic movement and feedback... collecting rich data from real-life experiments offers the opportunity to evaluate multiple variables across experiments possessing **high ecological validity***” (p. 841, emphasis ours). Yet, Brunswik’s lens model allows for ‘limited’ or ‘low’ ecological validities, albeit only in the sense that ecological validity describes the potential utility of a proximal cue (i.e., stimulus information) available to an organism to infer the state of a distal variable (e.g., object in the environment).

Thus, Brunswik did not use the concept of ecological validity to evaluate whether experimental research (or stimuli, tasks, and conditions) resembles or approximates ‘real life’ (as is done by e.g., Neisser, 1976, p. 33; Sonkusare et al., 2019, p. 1), nor did Brunswik imply that ‘more’ or ‘greater’ ecological validity warrants generalizability to the ‘real world’ (as implied by e.g., Matlin (1989, p. 8) and Smilek et al. (2006, p. 104)). Rather, Brunswik’s concern was that ‘ecological validities’ can be dramatically misrepresented in certain (lab-based) experiments (Brunswik, 1943, 1956b; see also Hammond & Stewart, 2001). As an example of this concern, Brunswik (1943) describes the experimental procedures used by researchers in his day to study learning behavior in rats. He wrote: “*Situations in which food can be found always to the right and never to the left, or always behind a black door and never behind a white one...*” (p. 261). Thus, the door color or direction in the experimental setting perfectly predicts food location. According to Brunswik, the variables studied in these experiments are often artificially ‘tied’ or ‘untied’ (Brunswik, 1955), which can result in either perfect correlations between variables, i.e.,

ecological validities of -1 or +1, or in uncorrelated variables, i.e., ecological validities of 0. However, in a ‘natural ecology’ (i.e., not the construction of an experimenter), the ecological validities between variables are rather somewhere in between. Thus, according to Brunswik (1943), the all-or-nothing reward schedules in these experiments are “... *not representative of the structure of the environment...*” (p. 261). To solve the limitations of such experimental designs, Brunswik devised his method of ‘representative design’ (Brunswik, 1955, 1956b). In Brunswik (1956a, p. 159) words:

“To obviate the intrinsic shortcomings of the artificial, ‘systematic’ designs ... I have advocated that in psychological research not only individuals be representatively sampled from well-defined ‘populations’ but also stimulus situations from well-defined natural-cultural ‘ecologies’; only by such representative design of experiments can the ecological generalizability of functional regularities of behavior and adaptation be ascertained.”

Simply put, representative design means that researchers first need to define a reference class of stimuli, tasks, and situations to which they intend to generalize a result. By sampling from this predefined set of conditions, the set of conditions towards which the generalization is intended is part of the experimental design. According to Brunswik (1956b, p. 53): “*Generalizability of results concerning the...variables involved must remain limited unless at least the range, but better also the distribution... of each variable, has been made representative of a carefully defined set of conditions.*”

THE REALITY OF “REAL-LIFE” NEUROSCIENCE

While Brunswik’s method of representative design has been discussed much more thoroughly elsewhere (Araujo et al., 2007; Dhami et al., 2004; Hammond & Stewart, 2001), here we primarily want to clarify a crucial difference between Brunswik’s representative design (1955, 1956b) and ST&M’s ecological approach (2019). Brunswik advocated that in order to justify claims about generalizability, one must first specify a reference class, or set of conditions toward which the generalization is intended. Thus, following Brunswik’s approach, experimental arrangements should be based on a well-defined set of conditions, that is, a representative sample of stimuli, tasks, and situations (i.e., in terms of their number, distribution, range, ecological validities, and intercorrelations of variables) which are deemed relevant to the context of functional behavior one is interested in. On the other hand, ST&M’s ecological approach (2019) primarily suggests to enhance the ‘ecological validity’ of experimental paradigms by conducting “*real-life experiments*” (p. 11) that include “*real-life situations*” (p. 11) and

“*real-life experiences*” (p. 11). However, these ‘real-life’-categories patently lack specificity. Without a more well-defined set of conditions, the notion of ‘real life’ invokes a potentially infinite range of stimuli, tasks, situations and conditions, which are logically and methodologically unfeasible to be represented in any formal manner. As Hammond, a former student and propagator of Brunswik’s research philosophy noted:

“The real trouble with introducing the terms ‘real world’ or ‘real life’ and the reason they should be abandoned is that they are simply low-grade escape mechanisms; their use makes it unnecessary to define the conditions toward which the generalization is intended. One need only assume (without evidence) that everyone knows what these terms entail. (Hammond & Stewart, 2001, pp. 7-8).”

ST&M (2019) ecological approach certainly has merit. ST&M’s review points to several limitations of certain lab-based experiments, and the authors identify areas of research where new technological advances may be particularly useful to overcome these limitations, such as the fields of social cognition and episodic memory. Indeed, there have been numerous theoretical and methodological developments which support ST&M’s claims about the importance of the social context and potential for social interaction on behavioral and neurobiological functioning (see also De Jaegher, Di Paolo, & Gallagher, 2010; Risko et al., 2016; Schilbach et al., 2013), as also highlighted in their review. However, we think that ST&M’s ecological approach may also benefit from Brunswik’s most crucial lesson, namely that researchers should first define the contexts of behaviors that they are interested in and then show how these are well represented in their experiments. Thus, instead of subsuming research on brain and behavior during social interactions under the broad category of “real social interactions” (Shamay-Tsoory & Mendelsohn, 2019, p. 11), emphasis should be on specific types of social interactions. For example, the interpersonal and behavioral dynamics of parent-child interactions (Reindl, Gerloff, Scharke, & Konrad, 2018) may differ fundamentally from interactions between romantic partners (Kinreich, Djalovski, Kraus, Louzoun, & Feldman, 2017) or from student-teacher interactions in a classroom (Dikker et al., 2017). Different types of social interactions may be representative of a person’s ‘natural ecology’ (as Brunswik might have put it) to various degrees and involve a diverse repertoire of cognitive and behavioral adaptations and responses. As such, we believe that specifying the interpersonal dynamics and the behavioral tasks and activities of social interactions is key to understanding their cognitive and neurobiological underpinnings. In this manner, we may begin to uncover the context-specific and the context-generic aspects of brain and behavior during ‘social interaction’ at large.

To conclude, we think that a full-fledged ecological approach to brain and behavior should not merely be a matter of employing more ‘ecologically valid’ and ‘real life’ stimuli and settings. The problem of such an approach is that it may foster misguided ideas about what ecological validity means and what is entailed by ‘real life’ (see Hammond, 1998 for a history). Nowadays, the concept of ecological validity is widely used as a catchall phrase to make intuitive claims about whether a given experiment approximate ‘real life’, or whether one’s findings generalize to the ‘real world’ (as previously pointed out by Hammond, 1998; Hammond & Stewart, 2001). Moreover, the concept of ecological validity is sometimes used a blunt weapon to criticize experiments, typically in the absence of a proper definition or specific set of criteria (for further discussion, see Dunlosky et al., 2009; Schmuckler, 2001). Therefore, to more adequately evaluate and discuss whether a given experimental arrangement has captured the relevant aspects of cognitive and behavioral functioning within a particular context, it would be more constructive if researchers would stop using ‘ecological validity’ as an easy substitute for ‘real life’. Instead, researchers need to be more explicit about their specific focus of inquiry and explain how and why certain experimental designs may or may not have captured the relevant characteristics of cognition and behavior that they are interested in. As Brunswik (1955, p. 204) reminds us:

“Mostly there is little technical basis for telling whether a given experiment is an ecological normal, located in the midst of a crowd of natural instances, or whether it is more like a bearded lady at the fringes of reality, or perhaps like a mere homunculus of the laboratory out in the blank.”

Chapter 4

Implying social interaction

Published as:

Holleman, G. A., Hessels, R. S., Kemner, C., & Hooge, I. T. C. (2020). Implying social interaction and its influence on gaze behavior to the eyes. *PloS one*, 15(2), 1-27.

Author contributions:

GH, RH, CK and IH designed the study. GH, RH, and IH collected the data. GH, RH, IH interpreted the data. GH drafted the manuscript. GH, RH, CK and IH finalized the paper.

ABSTRACT

Researchers have increasingly focused on how the potential for social interaction modulates basic processes of visual attention and gaze behavior. In this study, we investigated why people may experience social interaction and what factors contributed to their subjective experience. We furthermore investigated whether implying social interaction modulated gaze behavior to people's faces, specifically the eyes. To imply the potential for interaction, participants received either one of two instructions: 1) they would be presented with a person via a 'live' video-feed, or 2) they would be presented with a pre-recorded video clip of a person. Prior to the presentation, a confederate walked into a separate room to suggest to participants that (s)he was being positioned behind a webcam. In fact, all participants were presented with a pre-recorded clip. During the presentation, we measured participants' gaze behavior with an eye tracker, and after the presentation, participants were asked whether they believed that the confederate was 'live' or not, and why they thought so. Participants varied greatly in their judgements about whether the confederate was 'live' or not. Analyses of gaze behavior revealed that a large subset of participants who received the live-instruction gazed less at the eyes of confederates compared with participants who received the pre-recorded-instruction. However, for both the live-instruction group and the pre-recorded instruction group, another subset of participants gazed predominantly at the eyes. The current findings may contribute to the development of experimental designs aimed to capture the interactive aspects of social cognition and visual attention.

INTRODUCTION

Since the pioneering studies of Buswell (1935) and Yarbus (1967), many eye-tracking studies have shown that people have a strong tendency to look at other people's faces, especially their eyes (Birmingham et al., 2009; Henderson et al., 2005; Itier, Villate, & Ryan, 2007; Pelphrey et al., 2002; Walker-Smith, Gale, & Findlay, 1977). While previous studies have predominantly investigated how observers look at images and videos of people's faces (Kingstone, 2009), such 'non-responsive' stimuli may not adequately represent how people look at each other in the context of everyday social situations (for a review, see Risko et al. (2012)). As Risko et al. (2016) noted, many laboratory experiments uphold an *"illusory barrier between participants and the stimuli they are presented with...participants are not expected to interact with the faces presented to them or worry about what the stimuli might think of them"* (p. 70). In recent years, researchers in the field of social attention have increasingly focused on how visual attention and gaze behavior are modulated by the social context, for example, by the presence of 'live' people and potential interactions (Freeth et al., 2013; Laidlaw et al., 2011; Risko et al., 2012). The study of gaze behavior in the context of social interaction is important to better understand, for example, how and when gaze behavior may support the pickup of cognitive and emotional states (Beaudry, Roy-Charland, Perron, Cormier, & Tapp, 2014; Buchan et al., 2007; Neath & Itier, 2014), how gaze facilitates social learning and development (Franchak, Kretch, & Adolph, 2018; Gredebäck, Fikke, & Melinder, 2010; Yu & Smith, 2016), and how gaze supports and regulates speech and turn-taking behavior in conversations (Duncan & Fiske, 2015; Hessels et al., 2019; Ho et al., 2015; Kendon, 1967). Furthermore, understanding gaze behavior in the context of social interaction has several key implications for clinical research. For example, individuals scoring high on psychopathological traits of Autism Spectrum Disorder (ASD) and Social Anxiety Disorder (SAD) may display more 'atypical' gaze behavior during social interaction, such as gaze avoidance and a lack of 'eye contact' (Hessels, Holleman, et al., 2018; Nation & Penny, 2008). The current challenge for researchers in the field of social attention is to design experiments that capture the relevant aspects of gaze behavior in social situations, such as the presence of other people and the possibility to interact with them, while also maintaining a high level of experimental control (Caruana et al., 2017; Pfeiffer et al., 2013). In the present study, we investigated why people may experience social interaction and what factors contributed to their subjective experience. We furthermore investigated whether implying social interaction modulated gaze behavior to faces, specifically to the eyes. We will first discuss previous eye-tracking research on how the presence of other

people and the potential for interaction modulates basic processes of visual attention, and then we will outline the research questions and hypotheses of the present study.

Gaze behavior in context: social presence and potential interactions

In the last decade, several eye-tracking studies have shown that people look differently at ‘live’ people as compared with images and videos of people (see Foulsham et al., 2011; Freeth et al., 2013; Hessels et al., 2019; Laidlaw et al., 2011). Two prominent examples are given by Foulsham et al. (2011) and Laidlaw et al. (2011). Foulsham et al. (2011) investigated where participants looked while walking on a university campus. The experimenters compared participants’ gaze behavior during their walk on campus with participants’ gaze behavior while watching the first-person video-recordings of themselves and others walking on campus. Foulsham et al. (2011) discovered that when participants walked on campus, they were more likely to look away from other pedestrians on the street, especially when they approached upcoming pedestrians. Interestingly, when participants were presented with the first-person video-recordings, they continued to look at the pedestrians. Likewise, Laidlaw et al. (2011) demonstrated that participants looked less often and for shorter durations at a confederate sitting in a chair across the room in comparison with a confederate displayed on a computer screen in the same room.

The studies by Foulsham et al. (2011) and Laidlaw et al. (2011) both demonstrated that the tendency to look at other people, often assumed to be largely automatic and reflexive (for example, see Birmingham et al. (2009)), was markedly less when other people were physically present. These findings converge on the notion that the presence of other people can cause substantial changes in a wide variety of behaviors, feelings, physiological states, motives, and beliefs. Such changes in cognition and behavior are sometimes referred to as ‘social presence’ or ‘mere presence’ effects (Guerin, 1986; Latané, 1981; Myllyneva, Ranta, & Hietanen, 2015; Zajonc, 1965). Similarly, other eye-tracking studies have further demonstrated that gaze behavior can be modulated by social presence and potential interactions (Cañigueral & Hamilton, 2019; Freeth et al., 2013; Macdonald & Tatler, 2018; Risko et al., 2012). However, it is not always clear under what circumstances social presence may cause changes in people’s gaze behavior, or how gaze behavior may be modulated as a result. For example, Freeth et al. (2013) showed that when an interviewer engaged in ‘eye contact’ with interviewees, the participants who were presented with a ‘live’ interviewer gazed longer at the interviewer’s face (i.e., when the interviewer was physically present) in comparison with participants who

were presented with an interviewer displayed on a computer screen. In this interview-context, looking at the eyes and face of the other person may be necessary to adequately monitor and respond to the interviewer (see also (Hessels et al., 2019; Võ et al., 2012). However, in a waiting-room context, or while walking on campus (as in the experiments by Laidlaw et al. (2011) and Foulsham et al. (2011)), people may be more likely to avoid looking at other people in their physical proximity. In other words, different social and behavioral contexts, characterized by different social and behavioral norms, may produce particular patterns of gaze behavior associated with those contexts. For example, a phenomenon known as ‘civil inattention’ (Goffman, 2008) describes how people often ignore each other in public places, for example, in waiting-rooms, elevators, and on public transportation. In these situations, people may briefly glance at each other to ‘acknowledge’ each other’s presence but consequently look away and politely ‘ignore’ each other (see also Zuckerman, Miserandino, and Bernieri (1983)). Conversely, in some other situations it may be considered inappropriate not to look at and respond to other people. For example, it would be strange not to actively attend to the other person at a job interview or when having coffee with a friend. In other words, certain social contexts may afford interaction or not (Valenti & Gold, 1991), which likely influences how gaze behavior is allocated towards other people.

Implied social presence – the belief of being watched

Researchers have also discovered that certain beliefs and attitudes about the presence of others influences visual attention and gaze behavior, for example, when people believe they are being watched or observed by others (Nasiopoulos, Risko, & Kingstone, 2015). Several studies have shown that the implication of someone’s presence may already cause changes in how people allocate their gaze, for instance when people think they are being filmed (Cañigueral & Hamilton, 2019; Gobel et al., 2015; Risko & Kingstone, 2011). In a study by Gobel et al. (2015), participants were presented with several video clips of people’s faces under different viewing conditions. For half the videos, participants were instructed that they would simply look at the video clips, while for the other half, they were instructed that a webcam would also be recording them while watching the video clips. In the latter condition, Gobel et al. (2015) told participants that their recordings would later be watched back by the people displayed on screen. Gobel et al. also added a description of the social status of the people in the video clips. Interestingly, they discovered that when participants believed they were being filmed by a webcam, they looked less at the eyes of high-ranking targets, but not of the low-ranking targets. In this

study, Gobel et al. posited that gaze behavior to the eyes may be indicative of social rank signaling (see also Dovidio and Ellyson (1985)), similar to how some primate species and other animals visually communicate dominance and social status (Chance, 1967; Van Hooff, 1967). Studies in primatology have demonstrated that lower-ranking members of the social hierarchy often look at the dominant members of the group, except when the high-ranking individuals gaze directly back at them (Bernstein, 1970). However, Gobel et al. only prospectively implied social presence, meaning that participants were only told they would be evaluated by another person at a later point in time.

A more direct way of simulating social presence was achieved by Gregory and Antolin (2019). In this study, participants were led to believe that another person could see them via a so-called 'live' video-feed which was in fact a 'staged' video-feed. To investigate whether participants' beliefs about the possibility for social interaction modulated gaze behavior to faces, participants could receive one of three instructions prior to viewing: (1) they would be presented with a 'live' video-feed of a person with the possibility to interact; (2) they would be presented with a 'live' video-feed of a person without the possibility to interact; or (3) they would be presented with a video clip of a person. Crucially, all participants were shown a pre-recorded video clip of a confederate. The video clip displayed a person who sat in a chair and looked at a clipboard. At one point, the confederate turns her head and looks directly into the camera for approximately four seconds. Gregory and Antolin (2019) showed that when participants were instructed that the person displayed on the computer screen was 'live' and could potentially interact with them, participants gazed less at the person's face in comparison with participants who were told that the presentation was a pre-recorded clip. Gregory and Antolin (2019) also reported that when the person in the clip directly looked at the participant, which could be perceived as conveying the initiative to interact, participants were more likely to look at the face of the confederate regardless of the instructions received prior to the experiment. Gregory and Antolin (2019) posited that: *"Had the confederate prolonged this period of direct gaze, overall gaze avoidance might have been evident as the participants attempted to reduce the 'social risk' which results from making direct gaze with a stranger"* (p. 11). In other words, a prolonged period of direct gaze might result in overall gaze avoidance, because prolonged eye contact with strangers may be considered inappropriate or perceived as aggressive and threatening (Dovidio & Ellyson, 1985; Ellsworth, Carlsmith, & Henson, 1972; Hamilton, 2016; Kleinke, 1986).

Importantly, Gregory and Antolin (2019) demonstrated that participants can be led to believe that there is a potential to interact with a ‘live’ person displayed on a computer screen, even though this person was only displayed as a video clip. However, the video clip of the confederate in this study may have only minimally conveyed a potential to interact, which may have reduced the experience of being watched by another person, as well as the experience of potential interaction. Gregory and Antolin (2019) also reported that many participants in this study did not believe that the confederate was ‘live’, or that the confederate could really interact with them (as reported through several questionnaire items after the experiment). But why did some participants believe that they were looking at a ‘live’ person and others did not? And why did some participants believe there was a possibility for interaction and others did not? Previous studies have not always explicitly focused on why people may experience social interaction or why they perceive a potential for social interaction.

PRESENT STUDY

In the present study, we wondered why people may experience a potential for social interaction and what factors may contribute to this experience? To investigate this, we specifically wanted to imply social *interaction* instead of social *presence* (Gobel et al., 2015; Gregory & Antolin, 2019). Also, we investigated whether implying social interaction would also modulate gaze behavior to the eyes. We reasoned that staging a face-to-face encounter with a stranger would be the most suitable context for our research question. We wanted to create a situation in which a person appears to ‘look’ directly at the participant, and thereby minimize the potential ambiguity as to whether that person is actively attending the participant. Based on previous research, one might expect that if someone believes that another person can directly see them, some people may avoid that person’s gaze. The ‘social risk’-hypothesis, for example, posits that a prolonged period of direct gaze from a stranger results in gaze avoidance (Gregory & Antolin, 2019). We devised an experiment to investigate whether people also avoid someone’s gaze if they believe they are being watched by a ‘live’ stranger and, importantly, when they believe there is a potential for social interaction. In our study, participants were presented with video clips of confederates who displayed more ‘interactive’ behavior (e.g., several bids for social interaction) compared with the behavior exhibited by confederates in previous studies (cf. Gobel et al., 2015; Gregory & Antolin, 2019; Laidlaw et al., 2011). Prior to every presentation, one of four confederates walked into a separate room in order to suggest to participants that they were taking place behind a webcam. After the confederate entered

the separate room, some participants were instructed that they would be presented with the confederate via a ‘live-video connection’ and some participants were instructed that they would be presented with a pre-recorded clip of the confederate. In fact, regardless of the instructions, all participants were shown a pre-recorded video clip. During the presentation, we recorded participants’ eye movements with a high-end eye-tracker. This allowed us to measure gaze behavior to facial features with high temporal and spatial resolution. Afterwards, participants were asked whether they believed that they were presented with a ‘live’ person and why they thought so.

MATERIALS AND METHODS

Participants

Eighty-two participants (fifty females; $M_{age} = 31.69$ years, range 18 – 70, $SD = 11.66$) were recruited amongst visitors of the “Betweter Festival”, which is an annual science festival organized in Tivoli Vredenburg, a large concert venue in Utrecht. Based on previous studies (Gobel et al., 2015; Gregory & Antolin, 2019), we aimed to recruit a minimum of sixty participants. We reasoned this would be sufficient for our research questions and hypotheses. To compare, Gregory & Antolin’s (2019) study used a similar sample size to compare a pre-recorded-viewing group (28 participants) and a ‘two-way’-viewing group (34 participants). All participants reported normal or corrected-to-normal vision. The experiment was conducted in a large room where other activities were also going on. The room was noisy with moderately loud music and people.

Ethical considerations and informed consent

This study was approved by the Ethics Committee of the Faculty of Social and Behavioral Sciences of Utrecht University and is registered under protocol number FETC17-097. Written informed consent was obtained from each participant prior to the start of the experiment. The confederates who participated in this study also gave written informed consent for the publication of the stimulus video clips and illustrations of this material.

STIMULI AND APPARATUS

We aimed to imply social interaction via a staged ‘face-to-face’ interaction, albeit mediated through a computer screen. To achieve this, we recorded stimulus material in a two-way video setup devised by Hessels, Cornelissen, Hooge, and Kemner (2017). This setup

allowed us to create video clips of the confederates as if they were looking directly at each other while simultaneously looking into the camera. Similar recording techniques have previously been used by cinematographers to achieve a more engaging performance (e.g., actors, interviewees, models). We reasoned this would then create a more engaging viewing experience for the participants, as compared with watching a recording of a person who models in front a camera lens. We used four confederates as models (2 females, 2 males), and recorded two video clips of each confederate. Each clip displayed a frontal view of the confederate who exhibited various facial expressions which were meant to exhibit ‘responsive’ and ‘interactive’ behavior. We labeled the clips ‘video clip A’ and ‘video clip B’, respectively. Video clip A displays a confederate who looks directly into the camera, frowns, smiles, laughs, and then winks. Video clip B displays a confederate who first looks directly into the camera with a neutral expression, averts his/her gaze for a couple of seconds, and then shifts back to direct gaze. Finally, the confederate closes and opens his/her eyes for a second. The sequence of these events was the same for every confederate with slight variations in timing (see Figure 1 for snapshots from the video clips). Each video clip lasted approximately 30 seconds and was recorded with a Logitech webcam recording video frames of 800 by 600 pixels at 30Hz. The video frames were scaled to a resolution of 1024 by 768 pixels during the presentation. The rationale behind these social gaze cues and facial expressions was partly based on the literature. Direct gaze (‘eye contact’) and averted gaze are well documented as powerful social signals (Hamilton, 2016). Furthermore, we reasoned that smiling, frowning, laughing and winking would be appropriate to convey ‘interactive’ behavior within the context of this experiment. However, note that the goal of this study was not to investigate the influence of particular social gaze cues on gaze behavior, but rather to create some variation in the behavior of confederates to imply social interaction. Also, we did not include speech or audio in our video-recordings.

Figure 1. Snapshots of video clips. In video clip A (upper panels), the confederate first looks directly into the camera (**A0**). After 5 seconds, the confederate frowns and smiles and then resumes looking into the camera (**A5**). After 15 seconds, the confederate laughs and then resumes looking into the camera again (**A15**). Finally, after 25 seconds the confederate winks and smiles. In video clip B (lower panels), the confederate first looks directly into the camera with a neutral expression (**B0**). After 10 seconds (**B10**), the confederate averts his gaze to the right for about 5 seconds. Then, after 20 seconds, the confederate briefly closes his/her eyes and then looks back into the camera again (**B20**).

To record participants' eye movements, a Tobii TX300 eye tracker with an integrated 23-inch monitor (1920 by 1080 pixels; 60 Hz refresh rate) was used (Tobii Technology, Stockholm, Sweden). The Tobii eye-tracking system ran at 300 Hz and communicated with MATLAB (version R2012b, MathWorks Inc., Natick, MA, USA) running on a MacBook Pro (2.8 GHz i7 processor, OS X 10.9) via the Tobii SDK (*Tobii Pro Analytics SDK, version 3.0*). Stimulus presentation was done using the PsychToolbox (version 3.0.11; Brainard and Vision (1997)). The operator controlled the stimulus computer and eye tracker. The eye tracker was placed inside a box with a black cloth attached to the top and sides of the box (see Figure 2). The eye tracker was located at 65 cm from the participant. The video clips were centered on the screen of the Tobii and subtended 23.6° by 17.8° under the assumption that the viewing distance was approximately 62 cm. A chin- and forehead rest was used to keep the viewing conditions constant for every participant, minimize head movements, and ensure optimal data quality (Hessels, Cornelissen, Kemner, & Hooge, 2015; Niehorster, Cornelissen, Holmqvist, Hooge, & Hessels, 2018).

Figure 2. The eye-tracking setup consisted of a box placed on a cart. The box contained a Tobii TX300 eye tracker and presentation screen. A small desk with a keyboard was attached to the front of the box and a chinrest was mounted unto the front of the desk. In this illustration, the participant receives instructions from the operator. When the operator was ready, the participant would take place in the chinrest and the black cloth was draped over the participant. Reprinted from (Hooge, Holleman, Haukes, & Hessels, 2018) with permission.

GENERAL PROCEDURE

Participants were recruited at the festival site. The research setting consisted of an intake desk where visitors of the festival could sign up for participation. The experiment was promoted as follows: *“This experiment is about social interaction. We are not allowed to say more about it. Do you dare to step into this black box?”* Prior to the experiment, participants were given a brief information letter. Participants gave written informed consent and provided some basic demographic information. Every participant received a unique participant code which was coupled with a specific presentation procedure (see below). After participants signed up for the experiment, they were asked to line up in a queue and await their turn. At the research setting, the four confederates walked around in bright red sweaters to signal their participation in the experiment. Visitors and participants could talk to the confederates prior and after the experiment, but the confederates did not discuss the content of the experiment. At the start of every experiment, the operator

asked the first participant in line to come forward to be positioned in the eye-tracking booth. The operator then entered the unique code of the participant into a custom MATLAB script. Based on the participant code, every participant was coupled to one of the four confederates, combined with a specific instruction-type and video clip-type. This was done to ensure that we would have an approximately equal distribution of participants over the different confederates, instructions, and video clips at the end of the testing session. After the participant was positioned, the operator loudly asked the confederate assigned to the participant to get ready. The confederate then responded in an affirming way and walked past the participant and the other people present at the scene to step into the confederate booth (see Figure 3). The confederate's booth consisted of a wooden cubicle with a black curtain and was placed next to the eye-tracking booth. The confederate's booth was placed so that the current participant, as well as the other people waiting in line for their turn to participate (the audience), would be able to see the confederates walking in and out of the booth before and after the experiment.

Figure 3. Schematic top-view of the research setting. The participant is seated in the eye-tracking booth next to the operator. The confederate (dressed in a red sweater) walks into the confederate booth after being called for by the operator prior to each measurement. A queue of participants (the audience) is awaiting their turn after they signed up to participate at the intake desk. Note that this illustration is a simplified view. During the actual experiment, three more confederates dressed in red sweaters were walking around and the research setting was surrounded by other people, a bar, lights, as well as other installations and performances.

Calibration procedure

After the participant was positioned in front of the eye tracker and the confederate had stepped into the confederate's booth, the operator instructed the participant how to control the self-paced calibration procedure by using a keyboard press. The operator then draped a black cloth over the participant's head. Participants then conducted a self-paced 9-point calibration procedure (participant-controlled calibration has been shown to result in better accuracy and precision than operator- or system-controlled calibration, see Nyström, Andersson, Holmqvist, and Van De Weijer (2013)). After the calibration procedure, the operator visually assessed the calibration output. In some cases, the calibration output yielded calibration points either without any data, or with specific inconsistencies (e.g., dispersed gaze points around the calibration point). If necessary, these points were re-calibrated. After examining the calibration output of the Tobii SDK, and after recalibration if this was necessary, the operator instructed the participant to initiate the experiment by a spacebar press.

Instructions to participants

Approximately half of the participants received the *live-instruction* and the other half received the *pre-recorded-instruction*. Based on the unique participant code, participants received one of the instruction-types. The instruction appeared on the computer screen after the participant initiated the start of the experiment by a spacebar press (directly after the calibration procedure).

The *live-instruction* stated: “You are about to look at [name confederate] through a live-video connection. [Name confederate] can see you too via the webcam. There is no audio-connection, so you cannot hear each other. The experiment lasts approximately one minute. Press the spacebar to start”.

The *pre-recorded-instruction* stated: “You are going to look at a video clip of [name confederate]. The video clip is pre-recorded. The experiment lasts approximately one minute. Press the spacebar to start”.

After the participant read the instruction and pressed the space bar, the presentation started directly. Note that both the instructions did not explicitly task the participant to look at the face of the confederate or otherwise to behave in a certain way. Instead, the instructions were formulated in this way to set up the conditions of the experiment, namely to imply to participants whether the presentation of the confederate was ‘live’ or not and whether social interaction was possible.

Presentation procedure

Based on their participant code, participants could either receive video clip A or video clip B (see Figure 1) of the confederate to whom they were assigned. Each video clip presentation lasted approximately 30 seconds. When the video clip ended, a calibration-validation procedure was conducted. This part was included to assess the accuracy of the eye-tracking data. Hereafter, a final screen appeared which stated: *“End of the experiment”*. The participant then left the eye-tracking booth and was directed towards a desk by the operator. Here, the participant was asked to fill in a short questionnaire about their subjective experience during the experiment.

Questionnaire and debriefing after the experiment

The questionnaire at the end of the experiment contained the following text and questions: *“You have just seen [name confederate]. In this experiment, some people are shown a person via a live-video stream, and some people are shown a pre-recorded video of a person. However, the instructions you received prior to the experiment do not necessarily have to correspond to what you have seen, but it is possible. Now, do you think that you were looking at [name confederate] via a live-video connection, or that you have looked at a pre-recorded video clip? Tick the box to answer. (1) a live-video connection, or (2) a pre-recorded video clip”*. This was done to investigate whether participants believed whether the confederate was ‘live’ or not, regardless of the instructions they received prior to the presentation. Hereafter, participants were asked: *“How confident are you about your judgement? Please rate how confident you are by putting a cross on the confidence-scale. 0 means that you are completely not confident and 10 means you are completely confident”*. This questionnaire item consisted of a 10-point scale. Finally, the questionnaire contained one more open question asking participants to elaborate on why they chose their answers to the previous questions. After filling in the questionnaire the participant was dismissed. We did not debrief the participants directly after the experiment, because some participants might have talked to other visitors (and potential participants) about the experiment. Instead, we posted a general debriefing on the festival’s public webpage in which we revealed the setup of the experiment, as well as some preliminary results about how many people believed that the presentation was ‘live’ or not.

EYE TRACKING DATA: SIGNAL PROCESSING AND ANALYSIS

Eye-tracking data of the left eye were imported into MATLAB 2017a. Subsequently, we used custom MATLAB software to process the data according to the following steps:

1. **From raw data to fixations.** As a first analysis step, we classified fixations with a modified version of the adaptive velocity-threshold method devised by Hooge and Camps (2013). We chose this algorithm because it can classify fixations on slowly moving objects (a moving face in our case). To produce a velocity signal, we fitted data segments of seven subsequent data points in the position signal with a parabola (instead of three as in Hooge and Camps (2013)). The derivative of the fitted function was used to assign a velocity to the center-point (the fourth of the seven data points). This was done for all data points except the first and the last three of the eye-tracking data. The absolute velocity signal was computed by taking the magnitude of the combined horizontal and vertical eye-tracking signal. All data points having absolute velocities higher than the threshold (the average velocity plus three times the standard deviation of the velocity) were removed. This procedure was repeated until the velocity threshold converged to a constant value or until the number of repetitions reached 50. All episodes that survived the data removal were labeled as fixations. One may argue that we label smooth pursuit episodes as fixations because: 1) our visual stimuli contain slowly moving objects, and 2) our fixation classifier basically selects slow phases from the gaze-position signal containing both slow and fast phases. Note that this is intentional because in this study we define a fixation as a period during which a (moving) object in the world is projected to roughly the same location on the fovea (for a discussion on definitions of fixations, see Hessels, Niehorster, Nyström, Andersson, and Hooge (2018)). The product of the fixation classification is a list with subsequent fixations and their properties (fixation start time, fixation duration and mean fixation position per fixation).
2. **Determining positions of facial features in the video.** Our area of interest (AOI) analysis is based on a recently developed technique (Hessels, Benjamins, Cornelissen, & Hooge, 2018), which employs open-source face-detection and facial-landmark detection algorithms (Baltrušaitis, Robinson, & Morency, 2013; Baltrušaitis, Robinson, & Morency, 2016). This method has been extensively validated for eye-tracking data of observers watching videos of faces (Hessels, Benjamins, et al., 2018). In the current analysis, we used this technique to determine the position of the eyes, nose and mouth for every frame of the video clip.

3. **Assigning AOI labels to fixations.** For each fixation start, we determined the closest preceding video frame start in time. This video frame was linked to the fixation. For each frame, we have coordinates of the centers of facial features. To link a fixation to a facial feature (from now on called an AOI), we computed the distances from the fixation location to the centers of the facial features. The AOI label of the closest facial feature is assigned to the fixation. However, if this distance is larger than 120 pixels (3.05°) we labeled this fixation ‘Other’-AOI because the fixation is outside the eyes, nose and mouth AOI. This method has previously been described as the Limited-Radius Voronoi Tessellation (LRVT) method and has been extensively validated as an objective and noise-robust solution that allows for better between-group and cross-study comparisons of AOIs in face-scanning research (for more details on this method, see Hessels, Kemner, van den Boomen, and Hooge (2016)). The AOI span in the current study was 2.035° , reflecting the *“mean distance from each AOI cell center to the cell center of its closest neighbor”* (Hessels et al., 2016, p. 1701).
4. **From fixations to dwells.** We transformed the fixations into ‘dwells’ by determining the uninterrupted visits in AOIs, for example, see Holmqvist et al. (2011, pp. 189-190). From this, we also computed dwell time, that is the time from entering to leaving a specific AOI (eyes, nose, mouth). Note that the duration of saccades made within a specific AOI is included in the dwell time measure.

Relative total dwell time. Finally, we computed a relative total dwell time measure for each participant. This measure was computed by summing up the dwell times for each AOI (eyes, nose, mouth and ‘other’-AOI) and dividing them by the duration of the video clips. This was done because the video clips contained slight variations in timing. Thus, the relative total dwell time measure indicates what percentage of the total video-clip duration, approximately 30 seconds, a participant spent looking at one of the AOIs. For example, a person might spend 60% of the time looking at the eyes AOI, 20% looking at the nose AOI, 15% at the mouth AOI, and the remaining 5% at the ‘other’-AOI. The relative total dwell time measure will be used in all subsequent analyses to assess how much a person looked at one of the areas of the face.

RESULTS

Did participants believe there was a ‘live’ person or not?

Our first research question was to investigate why people experience a potential for social interaction and what factors contributed to this experience. Therefore, we first assessed whether participants believed there was a ‘live’ person or not. We analyzed participants’ responses to the questions given after the experiment. 1 person could not decide and was therefore excluded from the following analyses. Of the remaining 81 participants, 38 participants responded ‘live-video connection’ (46,9%) and 43 participants responded ‘pre-recorded clip’ (53,1%). Next, we tested whether the participants’ responses and the instructions they received prior to viewing the confederate were related to each other. In other words, we wanted to know whether instruction-type (‘live’ or ‘pre-recorded’) predicted how participants responded to the questionnaire items, even though all participants were told that the instructions did not necessarily correspond to the presentation. A chi-squared test conducted in JASP (JASP Team, 2018) showed that there was no significant association between participants’ responses and instruction-type ($\chi^2(1) = 3.466, p = 0.063$) at an alpha level of 0.05. The observed phi-coefficient was 0.207, which is considered a small-to-moderate effect size (Ferguson, 2009; Sullivan & Feinn, 2012). Although we could not show that instructions had a statistically significant effect on how participants responded, the instructions did seem to have some influence given the small-to-moderate effect size estimate. Moreover, looking at the proportions of participants whose responses aligned with the instruction-type, we observed that 57.8% of the participants who received the live-instruction answered ‘live-video connection’ and 62.8% of the participants who received the pre-recorded-instruction answered ‘pre-recorded clip’. Conversely, participants’ responses did not align with the received instruction-type. 42.2% of the participants who received the live-instruction answered ‘pre-recorded clip’ and 37.2% of the participants who received the pre-recorded-instruction answered ‘live-video connection’. As such, these results need to be interpreted with some caution. Next, we also tested whether participants’ responses were dependent on the video-clip type (video clip A vs. video clip B). A chi-squared test showed there was no significant association between participants’ responses and video clip-type ($\chi^2(1) = 0.302, p = 0.582$) at an alpha level of 0.05. The observed phi-coefficient was 0.061. The effect size estimate was very small and thus it is not very likely that video clip-type modulated participants’ responses to this questionnaire item. In Table 1, we depict the proportion of participants who answered ‘live-video connection’ or ‘pre-recorded clip’, categorized by instruction-type and video clip-type.

Table 1. Participants' responses (N = 81)

Condition	Response	
	Live-video connection	Pre-recorded clip
Live instruction (N=38)	27.2%	19.8%
<i>Video clip A</i>	14.8%	8.7%
<i>Video clip B</i>	12.4%	11.1%
Pre-recorded instruction (N=43)	19.7%	33.3%
<i>Video clip A</i>	7.4%	19.8%
<i>Video clip B</i>	12.3%	13.5%
Total (N=81)	46.9%	53.1%

Finally, we tested whether participants' responses depended on instruction-type and video clip-type combined. In other words, we tested whether participants' responses were a result of an interaction between instruction-type and video clip-type. A chi-squared test revealed that this was not the case ($\chi^2(3) = 5.674, p = 0.129$). The effect size estimate was 0.265 (here computed as a Cramer's *V* coefficient), which is considered a small-to-moderate effect size. Furthermore, participants who answered that the confederate was 'live' did not differ from those who answered 'pre-recorded' in their confidence ratings (measured on a scale from 1 to 10). Moreover, the confidence ratings did not substantially differ when we compared participants based on instruction-type and video clip-type.

Why did participants believe the person was 'live' or not?

In order to further investigate why participants experienced a potential for interaction and what factors contributed to this experience, we also analyzed participants' first-person responses to the open question after the experiment. Participants were asked to give a brief explanation as to why they answered that the person displayed on screen was 'live' or not. First-person responses were labeled and then grouped into several overarching categories by authors GAH and RSH. During this coding process, we found that some participants gave multiple reasons to support their answer. Therefore, we coded the responses separately for each reason provided. The full coding scheme is depicted in Table 2. Note that the reported numbers in Table 2 and in the text do not always refer to the number of participants, but instead to the number of responses given by participants that were coded under a specific label (these numbers are reported in *italics*).

1. **Attitudes.** In this category, we labeled answers that were suggestive of participants' attitudes and skepticism regarding the experiment (2), or when participants did not provide any specific reasons (4). For example, one participant responded: *"During the experiment I already questioned whether it was 'live' and after this question I doubt it even more"*. Another participant answered: *"No idea, it can both be true"*. Participants in this category also referred explicitly to the instructions given prior to the experiment (8), even though participants were specifically told that the instructions did not necessarily correspond to whether the presentation was 'live' or not. For example, one participant who answered 'pre-recorded clip' responded: *"This was described in the instruction"*. In total, 6 participants whose answers were labeled under 'Attitudes' answered 'live-video connection' and 8 participants answered 'pre-recorded clip'.

2. **Behavior of the confederate.** In this category, we labeled answers that explicitly referred to the confederate's behavior. Many participants who believed that the confederate was 'live' answered that they perceived synchrony and reciprocity with the confederate's behavior (14). For example, these participants wrote: *"There was action-reaction"*, *"It felt very natural, like she was interacting with me"*, *"He reacted when I laughed"*, and *"It was like she was mirroring me"*. Interestingly, many participants who commented on the confederate's behavior also answered 'pre-recorded clip'. This group either answered that the behavior of the confederate appeared to have no particular cause (12), or that they perceived 'strange' or 'inappropriate' timing in the behavior of the confederate (4). For example, one participant responded: *"She winked, and I winked back, but then she did not react as I expected"*. Similarly, other participants in this group wrote: *"His reaction / facial expression did not seem to have a cause"* and *"He smiled at a strange moment"*. Others also commented that the behavior of the confederate seemed rehearsed (11), for example, one of these participants answered: *"It looked fake, the way he moved his eyes looked planned and rehearsed"*. In total, 20 participants whose answers were labeled under 'Behavior of Confederate' answered 'live-video connection' and 24 participants answered 'pre-recorded clip'.

3. **Context.** In this category, we labeled answers of participants who commented on specific aspects of the research context (18), such as the fact that confederates walked in and out of the booth, the confederates clothing, or that they were talking with confederates prior to the experiment. For example, one participant answered that: *"I saw the person walking into the booth"* and another participant responded: *"I just talked to him, and he was wearing the same shirt"*. 12 participants in this 'Context' category answered 'live video-connection' and 6 participants answered 'pre-recorded clip'.

4. **Technicalities.** A small group of participants commented on several technicalities of the experimental setting (9). For example, one participant responded: *“It was bright in the room, not dark”*, mentioning the fact that the confederate booth from the outside seemed dark inside, while in the video the background of the room was illuminated. Furthermore, some participants commented on incongruencies and mismatches with the confederate or the room. For example, one participant wrote: *“He was sweating when I just saw him, but not in the video”*. From this category, 7 participants answered ‘pre-recorded clip’ and 2 participants answered ‘live video-connection’.

Table 2. Coding scheme for first-person responses (N = 81)

Categories / labels	Response	
	Live video-connection	Pre-recorded clip
Attitude (14)	6	8
Instructions (8)	3	5
Improbable / skeptical (2)	0	2
Random / no reason provided (4)	3	1
Behavior of confederate (44)	20	24
Good timing behavior (14)	13	1
Strange timing behavior (4)	2	2
No cause behavior (12)	1	11
Authentic / natural (2)	2	0
Rehearsed / planned (11)	1	10
Experience of eye contact (1)	1	0
No experience of eye contact (2)	0	2
Context (18)	12	6
Match confederates (14) (e.g., clothing, talking, location)	12	2
Mismatch confederates (4) (e.g., appearance, location)	0	4
Technicalities of setup (9)	2	7
Image quality (2)	1	1
Image too light (3)	1	2
No camera visible (1)	0	1
Room mismatch (1)	0	1

To conclude, we identified four main categories (see Table 2) under which we could label participants' first-person responses. These factors seem to be important for participants' experience of social interaction. The category: 'behavior of bonfederate' captured by far the largest part of participants' responses, showing that how participants experienced the behavior of the confederate was strongly associated with whether the participants responded that the presentation was 'live' or a 'pre-recorded clip'.

Eye-tracking data quality assessment

Prior to analyzing participants' gaze behavior, we first assessed eye-tracking data quality. Data quality is an important part of eye-tracking research and essential to the validity of study results (Holmqvist, Nyström, & Mulvey, 2012; Nyström et al., 2013). The following measures were used. First, we computed the percentage of data loss per participant (i.e., no gaze coordinate was reported). Second, we determined precision by computing root mean squared sample-to-sample deviation (RMS s2s) of the raw gaze position signals for every participant (Holmqvist et al., 2011). If eye-tracking recordings showed more than 25% data loss, or if the precision of the recordings exceeded 2 SD's above the mean RMS displacement, we excluded these recordings from subsequent analyses. Based on these criteria, 14 out of 82 participants were excluded (see Figure 4 for data loss, RMS s2s deviation and the exclusion criteria). Next, we computed the accuracy (i.e., systematic error) for the remaining 68 participants by measuring the difference between the average fixation location and the calibration point (averaged over participants and calibration points). The systematic error was found to be 0.65° on average (for more details, see Hooge et al. (2018)). We concluded that data quality for the remaining 68 participants was adequate for the subsequent analyses, because the average systematic error was approximately three times as low as the average AOI span of 2.035° (see Method section). Furthermore, data loss and RMS s2s deviation were low enough to allow for adequate fixation classification.

Figure 4. Visualization of data loss and RMS sample-to-sample (s2s) deviation. Every circle represents one participant. The dashed lines represent the exclusion criteria used; data loss > 25% on the y-axis, RMS s2s deviation (°) > 2 SD's above the mean on the x-axis. Open circles represent the included participants (N=68), closed circles represent the excluded participants (N=14).

Relative total dwell times on face AOIs

As stated previously, we wanted to investigate whether implying social interaction modulates gaze behavior to the eyes. However, first we analyzed how participants looked at the face of the confederate irrespective instruction-type and video clip-type. It is clear from Figure 5 that many participants gazed predominantly at the eyes AOI. Furthermore, as shown by the individual data points, there is a wide range in relative total dwell time per AOI across individuals. Some participants gazed approximately 70 – 100% of the time at the eyes, while other participants gazed at the eyes for 10 – 50 % of the time. Moreover, some participants gazed at the nose, mouth, or ‘other’– AOI for a relatively large amount of the time, while other participants only briefly looked at these AOIs, or not at all. Inter-individual variability in gaze behavior to facial features has been described in more detail in previous eye-tracking studies (Arizpe et al., 2017; Mehoudar, Arizpe, Baker, & Yovel, 2014; Peterson & Eckstein, 2013; Peterson, Lin, Zaun, & Kanwisher, 2016). In short, these studies show that face scanning behavior can be quite idiosyncratic and remain stable over time. Furthermore, face scanning behavior is also likely influenced

by particular task demands (e.g., face recognition, emotion recognition, free viewing).

Figure 5. Box and whisker plots for relative total dwell time (%) per Area of Interest. Boxplots are overlaid with individual data points which are jittered in the vertical direction to avoid overlap. Each dot represents data from one of 68 participants. Boxes cover the 25th to 75th percentile (inter-quartile range; IQR). The middle of the box represents the median. Whiskers extend from the 25th percentile and 75th percentile to cover all data points lying within 1.5 times the IQR from the 25th and 75th percentile respectively. Notice that the relative total dwell times (%) on the nose, mouth, and ‘other-AOI are not normally distributed. Many participants gazed at these AOIs only for a relatively small amount of the video clip duration (≈ 30 seconds), or did not look at these AOIs at all, hence the bounded and skewed distributions and the large clusters of datapoints at or near zero.

Does implied social interaction modulate gaze behavior to the eyes?

Our second research question was whether implying social interaction would modulate how people look at faces, specifically the eyes. As described previously, we expected that participants who believed that they were presented with a ‘live’ person would gaze less at the eye region of a face compared with participants who did not believe they were presented with a ‘live’ person. However, the first-person responses indicated that participants varied greatly in their beliefs about whether the confederate displayed on screen was ‘live’ or not and why they thought so. Therefore, we first analyzed relative total dwell time on the eyes AOI as a function of instruction-type. Subsequently, we analyzed

relative total dwell time on the eyes AOI as a function of participants' post-hoc beliefs. Note that participants who received the live-instruction are not necessarily the same as the participants who believed the confederate was 'live'. The numbers of participants per group are reported for the specific analyses in the following sections.

Gaze behavior to the eyes as a function of instruction-type

First, we compared gaze behavior at the eyes between participants who received the live-instruction and participants who received the pre-recorded-instruction. We expected that participants who received the live-instruction ($N = 30$) would gaze less at the eyes compared to the participants who received the pre-recorded-instruction ($N = 38$). In Figure 6 (upper panel), we visualized the individual data points of relative total dwell time on the eyes AOI for both groups. Then we computed the median relative total dwell time on the eyes AOI for both groups (see Figure 6, upper panel). This was done because the distribution of relative total dwell time to different areas of the face were both bounded and skewed for some AOIs. In such cases, the mean is not always a good measure to determine the central tendency of the distribution (for a discussion, see Rousselet, Pernet, and Wilcox (2017)). We found that the median relative total dwell time on the eyes AOI was 60.19% (Inter-Quartile-Range = 46.00% - 81.26%) for participants who received the live-instruction and 73.08% (IQR = 58.95% - 82.72%) for participants who received the pre-recorded-instruction. As is visible from Figure 6 (upper panel), many participants in both instruction-type groups looked predominantly at the eyes (approximately more than 70% of the time).

Next, we used a statistical tool called the shift function, illustrated in Figure 6 (lower panel). The shift function not only allows one to visualize and quantify whether two distributions differ, but also how they differ (for more details on the shift function, see Rousselet et al. (2017)). For example, if the relative total dwell time on the eyes AOI would be lower for participants in the live-instruction group, the decile differences should be negative and the shift function should indicate a straight line below the zero-difference line (horizontal dashed line). In our case, the shift function for independent groups indeed shows that the decile differences are negative for decile 1 until decile 5 (the median). However, the decile differences then become progressively less negative (decile 6 and 7) and shift back to the zero-difference line for decile 8 and 9. Deciles were computed using a quantile estimator (see Rousselet et al., 2017) for more details) and the error bars represent a 95% confidence interval computed by using bootstrapping (number of

samples was set to 2000). Finally, we computed a one-tailed Mann-Whitney-Wilcoxon test (MWW) in JASP (JASP Team, 2018) to test whether the median relative total dwell time on the eyes-AOI was significantly smaller for the live-instruction group compared with the pre-recorded instruction group. As stated, the distributions of relative total dwell times for some AOIs were skewed and bounded and we therefore used a non-parametric test. The median relative total dwell time on the eyes was not significantly smaller for the live-instruction group ($Mdn = 60.19\%$) than for the pre-recorded instruction group ($(Mdn = 73.08\%), W = 446.0, p = 0.064$). The effect size (here reported as a Rank-Biserial Correlation) was $r = -0.218$, which indicates a small-to-moderate negative correlation. Based on the shift-function and the effect size estimate (Kerby, 2014), we can infer that a larger subset of the live-instruction group gazed little at the eyes of the confederates compared with the pre-recorded instruction group. Yet, a subset of the live-instruction group looked predominantly at the eyes (i.e., approximately 70-100% of the total duration of the video clip). Since we found a p -value of 0.064, it is important to interpret these results with some caution. It could be argued that the sample sizes of the two instruction-type groups were too small to detect significant differences in gaze behavior at the eyes on a group-level. However, in Gregory & Antolin's (2019) study similar sample sizes were used to demonstrate a statistical difference in gaze behavior to faces between the pre-recorded-viewing group (28 participants) and the two-way-viewing group (34 participants). As such, the shift-function is particularly useful here, as it allows us to compare how the two groups differ.

Figure 6. Distributions of relative total dwell time on the eyes AOI as a function of instruction-type (upper panel) and the corresponding shift function (lower panel). Upper panel: Distributions of relative total dwell time on the eyes AOI. Participants are grouped by the instruction they received prior to the experiment (live-instruction group $N=30$, pre-recorded instruction group $N=38$). Data points are jittered vertically to prevent overlap and each dot represents data from one participant. Vertical lines mark the deciles of the distribution. The thicker vertical line represents the median (50th percentile). Lower panel: The shift function for independent groups (see Rousselet et al. (2017)). The differences between deciles are plotted as a function of deciles in the live-instruction group. The horizontal dashed line represents the zero-difference line. The vertical dashed line represents the median. The x-axis shows the deciles of the live-instruction group and the y-axis shows the decile differences in relative total dwell time on the eyes AOI. The error bars of the deciles represent a 95% confidence interval computed using a percentile bootstrap (number of samples was set to 2000). The decile differences in relative total dwell time on the eyes AOI for the live-instruction group are negative for decile 1 until decile 5 (the median). However, the decile differences then become progressively less negative (decile 6 and 7) and shift back to the zero-difference line for decile 8 and 9, as indicated by the zero-difference line.

Gaze behavior to the eyes as a function of beliefs

As a second approach to investigate whether implying social interaction modulated gaze behavior to the eyes, we grouped participants based on the responses they gave after the experiment. We compared the relative total dwell time on the eyes of participants who answered 'live-video connection' ($N = 33$) with the relative total dwell time on the eyes AOI of participants who answered 'pre-recorded clip' ($N = 34$). In Figure 7 (upper panel), we visualized the individual data points of total dwell times to the eyes AOI for both groups. The median relative total dwell time on the eyes AOI was 72.18% (IQR = 57.63% - 82.09%) for the group of participants who answered 'live-video connection' and 63.53% (IQR = 49.96% - 79.31%) for the group of participants who answered 'pre-recorded clip'. Again, we plotted the shift function to quantify how the two distributions differ (see Figure 7, lower panel). While the decile differences show a slight increase from decile 1 until decile 5, the group differences in relative total dwell time on the eyes AOI are much smaller than in the instruction-based comparison (see above). The error bars of the deciles also show consistent overlap with the zero-difference line, indicating no clear differences between the two groups.

As the median of relative total dwell time on the eyes AOI was larger for the group of participants who answered 'live-video connection' ($Mdn = 72.18\%$) than the median of relative total dwell time on the eyes AOI for the group of participants who answered 'pre-recorded clip' ($Mdn = 63.53\%$), a statistical analysis to verify whether participants who believed the confederate was live gazed less at the eyes would be superfluous. We conducted some follow-up tests to only look at participants whose beliefs also aligned with the instructions they received. First, we correlated the confidence ratings given by participants with regards to their responses with participants' relative total dwell time on the eyes AOI. The correlations between confidence ratings and relative total dwell time on the eyes AOI were $r = -0.06$ for the group who answered 'live video-connection' and $r = -0.11$ for the group who answered 'pre-recorded clip'. Finally, we visualized the distributions of the following two subgroups: 1) participants who received the live-instruction and who responded 'live-video connection' ($N=18$) and 2) participants who received the pre-recorded-instruction and who responded 'pre-recorded clip' ($N=23$), see supplementary materials. These follow-up analyses did not yield any additional insights.

Figure 7. Distributions of relative total dwell time on the eyes AOI as a function of participants self-reported ‘beliefs’ (upper panel) and the corresponding shift function (lower panel). Upper panel: Distributions of relative total dwell time (%) on the eyes AOI. Data points are jittered vertically to prevent overlap and each dot represents data from one participant. Participants are grouped according to their ‘belief’ in whether the presentation was ‘live’ or not (live-belief group $N=33$, rec-belief group $N=34$). Vertical lines mark the deciles of the distribution. The thicker vertical line represents the median (50th percentile). Lower panel: The shift function for independent groups. The differences between deciles are plotted as a function of deciles in the live-belief group. Decile differences in relative total dwell time are plotted on the y-axis. The horizontal dashed line represents the zero-difference line. The vertical dashed line represents the median. The x-axis shows the deciles of the live-belief group and the y-axis shows the decile differences in relative total dwell time.

From instruction to beliefs

Hitherto, we have analyzed gaze behavior to the eyes as a function of instruction-type and whether participants believed the other person was ‘live’ or not. However, one problem is that participants had to formulate their beliefs after the presentation was finished. As

Figure 8. Relative total dwell time on the eyes AOI over time as a function of instruction-type (left panel) and participants' beliefs (right panel). Lines depict average relative total dwell time (%) over time (in seconds). Shaded regions represent the standard error of the mean. The relative total dwell time on the eyes AOI is consistently lower for the live-instruction group compared with the pre-recorded instruction group. Conversely, the relative total dwell time on the eyes AOI is consistently higher for the live-belief group compared with the pre-recorded belief group.

such, we cannot know whether participants doubted whether the confederate was ‘live’ during the presentation itself, or whether they primarily formulated post-hoc beliefs. It could be that they did not question whether the confederate was ‘live’ or not during the presentation, and if some participants did question this (as some first-person reports indicate), they only likely did so after a substantial part of the presentation. We reasoned that, if participants realized that the confederate was not ‘live’ during the presentation, we might observe changes in gaze behavior over time, which are not readily visible in aggregated measures (as in Figures 6 and 7). To check this, we visualized relative total dwell time on the eyes AOI over the course of the presentation. As is visible from Figure 8, the temporal development of total dwell time on the eyes AOI seems to be representative of the aggregated endpoints. Relative total dwell time on the eyes AOI is consistently lower for the live-instruction group compared with the pre-recorded instruction group. Conversely, relative total dwell time on the eyes AOI is consistently higher for the live-belief group compared with the pre-recorded belief group. Note that there is considerable overlap between the two groups, as is visible from the shaded regions. All in all, we do not observe clear changes in gaze behavior to the eyes over the course of the presentation. The initial increase in relative total dwell time on the eyes AOI at the beginning of the presentation (approximately the first 5 seconds) is likely due to different fixation positions at the beginning of the recording.

DISCUSSION

We investigated why people experience the potential for social interaction and tested whether implying social interaction modulated gaze behavior to faces, specifically the eyes. Previous work has shown that people look differently at ‘live’ people compared with videos of faces (Laidlaw et al., 2011), or when they believe that they are being watched by others (Gregory & Antolin, 2019). However, these studies have primarily investigated how social presence, in the absence of social interaction, influences gaze behavior to faces. Our study was designed to overtly imply the potential for social interaction with another person and simultaneously measure how people look at faces. First, we will discuss our findings regarding participants’ experience of social interaction and then we will discuss our findings with regard to our hypotheses about gaze behavior to faces and eyes.

Implying social interaction: the importance of responsiveness and synchronicity

Our experiment was a novel attempt at implying the potential for social interaction. In previous work, potential for social interaction was only minimally implied, as confederates did not convey the initiative to engage in interaction (Gregory & Antolin, 2019; Laidlaw et al., 2011). Our goal was to uncover why social interaction may be experienced and how gaze behavior to the eyes was modulated as a result. Our research setting was structured in such a way that participants were led to believe that they would engage in social interaction with another person. Indeed, our paradigm seemed to have succeeded in conveying the potential for social interaction, as many people responded that they felt that the other person was ‘interacting’ and ‘responding’, even though participants were all presented with a pre-recorded video clip. Based on participants’ responses, half of the participants believed that they were presented with a person through a live-video connection. Participants’ responses were not strongly correlated with instruction-type or video-type, although some participants did base their judgements on the instruction they received prior to the experiment. Note that we only asked participants to judge whether they believed the other person was ‘live’ after the experiment. We found that many participants, when judging whether the confederate was ‘live’ or not, commented on the temporal and reciprocal aspects of the confederate’s behavior (e.g., timing of the gaze cues and facial expressions). Interestingly, while some participants reported that they perceived the confederate’s behavior to be ‘interactive’ (e.g., synchronicity with their own behavior), other participants perceived the confederate’s reactions to be strange and discontinuous. This clearly shows the importance of the subtle temporal interplay between the participants’ behavior and the perceived reciprocity and responsiveness of the confederate. As such, these first-person reports provide important insights regarding participants’ subjective experience of potential social interaction, and also to what extent potential interactions can be simulated in a non-interactive experimental setting.

Our first-person reports may be particularly relevant for researchers studying social cognition and gaze behavior during social interactions (Risko et al., 2016). In particular, the insights we derived from the first-person reports could be a starting point for the development of experimental designs aimed to simulate the experience of social interaction (Caruana et al., 2017; Pfeiffer et al., 2013; Schilbach, 2015). First-person reports indicated that the temporal and reciprocal aspects of the confederate’s behavior were particularly important for how people experienced the potential for interaction. We wondered how

one could further disentangle or model these aspects in order to optimize the experience of social interactions in experimental settings. One clever approach to disentangle these factors was achieved by Pfeiffer, Timmermans, Bente, Vogeley, and Schilbach (2011). In their so-called ‘non-verbal Turing test’, participants were asked to judge whether the behavior of a virtual avatar was controlled by a person or by a computer program. The ingenuity of this study was that the gaze direction of the virtual avatar was responsive to where the participant was looking. By systematically varying between gaze-following and gaze-aversion, Pfeiffer et al. (2011) demonstrated that participants’ judgement about the ‘humanness’ of the virtual avatar was dependent on the gaze congruency (i.e., gaze direction relative to person’s gaze) and gaze contingency (i.e., the temporal dynamics). Furthermore, judgements about ‘humanness’ were also dependent on participants’ beliefs about whether the virtual avatar would behave competitively or cooperatively. In other words, the simulated responsiveness of the virtual avatar influenced participants’ sense of social interaction. Brandi, Kaifel, Lahnakoski, and Schilbach (2019) also used a virtual avatar whose gaze behavior was responsive to the measured gaze behavior of the participant. Brandi and colleagues found that systematic changes in the responsiveness of the virtual avatar also influenced participants’ judgements about the ‘humanness’ and social agency of the virtual avatar.

Taken together, implying social interaction certainly seems possible. However, staging a full-fledged experience of social interaction may not be as straightforward as merely implying social presence. In our study, the first-person responses indicated that many participants perceived the confederate as being responsive to their own behavior. Yet, many participants also perceived the confederate’s behavior as being ‘out of sync’. While social presence and potential interactions may overlap in certain situations, social interaction likely constitutes a qualitatively different domain of the social context. The potential for social interaction likely depends on whether potential interactors overtly signal and engage in ‘interactive’ behavior (De Jaegher et al., 2010; Valenti & Gold, 1991). Importantly, we only used a limited set of possible social gaze cues and obviously excluded many social signals such as speech and gestures that are also typically used by people to interact (Duncan & Fiske, 2015). Furthermore, we primarily focused on why people experienced interaction with another person and whether implying social interaction modulated gaze behavior to the eyes.

Implying social interaction and its influence on gaze behavior to the eyes

Previous studies have demonstrated that the actual, imagined or implied presence of others modulated basic processes of visual attention and gaze behavior, for example, how people look at other people's faces (Gobel et al., 2015; Gregory & Antolin, 2019; Laidlaw et al., 2011). However, few studies have specifically addressed whether implying social interaction modulates gaze behavior to the eyes. Based on previous studies, we expected that prolonged direct gaze with a stranger would cause people to avoid looking at that person's eyes (Gregory & Antolin, 2019; Laidlaw et al., 2011). The 'social risk' hypothesis, for example, predicts that prolonged direct gaze may result in overall gaze avoidance (Gregory & Antolin, 2019), as prolonged direct gaze with a stranger is typically regarded as inappropriate, uncanny, or may even be perceived as aggressive and threatening (Dovidio & Ellyson, 1985; Ellsworth et al., 1972; Hamilton, 2016; Kleinke, 1986). In our study, we expected that participants who believed they were being looked at by a 'live' confederate would gaze less at the eyes of that person. However, our data only partly supports this hypothesis. We found that a relatively large subset of the live-instruction group gazed little at the eyes of the confederate compared with the pre-recorded-instruction group. However, many participants from the live-instruction group predominantly gazed at the eyes of the confederate.

In sum, we found only small differences in gaze behavior to the eyes between the live-instruction group and the pre-recorded-instruction group (as indicated by the shift function). Furthermore, when grouping participants based on their self-reported beliefs, the median relative total dwell time on the eyes AOI was numerically higher for the live-belief group participants compared with the pre-recorded belief group. This was against our initial expectations. However, as clarified by the first-person responses, participants who received the live-instruction were not necessarily the same participants who believed that the confederate was 'live', making this result harder to interpret. Moreover, confidence ratings did not correlate with participants' relative total dwell time on the eyes AOI, nor were they associated with instruction-type or video clip-type.

One could argue that our study does not have a large enough sample size. Yet, Gregory & Antolin's study (2019) used a similar number of participants and reported much larger effect sizes, as well as significant statistical differences between groups based on instruction-type. Although the study by Gregory & Antolin probably offers the best comparison for sample sizes given the overlap in research questions and hypotheses about

gaze behavior to eyes and faces, the experimental procedures and analyses also differ on several levels. For example, the confederate in the video clip in Gregory & Antolin's study did not generally look at the participants, but only did so for a brief moment. In our study, all video clips displayed confederates who directly looked at the observer (into the camera), which likely constitutes a rather different viewing experience. Importantly, we also used different methodological tools to map gaze behavior unto dynamic face stimuli. This allowed us to measure gaze behavior to specific facial features (e.g., eyes, nose, mouth), whereas Gregory & Antolin's study focused primarily on gaze behavior to the face in comparison with the background of the scene. For two smaller groups of participants, they report gaze behavior to the eyes. In our study, we focused on the entire distributions of relative total dwell time on the eyes AOI, and showed that participants varied greatly in how long they gazed at the eyes of the confederate, irrespective of the instruction-type, video clip-type or their self-reported belief. Indeed, other eye-tracking studies have reported large inter-individual differences in face scanning patterns. For example, several eye-tracking studies have indicated that how individual observers look at faces can be characterized by idiosyncratic face scanning patterns that are stable over time (see e.g., Arizpe et al., 2017; Mehoudar et al., 2014; Peterson & Eckstein, 2013). Simply put, some people may preferentially fixate the eyes while others generally fixate the nose or the mouth area. Furthermore, differences in gaze behavior to faces also seem to be modulated by specific task-demands, for example, during speaking, listening, free viewing, face recognition and emotion recognition, but also by cultural factors (Blais, Jack, Scheepers, Fiset, & Caldara, 2008). We did not specifically design this study in order to investigate to what extent our participants exhibited idiosyncratic biases in gaze behavior to faces, nor did we impose a specific task to participants. However, the extent to which gaze behavior to the eyes was modulated as a function of instruction-type could have been outweighed by these inter-individual differences.

CONCLUSION

We demonstrated a novel method for implying the potential for social interaction with another person. We have highlighted several factors that are likely to be important for implying social interaction, namely the behavior of the confederate, the attitudes of participants, and various contextual and technical aspects of the research setting. Furthermore, we found only marginal evidence supporting the 'social risk' hypothesis. In fact, participants varied greatly in gaze behavior to the eyes, irrespective of instruction-

type or participants' beliefs about the possibility to interact. Based on these results, as well as insights from previous research, we think that specific factors of the social context likely modulate how people allocate their gaze. Furthermore, we think that individual differences in gaze behavior to faces need to be more explicitly considered in future experiments, as they could be crucial to our understanding of how gaze behavior is allocated across different social contexts.

APPENDIX 1: FOLLOW-UP ANALYSIS

Follow-up analysis of gaze behavior to the eyes for subgroups (instruction-type aligned with participants' beliefs)

Figure 9. Distributions of relative total dwell times on the eyes AOI for two subgroups: instructions aligned with beliefs (lower panel). Upper panel: Distributions of relative total dwell time on the eyes AOI. Participants are grouped by the instruction they received aligned with their self-reported beliefs after the experiment. Live-group shows data points from participants who received the live-instruction and answered 'live video-connection' (N=18). Rec-group shows data points from participants who received the pre-recorded clip instruction and answered 'pre-recorded clip' (N=23). Data points are jittered vertically to prevent overlap and each dot represents data from one participant. Vertical lines mark the deciles of the distribution. The thicker vertical line represents the median (50th percentile). Lower panel: The shift function for independent groups.

The differences between deciles are plotted as a function of deciles in the live- group. The horizontal dashed line represents the zero-difference line. The vertical dashed line represents the median. The x-axis shows the deciles of the live-group and the y-axis shows the decile differences in total dwell time. The error bars of the deciles represent a 95% confidence interval which was computed by using a percentile bootstrap (number of samples was set to 2000).

Chapter 5

Gaze and speech in parent-child interactions

Holleman, G. A., Hooge, I. T. C., Huijding, J., Deković, M., Kemner, C., & Hessels, R. S. (in revision). Gaze and speech in parent-child interactions: The role of conflict and cooperation.

Author contributions:

GH, IH, JH, MD, CK and RH designed the study. GH collected the data. GH, IH and RH analyzed the data. GH, IH, and RH interpreted the data. GH drafted the manuscript. GH, IH, JH, MD, CK, and RH finalized the manuscript.

ABSTRACT

A primary mode of human social behavior is face-to-face interaction. In this study, we investigated the characteristics of gaze and its relation to speech behavior during video-mediated face-to-face interactions between parents and their preadolescent children. Parent-child dyads engaged in conversations about cooperative and conflictive family topics. We used a dual-eye tracking setup that is capable of concurrently recording eye movements, frontal video, and audio from two conversational partners. Our results show that children spoke more in the cooperation-scenario whereas parents spoke more in the conflict-scenario. Parents gazed slightly more at the eyes of their children in the conflict-scenario compared to the cooperation-scenario. Both parents and children looked more at the other's mouth region while listening compared to while speaking. Results are discussed in terms of the role that parents and children take during cooperative and conflictive interactions and how gaze behavior may support and coordinate such interactions.

INTRODUCTION

A primary mode of human social behavior is face-to-face interaction. This is the “central ecological niche” where languages are learned and most language use occurs (Holler & Levinson, 2019, p. 639). Face-to-face interactions are characterized by a variety of verbal and nonverbal behaviors, such as speech, gazing, facial displays, and gestures. Since the 1960s, researchers have extensively investigated the coordination and regulation of these behaviors (Duncan & Fiske, 2015; Kelly, Özyürek, & Maris, 2010; Kendon, 1967). A paramount discovery is that gaze and speech behavior are closely coupled during face-to-face interactions. Although some patterns of speech behavior during face-to-face interactions, such as in turn-taking, are common across different languages and cultures (Stivers et al., 2009), the role of gaze behavior in interaction seems to be culturally- as well as contextually-dependent (Foddy, 1978; Haensel et al., 2020; Haensel, Smith, & Senju, 2017; Hessels, 2020; Kleinke, 1986; Patterson, 1982; Rossano, Brown, & Levinson, 2009; Schofield, Parke, Castaneda, & Coltrane, 2008). Observational studies on gaze behavior during interaction have been conducted in many different interpersonal contexts, such as interactions between adults, parents and infants, parents and children, as well as clinical interviews and conversations with typically and atypically developing children (Argyle & Cook, 1976; Arnold, Semple, Beale, & Fletcher-Flinn, 2000; Ashear & Snortum, 1971; Berger & Cunningham, 1981; Cipolli et al., 1989; Kendon, 1967; Levine & Sutton-Smith, 1973; Mirenda, Donnellan, & Yoder, 1983). More recently, new eye-tracking techniques have been developed to measure gaze behavior of individuals during face-to-face interactions with higher spatial and temporal resolution (Hessels et al., 2019; Ho et al., 2015; S. L. Rogers, Speelman, Guidetti, & Longmuir, 2018). However, these techniques have not been used to study the relation between speech and gaze in parent-child conversations.

Parent-child interactions provide a rich social context to investigate various aspects of social interaction, such as patterns of verbal and nonverbal behavior during face-to-face communication. Parent-child interactions are crucial for children’s social, emotional, and cognitive development (Branje, 2018; Carpendale & Lewis, 2004; Gauvain, 2001). Also, the ways in which parents and children interact changes significantly from childhood to adolescence. In infancy and childhood, parent-child interactions play an important role in children’s socialization, which involves the acquisition of language and social skills, as well as the internalization of social norms and values (Dunn & Slomkowski, 1992; Gauvain, 2001). In adolescence, parent-child interactions are often centered around relational

changes in the hierarchical nature of the parent-child relationship, which typically consist of frequent conflicts about parental authority, child autonomy, responsibilities, and appropriate behavior (Laursen & Collins, 2009; Smetana, 2011). According to Branje (2018), *“parent-adolescent conflicts are adaptive for relational development when parents and adolescents can switch flexibly between a range of positive and negative emotions”* (p. 171). As children move through adolescence, they become more independent from their parents and start to challenge parents’ authority and decisions. In turn, parents need to react to these changes and renegotiate their role as a parent. How parents adapt to these changes (e.g., permissive, supporting, or authoritarian parenting styles) may have a significant impact on the social and emotional well-being of the child (Smokowski, Bacallao, Cotter, & Evans, 2015; Tucker, McHale, & Crouter, 2003).

In this study, we investigated gaze behavior and its relation to speech in the context of parent-child interactions. Specifically, we are interested in patterns of gaze and speech behavior as a function of cooperative and conflicting conversation topics, to which the parent-child relation is ideally suited. We focus primarily on gaze behavior because of its importance for perception – does one need to look at another person’s face in order to perceive certain aspects of it? – and its relation to speech in face-to-face interactions (Hessels, 2020; Holler & Levinson, 2019; Kendon, 1967). Although no eye-tracking studies to our knowledge have investigated the role of gaze behavior and its relation to speech in parent-child conversations, several studies have addressed the role of gaze behavior in face and speech perception and functions of gaze during conversational exchanges. Because these lines of research are directly relevant to our current study, we will briefly review important findings from this literature.

Where do people look at each other’s faces?

Faces carry information that is crucial for social interaction. By looking at other people’s faces one may explore and detect certain aspects of those faces such as facial identity, emotional expression, gaze direction, and cognitive state (Hessels, 2020; Jack & Schyns, 2017). A well-established finding, ever since the classic eye-tracking studies by Buswell (1935) and Yarbush (1967), is that humans have a bias for looking at human faces and especially the eyes (see e.g., Birmingham et al., 2009; Hessels, 2020; Itier et al., 2007). This bias already seems to be present in early infancy (Farroni, Csibra, Simion, & Johnson, 2002; Frank, Vul, & Johnson, 2009; Gliga, Elsabbagh, Andravizou, & Johnson, 2009). In a recent review, Hessels (2020) describes that where humans look at faces differs between,

for example, when the face is moving, talking, expressing emotion, or when particular tasks or viewing conditions are imposed by researchers (e.g., face or emotion recognition, speech perception, restricted viewing). Moreover, recent eye-tracking research has shown that individuals exhibit large but stable differences in gaze behavior to faces (Arizpe et al., 2017; Kanan, Bseiso, Ray, Hsiao, & Cottrell, 2015; Mehoudar et al., 2014; Peterson & Eckstein, 2013; Peterson et al., 2016). That is, some people tend to fixate mainly on the eye or brow region while others tend to fixate the nose or mouth area. Thus, what region of the face is looked at by an observer will likely depend on the conditions of the experimental context and on particular characteristics of the individual observer.

Previous eye-tracking research on gaze behavior to faces has mostly been conducted using static images or videos of faces presented to participants on a computer screen. However, some researchers have questioned whether gaze behavior under such conditions adequately reflects how people look at others in social situations, e.g., when there is a potential for interaction (Laidlaw et al., 2011; Risko et al., 2012; Risko et al., 2016). For example, Laidlaw et al. (2011) showed that when participants were seated in a waiting-room, they looked less at a confederate who was physically present compared to when that person was displayed on video-monitor. While this discovery has led researchers to question the presumed ‘automaticity’ of humans to look at faces and eyes, this situation may primarily pertain to *potential interactions*, e.g., situations where social interaction is possible but can be avoided, for example, in some public spaces and on the street (see also Foulsham et al., 2011; Hessels, Benjamins, et al., 2020; Hessels, van Doorn, et al., 2020; Rubo, Huestegge, & Gamer, 2020). Other studies have shown that, once engaged in *actual interaction*, such as in conversation, people tend to look at other people’s faces and its features (Freeth et al., 2013; Hessels et al., 2019; S. L. Rogers et al., 2018). One may thus expect that parents and children will also primarily look at each other’s faces during conversational interactions, but where on the face they will mostly look likely differs between individuals and may be closely related to what the face is doing (e.g., speaking, moving, expressing emotion) and to the social context of the interaction.

Where do people look at each other’s faces during conversations?

Already since the 1960s, observational studies (i.e., not using an eye tracker) of two-person conversations have shown that speakers tend to equally gaze at or away from listeners, whereas listeners gaze longer at speakers with only occasional glances away from the speaker in between (Duncan & Fiske, 2015; Kendon, 1967; Rossano et al., 2009). Yet,

the observational techniques used in these studies have limited reliability and validity to distinguish gaze direction at different regions of the face (Beattie & Bogle, 1982), which is of crucial importance for research on face-scanning behavior (see also Hessels, 2020, p. 869). Conversely, eye-tracking studies on gaze behavior to faces have used videos of talking faces, instead of actual interactions (e.g., Foulsham & Sanderson, 2013; Vatikiotis-Bateson et al., 1998; Võ et al., 2012). Most eye-tracking studies have therefore only been concerned with where people look at faces while listening to another person speak. Võ et al. (2012), for example, presented observers with close-up video clips of people being interviewed. They found that overall, participants gazed at the eyes, nose, and mouth equally often. However, more fixations to the mouth and fewer to the eyes occurred when the face was talking than when the face was not talking. This finding converges with the well-established finding that visual information from the human face may influence, or enhance how speech is perceived (Sumbly & Pollack, 1954). A well-known example is the McGurk effect (McGurk & MacDonald, 1976), where one's perception of auditory speech syllables can be modulated by the mouth and lip movements from a talking face.

Only recently have researchers begun to use eye-tracking technology to measure where people look at each other's faces when engaged in interactive conversational exchanges (Hessels et al., 2019; Ho et al., 2015; S. L. Rogers et al., 2018). Rogers et al. (2018), for example, used wearable eye trackers to measure where two people looked at each other while engaged in short "getting acquainted" conversations. They found that participants gazed away from the face of one's partner for about 10% of the total conversation duration when listening, and about 29% when speaking (cf. Kendon, 1967). When participants gazed at their partner's face, they looked primarily at the eyes and mouth region. Specifically, Rogers et al. (2018) reported that, on average, participants looked slightly more at the mouth area while listening compared to when they were speaking; a difference of approximately 5 percentage points of the time that they were looking at the other person's face. In a different eye-tracking study, Hessels et al. (2019) investigated gaze behavior of participants engaged in a face-to-face interaction with a confederate. They observed that when participants listened to a confederate's story, their gaze was directed at the facial features (e.g., eyes, nose, and mouth regions) for a longer total duration, as well as more often per second, compared to when speaking themselves. However, they did not find that participants looked proportionally longer at the mouth while listening compared to speaking, as in Rogers et al. (2018). One reason for this difference could be that participants in the Hessels et al. (2019) study did not need to exchange speaking

turns as they were specifically tasked to wait for the confederate to end his story. The small differences in these two studies may then be explained if turn-transitions are associated with looking at the mouth.

In sum, it has been well established that gaze to faces during conversations is dependent on speaker-state: who is speaking or who is being addressed. Based on eye-tracking studies with videos of talking faces, it has often been suggested that gaze will be directed more at the mouth while listening to someone speak, as looking at the mouth area may be beneficial (but not necessary) for speech perception (see e.g., Buchan et al., 2007; Vatikiotis-Bateson et al., 1998; Võ et al., 2012). Recent dual eye-tracking studies on the role of gaze behavior in two-person conversations (Hessels et al., 2019; S. L. Rogers et al., 2018) have found no, or only small, differences in gaze to specific facial features (e.g., eyes, mouth) during episodes of speaking and listening. We expect to observe a similar pattern for parents and children as well.

PRESENT STUDY

In this study, we investigated speech and gaze behavior during conversational interactions between a parent and their child. Parent-child dyads engaged in two conversations about potential disagreements (conflict) and agreements (cooperation) on common family topics, given their importance and frequent occurrence within the social context of the parent-child relationship (Branje, 2018; Dixon, Graber, & Brooks-Gunn, 2008; Laursen & Collins, 2004; Steinberg, 2001). We investigated (1) the similarities and differences between parents and children's speech and gaze behavior during face-to-face interaction, (2) whether patterns of speech and gaze behavior in parent-child conversations are related to the nature of the conversation (conflictive versus cooperative topics), and (3) whether gaze behavior to faces is related to whether someone is speaking or listening. To engage parents and children in conflictive and cooperative conversations, we used two age-appropriate semi-structured conversation-scenarios. This method, which is considered a 'gold standard' in the field, has extensively been used by researchers to assess various aspects of the parent-child relationship, e.g., attachment, interpersonal affect, relational quality, parental style, and child compliance (Aspland & Gardner, 2003; Ehrlich, Richards, Lejuez, & Cassidy, 2016; Scott, Briskman, Woolgar, Humayun, & O'Connor, 2011). To investigate the relation between speech and gaze in parent-child conversations, we needed a setup capable of concurrently recording eye movements and audio from two conversational partners with enough spatial accuracy to distinguish gaze to regions

of the face. To this end, we used a video-based dual eye-tracking setup by Hessels et al. (2017) that fulfills these criteria. Based on previous literature, we expected that parents and children on average looked predominantly at each other's faces, but that participants would exhibit substantial individual differences in what region of the face they looked at most (eyes, nose, mouth). Moreover, we expected that gaze behavior was related to whether subjects were speaking or listening. We may expect that when listening gaze is directed more at the mouth region, given its potential benefits for speech perception and turn-taking. Regarding the conflict and cooperative scenarios, we had no prior expectations.

METHOD

Participants

81 parent-child dyads (total $n=162$) participated in this study. All participants were also part of the YOUth study, a prospective cohort study about social and cognitive development with two entry points: *Baby & Child* and *Child & Adolescent* (Onland-Moret et al., 2020). The YOUth study recruits participants who live in Utrecht and its neighboring communities. The YOUth study commenced in 2015 and is still ongoing. For this study, a subset of participants from the first wave of the Child & Adolescent cohort (age range: 8–10 years) were recruited. A complete overview with descriptive statistics of the participants is given in the results section (Table 1). All participants received an information brochure at home in which this study was explained. Participants could then decide whether they wanted to participate in this additional study aside from the general testing program. All participants were included at their first visit to the lab and parents provided written informed consent for themselves as well as on behalf of their children. This study was approved by the Medical Research Ethics Committee of the University Medical Center Utrecht and is registered under protocol number 19-051/M.

Apparatus

A dual eye-tracking setup (see Figure 1a) was used to record gaze of two interactors simultaneously. Each person was displayed to the other by means of a monitor and a half-silvered mirror (see Figure 1b). The cameras behind the half-silvered mirrors were *Logitech* webcams (recording at 30 Hz at a resolution of 800 by 600 pixels). The live video-feeds were presented at a resolution of 1024 by 768 pixels in the center of a 1680 by 1050 pixels computer screen and concurrently recorded to disk. Two *SMI RED* eye trackers

running at 120Hz recorded participants' eye movements (see Figure 1b). A stimulus computer running *Ubuntu 12.04 LTS* handled the live video-connection and signaled to the eye-tracker computers to start and stop recording eye movements (for a more detailed explanation of this setup, see Hessels et al., 2017). See <http://www.royhessels.nl/research.html> for an example movie of the frontal videos and eye-movement data.

a

b

Figure 9. Overview of the dual eye-tracking setup. **1a)** Staged photographs of two interactors in the dual eye-tracking setup. **1b)** A schematic overview of the setup, reproduced from Hessels et al. (2018).

Audio was recorded using a set of *AKG C417-PP* Lavalier-microphones which were connected to a *Behringer Xenyx 1204-USB* audio panel. Each microphone was attached to the front of each setup (see the dashed orange circles on the left panels in Figure 1a). We

used Audacity v. 2.3.3 running on a separate computer (Ubuntu 18.04.2 LTS) to record audio. In a stereo recording the signal of the parent was panned to the left channel and the signal of the child was panned to the right channel. Upon recording start, a 100ms pulse was sent from the parallel port of the stimulus computer to the audio panel to be recorded. This resulted in a two-peak signal we used to synchronize the audio recordings to the beginning and end of the video and eye-tracking recordings. Audio recordings were saved to disk as 44100 Hz 32-bit stereo WAVE files. We describe in detail how the data were processed for the final analyses in the *Signal processing* section below.

Procedure

Upon entering the laboratory, a general instruction was read out by the experimenter (author GAH). This instruction consisted of a brief explanation of the two conversation-scenarios and the general experimental procedure. Participants were asked not to touch any equipment during the experiment (e.g., the screen, microphones, eye trackers). Because the experimenter needed to start and stop the video-feed after approximately five minutes for each conversation, he explained that he would remain present during the measurements to operate the computers. After the general instruction, participants were positioned in the dual eye-tracking setup. Participants were seated in front of one of the metal boxes at either end of the setup (containing the screens and eye trackers) such that their eyes were at the same height as the webcams behind the half-silvered mirrors using height-adjustable chairs. The distance of participants' eyes to the eye tracker was approximately 70 cm and the distance from eyes to the screen was approximately 81 cm. After positioning, the experimenter briefly explained the calibration procedure. The eye tracker of the parent was calibrated first using a 5-point calibration sequence followed by a 4-point calibration validation. We aimed for a systematic error (validation-accuracy) below 1° in both the horizontal and vertical direction (they are returned separately by *iViewX*). However, if for some reason, a sufficiently low systematic error could not be obtained, the experimenter continued anyway (see Section *Eye-tracking data quality and exclusion* how these recordings were handled). After calibrating the parent's eye tracker, we continued with the child's eye tracker. After the calibration procedure, the experimenter briefly repeated the task-instructions and explained that he would initiate the video-feed after a countdown. The experimenter repeated that he would stop recording the conversation after approximately five minutes. The experimenter did not speak or intervene during the conversation, only if participants questioned him directly, or when participants had changed their position too much. In the latter case, this was readily

visible from the *iViewX* software which graphically and numerically displays the online gaze position signals of the eye trackers. If participants slouched too much, the incoming gaze position signals would disappear or show abnormal values. In such instances, the experimenter would ask the participants to sit more upright until the gaze position signals were being recorded properly again.

Conflict-scenario

For the first conversation, children and their parents were instructed to discuss a family issue about which they have had a recent disagreement. The goal of conflict-scenario was to discuss the topic of disagreement and, if possible, come to possible solutions. To assist the participants in finding a suitable topic, the experimenter provided a list with common topics of disagreements between parents and children. The list included topics such as screen time, bedtime, homework, and household chores (see Appendix 1 for a complete overview). The main criteria for the conflict-scenario were that the topic should be about a recent disagreement, preferably in the last month. If no suitable topic could be found on the list or could be agreed upon, the parent and child were asked to come up with a topic of their own. Some parent-child dyads could not decide at all or requested to skip the conflict-task altogether. Note that for the final analyses, we only included dyads that completed both scenarios (see Figure 1 and Table 1 below). After participants had agreed on a topic, the experimenter explained that they should try to talk solely about the chosen topic for approximately 5 minutes and not digress.

Cooperation-scenario

For the second conversation, participants were instructed to plan a party together (e.g., birthday or family gathering). The goal of the cooperation-scenario was to encourage a pleasant and cooperative interaction between the parent and child. Participants were instructed to discuss for what occasion they want to organize a party and what kinds of activities they want to do. Importantly, participants had to negotiate the details and thus needed to collaborate to come up with a suitable party plan. Participants were instructed to discuss the party plan for approximately 5 minutes. Prior to the second conversation, the experimenter checked the participants' positioning in front of the eye trackers and whether the eye trackers and microphones were still recording properly. In some cases, the experimenter re-calibrated the eye trackers if participants had changed position, or if the eye tracker did not work for whatever reason. Note that we always started with

the conflict-scenario and ended with the cooperation-scenario because we reasoned this would be more pleasant for the children.

Debriefing

After the cooperation-scenario, the experimenter thanked the parents and children for their participation and the child received a small gift. The experimenter also asked how the participants had experienced the experiment, and if they were left with any questions about the goal of the experiment.

Signal processing

To prepare the eye-tracking, audio, and video signals for the main analyses, we conducted several signal processing steps (e.g., synchronization, classification). In the following sections, we describe these separate steps. Readers who do not wish to consider all the technical and methodological details of the present study may wish to proceed to the Results, Section *Main analyses*.

1. Synchronization of eye-tracking signals and video recordings

By using timestamps produced by the stimulus computer, the eye-tracking signal was automatically trimmed to the start and end of the experimental trial. Next, the eye-tracking signal was downsampled from 120 Hz to 30 Hz to correspond to the frame rate of the video. In the downsampling procedure, we averaged the position signals of four samples to produce a new sample. This caused the signal-to-noise ratio to increase by a factor of 2 ($\sqrt{4}$) due to the square root law.

2. Construction of Areas of Interest (AOI) and AOI assignment of gaze position

To determine where and when participants looked at each other's faces, we mapped gaze coordinates unto the frontal video-recordings. Because participants moved and rotated their faces and bodies during the conversations, we used an AOI construction method that can efficiently and effectively deal with an enormous number of images, namely the thousands of video frames produced in this experiment. This method consists of the fully automated Limited Radius Voronoi Tessellation procedure to construct Areas-of-Interest (AOIs) for facial features in dynamic videos (Hessels, Benjamins, et al., 2018; Hessels

et al., 2016). Briefly, this procedure assigns each gaze position to one of the four facial features (left eye, right eye, nose, or mouth) based on the closest distance to the facial feature. If this minimal distance exceeds the limited radius, gaze position was assigned to the background AOI (see Figure 2). This background area consists of the background and small parts of the upper body of the participant visible in the video. In our study, the LRVt-radius was set to 4° (200 pixels). The LRVt method is partly data-driven, resulting in smaller AOIs on the children's faces compared to AOIs on the parents' faces. We quantified the AOI size by the AOI span. The AOI span is defined as the mean distance from each AOI cell center to the cell center of its closest neighbor (see Hessels et al., 2016, p. 1701). The average AOI-span for parents' faces was 1.76° and an average AOI span for children's faces was 1.6° .

Figure 2. An example for computer-generated AOIs (Hessels, Benjamins, et al., 2018) for the left eye (L), right eye (R), nose (N), and mouth (M). The AOI for the left eye, for example, is the area closest to the left eye center but not further away from the center than the bounded radius of 4° (denoted with a red arrow). The background AOI (B) encompasses the background, the upper body of the participant and a small part of the top of the head.

3. From gaze data to dwells

After individual gaze samples were mapped unto AOIs (see previous section), we computed dwells (for a definition, see Holmqvist et al., 2011, p. 389). We operationalized a dwell as

the period from AOI entrance to exit. We rejected dwell candidates shorter than 120ms (or less than four consecutive video frames). For further details, see Hessels, Holleman, et al. (2018, p. 7).

4. From raw audio recordings to speaker categories

a) Trimming to prepare audio for synchronization with video and eye-tracking. In a self-written audio visualization script in MATLAB, we manually marked the timepoints of the characteristic two-peak synchronization pulse sent by the stimulus computer (see Section *Apparatus*), which indicated the start and stop of the two conversations at high temporal resolution (timing accuracy <1 ms). Then we trimmed the audio files based on these start-and-stop timepoints. Next, the trimmed stereo files (left channel – the parent, right channel – the child) were split into two mono signals, and the first conversation (conflict) and the second conversation (cooperation) were separated. Finally, the audio signal was downsampled to 1000 Hz and converted into an absolute signal. As a result, we produced four audio files per parent-child dyad.

b) Determination of speech samples. Speech episodes (as an estimator for who was speaking) were operationalized as follows. First, the absolute audio signal was smoothed with a Savitsky-Golay filter (order 4; window 500 ms). Then, samples were labelled silent when the amplitude was smaller than 1.2 times the median amplitude of the whole filtered signal. Then, we computed the standard deviation of the amplitude of the silent samples. Subsequently, a sample was labelled as a speech sample if the amplitude exceeded the mean plus 2 times the standard deviation of the amplitude of the silent samples.

c) Removing crosstalk. Because the microphones were in the same room, we observed crosstalk in the audio channels. That is, we sometimes heard that parent speech was present in the audio recording channel of the child and vice versa. The child's channel suffered more from crosstalk than the parent's channel. To deal with this, we first equalized the speech signals of the parent and child by making them on average equally loud (by using the average speech amplitudes of the single episodes) Then, we identified episodes of potential crosstalk by selecting the episodes that contained a signal in the channel of both speakers. We removed crosstalk with the following rule: We assigned a crosstalk episode to X if the amplitude of the signal in the X's channel was 3.33 times larger than in the Y's channel. The value 3.33 was derived empirically. If a crosstalk episode was assigned to the child, it was removed from the parent's channel and vice versa.

d) Determination of speech episodes. From the labelled samples, we determined speech episodes. Each speech episode is characterized by an onset and offset time, mean amplitude and duration.

e) Removing short speech and short silence episodes. We removed speech episodes shorter than 400 milliseconds followed by the removal of silence episodes shorter than 100 milliseconds.

f) Assigning speech labels to single samples. To link the speech signal to the eye-tracking signal in a later stage, we assigned speech labels to each sample of the speech signal of the parent-child dyad. For each timestamp, we produced a label for one of the following categories: parent speech, child speech, speech overlap and silence (no one speaks).

5

5. Combining speech and gaze behavior

The speech signal was combined with the gaze signal as follows. First, we upsampled the gaze signal (with AOI labels) from 30 Hz to 1000 Hz by interpolation to match the sampling frequency of the audio signal. Each sample in the combined signal contained a classification of speaker categories (child speaks, parent speaks, both are speaking, no one speaks) and gaze location on the face for both child and parent (eyes, nose, mouth, background). Not all recordings produced valid eye-tracking data (eye-tracking data loss), valid AOIs for all video frames (e.g., due to extreme head rotations construction of AOIs is impossible), or valid dwells (i.e., dwells longer than 120 milliseconds). These invalid cases were marked in our combined audio/gaze database. This is not necessarily problematic for data analysis, because we also conducted analyses on parts, or combinations of parts, of the data (e.g., speech analysis only).

6. Measures of speech and gaze

In this study, we mainly report relative measures of speech and gaze behavior because the total recording durations differed across dyads and conversations. We computed relative total speech durations as a descriptor of speech behavior and relative total dwell times as a descriptor of gaze behavior. To obtain relative measures, we determined the total duration for each speaker category (i.e., parent speech, child speech, overlap, no one speaks) and the total duration of dwells (with AOI labels eyes, nose, mouth, background) and then divided these durations by the total duration of the recording.

RESULTS

Signal processing

We first assessed the quality of the eye-tracking data, which is crucial for the validity of an eye-tracking study's results (Holmqvist et al., 2012). High-quality eye-tracking data is typically obtained when subjects are restrained with a chinrest/headrest to maintain equal viewing distance and minimize head movements. However, in the context of our face-to-face conversations, subjects could talk, gesture, move their face, head, and upper body. Although the dual eye-tracking setup used was specifically designed to allow for these behaviors, other eye-tracking studies have also demonstrated that such behaviors may negatively affect eye-tracking data quality (Hessels et al., 2015; Holleman, Hessels, Kemner, & Hooge, 2019; Niehorster et al., 2018). Moreover, young children may pose additional problems, such as excessive movement or noncompliance (Hessels & Hooge, 2019). We computed several commonly used eye-tracking data quality estimates, namely: *accuracy* (or systematic error), *precision* (or variable error), and *data loss* (or missing data).

First, we assessed accuracy. The average validation accuracy of parents' recordings was 0.98°. For the children's recordings it was 1.48°. We set an exclusion criterion of 1° for the 2d-validation accuracy. Second, we determined precision by computing the sample-to-sample root mean square displacement (s2s-RMS) of the gaze-position signal. We then divided the s2s-RMS values for every participant by the AOI span (see Section *Signal Processing*). This measure can range from 0 to infinity. A precision/AOI-span value of 1 means that precision is equal to the AOI span. In other words, a value of 1 means that the sample-to-sample variation of the gaze-position signal is equal to the average distance between AOIs. If the precision/AOI-span is larger than 1 this means that one cannot reliably map gaze position to an AOI. Therefore, we decided to exclude measurements in which the average precision/AOI span exceeded 1. Also, this measure accounts for differences between the recordings in the magnitude of the AOI spans in relation to the recorded gaze position. We also calculated periods of data loss – i.e., when the eye tracker did not report gaze position coordinates. Data loss is a slightly more complicated measure in the context of our study, given that data loss may coincide with talking and movement (Holleman et al., 2019). For example, it is well-known that some people gaze away more when speaking compared to when listening (Hessels et al., 2019; Kendon, 1967). Therefore, any exclusion based on data loss may selectively remove participants that spoke relatively more. For that reason, we did not exclude participants based on data loss but conducted separate sensitivity analyses for all our main findings as a function of a data loss exclusion criterion (see Appendix 3).

Figure 3. Flowchart of eye-tracking data quality assessment and exclusion criteria.

Table 2. Descriptive statistics of participants

	All participants		Audio only		Single parent ET + audio set	Single child ET + audio set
Participant	Parents	Children	Parents	Children	Parents	Children
Sample size	81	81	73	73	40	13
Age	$M = 42.11$ $SD = 5.01$	$M = 9.34$ $SD = 0.77$	$M = 42.30$ $SD = 5.18$	$M = 9.38$ $SD = 0.76$	$M = 42.50$ $SD = 5.33$	$M = 9.37$ $SD = 0.80$
Gender	♀ = 63 ♂ = 18	♀ = 55 ♂ = 26	♀ = 58 ♂ = 15	♀ = 48 ♂ = 25	♀ = 29 ♂ = 11	♀ = 9 ♂ = 4
Pair configuration	♀ ♀ = 43 ♀ ♂ = 13 ♂ ♀ = 20 ♂ ♂ = 5		♀ ♀ = 37 ♀ ♂ = 13 ♂ ♀ = 20 ♂ ♂ = 4		♀ ♀ = 20 ♀ ♂ = 8 ♂ ♀ = 8 ♂ ♂ = 4	♀ ♀ = 7 ♀ ♂ = 2 ♂ ♀ = 4 ♂ ♂ = 0

*♀♂ = child female + parent male ♂♀ = child male + parent female.

Based on the criteria for accuracy and precision, we determined how many measurements were suitable for further analyses. Figure 2 and Table 1 depicts an overview of the eye-tracking data quality assessment and how many participants were excluded for further analyses based on exclusion criteria described above. Out of 81 parent-child dyads who participated, 73 dyads completed the experiment (i.e., participated in both conversation-scenarios). Out of this set, we had eye-tracking data of sufficient quality for 40 parents and 13 children. Descriptive statistics of the participants are given in Table 1. Note that although the quality of the eye-tracking data is known to be worse for children, it was particularly problematic in our study as we needed data of sufficient quality for both conversations to answer our research questions. Although many more participants had at least one good measurement, applying our data quality criteria to both conversations for every parent-child dyad resulted in these substantial exclusion rates.

Main analyses

We present three main analyses in which we address (1) the similarities and differences between parents and children's speech and gaze behavior during face-to-face interaction, (2) whether patterns of speech and gaze behavior in parent-child conversations are related to the topics of conversation (conflictive versus cooperative), and (3) whether gaze behavior to faces is related to whether someone is speaking or listening. For all our figures and statistical descriptions, we used detailed visualizations and bootstrapping techniques

provided by Rousselet et al. (2017). Specifically, we used the Harrell-Davis estimator to compute 95% confidence intervals around the medians of each distribution with the MATLAB function *decilespbci*. The number of bootstrap samples was set to 2,000.

Speech behavior

In this section, we report parents' and children's speech behavior over the course of the two conversation-scenarios: conflict and cooperation. We wanted to know whether parents and children differed in how much they spoke when discussing potential (dis)agreements. For the analyses of speech behavior, we used the Audio-only dataset (see Table 1), which consisted of 73 parent-child dyads who completed both conversations. The average duration of the conflict-scenario was 281.25 seconds ($sd = 27.78$ seconds) and the average duration of the cooperation-scenario was 297.37 seconds ($sd = 28.69$ seconds).

Similarities and differences between parents and children

Figure 4 depicts parents' and children's relative total speech durations (i.e., how much they spoke as a percentage of the total conversation) across the two conversation-scenarios. We estimated relative speech durations of four speaker-categories: child speaks, parent speaks, both speak ('overlap'), and no one speaks ('none'). As is visible from Figure 4 (left panels) both parents and children varied substantially in how much they spoke in total over the course of the two conversations. For example, the range of relative speaking durations ranged from less than 5% for some individuals to nearly 50% of the total conversation duration. Overall, parents spoke for a longer total time compared to children, regardless of the conversation scenario. The median relative speaking duration for parents was 39.19%, 95% CI [37.69% – 40.53%] in the conflict-scenario and 34.64%, 95% CI [32.11% – 37.14%] in the cooperation-scenario, whereas the median relative speaking duration for children in the conflict-scenario was 18.51%, 95% CI [16.65% – 20.34%] and 23.59%, 95% CI [21.91% – 26.23%] in the cooperation-scenario.

Speech behavior as a function of conversation scenario

To compare how individual parents' and children's total speech durations differed between the two conversation-scenarios, we computed a difference-score for each participant by subtracting their relative total speech duration in the cooperation-scenario from their relative total speech duration in the conflict-scenario, see Figure 4 (right panel). A negative difference-score (i.e., value on the left side of the zero-difference line) means that the

participant spoke less in the conflict-scenario compared than in the cooperation-scenario and a positive difference-score (i.e., value on the right side of the zero-difference line) means that the participant spoke more in the conflict-scenario compared to how much they spoke in the cooperation-scenario. As is visible from Figure 4 (right panel), parents spoke more during the conflict-scenario than during the cooperation-scenario, as indicated by a positive median difference in relative total speaking duration of 2.91 percentage points (pp), 95% CI [1.42 pp – 4.27 pp]. Conversely, children spoke more in the cooperation-scenario compared to the conflict-scenario, as indicated by a negative median difference in relative total speaking duration of -5.32 pp, 95% CI [-7.15 pp – -3.60 pp]. Finally, there was slightly more silence (i.e., neither parent or child was speaking) during the conflict-conversation compared with the cooperation-scenario, as shown by the positive median difference-score of 2.50 pp, 95% CI [0.86 pp – 4.10 pp] for the ‘none’ speaker-category.

Figure 4. *Left panel.* Speech behavior of 73 parent-child dyads for the two conversation-scenarios (conflict and cooperation). Dark grey markers represent the relative total speech duration (as a percentage) of each participant during the conflict-scenario and light grey markers represent the relative total speech duration of each participant during the cooperation-scenario. The vertical orange stripes represent the median relative speech durations per speaker-category. *Right panel.* Difference-scores of speech behavior (conflict minus cooperation). Difference-scores were computed by subtracting participants' relative speech durations in the cooperation-scenario from the conflict-scenario for every speaker-category. Light grey markers represent individual difference scores (as percentage point difference in relative total duration). The orange markers represent the median difference-score of the relative speech durations and the error bars (barely visible) represent 95% confidence intervals of the median, both of which were obtained through bootstrapping using the MATLAB-function *decilespcbi* provided by Rousselet et al. (2017). The vertical dashed line represents a zero-difference line. Negative difference scores indicate that the participant spoke less in the conflict-scenario than in the cooperation-scenario (and vice versa for positive difference scores).

Gaze behavior

In this section, we report parents' and children's gaze behavior to facial features in relation to the conversation-scenarios. We analyzed whether and how parents and children differed in where they looked at the other's faces when discussing (dis)agreements. Based on previous eye-tracking studies on face scanning behavior (e.g., Mehoudar et al., 2014; Peterson et al., 2016; S. L. Rogers et al., 2018), we expected that parents and children will look predominantly at each other's faces, but that participants may exhibit large individual differences in where at the face is looked at most (eyes, nose, or mouth). We had no expectations regarding gaze behavior as a function of conversation-scenario. For the analyses of gaze behavior, eye-tracking data of 40 parents and 13 children were used (see Table 1).

Similarities and differences between parents and children

First, we analyzed the individual differences in parents' and children's gaze behavior to facial features. Figure 5 (upper panels) depicts relative total dwell times to face AOIs (e.g., eyes, nose, mouth) and the background AOI as a function of conversation-scenario (conflict and cooperation). As is visible from Figure 5, both parents (top left panel) and children (top right panel) varied greatly in the total duration that they looked at each other's faces, as indicated by the large range in relative total dwell times to different face AOIs (0 to approximately 50-75% of total looking time for the eyes and mouth AOI). This was regardless of the conversation scenario. Large individual differences in gaze behavior to facial features matches our expectation based on previous research. A smaller range in relative total dwell time to the background AOI was observed, ranging from 0 to approximately 15% of total available looking time. To investigate the consistency of participants' gaze behavior to facial features across the two conversations, we computed Spearman rank correlations for both parents' and children's relative total dwell time on face AOIs and the background AOI. For the parents, we found a high level of consistency in relative total dwell time across conversation scenarios for all AOIs (eyes AOI $\rho=0.73$, $p=3.72 \times 10^{-7}$; nose AOI $\rho=0.73$, $p=3.65 \times 10^{-7}$; mouth AOI $\rho=0.81$, $p=8.45 \times 10^{-9}$; background AOI $\rho=0.72$, $p=5.70 \times 10^{-7}$). For the children, we also found a high level of consistency in relative total dwell time across conversation scenarios for most AOIs (eyes AOI $\rho=0.80$, $p=0.001$; mouth AOI $\rho=0.78$, $p=0.002$; background AOI $\rho=0.78$, $p=0.002$), but slightly less so for the nose AOI ($\rho=0.43$, $p=0.140$). These Spearman correlations show that individuals were consistent in where they looked at on average on the other person's face across the two conversations. One difference that stands out from

Figure 5 is that parents had higher median relative dwell times for all face AOIs compared with those of the children. However, we cannot conclude that is a meaningful difference as data loss was generally much higher for the children than for the parents (see Appendix 2).

Gaze behavior as a function of conversation scenario

Next, we analyzed whether parents and children differed in what regions of each other face they looked at as a function of the conversation-scenario (conflict and cooperation). To investigate this, we compared the within-subject differences in gaze behavior to facial features of parents and children across the two scenarios. We computed the individual and median difference-scores in relative total dwell time per AOI, by subtracting participants' relative total dwell time to face AOIs in the cooperation-scenario from the conflict-scenario. As is visible from Figure 5, parents (lower left panel) gazed slightly more at the child's eyes AOI during the conflict-scenario, as indicated by a positive median difference-score of 4.15 percentage points (pp), 95% CI [1.28 pp – 7.78 pp] and a non-overlapping error bar with the zero-difference line. Also, it seemed that parents looked slightly more at the mouth AOI during the cooperation-scenario, as indicated by a negative median difference-score of -2.54 pp, 95% CI [-7.99 pp – 0.55 pp]. However, notice that the confidence interval overlaps slightly with the zero-difference line. For the children, no differences were observed between their relative total dwell times to face AOIs as a function of the conflict and cooperation scenario.

Gaze behavior to facial features and its relation to speech

In this section, we report parents' and children's gaze behavior to facial features in relation to speech behavior. We wanted to know where participants looked at the other's face during episodes of speaking and listening. Based on previous literature, we may expect that participants will gaze slightly more at the mouth when listening, and more at the eyes when speaking (S. L. Rogers et al., 2018). For these analyses, we computed relative total dwell times to AOIs as a function of speaker-state, by summing all dwells per AOI (eyes, nose, mouth, background) for all the classified episodes of self-speech (participant is speaking) and other-speech (other person is speaking) per conversation scenario (conflict, cooperation). Note that this excludes episodes of overlap or silence. Then we computed a relative measure of total dwell time per AOI category during self-speech and other-speech by dividing the total duration of dwells by the total duration of the conversation. Finally, we averaged relative total dwell times across the two scenarios.

Figure 6. Gaze behavior as a function of speaker-state. Distributions of gaze behavior to the facial feature and background AOIs for parents (left panels) and children (right panels). *Upper panels.* Relative total dwell times (as a percentage) as a function of self-speech and other-speech. Dark grey markers represent the relative total dwell time to face AOIs (eyes, nose, mouth) and the background AOI of each participant when speaking and light grey markers represent the relative total dwell time to facial features of one participant when the other person was speaking). The vertical orange stripes represent the median relative total dwell time per AOI. *Lower panels.* Distributions of difference-scores of relative total dwell time to the AOIs (self-speech minus other-speech). Difference-scores were computed by subtracting participants' relative total dwell time to face AOIs during other-speech from their relative total dwell time during speaking. Light grey markers represent individual difference scores. The orange markers represent the median difference-score of the relative total dwell times and the error bars (barely visible) represent 95% confidence intervals of the median, both of which were obtained through bootstrapping using the MATLAB-function *decilespcbi* provided by Rousselet et al. (2017). The vertical dashed line represents a zero-difference line. Positive difference scores indicate that the participant gazed more at particular AOI during episodes of self-speech. Negative difference scores indicate that the participant gazed more at a particular AOI during episodes of other-speech.

Figure 6 (upper panels) depicts participants' relative total dwell times to the AOIs as a function of self-speech and other-speech. As is visible from Figure 6 (upper panels), individual parents and children varied substantially in where they looked at the different face AOIs (eyes, nose, mouth, background), regardless of speaker-state. For most parents and children, relative total dwell times to face AOIs ranged somewhere between 0 and 50%, regardless of speaker-state, and some individuals gazed at a particular face AOIs for more than 50-75% of the time. Furthermore, total dwell times to the background-AOI were substantially lower on average, ranging between 0 and 18% of the total duration of available looking time during self-speech and other-speech. Again, parents had higher median relative dwell times for all face AOIs, but as stated previously, we cannot conclude that is a meaningful difference as data loss was generally higher for the children than for the parents (see Appendix 2).

Next, to compare how parents and children looked at the other's face during episodes of self-speech and other-speech we computed individual within-subject difference-scores by subtracting the relative total dwell times during other-speech from relative total dwell times during self-speech (see Figure 6, lower panels). Note that a negative difference-score indicates that the participant gazed more at a particular AOI during other-speech compared to self-speech, and a positive difference-score means the participant gazed more at a certain facial feature during episodes of self-speech. For both the single parent ($n=40$) and single child ($n=13$) data sets, a negative median difference-score was observed for gazing at the mouth AOI. We found a median difference-score in total dwell time on the mouth AOI for the parents of -7.64 percentage points (pp), 95% CI [-11.27 pp – -4.58 pp], and for the children the median difference-score of total dwell time on the mouth AOI was -6.28 pp, 95% CI [-14.12 pp – -2.39 pp]. Thus, overall, participants slightly gazed more at the mouth of the other when listening compared to when they were speaking themselves.

DISCUSSION

In this study, we investigated the role of gaze behavior to faces and its relation to speech of parents and their preadolescent children when engaged in conflictive and cooperative conversations about common family topics. We asked the following research questions: 1) What are the similarities and differences between parents and children's speech and gaze behavior during face-to-face interaction? 2) Are patterns of speech and gaze behavior in parent-child conversations related to the topics of conversation (conflictive versus cooperative)? 3) Is gaze behavior to faces during parent-child interaction related to who is speaking or listening?

Regarding our research questions on the similarities and differences in speech behavior between parents and children, and whether patterns of speech behavior were related to the conversation topic, we found clear differences in how much parents and children spoke across the two conversations. Overall, parents spoke more than children regardless of the conversation-scenario. Parents spoke more in the conflict-scenario compared with how much they spoke in the cooperation-scenario, while children spoke more in the cooperation-scenario compared to how much they spoke in the conflict-scenario. Finally, there was more silence during the conflict-scenario (neither the parent nor child was speaking).

The results on speech behavior clearly show that the two conversation-scenarios (conflict and cooperation) substantially influenced the dynamic of the interaction. This is likely due to how the two conversation-scenarios differ with regards to the role that the parent and the child take. When planning the party together (cooperation scenario), the parent and the child are more egalitarian partners, as they both recognize that their wishes and ideas carry the same weight. In the conflict situation, however, the topics that were chosen to discuss mostly concerned the child's behaviors that the parent considered undesirable (e.g., not cleaning one's own room, not listening to the parents, fighting with a brother or sister). In such situations, the parent tends to take a lead, assert their authority, and consequently speak more (Moed et al., 2015). The child, on the other hand, recognizes the parent's authority in such matters and mostly such conflicts are resolved by the child giving in to parent's demands (see Laursen & Collins, 2009 for a review). This submissive role is reflected in children speaking less in the conflict discussion. Moreover, the conflict discussion likely elicits more tension and uncomfortable feelings in children than in the cooperation scenario (Thomas et al., 2017), which might be an additional explanation for their smaller contribution to the conflict discussion with parents. As the nature of parent-child conflicts changes significantly across various stages of development (Dunn & Slomkowski, 1992; Laursen & Collins, 2009; Steinberg, 2001), one may expect that patterns gaze and speech behavior are different for parent-child conflicts in early or late childhood, or in early adolescence and late adolescence. Our participant sample consisted of preadolescent children (8-10 years) which somewhat precedes the relational changes in parental authority and child autonomy in adolescence. As such, we would expect that the contribution of children in the conversation would increase as they move through adolescence and the parent-child relationship becomes more egalitarian.

Regarding our research question on the similarities and differences in gaze behavior to faces of parents and children, we found substantial individual differences for both parents

and children in what region of the other's face was looked at most. In line with previous research with adults (Arizpe et al., 2017; Peterson et al., 2016; S. L. Rogers et al., 2018), we found that some parents and children looked most at the eyes while others looked at different regions of the face more equally (eyes, nose, and mouth). This was the case regardless of the conversation-scenario. While parents seemed to look more at the faces of their children than vice versa, this could be due to differences in data loss between parents and children (see Appendix 2).

Furthermore, we investigated whether gaze behavior to faces was related to the topic of the conversation. We did not find any differences in where children gazed at the parent's faces as a function of conversation topic. Interestingly, we did find that on average parents gazed more at the child's eyes during the conflict-scenario than in the cooperation-scenario. One reason may be that increased eye gaze asserts dominance and social status (Kleinke, 1986; Patterson, 1982). In this sense, increased gaze at the other person's eyes may serve as a nonverbal emphasis on a particular verbal message to persuade another person, or to press a particular response from that person (Timney & London, 1973). Thus, increased eye gaze from parents may signal authority while negotiating a family disagreement and persuade the child towards some goal or solution, at least in our conflict-scenario. However, increased gaze to the eyes has also been associated with the expression of affiliation and intimacy (Kleinke, 1986; Patterson, 1982). Thus, it could also be that parents looked more at the child's eyes in the conflict-scenario because they wanted to express more intimacy while negotiating a potentially conflicting topic of discussion. Our findings do not distinguish between these two potential explanations. If parents' increased gaze to the eyes of the child during the conflict conversation would indeed be the result of social control exercised by the parent, one may expect that the difference scores in gaze to eyes between the two conversations are correlated with some index of parental authority. If, on the other hand, increased gaze to the eyes would be an expression of intimacy, one may expect that the difference scores correlate with some measure of intimacy or interpersonal reactivity. Moreover, it would be interesting to investigate whether parents purposively exert authority or intimacy by means of increased gaze to the eyes when discussing a conflicting topic with their children.

Finally, we also investigated whether gaze to facial features in parent-child interactions is dependent on speaker-state, because previous eye-tracking studies suggest that what region of the face people look at during conversations depends on whether they are speaking or listening (S. L. Rogers et al., 2018). Both parents and children looked more at the

mouth region while listening than while speaking. This is in accordance with several non-interactive eye-tracking studies, which have shown that observers presented with videos of talking faces tend to gaze more at the mouth area, especially under noisy conditions or when tasked to report what is being said (Buchan et al., 2007; Vatikiotis-Bateson et al., 1998; Võ et al., 2012). Similar to S. L. Rogers et al. (2018)'s interactive eye tracking study, we found that both parents and children gazed slightly more at the mouth when the other person spoke compared to when they were speaking themselves, although these differences were small on average (i.e., approximately 5-10% of total looking time), but see also Hessels et al. (2019). In our study, we found that differences in mouth-looking between speaking and listening were always (or almost always) in the same direction, although the magnitude of this difference varied between individuals. Increased gaze at the mouth when listening to speech may be explained by the fact that people look at the visual cues from mouth and lip movements to support speech perception (Sumby & Pollack, 1954; Vatikiotis-Bateson et al., 1998).

Our study is one of the first to use eye-tracking to study gaze behavior to faces and its relation to speech in parent-child interactions. Previous research has often been conducted using observational techniques (i.e., which lack in reliability and validity to distinguish gaze to specific facial features), or has been limited to non-interactive eye-tracking procedures with photographs and videos of faces (Risko et al., 2016). As such, the main contribution of our study is empirical, by describing patterns of gaze and speech behavior during different types of conversations (conflict or cooperation). Moreover, the specific interpersonal context (parent-child interaction) is relevant for child development and has not been studied in this manner before. We focused on how aggregate speech and gaze behavior (e.g., total speech durations and dwell times) differed between parents and children and as a function of conversation topic and speaker-state. In future research, it will be useful to investigate the moment-to-moment characteristics of speech and gaze during parent-child interactions, for example, by looking at transitions between face AOIs as a function of speaker-state and conversation-scenarios. Also, it may be interesting to investigate to what extent patterns of verbal and nonverbal behavior are indicative of parents' conversational style and conflict resolution strategy (Beaumont & Wagner, 2004; Moed et al., 2015), and how this is related to children's adjustment, emotional reactivity, and social competence (Junge, Valkenburg, Deković, & Branje, 2020; Moed, Dix, Anderson, & Greene, 2017). For example, functional conflict resolution is typically characterized by validation, support, listening, and expressing positive or neutral affect, whereas dysfunctional conflict styles consist of negative affect, criticism, and hostility

(Laursen & Hafen, 2010; Moed et al., 2017). We showed that parents look longer at the eyes during conflict-scenario, but to discern whether this could also be exemplary of either functional or dysfunctional conflict resolution strategies would require in depth sequential analyses of speech content, voice affect, gaze, and facial expressions. Although this may seem like a daunting task, it may be a worthwhile approach to provide new and crucial insights into distinguishing different types of conflict styles in parent-child interactions.

One limitation of this study is the fact that we had to exclude many participants based on data quality criteria (see Figure 3). Given the nature of our study, we wanted participants to be relatively unrestrained during face-to-face conversations. However, this came at the cost of lower data quality, especially for the children. Perhaps if participants would have had more time to practice and get experienced with eye-tracking this may have improved data quality. However, this would consequently require even more time-investment by both parents, children, and researchers. Therefore, we suggest that researchers who aim to conduct eye-tracking studies with children should be aware of these additional difficulties associated with using dual eye-tracking technology.

APPENDIX 1: CONVERSATION TOPICS

Transportation, using bicycle or car
Money or valuable things
Not being at home, or being too much at home
Table manners, being rude
Bed time
Screen time, TV, computer, phone
Bullying
Behavior in class or at school
Privacy
School performance
Different opinions
Future plans
Gossip, secrets
Being annoying
Associating with others
Being on time, cancelling appointments
Leisure time
(Not) doing something when asked
Cleaning room
Appearance, clothing
School grades
Being honest
Homework
Household chores
(Not) sharing problems
Other...

APPENDIX 2: THREE TYPES OF DATA LOSS

Figure 7. Three types of data loss: no ET data, no AOI data, and short dwells. Left panels depict data loss measures for the parents and right panels for the children. Relative total duration (as a percentage) without eye-tracking data (i.e., no gaze position recorded), without AOI data, or with dwells shorter than 120 ms (i.e., invalid dwells) are depicted on the x-axis. Upper panels depict data loss measures as a function of conversation-scenario (conflict and cooperation). Lower panels depict data loss measures as a function of speaker-state (self-speech and other-speech). Each marker represents data from one participant. The vertical orange stripes represent the median relative total dwell time per data loss type.

Data loss is a problem in eye-tracking research and in our study it was expected due to relatively unrestrained behavior of the participants. Here we are interested in what were the most important sources of data loss, and how they were related to our main analyses. We identified at least three types of data loss in our study, namely: 1) samples with no eye-tracking data (i.e., no gaze position recorded); 2) samples with no AOI labels assigned to all video frames (i.e., no AOI data), and samples that did not count as a dwell (i.e.,

shorter than 120 ms). As is visible from Figure A1, we found that missing eye-tracking data (i.e., no gaze position recorded) was the most prominent source of data loss in our study. Moreover, we found that total amount of time without eye-tracking data was much higher for children than parents. This may explain why relative total dwell times to the facial and background AOIs were generally higher for parents than for children (see Results, Section: Gaze behavior).

APPENDIX 3: DATA LOSS EXCLUSION CRITERIA DO NOT AFFECT MAIN FINDINGS

As stated, we expected that eye-tracking data loss was especially likely to occur while speaking. Therefore, we did not exclude participants based on a general data loss criterion, as this may selectively exclude participants that spoke relatively more. Therefore, we needed to validate that not excluding participants based on a data loss criterion did not affect our main findings. In this section, we present two separate sensitivity analyses for two of our main findings.

Figure 8. Relative total dwell time difference to the eyes AOI between conflict and cooperation scenario. In this case, we validated whether the positive median percentage point difference in parents’ total dwell time to the eyes AOI was affected by data loss. Eye-tracking data loss was operationalized as the relative total duration during which no eye-tracking data was recorded. Participants were excluded if data loss exceeded the value on the x-axis for both conflict and cooperation conversations. Lower values of the data loss exclusion criterion represent a stricter exclusion policy (at 0% all participants are excluded and at 100% no participants are excluded). The solid black line represents the median percentage point difference, and the dashed black lines represent the 95% confidence intervals of the median, both of which were obtained through bootstrapping using the MATLAB-function *decilespcbi* provided by Rousselet et al. (2017).

The first sensitivity analysis pertains to our finding that parents looked more at the eyes AOI during the conflict-scenario. We validated whether this finding was not affected by exclusion of participants based on a data loss criterion. Figure A2 depicts an outcome

measure as a function of the data loss criterion. The data loss criterion used here was the percentage of time during listening without eye-tracking data, as this does not include data loss due to excessive looking away (as in speaking). On the y-axis the percentage point difference in total dwell time at the eyes AOI is depicted. On the x-axis a data loss criterion is depicted (i.e., as a % of missing ET data). As is visible from this figure, the positive median percentage point difference in looking at the eyes AOI between conflict and cooperation (see Results, Section Gaze behavior) was not affected by a data loss criterion, as its median and 95% confidence intervals were consistently above zero. In other words, the direction of the percentage point difference is not related to the data loss exclusion criterion. Thus, the finding that parents looked more at their children's eyes in the conflict conversation does not depend on whether we exclude participants based on data loss.

The second sensitivity analysis pertains to our finding that both parents and children looked more at the mouth AOI during episodes of other-speech. We validated whether this finding was not affected by exclusion based on a data loss criterion. Figure A3 depicts the percentage point difference in total dwell time to the mouth AOI between self-speech and other-speech episodes for the parents (left panel) and children (right panel) as a function of the same data loss exclusion criterion. On the y-axis the percentage point difference in total dwell time at the mouth AOI is depicted. On the x-axis a data loss criterion is depicted (i.e. as a percentage of missing ET data). As is visible from this figure, the negative median percentage point difference in looking at the mouth AOI between self-speech and other-speech (see Results, Section Gaze behavior) was not affected by a data loss criterion, as its median and 95% confidence intervals were consistently below zero. Thus, the direction of the percentage point difference is not related to the data loss exclusion criterion. Thus, the finding that participants looked more at the other's mouth during other-speech does not depend on whether we exclude participants based on data loss.

Figure 9. Relative total dwell time difference to the mouth AOI between self-speech and other-speech. In this case, we validated whether the negative median percentage point difference in parents' (left panel) and children's (right panel) relative total dwell time to the mouth was affected by data loss. Eye-tracking data loss was operationalized as the relative total duration during which no eye-tracking data was recorded during periods of other-speech. Participants were excluded if the eye-tracking data loss during periods of other-speech exceeded the value on the x-axis (eye-tracking data loss was averaged over the two conversations). Lower values of the data loss exclusion criterion represent a stricter exclusion policy (at 0% all participants are excluded and at 100% no participants are excluded). The solid black line represents the median percentage point difference, and the dashed black lines represent the 95% confidence intervals of the median, both of which were obtained through bootstrapping using the MATLAB-function *decilespci* provided by Rousselet et al. (2017).

Chapter 6

Gaze and turn-taking in parent-child conversations

Holleman, G. A., Hooge, I. T. C., Huijding, J., Deković, M., Kager, R., Kemner, C., & Hessels, R. S. (in preparation). Gaze and turn-taking in parent-child conversations.

Author contributions:

GH, IH, JH, MD, RK, CK, and RH designed the study. GH collected the data. GH, IH and RH analyzed the data. GH, IH, and RH interpreted the data. GH drafted the manuscript. GH, IH, JH, MD, RK, CK, and RH finalized the manuscript.

ABSTRACT

Previous studies have shown that gaze – where one looks and when – plays an important role in the regulation and coordination of turn-taking behavior in face-to-face interactions. Recently, researchers in the field of social attention have used modern eye-tracking technology to measure gaze during face-to-face conversations. We investigated the role of gaze behavior in turn-taking during parent-child conversations. Specifically, we focused on what regions of the face are looked at during turn transitions, and whether gaze to facial features was dependent on turn transition direction and topic of conversation. We used a dual eye-tracking setup to measure gaze in interaction. Eight parent-child dyads engaged in two brief conversations about two topics: a ‘conflict’ topic and a ‘cooperation’ topic. We did not find that turn transition direction, pre- and post-turn transition, and topic of conversation modulated where parents and children looked at each other’s faces during turn transitions. Data loss was especially large during turn transitions, especially for the children, which makes it difficult to draw solid conclusions about what regions of the face are looked at during turn-taking. We highlight several solutions to overcome of limitations of eye-tracking technology in this area of research.

INTRODUCTION

Human conversations are typically structured into alternating turns between speaker and listener, a phenomenon referred to as turn-taking. Researchers have shown that, in general, human face-to-face conversations can be characterized by a set of social norms and behaviors, e.g., avoiding simultaneous speech, monitoring and responding to the other, and using various ‘turn-allocation techniques’ (for an overview, see Duncan & Fiske, 2015; Kendon, 1967; Sacks, Schegloff, & Jefferson, 1978). To engage in smooth conversational interactions, researchers have posited that interlocutors need to attend to what the other is saying and doing, anticipate when a turn approaches completion, signal turn-yielding cues to the other (e.g., prosody, gaze), and prepare and articulate responses at the appropriate moment (Corps, Gambi, & Pickering, 2018; Duncan, 1972; Holler & Levinson, 2019). Researchers have discovered that the human turn-taking system is highly efficient: turn transitions between interactants occur with average inter-turn intervals between 0 and 200 ms (i.e., between-speaker gaps) and, on average, less than 5% of the total speech stream consists of overlap between two or more simultaneous speakers (Levinson, 2016). These aspects of turn-taking seem to be general across many different languages and cultures, although there are some exceptions (see Stivers et al., 2009 for a review).

How do humans achieve the rapid and generally smooth turn-taking behavior in face-to-face conversations? Researchers have extensively investigated the role of various ‘signals’ and ‘cues’ that allow interlocutors to indicate and predict when a turn transition is forthcoming, such as prosodic (e.g., intonation, pitch, duration, pauses) and syntactic (e.g., sentence- and word structure) information available in speech utterances (Gravano & Hirschberg, 2011). Also, nonverbal behavior, such as gaze, head movements, and gestures play an important role in the regulation and coordination of turn-taking behavior in face-to-face conversations (Duncan & Fiske, 2015; Kelly et al., 2010; Kendon, 1990). In this article, we will focus specifically on the role of gaze behavior to faces in conversational turn-taking during parent-child interactions. First, we will briefly review relevant literature on the role of gaze in face-to-face conversations. Second, we outline the reasons behind our focus on parent-child interactions. Third, we will present the research questions of the current study.

The role of gaze behavior in face-to-face conversations

It has long been known that gaze behavior plays an important role in human face-to-face conversations (Argyle & Cook, 1976; C. Goodwin, 1981; Kendon, 1967). In Kendon's classic study (1967), it was observed that speakers were more likely to end their speaking turn by gazing at the other person, whereas just at the start of one's turn speakers were more likely to gaze away from the other. Kendon (1967) posited that 1) by gazing at the other person at the end of an utterance, the speaker can monitor the other person's behavior (e.g., facial expression, attention of the listener), and 2) by means of a prolonged gaze at the end of one's turn the speaker signals to the other that he expects a response, and thus transfers the turn to the next speaker. In a cross-cultural investigation by Stivers and colleagues (2009), it was shown that, in 9 out of 10 different languages, responses to yes/no questions were generally faster when the speaker was gazing at the recipient while asking a question. Other studies on gaze behavior in face-to-face interactions have also shown that patterns of gaze can reflect different contextual and interpersonal aspects of human interactions, such as the interactants' interpersonal affiliation, intimacy, motives, tasks, and goals (Argyle & Dean, 1965; Behrens & Kret, 2019; Foddy, 1978; Kendon, 1967; Kleinke, 1986; Levine & Sutton-Smith, 1973). One example of this was given by Foddy (1978). Foddy investigated whether patterns of speech and gaze behavior would be reflective of the 'motivational orientation' of the interactants in a negotiation context. Dyads engaged in either a cooperative-negotiation or a competitive-negotiation game. Foddy showed that competitive and cooperative negotiation games differently modulated patterns speech and gaze behavior (e.g., average duration and frequency). During the cooperation-task, participants' average duration of gazes and mutual gazes were longer compared with the competition-task, although this was irrespective of average frequency of gazes. Also, cooperators spoke more frequently than competitors. Foddy (1978) concluded that, indeed, the cooperative and competitive dynamics, elicited by the negotiation games, modulated speech and gaze behavior in competitive and cooperative negotiations.

From observational studies to dual eye-tracking

In (early) observational studies (e.g., Argyle & Cook, 1976; Beattie & Bogle, 1982; Foddy, 1978; Kendon, 1967), researchers have often studied gaze behavior in face-to-face conversations by manually coding interactants' gaze behavior from video-recordings. More recently, researchers in the field of social attention have started to use modern eye-tracking technology to study gaze in interaction at a higher spatial and temporal

resolution (Haensel et al., 2020; Hessels et al., 2019; Ho et al., 2015; S. L. Rogers et al., 2018). Ho et al. (2015), for example, used wearable eye-tracking glasses to measure gaze behavior of student dyads engaged in two word-guessing games. This study corroborated findings from earlier observational studies (cf. Duncan & Fiske, 2015; Kendon, 1967), namely that speakers tended to gaze at the other at the end of their turn (i.e., to signal an upcoming turn transition), and that speakers averted gaze prior to the beginning of their turn. Specifically, Ho et al. (2015) reported that the average lag between looking at the other person until that person started speaking was approximately 400 ms. Also, they found that speakers tend to avert their gaze just after the beginning, or sometimes even slightly before of a speaking turn, and did not gaze back at the other until 700 ms on average after they began speaking. However, one limitation of some wearable eye trackers is that they can lack the spatial accuracy necessary to estimate gaze directed at specific facial features (Nichorster et al., 2020). Yet, from the face-processing and eye tracking literature, it is well-known that looking at specific regions of a face (e.g., eyes, mouth) may be potentially relevant for aspects of face perception, emotion recognition, and speech comprehension (for a recent review, see Hessels, 2020). For example, Buchan et al. (2007) showed that observers presented with videos of talking faces gazed closer at the mouth region while listening to someone speak, and even closer when multi-talker noise was added to the audio. This indicates that looking at someone's mouth may be beneficial for speech perception and comprehension (see also Vatikiotis-Bateson et al., 1998). In a different eye-tracking study, Rogers et al. (2018) used wearable eye-trackers to study gaze behavior to faces in short 'getting acquainted' conversations between undergraduate students. They reported that, on average, gaze behavior to the mouth region was slightly increased when people were listening to someone talking compared to when they were speaking themselves (see also Holleman et al. submitted).

Could it also be the case that looking at specific regions of face may be indicative of aspects of regulation and coordination in turn-taking behavior? Do speakers and listeners look at specific regions of the other person's face prior and after switching turns? While observational studies and some eye-tracking studies have shown that looking at and away from someone is related to turn-taking behavior, no eye-tracking studies have hitherto directly investigated what the regions of the faces are looked during turn transitions, and whether this is coupled with specific aspects of turn-taking, e.g., signaling and monitoring upcoming turn transitions. Hessels et al. (2017) developed a dual eye-tracking setup with video remote eye trackers to measure gaze behavior of two people simultaneously at high spatial resolution (see also Hessels, Holleman, et al., 2018; Hessels

et al., 2019). This dual eye-tracking setup could be particularly relevant for the study of gaze behavior during conversational interaction because this setup allows one to estimate gaze at specific facial features, which may be closely related to visual processing of signals and cues that contribute to smooth turn-taking behavior. For example, looking at someone's eyes may signal the end of one's turn and looking at someone's mouth may be beneficial for anticipating an upcoming turn transition. In a previous study (Holleman et al., submitted), we used this dual eye-tracking setup to study gaze behavior of parents and children while engaged in video-mediated face-to-face conversations. Specifically, the goal of that study was to investigate what regions of the face are looked at by parents and children during episodes of speaking and listening, and how this was modulated by the topic of conversation (i.e., conflict or cooperation). In that study, parents and children engaged in two brief conversations about family issues: a 'conflict' topic (e.g., cleaning room) and a 'cooperation' topic (e.g., birthday party). We found that both parents and children looked more at the other's mouth region while listening compared to while speaking. Also, parents gazed more at the eyes of their children in the conflict-scenario compared to the cooperation-scenario. In the current study, we used a subset of this rich multimodal dataset to investigate gaze behavior to facial features during turn transitions in parent-child conversations.

PRESENT STUDY

In the present study, we focused on the relation between gaze behavior to facial features and turn-taking in parent-child conversations. As discussed above, gaze behavior is important for the regulation and coordination of turn-taking. However, no eye-tracking studies have focused on what specific regions of the face are looked at during turn-taking in face-to-face conversations. Other studies on the role of gaze in face-to-face interactions have shown that gaze behavior to faces may further be dependent on task-demands (e.g., listening to someone speak, see Buchan et al., 2007; Hessels et al., 2019; S. L. Rogers et al., 2018), or may be modulated by interpersonal and contextual factors such as motivational orientation and interpersonal affiliation (Foddy, 1978; Holleman et al., submitted; Jarick & Kingstone, 2015; Kleinke, 1986). We are concerned specifically with parent-child interactions, because the parent-child relationship is a relevant interpersonal context of social interaction, for example (e.g., for social and linguistic development, see Gauvain, 2001). From infancy to adolescence, the nature of the parent-child relationship is characterized by various relational changes in dependence, authority, compliance, and autonomy, although these changes differ across different ages and stages in development

(Branje, 2018; De Goede, Branje, & Meeus, 2009; Dunn & Slomkowski, 1992). Relevant to our current study is that interactions between parents and (pre)adolescent children consist of frequent, low-level conflicts in which children challenge the parents' authority, and parents try to maintain authority while also providing emotional support (Smokowski et al., 2015; Tucker et al., 2003).

In this study, we investigated what regions of the face are looked at during turn transitions in conversations between parents and their preadolescent children. Also, we wondered whether conflict and cooperation dynamics may affect how parents and children look at each other during turn transitions. To this end, we formulated the following research questions: 1) What regions of the face do parents and children look prior, during, and after turn transitions? We expected that speakers will gaze at the face of the other person prior to a turn transition and will gaze away around the beginning of a turn, while listeners are more likely to gaze at the speaker's face more consistently (Duncan & Fiske, 2015; Kendon, 1967). We expected that the eyes and the mouth will be relevant when signaling to and monitoring each other during turn transitions (Holleman et al., submitted; S. L. Rogers et al., 2018). 2). For example, the mouth area may be relevant to monitor the speaker, facilitating the perception of speech and verbal turn-yielding cues to indicate and predict when a turn transition is forthcoming, such as prosodic and syntactic information (Gravano & Hirschberg, 2011). Furthermore, looking at the eyes may be relevant for signaling upcoming turn-transitions; as a nonverbal emphasis on a verbal message, for example, at the end of one's turn, when asking a question, or to pressure for response (Timney & London, 1973). Additionally, we have the following explorative questions: 2) Is gaze behavior at specific facial features dependent on the transition direction (i.e., from parent to child or vice versa)? 3) Is gaze behavior to facial features during turn transitions indicative of the topic of conversation (i.e., conflict and cooperation). Patterns of gaze behavior to faces during turn-transitions could be informative to better understand interpersonal dynamics of parent-child interactions. In our previous study, for example, we found that parents gazed more at the eyes of their children in the conflict conversation over the course of the whole conversation. In that study, we posited that increased eye gaze could either be indicative of how parents assert their dominance and social control, or how they convey empathy and parental affect, as increased eye gaze has been related as an expression of intimacy and interpersonal affiliation (Kleinke, 1986; Patterson, 1982). In the current study we therefore wondered whether the regions of the face are looked at by parents and children during turn transitions are also dependent on who is yielding the turn and the topic of conversation (conflict or cooperation).

METHOD

Participants

All participants were part of the YOUTh study, a prospective cohort study about social and cognitive development with two entry points: *Baby & Child* and *Child & Adolescent* (Onland-Moret et al., 2020). The YOUTh study recruits participants who live in Utrecht and its neighboring communities. The YOUTh study commenced in 2015 and is still ongoing.

We used a previously collected dataset (Holleman et al., submitted) from a subset of YOUTh participants aged around 8-10 years. Data collection for this previous study was approved by the Medical Research Ethics Committee of the University Medical Center Utrecht and is registered under protocol number 19-051/M and further approval for the re-use of the dataset in this study was granted by the YOUTh Executive Board.

The multimodal dataset consists of audio recordings, frontal videos, and gaze position (measured with an eye tracker) of both parent and child, obtained during parent-child conversations. To estimate gaze behavior to facial features during turn transitions, and to investigate whether gaze to facial features differed as a function of transition direction and topic of conversation, we selected data that fulfilled the following criteria: 1) two completed conversations for the parent-child dyad, and 2) the eye-tracking data of both parent and child was of sufficient quality in terms of accuracy and precision. Sufficient quality for accuracy was operationalized as a systematic error of the gaze position signal below 1°. The criterion for precision (or variable error) was a sample-to-sample deviation of the gaze signal divided by the Area-of-Interest (AOI) span was below 1 (Holleman et al. submitted). If the precision/AOI-span ratio is larger than 1, this means that the sample-to-sample deviation is larger than the inter-AOI distance and one cannot therefore reliably map participants' gaze position signal unto a specific facial feature.

We set no cut-off criterion for data loss (i.e., missing data), which is an important characteristic of eye-tracking data quality. Data loss is a complicated measure to interpret in the context of our study, given that data loss may coincide with talking and movement (Holleman et al., 2019). For example, it is well-known that speakers gaze away more from the other person while listeners gaze at the speaker more consistently (Hessels et al., 2019; Kendon, 1967). Therefore, any exclusion based on data loss may selectively remove participants that spoke relatively more.

Our data quality criteria resulted in a final selection of 8 parent-child dyads (n=16). The average 2d-validation accuracy was 0.74° for children and 0.69° for parents. The average precision (s2s-RMS/AOIspan) was 0.5 for children and 0.17 for parents. The average percentage of data loss was 35.7% for children and 12.35% for parents. A complete overview with descriptive statistics of the parent-child dyads (i.e., age, gender, dyad composition) and eye-tracking data quality estimates per participant are given in Table 1. Note that c1 refers to the conflict conversation and c2 refers to the cooperation conversation.

Table 1. Descriptive statistics of parent-child dyads and eye-tracking data quality estimates

<i>Child</i>	1	2	3	4	5	6	7	8
<i>Age (years)</i>	9.7	10.4	8.25	8.75	10.9	10.2	8.9	9.25
<i>Sex</i>	♀	♀	♀	♀	♀	♀	♀	♂
<i>Accuracy (deg)</i>	0.9	0.52	0.84	0.62	0.94	0.32	0.83	0.95
<i>Precision/AOIspan</i>								
<i>c1</i>	0.33	0.25	0.36	0.69	0.79	0.35	0.37	0.92
<i>c2</i>	0.37	0.30	0.55	0.75	0.57	0.44	0.24	0.94
<i>Data loss (%)</i>								
<i>c1</i>	25.9	43.1	12.1	60.8	39.0	10.1	46.7	47.6
<i>c2</i>	65.4	52.8	24.3	51.9	27.7	7.12	24.6	63.9
<i>Parent</i>	1	2	3	4	5	6	7	8
<i>Age (years)</i>	39	49	48	45	43	44	35	44
<i>Sex</i>	♀	♂	♀	♂	♀	♀	♀	♀
<i>Accuracy (deg)</i>	0.55	0.69	0.39	0.58	0.71	0.97	0.87	0.82
<i>Precision/AOIspan</i>								
<i>c1</i>	0.31	0.18	0.12	0.06	0.13	0.23	0.19	0.12
<i>c2</i>	0.18	0.35	0.16	0.07	0.12	0.24	0.16	0.12
<i>Data loss (%)</i>								
<i>c1</i>	17.2	13.8	16.9	6.32	4.51	22.3	6.5	11.2
<i>c2</i>	12.5	19.4	35.4	10.7	1.68	13.0	10.88	16.9

Apparatus

The dataset collected by Holleman et al. (submitted) consisted of audio, video, and gaze position recordings of parents and their children engaged in video-mediated face-to-face conversations. Parents and their children could see and converse with each other through a live video-connection (see Figure 1a). Each person was displayed to the other by means of a monitor and a half-silvered mirror. Two remote eye trackers placed inside the

metal boxes located beneath the monitor screens recorded participants' eye movements (Figure 1b), and two Lavalier microphones attached to the front of each setup recorded participants' speech (see Figure 1a). A full description of the technical details about the dual eye-tracking setup and procedure are given in Hessels et al. (2017) and Holleman et al. (submitted).

Procedure

Parents and their children were invited to the laboratory and the experimenter read them a general instruction about the task and procedure of the experiment. The experimenter also told participants that they should ask questions about the experiment if they did not understand particular aspects of the instruction and experimental procedure. Each parent-child dyad was asked to engage in two brief conversations about common family topics. We used two age-appropriate semi-structured conversation tasks which have been extensively used by researchers to assess various aspects of the parent-child relationship, e.g., attachment, interpersonal affect, relational quality, parental style, and child compliance (Aspland & Gardner, 2003; Ehrlich et al., 2016; Hutchinson et al., 2019). In our version, we had a 'conflict' and a 'cooperation' conversation. The two conversations are designed to elicit interactions between parents and children that are representative of discussions about common family issues.

After the general instruction, participants were seated in front of each end of the dual eye-tracking setup such that their eyes were at the same height as the webcams. This was done to ensure that participants had the impression of looking each other in the eye (see Hessels et al., 2017). The distance of participants' eyes to the eye tracker was approximately 70 cm and the distance from the eyes to the screen was approximately 81 cm. After positioning, a 5-point calibration was conducted followed by a 4-point calibration validation. After the calibration procedure, the experimenter briefly repeated the task-instructions (see below) and explained that he would initiate the video-feed after a countdown. The experimenter explained he would stop recording the conversation after approximately five minutes. The experimenter did not speak or intervene during the conversations, only if participants questioned him directly, or when participants had changed their position too much.

a **b**

Figure 1. Overview of the dual eye-tracking setup. 1a) Staged photographs of two interactors in the dual eye-tracking setup, reproduced from Holleman et al. (submitted), with permission. 1b) A schematic overview of the setup, reproduced from Hessels et al. (2018), with permission. The cameras behind the half-silvered mirrors recorded video at 30 Hz at a resolution of 800 by 600 pixels). The live video-feeds were presented at a resolution of 1024 by 768 pixels in the center of a 1680 by 1050 pixels computer screen and concurrently recorded to disk. Two *SMIRED* eye trackers running at 120Hz recorded participants' eye movements (see Figure 1b). Audio was recorded with Lavalier-microphones (one for each participant) which were connected to an audio mixing console. Each microphone was attached to the front of each setup (see the dashed orange circles on the left panels in Figure 1a). To record the audio, we used *Audacity v. 2.3.3* which ran on a separate computer.

Conflict-scenario

In the first conversation, parent-child dyads were instructed to have brief discussion about a topic of disagreement and, if possible, discuss how they may resolve the disagreement. Prior to the experiment, parent-child dyads were presented with a list with common topics of potential disagreements that occur within family life. The list included topics such as bedtime, screen time, listening to parents, and household chores. After the parent-child dyads agreed on a topic, the experimenter instructed the participants to solely talk about this topic for the next 5 minutes during the experiment.

Cooperation-scenario

In the second conversation, parent-child dyads were instructed to discuss a party plan together (e.g., birthday, family party). The idea behind this task was to elicit a more egalitarian and pleasant interaction between the parent and child as compared with the first conversation. Participants were instructed to think about a party that they would like to organize, and discuss, for example, what the occasion for the party would be, who they would want to invite to the party, and what kind of activities they would want to organize. After this final conversation, the experimenter thanked the parents and their children for their participants and the child was rewarded with a small souvenir.

Data preparation and coding

As stated, we investigated where parents and children looked at each other's faces during turn transitions in conversational interactions. To this end, we needed 1) high-resolution eye tracking data of parents and children to determine what region of the other's face they looked at for each point in time, 2) audio recordings of parents and children's conversations to determine who was speaking and when (referred to as speaker turns), and 3) specific timepoints at which turn transitions occurred. From our previous study (Holleman et al., submitted), we already had the eye-tracking data mapped unto Areas-of-Interest (AOIs). Although we had audio recordings from the parent-child conversations, we still needed to manually derive speaker turns from the audio recordings, from which we could then derive turn transitions. First, we give a brief recap of how we mapped eye-tracking data unto the AOIs. It is described in more detail in Holleman et al. (submitted). Next, we describe how we manually coded speaker turns from the audio recordings of parent-child conversations. After this, we describe how we determined turn transitions from the classified turns. Finally, we describe how we combined the eye-tracking data with the

turn transition events and describe what measures were used to assess participants' gaze behavior during turn transitions as a function of transition direction, question-response sequences, and topic of conversation.

Estimating gaze to facial features

Parents and children's gaze behavior was mapped unto AOIs for the facial features: the eyes, nose, and mouth region. For this, we used an automatic AOI construction method which finds facial landmarks for each frame of the frontal video recorded in the dual eye-tracking setup (Baltrušaitis et al., 2016; Hessels, Benjamins, et al., 2018). Each sample of the gaze-position signal is then assigned to a facial-feature AOI if its coordinates fall within a specified radius around the facial feature's AOI center (Hessels et al., 2016). If the measured gaze position falls outside of all radii, gaze position is assigned to a 'background' AOI (see Figure 2). This background area consists of the background and small parts of the upper body of the participant visible in the video. In our study, the radius was set to 4° ³ (200 pixels). This method is called Limited-Radius-Voronoi-Tessellation (LRVT; see Hessels et al., 2016). The LRVT method is partly data-driven, resulting in smaller AOIs on the children's faces compared to AOIs on the parents' faces. We quantified the AOI size by the AOI span. The AOI span is defined as the mean distance from each AOI cell center to the cell center of its closest neighbor (see Hessels et al., 2016, p. 1701). After individual gaze samples were mapped unto AOIs (see previous section), we computed dwells (Holmqvist et al., 2011). We operationalized a dwell as the period which gaze position enters the AOI entrance until it exits the AOI for a minimum duration of 120 ms.

³ All visual angles in our study are reported under the assumption that participants were seated at approximately 81 cm distance from the screen and in the center of the camera image.

Figure 2. A schematic example of the automatically-generated facial feature AOIs (Hessels, Benjamins, et al., 2018) for the left eye (L), right eye (R), nose (N), and mouth (M). The AOI for the left eye, for example, is the area closest to the left eye center but not further away from the center than the bounded radius of 4° (denoted with a red arrow). The background AOI (B) encompasses the background, the upper body of the participant and a small part of the top of the head. Reproduced from Holleman et al. (submitted), with permission.

Classifying turn-taking behavior

The audio recordings of parent-child conversations were trimmed at the start and end of the two conversations. These start and endpoints coincided with the start and stop of the eye-tracking and video recordings. For every conversation, we determined speaking turns in *ELAN v. 6.0*, a software package for annotation of audio and video files (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). Author GH manually marked the beginning and endpoints of speaker turns for both conversations. Our definition of a single turn is a period where one person is speaking until a next speaker takes over. Speaker turns are operationalized as the period between the beginning and end of a speaker's sequence of utterances which together constitute the speaker's turn. In other words, turns are stretches of one or more utterances that are not interrupted by another speaker. Note that our operational definition of a single turn includes all within-speaker gaps ('pauses') and excludes between-speaker gaps. Also, fillers and backchanneling ('uh-

huhs') do not constitute a separate turn or count as interruptions of the other's speaking turn. In the case of a simultaneous start, the turn was assigned to the participant who ends up 'holding the floor'. For each turn, we coded whether the turn ended with a question or not. In some conversations, it happened that the operator spoke during the conversation, for example when one of the participants asked a question directly to the operator, or when the operator instructed participants to stay on topic or asked the child to sit more upright. In these cases, these events were coded separately with an 'operator' label. This was done to later remove the participant-to-operator transitions (and vice versa) from the analyses.

Classifying turn transitions periods

From the annotated turns, we computed turn transitions periods; the moments when turn transitions between speakers occur. Turn transition periods in this study are operationalized as the period between the beginning and ends of gaps (between-speaker intervals) or overlaps. In Figure 3, our operational model of turn-taking behavior is depicted. In Figure 3a, Speaker 1 (Sp1) is the parent (white bar) and speaker 2 (Sp2) is the child (grey bar). T1 I in the white bar indicates the parent's first turn and T1 in the grey bar indicates the child's first turn. In Figure 3a, turn transition periods (*ttp*) are illustrated in the case of gaps (G). In figure 3b, depicts turn transition periods (*ttp*) in the case of overlaps (O). In Figure 3c, the case of a simultaneous start (SS) is depicted. In this example, the first turn T1 is assigned to the parent because the parents ends up 'holding the floor'. All turn transition periods are characterized by an onset and offset time, a central tendency (a middle point between the onset and offset time), and a total duration. We also coded the transition direction for each turn transition as being from parent to child or from child to parent, and whether the turn transition was preceded by an explicit question by one of the participants.

Figure 3. Model of turn-taking behavior. Speaker 1 (Sp1) is the parent (white bar) and speaker 2 (Sp2) is the child (grey bar). T1 is the parent’s first turn. T1 in the grey bar is the child’s first turn. We are interested in the moments in time when turn transitions between speakers occur; the ‘turn transition period’. In Figure 3a, turn transition periods (*ttp*) are illustrated in the case of gaps (between-speaker intervals, indicated by G). In Figure 3b, depicts a turn transition period (*ttp*) in the case overlaps (overlapping speakers, indicated by O). In Figure 3c, the case of a simultaneous start (SS) is depicted. In this example, the first turn T1 is assigned to the Speaker 1 who ends up ‘holding the floor’ after a simultaneous start.

Estimating gaze behavior to faces during turn transitions

For each turn transition period, we computed the proportion of gaze on the facial features AOIs and the background-AOI. We did this for a time window ranging from -2500 to +2500 ms around the turn transition center point. We also computed the proportion of data loss. Data loss may stem from three sources, namely: 1) samples with no eye-tracking data, 2) samples with no AOI labels assigned to a video frame, and 3) samples that did not count as a dwell (i.e., shorter than 120 ms). Missing eye-tracking data (i.e., no gaze position recorded) was the most prominent source of data loss in our study. As stated previously, the total amount of time without eye-tracking data was higher for children than parents (see also Table 1 above).

RESULTS

Number and duration of turn transitions

In this study we investigated what regions of the face are fixated during turn transitions. However, we first examined whether the number and duration of turn transitions between parents and children differed as a function of transition direction and topic of conversation. This is relevant, because it could have affected how the analyses of gaze to facial features during turn transition periods was approached. If, say, the parent to child turn transitions are substantially different (i.e., longer or shorter) from child to parent turn transitions, we might have to perform separate analyses.

In Figure 4, we visualized distributions of turn transition durations of 8 parent-child dyads for the two conversations: conflict and cooperation. Most transition durations had positive values, irrespective of transition direction (parent to child or vice versa). In other words, most turn transitions consisted of between-speaker intervals, or gaps (cf. Heldner & Edlund, 2010; Ten Bosch, Oostdijk, & Boves, 2005). Overall, we found that more turn transitions occurred in the cooperation conversation compared to the conflict conversation (lower panels). Numerically, turn transition durations were shorter for child to parent transitions (conflict: median = 0.318 seconds [IQR 0.146 s – 0.697 s]; cooperation: median = 0.221 seconds [IQR 0.065 s – 0.511 s]) than parent to child transitions (conflict: median = 0.429 seconds [IQR 0.167 s – 0.890 s]; cooperation: median = 0.329 seconds [IQR 0.125 s – 0.6830 s]). This suggests that parents reacted faster than children. Also, the median turn transition duration was numerically shorter in the cooperation conversation, regardless of the turn transition direction. This suggests that parents and children on average reacted faster in the cooperation conversation.

Figure 4. Distributions of turn transition durations. *Upper panels:* turn transitions from the conflict conversation. *Lower panels:* turn transitions from the cooperation conversation. The left panels depict parent to child turn transitions. The right panels depict child to parent turn transitions. In this case, turn transition durations of 0 indicate ‘perfect’ turn transitions with no gap or overlap.

We conducted Bayesian statistical analyses in JASP *v.* 0.14 (JASP Team, 2018) to test whether conversation topic and transition direction modulated the number and durations of turn transitions. Below, we report Bayes’ factors (BF10 and BFm) to quantify the likelihood ratio of different statistical models. Briefly, BF10 represents the ratio between the likelihood of the alternative hypotheses given the data and the likelihood of the null hypothesis given the data. For details, we refer the reader to the JASP website at <https://jasp-stats.org/>. We considered a Bayes’ factor of 10–100 as strong support for a model, 3–10 as moderate support and 1–3 as anecdotal support (Jarosz & Wiley, 2014, p. 8). First, we conducted a Bayesian paired-sampled t-test to compare whether the

number of turn transitions differed between the two conversation topics (conflict and cooperation). The Bayes' factor (BF_{10}) for the hypothesis that the number of transitions differed between conversation topic was 145. Second, we conducted a within-dyads Bayesian repeated-measures ANOVA to compare whether turn transition durations were modulated by the topic of conversation, and by the turn transition direction. The model that yielded the highest Bayes' factor was the 'conversation topic' ($BF_m = 2.874$). The next two highest Bayes' factors were obtained for the 'transition direction' model ($BF_m = 0.172$) and the 'conversation topic' x 'transition direction' model ($BF_m = 1.441$). Given that all Bayes' factors were below 3, we concluded that turn transition durations were not substantially modulated by topic of conversation and transition direction. Only the number of transitions seemed to be dependent on the conversation topic. Because turn transition durations were not substantially modulated by topic of conversation and transition direction, we did not carry out separate analyses for different turn transitions durations in relation to gaze behavior (see section: Gaze behavior to facial features during turn transitions).

Next, we examined whether turn transition durations that included a question were different in their number and median duration from turn transitions that did not include a question. Out of all turn transitions (total = 1338), 333 turn transitions (29%) contained a question at the end of the turn preceding the turn transition across all parent-child dyads and regardless of conversation topic. In total, parents asked 268 questions whereas children asked 65 questions. Interestingly, while the total number of questions asked by parents was quite similar across the two conversations (conflict: $n=129$; cooperation: $n=139$), children asked more questions in the cooperation conversation ($n=46$) compared to the conflict conversation ($n=19$). Because the total number of questions asked by children was so small, we decided to refrain from further analyses to examine whether turn transitions durations with a question were different in terms of, e.g., turn transition direction and topic of conversation.

Gaze behavior to facial features during turn transitions

As outlined in the introduction, previous studies have shown that gaze behavior is closely coupled with the regulation and coordination of turn-taking. For example, speakers may look at the other at the end of turns to signal upcoming turn transitions, while listeners may look at the other to anticipate one's next turn (Duncan & Fiske, 2015; Ho et al., 2015; Kendon, 1967). Here, we investigated whether participants' gaze behavior to faces

during turn-taking would be modulated by 1) the turn transition direction (parent to child, child to parent), 2) temporal aspects of turn transitions, namely the time periods just prior to and after the turn transitions, and 3) the type of conversation, (conflict or cooperation). To this end, we examined what regions of the face (eyes, nose, mouth) are looked at by parents during turn transition periods. We computed the average proportion of gaze to all AOIs before (-2500 to -1 ms), and after (+1 to 2500 ms) turn transition periods. Note that a value of 0 represents the center of the turn transition. This was done for both turn transition directions and for both conversations.

We conducted a Bayesian repeated-measures ANOVA in JASP *v.* 0.14 (JASP Team, 2018) to test whether turn-direction, pre-and-post turn transition periods, and topic of conversation, predicted gaze behavior to facial features during turn transitions. We used four model terms: 1) an AOI-category with five levels (eyes, nose, mouth, background, data loss), 2) the transition direction with two levels (from parent to child and from child to parent), 3) the timing relative to the turn transition with two levels (pre: -2500 ms to -1 and post: +1 to 2500 ms), and 4) conversation topic with two levels (conflict and cooperation). Note that 0 represents the center of the turn transition. We furthermore included the two-way interaction terms with the AOI-category. The best model was the 'AOI-only' model, as indicated by a Bayes' Factor of 106.61 (BFm). This model was (more than) ten times more likely than all other considered models, for example, the 'AOI x pre- and post-turn transition' model (BFm = 2.477), the 'AOI x conversation topic' model (BFm = 2.445), and the 'AOI x transition direction' model (BFm = 2.379). In short, we found no substantial differences in gaze behavior to the AOIs as a function of turn transition direction, timing relative to the turn transitions, and conversation topic. Therefore, we will only depict average proportion of gaze to facial feature AOIs for eight dyads, irrespective of turn transition direction, pre- and post turn transition periods, and conversation topic.

In Figure 5, we summarize our results by showing proportions of gaze to facial feature AOIs, background-AOI, and proportion of data loss for every parent-child dyad. We computed the average proportion of gaze to facial features, background, and data loss for both the parent and the child by averaging across all turn transitions obtained from both conversations. This was done because the previous analysis showed that gaze behavior to facial features did not differ substantially across transition direction, prior and after turn transitions, and conversation topic. Overall, participants varied greatly in where they looked at the other's face. Some parents and children gazed more at the eyes, while

others gazed more at the mouth, or gazed more equally across different regions of the face during turn transitions. This converges with our analyses from the previous study in which we showed substantial individual variability in what regions of the face were looked at by both parents and children across the course of the conversation (Holleman et al., submitted).

As in our previous study (Holleman et al., submitted), we found that especially the children's proportion of data loss (DL) was higher compared to the average proportions of data loss of parents, although this was not consistent for all parent-child dyads. When we compared the proportion of data loss during turn transitions to the average proportions of data loss across the conversations itself, we found that data loss was much higher during turn transitions. For most children (except for children of dyad 3 and dyad 6), average proportions of data loss during turn transitions ranged between 0.6 and 0.8. However, proportions of data loss of children across the entire conversations ranged from 0.1 to 0.6. Likewise, proportions of data loss of parents during turn transitions ranged from 0.1 to 0.5 whereas the proportions of data loss across conversations ranged from 0.02 to 0.22 (see Table 1). In the discussion, we will further address this problem and outline some possible solutions.

◀ **Figure 5.** Gaze behavior to facial features for all parent-child dyads during turn transitions. Every panel represents one dyad (N=8). Blue bars represent the proportion of parents' gaze to face-AOIs (eyes, nose, mouth), background-AOI (NON), and proportion of data loss (DL) averaged across all turn transitions and averaged across the two conversations. The yellow bars represent the average proportion of children's gaze to face-AOIs, background-AOI and proportion of data loss.

DISCUSSION

In this study, we investigated the relation between gaze behavior to faces and turn-taking in parent-child conversations using a recently developed dual eye-tracking setup. As discussed in the introduction, it is well known from the observational literature on face-to-face conversations that gaze behavior plays an important role in the regulation and coordination of turn-taking. A well-established observation is that when speakers reach the end of their turn, they are more likely to gaze at the other person, while at the start of a turn they tended to gaze away from the other (Duncan & Fiske, 2015; Ho et al., 2015; Kendon, 1967). Several eye-tracking studies have also shown that people are more likely to fixate the mouth region when listening to someone speak (Buchan et al., 2007; Holleman et al., submitted; S. L. Rogers et al., 2018; Vatikiotis-Bateson et al., 1998), which suggests that fixating the mouth region is related to speech perception and comprehension. In this study, we posited that looking at facial features may also be indicative of whether a speaker signals an upcoming turn transition, or whether a listener monitors the speaker for an upcoming turn transition.

We asked the following questions: 1) What regions of the face do parents and children look at during turn transition periods? 2) Is gaze behavior at specific facial features dependent on the transition direction (from parent to child or vice versa)? 3) Is gaze behavior to facial features during turn transitions indicative of the topic of conversation (conflict and cooperation)? To estimate gaze behavior to facial features during turn-taking, we used a multimodal dataset from a previous study (Holleman et al., submitted), which consisted of audio recordings, frontal videos, and gaze position (obtained with a dual eye-tracking setup) of parents and their children engaged in conflict and cooperation conversations. We first briefly recap the results regarding the number and duration of turn transitions after which we will discuss our results regarding gaze behavior during turn transitions. As this study represents a first attempt to study gaze behavior to facial features during turn transitions using a dual eye-tracking setup, we will consider several limitations of the current approach and outline some possible solutions.

Summary of results

We first conducted several analyses regarding the number and durations of turn transitions. Interestingly, the total number of turn transitions was larger for cooperation conversations than for the conflict conversations. Furthermore, it seemed that for both parents and children, the median turn transition durations were shorter in the cooperation conversation than in the conflict conversation. Also, we found that parents asked most of the questions, regardless of conversation topic, while children asked more questions in the cooperation conversation than they did in the conflict conversation. The fact that parents asked most of the questions shows that, overall, parents tended to take the lead during the conversations. Moreover, the greater number of turn transitions in the cooperation conversation compared to the conflict conversation indicates that the different topics elicit different dynamics in the role that parents and children take within the conversation.

Regarding our main research goal, we did not find that gaze to facial features (eyes, nose, mouth) differed between periods prior and after the turn transition, or that gaze to facial features was modulated by the turn transition direction (i.e., from parent to child or vice versa). Moreover, no substantial differences were found in proportions of gaze to facial features as a function of the conversation topic. Overall, our results suggest that some individuals gazed more at the eyes or mouth on during turn transitions, while others gazed at different facial features more equally. Such idiosyncratic face scanning patterns have been reported previously by eye-tracking studies, and seem to be stable individual differences in what regions of the face people tend to fixate most during face perception and face recognition tasks (Arizpe et al., 2017; Mehoudar et al., 2014; Peterson & Eckstein, 2013; Peterson et al., 2016). In our previous study, we found evidence for individual consistency in what regions of the face parents and children looked at across the two conversations (Holleman et al., submitted). Thus, it could very well be that gaze behavior to faces during turn transitions is likewise idiosyncratic. However, from our previous study (Holleman et al., submitted) we also knew that the average quality of dual eye-tracking data was far from optimal. In that study, for which we conducted an extensive analysis of data quality, we could not use over 50% of all the data because a lot of data was of insufficient quality to estimate gaze to facial features. Also, we knew that children's data quality was much lower than parents' data in terms of its overall proportion of data loss, as well as measures of precision and accuracy. In the current study, the proportions of data loss during turn transitions were large for many participants, especially for the children. As such, we could not draw solid conclusions about what regions of the face are looked at most during turn transitions.

Technical limitations of eye tracking to study gaze in face-to-face interactions

One potential reason for the large proportions of data loss is that gaze behavior during turn-taking behavior is characterized by looking at and away from each other at the end and beginning of turns (Duncan & Fiske, 2015; Kendon, 1967). In the dual eye-tracking setup used in this study, gazing away from the other person's face could potentially be recorded as gaze directed at the background area (i.e., not looking at the face), but if participants turned their heads away too much away from the screen the eye tracker could no longer track participants' gaze position. Also, a gaze shift back towards the screen after the eye tracker loses track of the gaze position signal does not always coincide with a smooth and instantaneous recovery of the gaze position signal (Hessels et al., 2015; Niehorster et al., 2018). Such problems occur specifically for remote eye trackers (i.e., eye trackers positioned at a distance from the participant). Other researchers have used wearable eye-tracker systems to study gaze behavior during face-to-face interactions. While wearable eye-trackers do not necessarily suffer from the same problem of losing track of someone's gaze due to looking at and away from the other person, wearable eye-trackers have different limitations, such as a lack of accuracy to distinguish gaze position on different regions of the face. Also, slippage of the head-worn eye tracker may occur when people are speaking or smiling for example, which can result in a loss of accuracy (Niehorster et al., 2020). Clearly, face-to-face interactions are not the most ideal conditions for most eye-tracking systems. Consequently, these technical limitations thereby constrain what kind of research questions can feasibly be investigated by researchers interested in gaze behavior in interactive situations (for a recent review, see Valtakari et al., 2021). In our case, to better estimate what regions of the face are looked at during turn transitions periods in face-to-face interactions, we need an eye tracker which is not only robust against some face and head movements but is also quick to recover gaze position after face and head movements during turn transitions.

Due to the high proportions of data loss in the current study, we cannot draw firm conclusions about the role of gaze to facial features during turn transitions. We cannot claim, for example, that the large proportions of data loss were the result of looking away the other during the turn transitions periods, or whether data loss was due to other causes. To assess to what extent data loss was the result of looking at and away from the other person during turn transitions, additional analyses and coding of the video-recordings will be necessary. One approach would be to visually check whether loss of the gaze position signal coincides with face and head movements. However, manual coding of participants'

gaze direction from frames of frontal video recordings is an incredibly tedious job. A single video recording of one person in a 5-minute conversation may consist of around 9000 to 10,000 frames. We had video-recordings of two people, who had two conversations, and 8 dyads in total, meaning a total of 32 video-recordings need to be manually coded, including at least an additional coder to establish inter-rater reliability. Instead, a more efficient approach would be to use automated face-detection and head pose-detection methods (Baltrušaitis et al., 2013; Baltrušaitis et al., 2016). These methods can provide additional objective estimates of face and head orientations, which could be used to label data loss as resulting from the behaviors of interest or not. Additional estimates of face and head orientations would then make it possible to identify parts of the data where gaze behavior to the other's face is correctly recorded in and around each turn transition. Combined with an eye-tracking system that is robust against face and head movements, this would make it more feasible to address our current research questions.

Future directions for the study of gaze behavior in parent-child interaction

Our study is the first to use a dual eye-tracking setup with remote eye trackers to study the relation between gaze to facial features and turn-taking in parent-child interactions. As outlined above, there are substantial technical problems to overcome when using eye tracking technology to investigate gaze behavior in face-to-face interactions. However, one reason to pursue the study of gaze behavior in parent-child interactions using eye-tracking technology, despite of its challenges, is to better understand how parents and children engage with each other across development and how aspects of parent-child interactions at the micro-scale (e.g., sequences of verbal and nonverbal behavior) are related to parent-child dynamics at the macro-level (e.g., relationship quality). In our previous study (Holleman et al. submitted), we found that some parents looked more at the eyes of their children in the conflict conversation compared to the cooperation conversation. We posited that gaze behavior to facial features (e.g., gazing at the eyes) during parent-child interactions could be related to how parents exert authority and express affect, which could influence children's emotional reactivity and social competence (Junge et al., 2020; Moed et al., 2017; Woody et al., 2020). To further investigate whether such gaze patterns are indeed related to parents' emotional affect and parental authority, we need additional analyses of parents and children's verbal and nonverbal behavior. As such, further analyses of parents' and children's verbal and nonverbal behavior may be used to study what conflict resolution strategies are adopted by parents and children

when discussing conflicts of interest (Beaumont & Wagner, 2004; Moed et al., 2015), and how parents' and children's verbal and nonverbal behavior is related to functional and dysfunctional conflict styles. A recent eye-tracking study by Woody et al. (2020) gives an example of how such aspects of parent-child interactions can be combined. Woody and colleagues used wearable eye-trackers in combination with facial affect detection algorithms to investigate whether young adolescents' gaze aversion was modulated by displays of maternal affect (i.e., facial expressions) during conflict discussions. They found that adolescents' gaze aversion (looking away from the mother) increased with displays of both positive and negative maternal affect, but not neutral affect. Woody et al. posited that adolescents' affect-biased attention to their mothers during conflict discussions are consistent with more general approach-avoidance models (Adams, Ambady, Macrae, & Kleck, 2006; Adams & Kleck, 2005).

In sum, the use of dual eye-tracking technology, in combination with algorithms to classify face and pose, speech content, and voice affect, could be the key to further understanding parent-child interactions in terms of, for example, parents' and children's conversational style and conflict resolution strategies (Beaumont & Wagner, 2004; Dixon et al., 2008; Thomas et al., 2017), as well as individual and interpersonal differences in emotion regulation and social competence (Hutchinson et al., 2019; Junge et al., 2020; Moed et al., 2017; Moed et al., 2015; Speer, Cook, McMahon, & Clark, 2007; Woody et al., 2020).

CONCLUSIONS

In this study, we investigated what facial features are looked at during turn transitions in parent-child conversations. We cannot claim that gaze behavior to facial features is related to specific aspects of turn-taking behavior, for example, when signaling upcoming turn transitions or when monitoring the other to anticipate upcoming turn transitions. We have outlined several solutions to some of the technical problems of dual eye-tracking technology to measure gaze to faces during turn-taking. Future research in this area may also benefit from automatic analyses of head and face orientations, in addition to eye tracking recordings, to codify data loss as resulting from the behaviors of interest, e.g., looking at and away from another person's face during turn-taking.

Chapter 7

General discussion

The primary topic of the present dissertation is gaze behavior in human social interaction, specifically face-to-face interactions. As discussed in **Chapter 1** (General Introduction), a dominant approach in the field of social attention – a field of study specifically concerned with attentional processes in relation to other social agents – has been to use experimental laboratory paradigms designed to investigate how individuals look at, process, and respond to ‘nonresponsive’ stimuli presented to them on a computer screen (e.g., images and videos of faces). This approach has drawn a lot of criticism in the past two decades or so. It has been argued that a narrow focus on socio-cognitive and behavioral processes in the absence of interaction (e.g., face perception, emotion recognition, and joint attention) has resulted in a limited understanding of social cognition and behavior as it occurs in ‘real-life’ (Risko et al., 2016; Schilbach et al., 2013; Shamay-Tsoory & Mendelsohn, 2019). Moreover, it is argued that laboratory experiments using images and videos of faces are low in ‘ecological validity’. The current challenge for the field of social attention then is to design experiments that will capture the relevant aspects of visual attention and gaze behavior in socially interactive situations, while also maintaining a high level of experimental control.

Given the recent ‘real-world or the ‘lab’-discussion within the field of social attention, I wanted to be able to defend the use of different eye-tracking procedures and experimental setups against potential criticism typically wielded against experimental laboratory paradigms. As such, **Part 1** of this dissertation (Chapter 2 and 3) consisted of a conceptual and historical analysis of the so-called ‘real world vs. the lab’-dilemma in psychological science. Specifically, I presented an extended critique of the concept of ecological validity and its role in the ‘real-world vs. the lab’-discussion. The ideas and conclusions presented in the first conceptual part have shaped the experiments I conducted for the second empirical part of this dissertation. In **Part 2**, I presented three eye-tracking studies in which I investigated different aspects of gaze behavior in face-to-face interactions (Chapter 4, 5, and 6). I will briefly summarize these studies and address their limitations. Finally, I will outline some directions and implications for future research based on the chapters presented in this dissertation.

PART 1: CONCEPTUAL ANALYSIS OF THE ‘REAL WORLD VS. THE LAB’-DILEMMA: A CRITIQUE OF THE TERM ECOLOGICAL VALIDITY

In **Chapter 2**, I have presented a conceptual and historical analysis of the so-called ‘real world vs. the lab’-dilemma in psychological science. Specifically, I focused on how researchers over the past 70 years have often used the term ‘ecological validity’ to address the correspondence between human behavior and cognition as studied in the ‘laboratory’ and human behavior and cognition as it supposedly occurs in the ‘real world’ (Koehler, 1996; Neisser, 1976; Osborne-Crowley, 2020; Shamay-Tsoory & Mendelsohn, 2019; Smilek et al., 2006). What I have termed the ‘real-world approach’ is a strong emphasis on the idea that, to bridge the gap between the ‘laboratory’ and the ‘real world’, experimental paradigms should have ‘high’ or ‘greater’ ecological validity. An experimental paradigm is said to be **ecologically valid** if the stimuli, tasks, situations, conditions, etc. are **natural** and **complex**. If so, you may assume that your study yields results and conclusions that are generalizable to the ‘real world’. Conversely, an experimental paradigm is said to be **ecologically invalid** if the stimuli, tasks, situations, conditions, etc. are **artificial** and **simple**. If so, one should be skeptical about the extent to which a study yields results and conclusions that one may generalize to the ‘real world’. In fact, according to the dominant view, it is quite likely that a study yielded nothing more than just another ‘laboratory curiosity’.

As made clear in Chapter 2, the ‘real-world approach’ and its emphasis on ecological validity has several problems. One problem is that definitions of ecological validity vary greatly between researchers and between different subfields of psychology. As such, it is not always clear what is meant by this term nor how ‘more’ or ‘greater’ ecological validity should be achieved. One reason for this is that the term ‘ecological validity’, as it is used to address the correspondence between human behavior and cognition as studied in the ‘laboratory’ and human behavior and cognition as it supposedly occurs in the ‘real world’, has become so general that it could potentially refer to anything. Its indiscriminate usage creates the problem that researchers may have different ideas and criteria for what exactly they mean with terms like ‘ecologically valid’, ‘natural’, and ‘complex’ (and the same goes for their antonyms). Thus, as general ascriptions of a given study, experimental setup, or laboratory paradigm, such terms do not provide an adequate specification of the (potential) range, distribution, and boundary conditions of the variables under study (Dunlosky et al., 2009; Hammond, 1998; Hammond & Stewart, 2001; Kruglanski,

1975). Consequently, discussions about ‘ecological validity’, representativeness, and generalizability have often been confused.

In **Chapter 3**, I commented in more detail on the historical roots of ‘ecological validity’ and ‘representative design’. These concepts were coined by Egon Brunswik (1903 – 1955) but have become misunderstood and confused with each other over the past decades. At least since Ulric Neisser’s influential book *Cognition and Reality* (1976), the term ‘ecological validity’ has been used increasingly to address the correspondence between human behavior and cognition as studied in ‘laboratory’ experiments and human behavior and cognition as it manifests in the ‘real world’. My commentary was written to remind researchers in psychology about the original meaning of these concepts – ecological validity and representative design – and their relation to psychological theory. Specifically, I discussed how these ‘Brunswikian’ ideas may benefit current research practices and discussions in the field of social attention. Interestingly, this commentary sparked another commentary published more recently by Kihlstrom (2021), who provided additional historical details about interpretations of ecological validity and how these are related to the current popular usage of this term. Kihlstrom discussed that in the early 1960s, the psychiatrist and social psychologist Martin Orne grounded his idea of ‘demand characteristics’ in a Brunswikian framework, namely Brunswik’s lens model of perception. According to Orne (1962), demand-characteristics refer to the “*totality of cues which convey an experimental hypothesis to the subject*” (Orne, 1962, p. 779). Orne used Brunswik’s lens model to investigate whether subjects in experiments pick up information that reveal the goals of the experimenter (e.g., details about the instruction, experimental setting and procedure, behavior of experimenter, etc.) which may then influence the behavior of the subject and outcome of the experiment. Kihlstrom’s (2021) commentary explains that Orne revised the concept of ‘ecological validity’ as “*appropriate generalization from the laboratory to nonexperimental situations*” (p. 776), shifting Brunswik’s original meaning of ‘ecological validity’ as the potential utility of proximal cues used by organisms to make inferences about the state of the environment (Brunswik, 1949, 1956b) to the ‘ecological validity of experiments’ (Orne, 1962).

Conclusions and implications of Part 1

In **Part 1** of this dissertation, I have argued that the contemporary framework of the ‘real world approach’ and its emphasis on ‘ecological validity’ is based on several questionable assumptions, and eventually fails to address the more fundamental problem of generalizability and representativeness. Instead, a more constructive way to uncover the

context-specific and context-generic principles of human cognition and behavior requires one to try to specify the particular social and behavioral contexts in which one is interested. My arguments are based on several ideas from the history of psychology: Egon Brunswik's methodological program of *representative design* (1956b), Roger Barker's concept of *behavioral setting* (1968), and James Gibson's concept of *affordances* (1979/2014). What these ideas have in common is a specific emphasis on the environment and its relation to cognitive and behavioral functioning, albeit in different ways. What matters for the purpose of moving beyond the 'real-world vs. the lab'-discussion is that more emphasis on a description, or theory of the environment, is needed to adequately contextualize human cognition and behavior as a function of organism-environment relationships (Hammond & Stewart, 2001; Heft, 2001, 2013). For the field of social attention, I have proposed that many so-called 'real world' studies of social attention actually show that gaze behavior to others may vary substantially across different social situations. Thus, there is no such thing as 'real-world' social attention. Instead, in order to build a more sophisticated theory of 'social attention', researchers first need to specify the social context of human interaction which they are interested. I have argued that this will be a more constructive way to understand gaze behavior as a function of different social contexts, and thereby uncover the context-generic and context-specific aspects of social attention at large.

PART 2: USING EYE-TRACKING TO STUDY GAZE BEHAVIOR IN HUMAN INTERACTIONS

In the second part of this dissertation, I described three eye-tracking studies in which I investigated different aspects of gaze behavior in face-to-face interactions. From a large tradition of observational (ethological) studies on the role of gaze behavior in human social interaction (Argyle & Cook, 1976; Duncan & Fiske, 2015; Goffman, 1967a, 1967b; Kendon, 1967, 1990; Rossano et al., 2009), combined with recent eye-tracking studies (Gregory & Antolin, 2019; Hessels, 2020; Hessels et al., 2017; Hessels, Holleman, et al., 2018; Hessels et al., 2019; Ho et al., 2015; Laidlaw et al., 2011; Macdonald & Tatler, 2018; Rubo et al., 2020), it has become increasingly clear that to better understand the role of gaze behavior in (potential) social interactions, many aspects of the social context need to be considered. For example, people may look at each other differently when there is a potential for social interaction with others, or when they are engaged in actual interactions. Moreover, the social and behavioral context of human interactions, e.g., where people are in the world, what they are doing, and with whom, is likely to be

relevant for how humans look at and interact with others in different social situations. In this dissertation, I therefore decided to focus on different aspects of gaze behavior in face-to-face interactions in two different social contexts: 1) potential social interactions with a stranger (Chapter 4), and 2) face-to-face conversations between parents and their children (Chapter 5 and 6).

In **Chapter 4**, I presented an experiment designed to imply a potential social interaction with a stranger. The goal of this study was to investigate why people believed they were interacting with another person or not, and if participants' gaze behavior to the eyes was related to whether they believed the other person could see and interact with them. Half of the participants were instructed that they would be presented with a 'live' person and the other half were instructed that they would be presented with a pre-recorded video clip of a confederate. Prior to the experiment, a confederate walked into a concealed booth to suggest s/he was going to be at the 'other side' of the 'live' video-connection. This 'Turingesque' paradigm⁴ allowed us to compare how people look at each other when they believed that there was a potential for interaction or not. Analyses of participants' responses obtained after the experiment indicated that participants used a variety of cues to judge whether the person was 'live' or not. Some participants perceived the confederate's behavior as interactive and reciprocal, while others perceived the confederate's behavior as discontinuous and 'out of sync'. We found only partial evidence for the 'social risk hypothesis', which posits that people would be more likely to avoid each other's eyes when confronted with a stranger because direct gaze can be perceived as unpolite, uncanny, or even threatening (Gregory & Antolin, 2019; Laidlaw et al., 2011). We found that, although some participants looked less at the eyes of a stranger if they were instructed that the presentation was live, other participants still predominately looked at the eyes, regardless of the received instruction or their post-hoc beliefs. This chapter showed that social interaction can be implied but is much less straightforward as implying social presence. This 'Turingesque' experiment could be further developed to study the relation between participants' attitudes and beliefs about the context of potential social interactions and how such attitudes and beliefs influence gaze behavior to faces (Pfeiffer et al., 2011).

⁴ The mathematician and computer scientist Alan Turing (1912 – 1954) proposed a test to address the question "can machines think" (Turing, 1950). He wondered under what conditions humans would ascribe "human-like intelligence" to an agent. In his test, a human engages in a verbal interaction with another human and a computer and has to judge with whom he is interacting. In my version, participants had to judge whether they were interacting with a 'live' person or that they were presented with a pre-recorded videoclip of a person.

In **Chapter 5** and **Chapter 6**, I presented two studies in which I investigated gaze behavior to faces in parent-child conversations. I used a recently developed dual-eye tracking setup which allows for objective measurement of gaze behavior to faces of two interactors simultaneously in high spatial and temporal resolution (Hessels et al. 2017; 2018; 2019). My goal was to study different aspects of gaze behavior in parent-child interactions, with the potential to generate new insights into the role of gaze as part of the interactive dynamics of this particular social context. These chapters are informed by (early) observational studies on gaze behavior in face-to-face conversations (Duncan & Fiske, 2015; Kendon, 1967), developmental studies on parent and child behavior (Branje, 2018; Dunn & Slomkowski, 1992; Mastrotheodoros, Van der Graaff, Deković, Meeus, & Branje, 2020; Thomas et al., 2017), and modern eye-tracking studies on face scanning, speech perception, and turn-taking (Hessels et al., 2017; Hessels, Holleman, et al., 2018; Hessels et al., 2019; Peterson et al., 2016; S. L. Rogers et al., 2018).

In **Chapter 5**, I used a semi-structured conversation paradigm to engage parents and children in two brief conversations about common family topics of discussion. I provided detailed descriptions of how gaze and speech behavior during face-to-face conversations between parents and children were modulated by the topic of conversation and during episodes of speaking and listening. Parents spoke relatively more during the conflict conversation compared to the cooperation conversation, whereas children spoke relatively more during the cooperation conversation compared to the conflict conversation. Regarding patterns of gaze behavior to facial features, I found that parents looked relatively more at the eyes of the child during the conflict discussion compared with the cooperation conversation. We also found that both parents and children looked relatively more to the mouth region when the other was speaking compared to when they were speaking themselves. However, there were also considerable individual differences in participants' gaze behavior to facial features, regardless of the participant-role and topic of conversation. Both parents and children varied greatly in how much they looked at different regions of the face (eyes, nose, mouth). Some individuals looked primarily at the eyes or mouth region, while others gazed at different facial features more equally over the course of the conversations. We also found that these individual differences were consistent across the two conversations, suggesting that individuals also exhibit stable, idiosyncratic face scanning patterns in face-to-face interactions (Arizpe et al., 2017; Peterson et al., 2016; S. L. Rogers et al., 2018).

In **Chapter 6**, I investigated the relation between gaze and turn-taking behavior during parent-child conversations. Using a selection of the data presented in Chapter 5, here I focused on where parents and children looked at each other's faces during turn transitions, i.e., when speakers and listeners switch roles. I showed that more turn transitions occurred in the cooperation conversation compared to the conflict conversation. Also, both parents and children were generally faster in responding to the other in the cooperation conversation. Regarding gaze behavior to faces, we found that individuals differed substantially in what regions of the face were looked at most during turn transitions, irrespective of the transition direction or the topic of conversation. As in Chapter 5, I suggested that this may further reflect the individual variability of face scanning behavior. However, due to high proportions of data loss, especially for the children, I could not draw conclusions as to whether gaze to specific facial features was related to aspects of turn taking, i.e., signalling to and monitoring each other during when switching between speaking turns. One likely reason for the large proportions of data loss is that interlocutors may look at and away from each other at the end and beginning of turns (Duncan & Fiske, 2015; Kendon, 1967). If participants turned their heads away too much away from the screen, the eye tracker could no longer track participants' gaze position. Also, a gaze shift back towards the screen after the eye tracker loses track of the gaze position signal does not always coincide with a smooth and instantaneous recovery of the gaze position signal (Hessels et al., 2015; Niehorster et al., 2018). As such, large parts of the data were of insufficient quality to be adequately used for my analyses.

Conclusions and implications of Part 2

In **Part 2** of this dissertation, I presented three eye-tracking studies in which I investigated different aspects of gaze behavior in face-to-face interaction. I focused on two different social contexts: potential interactions with a stranger (Chapter 4) and parent-child conversations (Chapters 5 and 6). These studies show how eye-tracking technology provides a tool to objectively measure gaze behavior which can be used to investigate aspects of gaze and how these are related to visual perception, social attention, and face-to-face communication in different social situations. However, there are also many challenges and limitations of eye-tracking when applied to the context of human interaction (for a recent review, see Valtakari et al., 2021). For example, as was clear from Chapter 5 and Chapter 6, it was especially difficult to measure gaze behavior during parent-child conversations. A lot of eye tracking data was lost during the measurements and many recordings were lacking in precision and accuracy. Clearly, there is a lot of room

for technical improvement to achieve better measurements of gaze behavior during full-fledged face-to-face interactions.

Nevertheless, assuming that the technological and methodological limitations discussed above could be overcome, how could the eye-tracking paradigms that I introduce benefit future studies on the role of gaze behavior during parent-child interactions? Firstly, a more fine-grained analysis of parents' and children's gaze behavior could shed new light on some of the interpersonal dynamics of parent-child relationships. Observational techniques (i.e., manual coding from video-recordings) used to estimate gaze position may lack the reliability and precision to distinguish specific aspects of gaze in interaction, e.g., what regions of the face are looked at and how these are related with other behaviors, such as speaking, listening, and turn-taking. Secondly, many studies have relied heavily on self-reports to assess the parent-child relationship, e.g., parents' and children's perceptions about their relationship quality, the intensity and frequency of parent-child conflicts across different ages and stages of development, etc. While studies using self-reports have provided valuable insights into the general structure and relational changes of parent-child interactions (Branje, 2018; Mastrotheodoros et al., 2020; Smetana, 2011), such methods do not directly investigate the behavioral and interpersonal dynamics of parent-child interactions in the 'heat of the moment'. According to (Branje, 2018, pp. 172-173): *"Short-term fluctuations in behavior are often ignored or attributed to lower reliability and measurement error. However, since moment-to-moment interactions are the proximal engines of development... these fluctuations may be a key feature of relationships and relationship development... investigating variability during conflict interactions between parents and children may expose fundamental relational processes.* In my view, the use of dual eye-tracking technology, in combination with algorithms to classify face and pose, speech content, and voice affect, could be the key to further understanding parent-child interactions in terms of, for example, parents' and children's conversational style and conflict resolution strategies (Beaumont & Wagner, 2004; Dixon et al., 2008; Thomas et al., 2017), as well as individual and interpersonal differences in emotion regulation and social competence (Hutchinson et al., 2019; Junge et al., 2020; Moed et al., 2017; Moed et al., 2015; Speer et al., 2007; Woody et al., 2020).

Furthermore, the eye-tracking setups described in **Part 2** may be applied to study social and communicative impairments of individuals with autism spectrum disorder (ASD) and social anxiety disorder (SAD). In the case of ASD, it has often been suggested that individuals with ASD show deficits in social interaction and communication which

are characterized by a failure to initiate or respond to bids for social interaction, and exhibit abnormal nonverbal and verbal communication, such as a lack of eye-contact, or a failure to engage in normal turn-taking in conversations (DSM-5, American Psychiatric Association, 2013). Several eye-tracking studies have shown that adults, adolescents, school-aged children, and toddlers exhibit reduced looking time to the eyes (Hernandez et al., 2009; Jones & Klin, 2013; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Rice, Moriuchi, Jones, & Klin, 2012; Sterling et al., 2008). However, other eye-tracking studies using videos and images of people's faces have produced inconsistent results in reporting reduced looking time to the eyes (for a review, see Speer et al., 2007). One explanation for these inconsistencies is that studies with videos and images lack important aspects of a 'social context' (e.g., social presence and the potential to interact). As such, more interactive experimental setups thus provide a way to simulate such aspects in experimental situations to more closely investigate how and when the deficits in social interaction and communication of individuals with ASD occur. Likewise, interactive eye-tracking can be used to study individuals with social anxiety (SAD). Individuals with SAD may experience increased anxiety in the presence of other people and may exhibit 'inadequate eye contact' or 'gaze aversion' when interacting with other people (DSM-5, American Psychological Association, 2013). Several eye-tracking studies with images and videos of faces have shown that individuals with SAD look less at facial features, especially the eyes, compared to healthy individuals. Specifically, these studies have suggested that individuals with SAD look less often, and for shorter durations at the eyes (Horley, Williams, Gonsalvez, & Gordon, 2004; Moukheiber et al., 2010; Wieser, Pauli, Alpers, & Mühlberger, 2009).

Interactive eye-tracking setups could be used to test different models of abnormal gaze behavior related to ASD and SAD (Hessels, Holleman, et al., 2018; Kleberg et al., 2017; Kleberg, Högström, Sundström, Frick, & Serlachius, 2021). For example, the 'gaze indifference' model of ASD posits that the lack of eye contact (i.e., reduced looking time at the eyes) of individuals with ASD is because that these individuals are less interested in, or at least process differently the socio-emotional information conveyed by other people's faces. However, the 'gaze aversion' model of ASD suggest that individuals with ASD experience increased arousal when viewing faces, and thus actively avert their gaze away from other people's faces. In a similar vein, the 'vigilance-avoidance' model posits that individuals with social anxiety (SAD) are faster to gaze at the eyes ('vigilance'), but then quickly avert their gaze away from the eyes ('avoidance') because perceived eye contact is aversive to them. My colleagues and I have already used the dual eye-tracking setup

from Chapter 5 and Chapter 6 to investigate whether gaze behavior to facial features were related with traits of ASD and SAD in a constrained interactive setting, i.e., not having a conversation (Hessels, Holleman, et al., 2018). Also, we conducted this study with healthy adults from a general population, not with actual clinical samples. We found that traits of ASD and SAD were indeed differently related to traits of ASD and SAD, corroborating previous hypotheses of face scanning behavior in ASD and SAD. Future research on this topic could benefit from the eye-tracking experiments described in the current thesis. For example, the ‘Turingesque’ paradigm to imply social interaction (Chapter 4) could be used to contribute to resolving the inconsistencies about reduced looking time to eyes in the ASD literature. As discussed, one potential explanation for this inconsistency in studies with images and videos of faces is that such experiments lack important features of the social context. Using our experimental procedure (i.e., staging a ‘live’ social interaction), one could thus investigate whether individuals with ASD will exhibit reduced looking time at another person’s eyes (or not) dependent on whether they believe the other person is ‘live’ (or not). Furthermore, the ‘conflict and cooperation’ paradigm used in Chapter 4 and 5 may be effective to differentiate between the different models of autism and social anxiety, specifically ‘gaze aversion’ and ‘gaze indifference’. For example, one may expect that, if gaze avoidance in ASD and SAD individuals is best characterized by ‘gaze aversion’ (i.e., looking away to reduce arousal), this may be especially pronounced in conflict-like interactions, which are designed to elicit additional levels of arousal. However, in the case of ASD, if abnormal gaze is the result of ‘gaze indifference’, one would not necessarily expect that ‘conflict’ or ‘cooperation’ topics are related to how much individuals with ASD look at the eyes of the other person or not. Instead, one may expect more equal levels of gaze avoidance, irrespective of conversation topic.

FINAL CONCLUSION AND ADVICE

In this dissertation, I investigated the role of gaze behavior during face-to-face interactions using new eye-tracking paradigms. However, developing new eye-tracking experiments to study the role of gaze behavior during face-to-face interactions is not an easy task. One is faced with a host of theoretical, methodological, and technical problems. Yet, I hopefully have also shown how this technology may be used to study different aspects of gaze in human interaction and lead to potential applications in the fields of social attention, parent-child interaction, turn-taking, and clinical assessment. As such, I conclude by offering some general advice for future ‘interactive’ eye-tracking researchers based on the

past (more than) four years of thinking about these problems. I have argued that there is no such thing as ‘real world’ cognition and behavior. Human cognition and behavior are for a large part context-dependent. For the field of social attention, I have suggested that, instead of wondering whether a given experimental paradigm really captures ‘real-world’ social attention or a ‘real-world’ social interaction, one should first ask what is meant by these terms. As such, one could ask: what social context is relevant to study gaze behavior in interaction? What are the relevant interactive aspects of the social context, and how are these related with the behaviors of interest? How does one represent these interactive features into an experimental paradigm, while still making it technically feasible to obtain good measurements of gaze behavior? Finally, there are many technical limitations to eye-tracking. As such, one has to be realistic about the potential of this technology (Hessels, Niehorster, Holleman, Benjamins, & Hooge, 2020). There are trade-offs to be made between experimental control, quality of the recordings, and the representativeness of human behavior in experimental research.

Chapter 8

Nederlandse samenvatting

Een belangrijke vorm van menselijk sociaal gedrag is *face-to-face* interactie. Face-to-face interacties bestaan uit verschillende verbale en non-verbale gedragingen, bijvoorbeeld spraak, oogcontact, gezichtsuitdrukkingen, gebaren en lichaamshouding. Tezamen faciliteren en reguleren deze gedragingen interpersoonlijke communicatie (Hessels, 2020; Holler & Levinson, 2019; Kendon, 1967, 1990). Binnen het onderzoek naar ‘sociale aandacht’ (een onderzoeksgebied op het snijvlak tussen experimentele, cognitieve, en sociale psychologie) is een belangrijke bevinding dat menselijk kijkgedrag – waar men kijkt, hoe lang, en wanneer – niet alleen belangrijk is voor hoe mensen de wereld visueel waarnemen, maar dat het zichtbare kijkgedrag van een persoon ook een belangrijke bron van informatie kan zijn voor anderen. Zo kan iemands kijkrichting informatief zijn over waar die persoon zijn of haar aandacht op heeft gericht. Deze dubbele functie van kijkgedrag – zowel voor de visuele waarneming en als sociaal signaal naar anderen – is interessant omdat dit in bepaalde omgevingen tot ongemakkelijke situaties kan leiden. Stel je voor dat je in een trein tegenover een andere persoon zit. Door naar deze persoon te kijken kan je observeren wat deze persoon aan het doen is, maar het recht blijven aankijken van de persoon in kwestie kan door deze persoon ook worden opgevat als onbeleefd of zelfs als bedreigend (Dovidio & Ellyson, 1985; Hamilton, 2016). Zo beschrijft Goffman (2008) bijvoorbeeld hoe mensen in publieke ruimtes elkaar kort aankijken, als blijk van erkenning van de ander zijn aanwezigheid, om elkaar vervolgens ‘beleefd te negeren’ door elkaar niet meer aan te kijken. Dit fenomeen noemt Goffman ‘civil inattention’. Kijkgedrag is dus ook nauw verbonden met sociale normen.

Onderzoekers die ‘sociale aandacht’ bestuderen zijn onder andere geïnteresseerd in de sociale functie van kijkgedrag tijdens interacties met anderen. Zo is er onderzocht hoe kijkgedrag bepaalde aspecten van sociale interacties kan ondersteunen, zoals samen naar een object in de omgeving kijken (joint attention), en het wisselen van beurten tijdens een gesprek (turn-taking). Onderzoek naar ‘sociale aandacht’ wordt meestal gedaan met behulp van *experimentele paradigma’s*. Deze experimentele paradigma’s bestaan uit een set van specifieke procedures uitgevoerd in een testomgeving of laboratorium. In zo’n laboratorium worden proefpersonen bijvoorbeeld voor een computerscherm geplaatst waarop vervolgens visuele ‘stimuli’ worden gepresenteerd, zoals afbeeldingen of filmclips van menselijke gezichten. De proefpersonen worden voorafgaand aan het experiment geïnstrueerd om een taak uit te voeren, bijvoorbeeld om naar een serie verschillende gezichten te kijken, om aan te geven welke emoties de gezichten uitdrukken, of om zo snel mogelijk op een knop te drukken zodra ze een gezicht herkennen. Zie Figuur 1 voor een voorbeeld.

Figuur 1. Verschillende gezicht-stimuli die veel worden gebruikt in experimenten naar emotie-herkenning. Een proefpersoon krijgt plaatjes van gezichten te zien op een computerscherm en moet aangeven welke emoties er worden uitgedrukt. Met behulp van een eye tracker wordt er vervolgens bepaald waar de proefpersoon op het gezicht heeft gekeken en of patronen in kijkgedrag op delen van het gezicht emotie-afhankelijk zijn. Welke emoties denk jij dat er worden uitgedrukt?⁵.

Hoewel experimentele paradigma's zoals weergegeven in Figuur 1 veel zijn gebruikt om kijkgedrag en 'sociale aandacht' te onderzoeken, zijn er ook een aantal onderzoekers die stevige kritiek hebben geuit over het gebruik van zulke experimentele paradigma's (Kingstone, Smilek, Ristic, Friesen, & Eastwood, 2003; Kompatsiari et al., 2018; Risko et al., 2016; Schilbach et al., 2013; Shamay-Tsoory & Mendelsohn, 2019). Volgens deze onderzoekers zijn de procedures die gebruikt worden, zoals het laten kijken van proefpersonen naar afbeeldingen en filmclips van menselijke gezichten, niet geschikt om de sociale functie van kijkgedrag tijdens sociale interacties te onderzoeken. Als gevolg stellen deze onderzoekers dat de conclusies van dit soort onderzoek dan ook niet te generaliseren zijn naar hoe mensen naar elkaar kijken en met elkaar interacteren in de 'echte wereld'. In reactie hierop hebben onderzoekers gepoogd om sociale aandacht en kijkgedrag in de 'echte wereld' te onderzoeken, bijvoorbeeld door experimenten met meer 'ecologische validiteit' uit te voeren waarin er meer oog is voor de rol van de sociale context (Osborne-Crowley, 2020; Risko et al., 2012; Risko et al., 2016; Shamay-Tsoory & Mendelsohn, 2019; Smilek et al., 2006).

⁵ **Antwoorden bij Figuur 1:** A. Treurig. B. Blij. C. Verbaasd. D. Boos. E. Verveeld, of misschien toch te veel gedronken?

De discussie over de validiteit en generaliseerbaarheid van conclusies gebaseerd op de ‘conventionele’ laboratoriumexperimenten, en de daaropvolgende reactie om het lab te willen verlaten om in de ‘echte wereld’ onderzoek te doen, is echter allesbehalve uniek voor experimenteel onderzoek naar sociale aandacht. Deze zogenaamde ‘echte wereld of het lab’-discussie heeft een lange geschiedenis in de psychologie en heeft herhaaldelijk voor heftige discussies gezorgd over hoe wetenschappers het beste onderzoek kunnen doen naar menselijke cognitie en gedrag (Banaji & Crowder, 1989; Kingstone et al., 2008; Koehler, 1996; Neisser, 1976, 1991). Omdat ik niet alleen geïnteresseerd ben in de rol van kijkgedrag tijdens sociale interacties, maar ook in de theoretische en empirische grondslagen van de experimentele psychologie, bestaat dit proefschrift uit twee delen. In **Deel 1** ga ik dieper in op het theoretische probleem van het ‘echte wereld of het lab’-dilemma, en daarin specifiek het begrip *ecologische validiteit* en de verschillende interpretaties daarvan. Dit deel bestaat uit een conceptuele en historische analyse van het begrip ecologische validiteit binnen de psychologische wetenschapsliteratuur.

In **Hoofdstuk 2** beschrijf ik hoe het begrip ‘ecologische validiteit’ de afgelopen 70 jaar veelvuldig gebruikt is om uitspraken te doen over de mate waarin menselijk gedrag in een experimentele lab omgeving wel of niet overeenkomt met menselijk gedrag in de ‘echte wereld’ (Koehler, 1996; Neisser, 1976; Osborne-Crowley, 2020; Shamay-Tsoory & Mendelsohn, 2019; Smilek et al., 2006). Wat ik de ‘echte wereld aanpak’ heb genoemd, beschrijft de stelling van psychologen dat er experimenten met ‘meer ecologische validiteit’ nodig zijn om ervoor te zorgen dat de conclusies van experimenteel onderzoek te generaliseren zijn naar menselijk gedrag in de ‘echte wereld’. Ik beargumenteer dat het begrip ‘ecologische validiteit’ voor een hoop conceptuele verwarring heeft gezorgd. Het begrip ecologische validiteit wordt door wetenschappers veelal niet gedefinieerd, ondanks het veelvuldige gebruik van dit begrip in de wetenschappelijke literatuur. Verder lijkt er geen consensus te bestaan over specifieke criteria om te bepalen wat wel of niet ‘ecologisch valide’ is (Dunlosky et al., 2009; Schmuckler, 2001). In **Hoofdstuk 3** ga ik dieper in op de betekenis van het begrip ecologische validiteit, zoals uiteengezet door de Hongaars-Amerikaanse psycholoog Egon Brunswik (1903 – 1955). Net als Hammond (1998) concludeer ik dat Brunswik’s definitie van ecologische validiteit is verstrengeld geraakt met Brunswik’s ideeën over ‘representatief ontwerp’ (Brunswik, 1949, 1955, 1956a; Hammond & Stewart, 2001). Brunswik’s definitie gaat over de ‘ecologische validiteit van cues’ – de betrouwbaarheid van de beschikbare zintuigelijke informatie met betrekking tot de toestand van de wereld. Volgens Brunswik is de wereld zoals die waarneembaar is voor een organisme vol onzekerheden. Het organisme neemt de wereld niet ‘direct’ waar,

maar indirect met behulp van verschillende ‘cues’ die mogelijk bruikbare en betrouwbare informatie verschaffen over de toestand van de wereld. De verwarring van Brunswik’s definitie met de modernere interpretatie van de ‘ecologische validiteit van experimenten’ – de mate waarin een experiment lijkt op de ‘echte wereld’ – is te herleiden naar het werk van Martin T. Orne (Kihlstrom, 2021; Orne, 1962).

In **Deel 1** van dit proefschrift concludeer ik dat het moderne gebruik van het begrip ‘ecologische validiteit’, dat vaak wordt gebruikt om iets te zeggen over de mate waarin een experiment lijkt op de ‘echte wereld’, conceptueel en methodologisch inadequaat is om het probleem van representativiteit en generaliseerbaarheid te benaderen. In de woorden van Hammond (1998)⁶:

“Er bestaat natuurlijk niet zoiets als een “echte wereld”. De “echte wereld” wordt niet gedefinieerd en heeft geen specifieke eigenschappen. Termen zoals de “echte wereld” worden enkel gebruikt in de afwezigheid van een theorie over de relevante taak, omgeving of context. Als gevolg is er geen referentiekader op basis waarvan men uitspraken over generaliseerbaarheid kan doen”.

Het probleem van het gebruik van termen zoals de ‘echte wereld’ of het ‘echte leven’ is dus dat deze termen ervoor zorgen dat er geen specifieke criteria opgesteld worden met betrekking tot de randvoorwaarden van de generalisatie. Als gevolg zijn uitspraken over generaliseerbaarheid te breed om empirisch te toetsen. Een constructievere aanpak om de context-specifieke en context-generieke aspecten van menselijke cognitie en gedag te onderzoeken is om eerst een sociale context goed te beschrijven, om deze vervolgens te representeren in een experimentele context.

In **Deel 2** van dit proefschrift beschrijf ik twee eye tracking experimenten waarin ik verschillende aspecten van kijkgedrag onderzoek tijdens face-to-face interacties. Eye tracking is een techniek waarmee men de kijkrichting van mensen kan bepalen. Deze techniek is gebruikt om kijkgedrag in verschillende sociale situaties te onderzoeken, namelijk 1) kijkgedrag op gezichten in een gesimuleerde face-to-face interactie en 2) kijkgedrag op gezichten tijdens face-to-face gesprekken tussen ouders en hun kinderen.

Hoofdstuk 4 beschrijft een eye tracking experiment waarin er een potentiële sociale interactie wordt gesimuleerd. In deze studie werd er onderzocht of het mogelijk was

⁶ Citaat van Kenneth Hammond (1998) in het Engels: “There is, ofcourse, no such thing as a “real world”. It has been assigned no properties and no definition; it is simply used because of the absence of a theory of tasks or other environments, and thus does not responsibly offer a frame of reference for the generalization”.

om een face-to-face ontmoeting met een vreemdeling te simuleren, en of de subjectieve ervaring van interactie van de proefpersoon gerelateerd was aan het kijkgedrag naar gezichten, specifiek naar de ogen. Ook werd er onderzocht welke factoren bijdragen aan de subjectieve ervaring van sociale interactie. In het experiment werd de helft van de proefpersonen verteld dat ze een andere persoon via een 'live' videoverbinding zouden gaan ontmoeten (de 'live'-conditie). De andere helft werd uitgelegd dat ze een videoclip van een persoon zouden gaan ontmoeten (de 'videoclip'-conditie). Voorafgaand aan iedere sessie met een deelnemer liep er een van de proefleiders een kamertje in om de suggestie te wekken dat deze persoon ook deelnam aan het sociale interactie experiment. Met behulp van deze 'sociale Turing-test' onderzochten we waar op het gezicht van de ander mensen keken, of dit afhankelijk was van de uitleg die voorafgaand aan het experiment werd gegeven, en of dit afhankelijk was van of de deelnemers ook daadwerkelijk geloofden dat de andere persoon 'live' was. In werkelijkheid kregen alle deelnemers een videoclip van een persoon te zien, en was er dus geen 'echte' interactie. Volgens de 'social risk hypothesis' zijn mensen geneigd zijn om oogcontact met een vreemde te vermijden omdat dit als onbeleefd of bedreigend ervaren kan worden (Gregory & Antolin, 2019; Laidlaw et al., 2011). Gebaseerd op eerder onderzoek werd er dan ook verwacht dat mensen in de 'live'-conditie minder naar de ogen zouden kijken dan mensen in de 'video'-clip conditie. De resultaten van het gemiddelde kijkgedrag wezen uit dat er een groter aandeel van proefpersonen was in de 'live'-conditie die weinig naar de ogen keken vergeleken met proefpersonen in de 'videoclip'-conditie. Een aantal proefpersonen in de 'live'-conditie keken daarentegen bijna continue naar de ogen. Statistische analyses lieten zien dat het gemiddelde relatieve kijkgedrag naar de ogen niet significant verschilde tussen de twee condities. Verder wezen de analyses van de vragenlijsten uit dat proefpersonen verschillende antwoorden en redenties gaven over waarom ze dachten dat de andere persoon 'live' aanwezig was of niet. Sommige proefpersonen vonden dat het net leek alsof de persoon op het scherm met hen interacteerde, bijvoorbeeld omdat de persoon die ze zagen naar ze lachte, knipoogde, en ze aankeek op een manier die suggereerde dat de persoon op het scherm reageerde op hun eigen gedrag. Andere proefpersonen, die niet geloofden dat de persoon 'live' was, vonden juist dat de interactie niet synchroon verliep, en gaven aan dat het gedrag van de persoon op het scherm willekeurig leek te zijn, of juist overduidelijk geacteerd.

In **Hoofdstuk 5** en **Hoofdstuk 6** werd kijkgedrag tussen ouder en kind tijdens alledaagse gespreksituaties onderzocht. Voor dit onderzoek werd er een interactie-opstelling met eye trackers gebruikt waarmee het kijkgedrag van twee personen tegelijkertijd kon worden

gemeten (Hessels et al., 2017). Het doel van dit onderzoek was om nieuwe inzichten te verwerven in de rol van kijkgedrag tijdens deze specifieke sociale context. Deze studies bouwen voort op observationeel onderzoek naar de rol van kijkgedrag in face-to-face gesprekken (Duncan, 1972; Duncan & Fiske, 2015; Kendon, 1967, 1990), onderzoek naar de ontwikkeling van ouder-kind relaties (Branje, 2018; De Goede et al., 2009; Dunn & Slomkowski, 1992; Junge et al., 2020; Mastrotheodoros et al., 2020; Thomas et al., 2017), en moderne eye tracking experimenten over de rol van kijkgedrag tijdens gezichtsperceptie, spraakperceptie, en turn-taking (Hessels et al., 2019; Ho et al., 2015; Peterson et al., 2016; S. L. Rogers et al., 2018).

In **Hoofdstuk 5** werden er twee semigestructureerde gespreksituaties gebruikt waarin ouders en kinderen gesprekken moesten voeren over hun gezinsleven. In het eerste gesprek discussieerden ze over een onderwerp dat voor conflict zorgt, bijvoorbeeld, dat het kind beter moet luisteren, of z'n kamer moet opruimen. In het tweede gesprek moesten de ouder en het kind samen een feest plannen, bijvoorbeeld voor de verjaardag van het kind. Er werd onderzocht of kijkgedrag op gezichten tijdens deze gesprekken gerelateerd was aan de twee gespreksituaties: conflict en coöperatie. De resultaten wijzen uit dat ouders relatief meer praatten in het conflict-gesprek dan in het coöperatie-gesprek, en de kinderen praatten relatief meer in het coöperatie-gesprek dan in het conflict-gesprek. Verder bleek dat ouders meer naar de ogen van de kinderen keken in het conflict-gesprek ten opzichte van het coöperatie-gesprek. Ook bleek dat zowel ouders als kinderen meer naar de mond keken als de ander sprak dan wanneer ze zelf spraken. Echter, de individuele verschillen in spraak- en kijkgedrag waren substantieel. Sommige proefpersonen keken voornamelijk naar de ogen, de neus, of de mond, terwijl anderen hun blik gelijkmatiger naar verschillende delen van het gezicht richtten. Deze individuele verschillen waren consistent over de gesprekken, en deze bevindingen convergeren met de resultaten uit eerder onderzoek over idiosyncratische karakteristieken van kijkgedrag op gezichten (Arizpe et al., 2017; Peterson et al., 2016; S. L. Rogers et al., 2018).

In **Hoofdstuk 6** is onderzocht hoe kijkgedrag op gezichten gerelateerd is aan turn-taking (het wisselen van beurt tijdens gesprekken) tijdens ouder-kind interacties. De gegevens van een deel van de ouder-kind koppels van Hoofdstuk 5 zijn gebruikt om te onderzoeken waar op het gezicht ouders en kinderen kijken wanneer sprekers en luisteraars van rol wisselen. De resultaten wezen uit dat er meer beurtwisselingen plaatsvonden tijdens de coöperatie-gesprekken dan tijdens de conflict-gesprekken. Verder waren zowel de ouders als de kinderen gemiddeld sneller in het reageren op de ander tijdens een beurtwisseling

in het coöperatie-gesprek vergeleken met het conflict-gesprek. Met betrekking tot het kijkgedrag op delen van het gezicht tijdens beurtwisselingen vonden we dat er grote individuele verschillen waren in welke delen van het gezicht het meest bekeken werden. Deze patronen waren onafhankelijk van de transitie-richting (beurtwisselingen van ouder naar kind en vice versa) en gespreksonderwerp (conflict en coöperatie). Net als in Hoofdstuk 5 vonden we grote individuele verschillen in kijkgedrag op gezichten tijdens face-to-face interacties. Een probleem was echter dat er veel eye tracking data niet van voldoende kwaliteit waren tijdens de beurtwisselingen om te analyseren waar de ouders en kinderen op het gezicht van de ander keken. Dit was voornamelijk het geval bij de eye tracking data van kinderen. Een mogelijke oorzaak van dit probleem is dat tijdens de periodes waarin beurtwisselingen plaatsvonden, proefpersonen wegstijven van de ander nadat ze de beurt overnemen (Duncan & Fiske, 2015; Kendon, 1967). Als proefpersonen te veel wegstijven van het scherm kan de eye tracker niet langer de kijkrichting bepalen. Daarnaast kan een kijkbeweging terug naar de andere persoon (bijvoorbeeld aan het einde van een beurt) niet altijd meteen weer worden opgepikt door de eye tracker, waardoor er signaalverlies kan ontstaan (Hessels et al., 2015; Niehorster et al., 2018). Mede hierom konden er geen duidelijke conclusies getrokken worden over of het kijkgedrag op specifieke delen van het gezicht (bv. ogen, neus, mond) gerelateerd was aan aspecten van turn-taking, bijvoorbeeld, het signaleren en monitoren van aankomende beurtwisselingen.

In **Deel 2** van dit proefschrift is beschreven hoe eye tracking technologie gebruikt kan worden om kijkgedrag tijdens sociale interacties te onderzoeken. Echter, er zijn ook nog vele beperkingen aan deze technologie om kijkgedrag tijdens sociale interacties te onderzoeken (zie ook Valtakari et al. 2021 voor een uitgebreide bespreking). Uit Hoofdstuk 5 en 6 bleek bijvoorbeeld dat er veel eye tracking data van onvoldoende kwaliteit waren om antwoord te kunnen geven op de onderzoeksvragen, omdat er verschillende vormen van signaalverlies optraden tijdens de metingen. Deze problemen zijn onder andere te herleiden tot de relatie tussen bepaalde karakteristieken van het gedrag dat onderzocht werd (spreken, luisteren, beurtwisselingen), en de technische beperkingen van de eye tracking apparatuur.

Conclusie en advies

In dit proefschrift werd menselijk kijkgedrag tijdens face-to-face interacties onderzocht met behulp van eye tracking technologie. Het gebruik van eye tracking technologie om onderzoek te doen naar kijkgedrag tijdens sociale interacties gaat gepaard met verschillende theoretische, methodologische en technische problemen. Daarom sluit ik hier af met wat pragmatisch advies voor andere onderzoekers.

Er bestaat niet zoiets als ‘echte wereld’ cognitie en gedrag. Menselijk gedrag is grotendeels context afhankelijk. Vanwege de grote verscheidenheid aan potentiële sociale context waarin de mens zich kan bevinden zorgt een paraplu-term zoals ‘echte wereld’ aandacht of ‘echte wereld’ gedrag niet voor de benodigde specificatie om de context-specifieke en context-generieke aspecten van menselijk gedrag adequaat te beschrijven, laat staan om er experimenteel onderzoek naar te doen. Binnen het onderzoeksveld naar sociale aandacht en cognitie heb ik daarom beargumenteerd dat het niet zinvol is om altijd maar af te vragen of we met een experiment wel echt ‘echte wereld’ gedrag kunnen onderzoeken. In plaats daarvan zou men zich af kunnen vragen wat er precies bedoeld wordt met termen zoals ‘echte wereld’ aandacht, en ‘echte wereld’ interacties. Voor welke sociale context zijn bepaalde aspecten van (kijk)gedrag relevant om cognitief functioneren te onderzoeken? Welke relevante aspecten van de sociale context zouden gerelateerd kunnen zijn aan patronen van kijkgedrag en sociale aandacht? En hoe kan men vervolgens deze aspecten representeren in een experimentele context waarin het technisch haalbaar blijft om metingen van kijkgedrag te verrichten met voldoende kwaliteit om uiteindelijk de onderzoeksvragen te beantwoorden. Omdat er methodologische en technische beperkingen zijn aan de toepassing van eye tracking technologie om kijkgedrag tijdens sociale interacties te onderzoeken, zal men realistisch moeten zijn over de potentie van deze technologie in het beantwoorden van bepaalde onderzoeksvragen. Onderzoekers zullen afwegingen moeten maken tussen de mate van experimentele controle, de kwaliteit van de metingen, en de representativiteit van menselijk gedrag in experimenteel onderzoek.

Chapter 9

Dankwoord

Allereerst wil ik mijn promotoren Chantal Kemner en Maja Deković bedanken voor de mogelijkheid om vier jaar onderzoek te mogen doen bij YOUth en het Consortium on Individual Development. Ik heb enorm genoten van deze periode waarin ik samen heb kunnen werken met fantastische mensen en mezelf kon ontwikkelen als onderzoeker. Chantal, bedankt dat jij me de ruimte hebt gegeven om naast de werkzaamheden voor YOUth ook mijn eigen interesses te volgen. Je vertrouwde erop dat dit mooie resultaten zou opleveren. Ook wil ik je bedanken voor de mogelijkheid om aan te blijven als postdoc en nog een tijdje verder te mogen filosoferen over psychologie en wetenschap. Maja, hoewel je zelf vindt dat jouw rol beperkt is geweest tijdens mijn promotietraject, wil ik je hier hartelijk bedanken voor je enthousiasme, hulp en inspiratie bij het koppelen van het eye tracking onderzoek aan het ouder-kind onderzoek, en je hulp bij het schrijven van de artikelen voor Hoofdstuk 5 en Hoofdstuk 6.

Vervolgens wil ik mijn copromotoren Roy Hessels en Ignace Hooge bedanken. Jullie hebben mij samen de afgelopen (meer dan) vier jaar begeleid, opgeleid, afgericht, en daar waar nodig er doorheen gesleept. Het is een voorrecht om met jullie samen wetenschap te beoefenen en ik heb onwijs veel van jullie geleerd. Roy, jij dacht dat onderzoek doen wel eens iets voor mij zou kunnen zijn, en inderdaad, de iets uit de hand gelopen masterthesis bleek later slechts de proloog van het proefschrift dat nog zou volgen. Ik wil je bedanken voor de wijze waarop je me de afgelopen jaren uit de modder wist te trekken als ik het even niet meer zag zitten (bijvoorbeeld na de ATL op het vliegveld van Alicante). Daarnaast wil ik je ook bedanken voor alle leuke momenten naast het werk, van filosofische discussies en jamsessies tot mountainbike ritjes over de Utrechtse Heuvelrug. Ignace, ik wil je bedanken voor de wijze waarop je me altijd wist uit te dagen en te motiveren om het beste uit mezelf te halen. Daarnaast zorgde je ervoor dat na al het harde werk (bijvoorbeeld na het draaien van de experimenten, programmeren van dataverwerkers, en het schrijven en herschrijven van artikelen) er ook altijd ruimte werd gemaakt voor een flinke dosis gezelligheid om de behaalde prestaties te vieren. Meestal in de vorm van de befaamde bier & bbq-sessies bij jou en Jenny thuis. Samen met collega's en vrienden naar muziek luisteren, gekke biertjes proeven, en experimenteren met een arsenaal aan sauzen en sambal voor op de broodjes worst en hamburgers. Bedankt voor jullie begeleiding, wijze lessen, humor, geduld en vertrouwen.

Verder wil ik mijn collega's en oud-collega's van het Kinder Kennis Centrum en het YOUth onderzoek bedanken die mij hebben geholpen bij de opzet en logistiek van mijn experimenten: Lilli, Gwen, Juliette, Caroline, Jacco, Femke, Liset, Ivonne, Leon,

Alicia, Mark, Ron, Marije, Coosje, en Jacobine. Ook wil ik mijn co-auteurs Jorg en René bedanken voor hun hulp en samenwerking bij Hoofdstuk 5 en Hoofdstuk 6. Veel dank ook aan alle proefpersonen van mijn experimenten, met name de ouders en kinderen die hebben deelgenomen aan het onderzoek.

Vervolgens wil ik mijn andere collega's van de Universiteit Utrecht bedanken. Jeroen, bedankt voor jouw hulp en gezelligheid de afgelopen jaren. Samen naar congressen (ECEM, ECVP), samen naar concerten (Opeth, De Witte Buffel), en samen een paar uur staan verkleumen bij een fjord in Noorwegen bij -15° wachtend op de bus. Jan en Ans, bedankt voor alle inspirerende discussies over wetenschap en aanverwante zaken onder 't genot van een trappist met een kaasje of een knabbeltje. Hopelijk zetten we dit nog lang voort. Verder wil ik mijn collega-PhD kandidaten Yentl, Niilo, Bauke, Martijn, Remo, Paul, Zsafia, en Yun bedanken voor de gezellige en inspirerende momenten. Ook veel dank aan mijn oud-docenten tijdens de bachelor en master, Stella, Chris Paffen, Stefan, Albert, Joke en Maarten, voor jullie interesse in mijn academische ontwikkelingen en de leuke gesprekken die we de afgelopen jaren hebben gehad. Anouk, Krista, Martijn, Carlijn, Nico, Nathan, Gini, Annemarie, Sjoerd en Marnix, collega-docenten bij onder andere Introductie tot Cognitiewetenschappen en Experimenteren & Registreren I & II. Bedankt voor de prettige samenwerking tijdens het lesgeven. Leon, Chris Janssen en Chris Dijkerman (de Hellmen), bedankt voor de gezellige en muzikale momenten. Bij de volgende retraite gaan we hopelijk weer los! Verder ook dank aan alle andere collega's die ervoor zorgen dat de afdeling Psychologische Functieleer een fijne plek is om te werken.

Thanks to dr. Esther Kaufman, prof. dr. Mandeep Dhani and the Brunswik Society, for inviting me to present my work.

Dank aan Jessie Waalwijk en de organisatie van het Betweterfestival voor de mogelijkheid om in 2017 en 2018 een experiment te kunnen doen. Dank aan alle festivalgangers die meededen met onze experimenten.

Verder wil ik een aantal mensen in Utrecht en Tilburg bedanken. **Utrecht:** Thanks to David, for being my friend and musical companion, and for our many memorable adventures throughout the years. Ook dank aan Anne, Paul, en Max (en de Graaicultuur) waarmee menig pubquiz is verloren, en waarmee menig flesje whiskey soldaat is gemaakt tijdens onze pseudo-intellectuele proeverijen. Terrence en Vasilios, Super Saiyan and Greek Rock God, voor als je naar de next level wil gaan. Ook veel dank aan het illustere muziekgenootschap Bikavar: Miranda, Wim, John, Vincent, Elselien, Mario, Anja,

Annemiek, en Mariël, voor swings en de grooves in en rondom Utrecht de afgelopen jaren. **Tilburg:** Ik wil graag mijn familie bedanken. Mijn moeder Annabelle, vader Frederik, zus Roos, en broer Koen. Bedankt voor jullie liefde en steun. Ik hou heel veel van jullie! Patrice, Goof, Daan, Juup, ook van jullie! Tilburgse vrienden: Jules, fijn om je weer zo vaak te zien. Wanneer gaan we nou ook alweer precies stoppen met roken? Tinus, bedankt voor het behouden van ons contact over de jaren heen. The force is strong with this one. Goof, bedankt voor je vriendschap en broederschap. Die drie weken bij jou in Curaçao in 2019 waren precies wat ik nodig had.

En dan mijn twee grote liefdes, Frédérique en Livia. Frédérique, wie had gedacht dat zo'n avondje in de Voorstraat zo uit de hand zou lopen? Nu wonen we saampjes in een echt huis-huis in de Korvel in Tilburg, inclusief Liflafje, futuristische badkamerspiegel, epoxy-infused furniture, en een stroboscopische wolk-lamp. Lekker daten. Bedankt voor je liefde en steun in de afgelopen periode en voor het dealen met mijn eigenaardigheden. Livia, wat gezellig dat we nu met z'n drietjes ons eigen pinguïnnestje hebben in het nieuwe huisje. Ik kan niet wachten totdat ik weer met jou een rondje ga fietsen of samen ga stunts in het Spoorpark met step en skateboard. Zullen we daarna nog een ijsje eten?

Gijs Holleman

27 September 2021

Chapter 10

List of publications

In this dissertation, published:

Holleman, G. A., Hessels, R. S., Kemner, C., & Hooge, I. T. C. (2020). Implying social interaction and its influence on gaze behavior to the eyes. *PLoS one*, 15(2), 1-27.

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The ‘real-world approach’ and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, 11, 721.

Holleman, G. A., Hooge, I. T. C., Kemner, C., Hessels, R. S. (2021). The reality of “real-life” neuroscience: A commentary on Shamay-Tsoory and Mendelsohn (2019). *Perspectives on Psychological Science*, 16(2), 461–465.

In this dissertation, not published:

Holleman, G. A., Hooge, I. T. C., Huijding, J., Deković, M., Kemner, C., & Hessels, R. S. (in revision). Gaze and speech in parent-child interactions: The role of conflict and cooperation.

Holleman, G. A., Hooge, I. T. C., Huijding, J., Deković, M., Kager, R., Kemner, C., & Hessels, R. S. (in preparation). Gaze and turn-taking in parent-child conversations.

Not in this dissertation:

Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T.C., & Kemner, C. (2018). Eye contact takes two—autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology*, 9(2), jep-062917.

Hooge, I. T.C., Holleman, G. A., Haukes, N. C., & Hessels, R. S. (2019). Gaze tracking accuracy in humans: One eye is sometimes better than two. *Behavior Research Methods*, 51(6), 2712-2721.

Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T.C., & Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition*, 184, 28-43.

Hessels, R. S., van Doorn, A. J., Benjamins, J. S., Holleman, G. A., & Hooge, I. T. C. (2020). Task-related gaze control in human crowd navigation. *Attention, Perception, & Psychophysics*, 82(5), 2482-2501.

Hessels, R. S., Benjamins, J. S., van Doorn, A. J., Koenderink, J. J., Holleman, G. A., & Hooge, I. T. C. (2020). Looking behavior and potential human interactions during locomotion. *Journal of Vision*, 20(10), 5-5.

Hessels, R. S., Niehorster, D. C., Holleman, G. A., Benjamins, J. S., & Hooge, I. T. C. (2020). Wearable technology for “real-world research”: realistic or not? *Perception*, 49(6), 611-615.

Chapter 11

Curriculum Vitae

Gijs Holleman was born on May 22nd 1991 in Tilburg, the Netherlands. After graduating high school in 2009, he started his Psychology Bachelor at Utrecht University in 2010. During his Bachelor, he was a board member of *Brainwave*, a student association dedicated to organizing activities, workshops, and symposia on topics in cognitive, applied, and neuropsychology. During these years, he also worked in two forensic psychiatric hospitals (in

Utrecht and Almere) as an assistant in a music therapy program for patients. In 2015, he worked as a research-assistant at the Utrecht Medical Center in the Division of Psychiatry and Neuroscience. In 2015, he started his Master in Applied Cognitive Psychology and received his degree in 2016. After graduation, he worked at Triple (Alkmaar) as a data analyst and UX researcher, and as a research-assistant in the KinderKennisCentrum and YOUTh study (Utrecht). In 2017, he started his 4-year PhD-trajectory. Currently, Gijs is employed as a postdoctoral researcher at the Department of Experimental Psychology at Utrecht University.

Chapter 12

Bibliography

- Aanstoos, C. M. (1991). Experimental psychology and the challenge of real life. *American psychologist*, 1(46), 77.
- Adami, C. (2002). What is complexity? *BioEssays*, 24(12), 1085-1094.
- Adams, R. B., Ambady, N., Macrae, C. N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation and emotion*, 30(2), 177-186.
- Adams, R. B., & Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, 5(1), 3.
- Adolph, K. E. (2019). Ecological validity: mistaking the lab for real life. In R. Sternberg (Ed.), *My Biggest Research Mistake: Adventures and Misadventures in Psychological Research* (pp. 187-190). New York: Sage.
- Alderman, N., Burgess, P. W., Knight, C., & Henman, C. (2003). Ecological validity of a simplified version of the multiple errands shopping test. *Journal of the International Neuropsychological Society*, 9(1), 31-44.
- Ancel, A., Beaulieu, M., & Gilbert, C. (2013). The different breeding strategies of penguins: a review. *Comptes Rendus Biologies*, 336(1), 1-12.
- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science*, 8(1), 3-9.
- Araujo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology*, 19(1), 69-78.
- Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge, England: Cambridge University Press.
- Argyle, M., & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, 289-304.
- Arizpe, J., Walsh, V., Yovel, G., & Baker, C. I. (2017). The categories, frequencies, and stability of idiosyncratic eye-movement patterns to faces. *Vision research*, 141, 191-203. doi:10.1016/j.visres.2016.10.013
- Arnold, A., Semple, R. J., Beale, I., & Fletcher-Flinn, C. M. (2000). Eye contact in children's social interactions: What is normal behaviour? *Journal of Intellectual and Developmental Disability*, 25(3), 207-216.
- Ashcraft, M., & Radvansky, G. (2009). *Cognition* (5th ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.
- Ashear, V., & Snortum, J. R. (1971). Eye contact in children as a function of age, sex, social and intellectual variables. *Developmental psychology*, 4(3), 479.
- Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: An introductory review. *Child and Adolescent Mental Health*, 8(3), 136-143.
- Association, A. P. (2013). DSM 5. *American Psychiatric Association*, 70.
- Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2013). *Constrained local neural fields for robust facial landmark detection in the wild*. Paper presented at the Proceedings of the IEEE international conference on computer vision workshops.
- Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). *Openface: an open source facial behavior analysis toolkit*. Paper presented at the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV).
- Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American psychologist*, 44(9), 1185.
- Barker, R. G. (1966). On the nature of the environment. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik*. New York: Holt, Rinehart and Winston.
- Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, CA: Stanford University Press.

- 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. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(2), 241-251.
- Beattie, G. W., & Bogle, G. (1982). The reliability and validity of different video-recording techniques used for analysing gaze in dyadic interaction. *British Journal of Social Psychology*, 21(1), 34-35.
- Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cognition & emotion*, 28(3), 416-432.
- Beaumont, S. L., & Wagner, S. L. (2004). Adolescent-parent verbal conflict: The roles of conversational styles and disgust emotions. *Journal of language and social psychology*, 23(3), 338-368.
- 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.
- Berger, J., & Cunningham, C. C. (1981). The development of eye contact between mothers and normal versus Down's syndrome infants. *Developmental psychology*, 17(5), 678.
- Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American psychologist*, 37(3), 245.
- Bernstein, I. S. (1970). Primate status hierarchies. *Primate behavior: Developments in field and laboratory research*, 1, 71-109.
- Bindemann, M., Burton, A. M., Hooge, I. T. C., Jenkins, R., & De Haan, E. H. (2005). Faces retain attention. *Psychonomic Bulletin & Review*, 12(6), 1048-1053.
- Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision research*, 49(24), 2992-3000.
- Birmingham, E., Ristic, J., & Kingstone, A. (2012). Investigating social attention: A case for increasing stimulus complexity in the laboratory. *Cognitive neuroscience, development, and psychopathology: Typical and atypical developmental trajectories of attention*, 251-276.
- Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS One*, 3(8), e3022.
- Blanco-Elorrieta, E., & Pykkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in cognitive sciences*, 22(12), 1117-1126.
- Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, 10, 433-436.
- Brandi, M.-L., Kaifel, D., Lahnakoski, J. M., & Schilbach, L. (2019). A naturalistic paradigm simulating gaze-based social interactions for the investigation of social agency. *Behavior Research Methods*, 1-12.
- Branje, S. (2018). Development of parent-adolescent relationships: Conflict interactions as a mechanism of change. *Child Development Perspectives*, 12(3), 171-176.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American psychologist*, 32(7), 513.
- Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological review*, 50(3), 255.
- Brunswik, E. (1949). Remarks on functionalism in perception. *JOURNAL OF PERSONALITY*, 18, 56-65.
- Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological review*, 62(3), 193.
- Brunswik, E. (1956a). Historical and thematic relations of psychology to other sciences. *The Scientific Monthly*, 83(3), 151-161.

- Brunswick, E. (1956b). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, 2(1), 1-13.
- Buswell, G. T. (1935). *How people look at pictures: a study of the psychology and perception in art*. Chicago: Chicago University Press.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological bulletin*, 54(4), 297.
- Cañigueral, R., & Hamilton, A. F. d. C. (2019). Being watched: Effects of an audience on eye gaze and prosocial behaviour. *Acta psychologica*, 195, 50-63.
- Caro, T., & Stankowich, T. (2015). Concordance on zebra stripes: a comment on Larison et al.(2015). *Royal Society open science*, 2(9), 150323.
- Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and brain sciences*, 27(1), 79-96.
- Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017). Simulating social interactions for the experimental investigation of joint attention. *Neuroscience & Biobehavioral Reviews*, 74, 115-125.
- Chance, M. R. (1967). Attention structure as the basis of primate rank orders. *Man*, 2(4), 503-518.
- Cipolli, C., Sancini, M., Tuozi, G., Bolzani, R., Mutinelli, P., Flamigni, C., & Porcu, E. (1989). Gaze and eye-contact with anorexic adolescents. *British journal of medical psychology*, 62(4), 365-369.
- Cohen, G., & Conway, M. A. (2007). *Memory in the real world*: Psychology Press.
- Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, 55(2), 230-240.
- Danziger, K. (1994). *Constructing the subject: Historical origins of psychological research*: Cambridge University Press.
- Davids, K. (1988). Ecological validity in understanding sport performance: some problems of definition. *Quest*, 40(2), 126-136.
- De Goede, I. H., Branje, S. J., & Meeus, W. H. (2009). Developmental changes in adolescents' perceptions of relationships with their parents. *Journal of Youth and Adolescence*, 38(1), 75-88.
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in cognitive sciences*, 14(10), 441-447.
- De Waal, F. B. (1989). *Peacemaking among primates*: Harvard University Press.
- Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological bulletin*, 130(6), 959.
- Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., . . . Ding, M. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, 27(9), 1375-1380.
- Dixon, S. V., Graber, J. A., & Brooks-Gunn, J. (2008). The roles of respect for parental authority and parenting practices in parent-child conflict among African American, Latino, and European American families. *Journal of Family Psychology*, 22(1), 1.
- Dovidio, J. F., & Ellyson, S. L. (1985). Pattern of visual dominance behavior in humans. In *Power, dominance, and nonverbal behavior* (pp. 129-149): Springer.

- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, 6(5), 509-540.
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2), 283.
- Duncan, S., & Fiske, D. W. (2015). *Face-to-face interaction: Research, methods, and theory*: Routledge.
- Dunlosky, J., Bottiroli, S., & Hartwig, M. (2009). Sins committed in the name of ecological validity: A call for representative design in education science. In *Handbook of metacognition in education* (pp. 442-452): Routledge.
- Dunn, J., & Slomkowski, C. (1992). Conflict and the development of social understanding.
- Eaton, W. O., & Clore, G. L. (1975). Interracial imitation at a summer camp. *Journal of personality and social psychology*, 32(6), 1099.
- Edmonds, B. (1995). What is Complexity?-The philosophy of complexity per se with application to some examples in evolution. In *The evolution of complexity*: Kluwer, Dordrecht.
- Ehrlich, K. B., Richards, J. M., Lejuez, C., & Cassidy, J. (2016). When parents and adolescents disagree about disagreeing: Observed parent-adolescent communication predicts informant discrepancies about conflict. *Journal of research on adolescence*, 26(3), 380-389.
- Ekman, P. (1976). Pictures of facial affect. *Consulting Psychologists Press*.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings* (Vol. 11): Elsevier.
- Ellsworth, P. C., Carlsmith, J. M., & Henson, A. (1972). The stare as a stimulus to flight in human subjects: A series of field experiments. *Journal of personality and social psychology*, 21(3), 302.
- Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, 99(14), 9602-9605.
- Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532.
- Foddy, M. (1978). Patterns of gaze in cooperative and competitive negotiation. *Human Relations*, 31(11), 925-938.
- Foulsham, T., & Sanderson, L. A. (2013). Look who's talking? Sound changes gaze behaviour in a dynamic social scene. *Visual Cognition*, 21(7), 922-944.
- Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Res*, 51(17), 1920-1931. doi:10.1016/j.visres.2011.07.002
- Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant-caregiver social looking during locomotor free play. *Developmental science*, 21(4), e12626.
- Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, 110(2), 160-170.
- Fredericksen, R., Bex, P. J., & Verstraten, F. A. (1997). How big is a Gabor patch, and why should we care? *JOSA A*, 14(1), 1-12.
- Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *PLoS One*, 8(1), e53286. doi:10.1371/journal.pone.0053286
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3), 490-495.

- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4), 694.
- Gauvain, M. (2001). *The social context of cognitive development*: Guilford Press.
- Gell-Mann, M. (1995). What is complexity? Remarks on simplicity and complexity by the Nobel Prize-winning author of *The Quark and the Jaguar*. *Complexity*, 1(1), 16-19. doi:10.1002/cplx.6130010105
- George, N., Driver, J., & Dolan, R. J. (2001). Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing. *Neuroimage*, 13(6), 1102-1112.
- Gibson, J. J. (1950). *The perception of the visual world*. Cambridge, Boston: Houghton Mifflin Company, The Riverside Press.
- Gibson, J. J. (1970). On the relation between hallucination and perception. *Leonardo*, 3(4), 425-427.
- Gibson, J. J. (2014). *The ecological approach to visual perception: classic edition*. New York: Psychology Press, Taylor & Francis Group.
- Gillis, J., & Schneider, C. (1966). The historical preconditions of representative design. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 204-236). New York: Holt, Rinehart & Winston, Inc.
- Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, 14(5), 550-562.
- Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, 136, 359-364.
- Goffman, E. (1967a). Interaction ritual: Essays on face-to-face interaction.
- Goffman, E. (1967b). On face-work. *Interaction ritual*, 5-45.
- Goffman, E. (2008). *Behavior in public places*: Simon and Schuster.
- Goodwin, C. (1981). Conversational organization. *Interaction between speakers and hearers*.
- Goodwin, C. J. (2015). *A history of modern psychology* (5 ed.). Hoboken, New Jersey: John Wiley & Sons.
- Gravano, A., & Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25(3), 601-634.
- Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: a longitudinal study of gaze following during interactions with mothers and strangers. *Developmental science*, 13(6), 839-848.
- Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological bulletin*, 83(2), 314.
- Gregory, N. J., & Antolin, J. V. (2019). Does social presence or the potential for interaction reduce social gaze in online social scenarios? Introducing the "live lab" paradigm. *Quarterly Journal of Experimental Psychology*, 72(4), 779-791. doi:10.1177/1747021818772812
- Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of experimental social psychology*, 22(1), 38-77.
- Haensel, J. X., Danvers, M., Ishikawa, M., Itakura, S., Tucciarelli, R., Smith, T. J., & Senju, A. (2020). Culture modulates face scanning during dyadic social interactions. *Scientific reports*, 10(1), 1-11.
- Haensel, J. X., Smith, T. J., & Senju, A. (2017). Cultural differences in face scanning during live face-to-face interactions using head-mounted eye-tracking. *Journal of vision*, 17(10), 835-835.
- Hamilton, A. F. d. C. (2016). Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1686), 20150080.
- Hammond, K. R. (1998). Ecological validity: Then and now. Available online at: <http://www.brunswik.org/notes/essay2.html> (accessed April 1, 2020).

- Hammond, K. R., & Stewart, T. R. (2001). *The essential Brunswik: Beginnings, explications, applications*. New York: Oxford University Press.
- Hatfield, G. (2002). Psychology, philosophy, and cognitive science: Reflections on the history and philosophy of experimental psychology. *Mind & language*, 17(3), 207-232.
- Hefft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Hove, East Sussex: Psychology Press, Taylor & Francis Group.
- Hefft, H. (2013). An ecological approach to psychology. *Review of General Psychology*, 17(2), 162-167.
- Heinrichs, R. W. (1990). Current and emergent applications of neuropsychological assessment: Problems of validity and utility. *Professional Psychology: Research and Practice*, 21(3), 171.
- Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4), 555-568.
- Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Mem Cognit*, 33(1), 98-106.
- Hernandez, N., Metzger, A., Magné, R., Bonnet-Brilhaut, F., Roux, S., Barthelemy, C., & Martineau, J. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia*, 47(4), 1004-1012.
- Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review*, 1-26.
- Hessels, R. S., Benjamins, J. S., Cornelissen, T. H. W., & Hooge, I. T. C. (2018). A validation of automatically-generated Areas-of-Interest in videos of a face for eye-tracking research. *Frontiers in psychology*, 9, 1367.
- Hessels, R. S., Benjamins, J. S., van Doorn, A. J., Koenderink, J. J., Holleman, G. A., & Hooge, I. T. C. (2020). Looking behavior and potential human interactions during locomotion. *Journal of vision*, 20(10), 5-5.
- Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2017). Gaze behavior to faces during dyadic interaction. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 71(3), 226-242. doi:10.1037/cep0000113
- Hessels, R. S., Cornelissen, T. H. W., Kemner, C., & Hooge, I. T. C. (2015). Qualitative tests of remote eyetracker recovery and performance during head rotation. *Behavior Research Methods*, 47(3), 848-859.
- Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2018). Eye contact takes two—autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology*, 9(2), jep. 062917.
- Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T. C., & Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition*, 184, 28-43.
- Hessels, R. S., & Hooge, I. T. C. (2019). Eye tracking in developmental cognitive neuroscience—The good, the bad and the ugly. *Developmental cognitive neuroscience*, 40, 100710.
- Hessels, R. S., Kemner, C., van den Boomen, C., & Hooge, I. T. C. (2016). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods*, 48(4), 1694-1712.
- Hessels, R. S., Niehorster, D. C., Holleman, G. A., Benjamins, J. S., & Hooge, I. T. C. (2020). Wearable Technology for “Real-World Research”: Realistic or Not? In: SAGE Publications Sage UK: London, England.
- Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society open science*, 5(8), 180502.

- Hessels, R. S., van Doorn, A. J., Benjamins, J. S., Holleman, G. A., & Hooge, I. T. C. (2020). Task-related gaze control in human crowd navigation. *Attention, Perception, & Psychophysics*, 1-20.
- Higginson, C. I., Arnett, P. A., & Voss, W. D. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Clinical Neuropsychology*, 15(3), 185-204.
- Hilton, W. (1920). *Applied Psychology: Driving Power of Thought*. San Francisco: Applied Psychologist Press.
- Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: gaze signaling during dyadic interactions. *PLoS One*, 10(8), e0136905.
- Hoc, J.-M. (2001). Towards ecological validity of research in cognitive ergonomics. *Theoretical issues in ergonomics science*, 2(3), 278-288.
- Hogarth, R. M. (2005). The challenge of representative design in psychology and economics. *Journal of Economic Methodology*, 12(2), 253-263.
- Holleman, G. A., Hessels, R. S., Kemner, C., & Hooge, I. T. C. (2019). *Eye Tracking During Interactive Face Perception: Does Speech Affect Eye-Tracking Data Quality?* Paper presented at the *European Conference on Visual Perception [ECVP]*, Leuven, Belgium.
- Holleman, G. A., Hessels, R. S., Kemner, C., & Hooge, I. T. C. (2020). Implying social interaction and its influence on gaze behavior to the eyes. *PLoS One*, 15(2), e0229203.
- Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (in press). The reality of 'real-life' neuroscience: A commentary on Shamay-Tsoory & Mendelsohn. *Perspectives on Psychological Science*.
- Holleman, G. A., Hooge, I. T. C., Huijding, J., Deković, M., Kemner, C., & Hessels, R. S. (submitted). Gaze and speech in parent-child interactions: the role of conflict and cooperation.
- Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in cognitive sciences*, 23(8), 639-652.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Holmqvist, K., Nyström, M., & Mulvey, F. (2012). *Eye tracker data quality: what it is and how to measure it*. Paper presented at the Proceedings of the symposium on eye tracking research and applications.
- Hooge, I. T. C., & Camps, G. (2013). Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. *Frontiers in psychology*, 4, 996.
- Hooge, I. T. C., Holleman, G. A., Haukes, N. C., & Hessels, R. S. (2018). Gaze tracking accuracy in humans: One eye is sometimes better than two. *Behavior Research Methods*, 1-10.
- Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2004). Face to face: visual scanpath evidence for abnormal processing of facial expressions in social phobia. *Psychiatry research*, 127(1-2), 43-53.
- Hutchinson, E. A., Rosen, D., Allen, K., Price, R. B., Amole, M., & Silk, J. S. (2019). Adolescent gaze-directed attention during parent-child conflict: The effects of depressive symptoms and parent-child relationship quality. *Child Psychiatry & Human Development*, 50(3), 483-493.
- Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. *Neuropsychologia*, 45(5), 1019-1028.
- Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual review of psychology*, 68, 269-297.
- Jarick, M., & Kingstone, A. (2015). The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Frontiers in psychology*, 6, 1423.
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7(1), 2.

- Jenkins, J. J. (1974). Remember that old theory of memory? Well, forget it. *American psychologist*, 29(11), 785.
- Johnston, P., Molyneux, R., & Young, A. W. (2014). The N170 observed 'in the wild': robust event-related potentials to faces in cluttered dynamic visual scenes. *Social cognitive and affective neuroscience*, 10(7), 938-944.
- Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, 504(7480), 427-431.
- Junge, C., Valkenburg, P. M., Deković, M., & Branje, S. (2020). The building blocks of social competence: contributions of the Consortium of Individual Development. *Developmental cognitive neuroscience*, 45, 100861.
- Kanan, C., Bseiso, D. N., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision research*, 108, 67-76.
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological science*, 21(2), 260-267.
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, 26, 22-63.
- Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters* (Vol. 7): CUP Archive.
- Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3, 11. IT. 13.11.
- Kihlstrom, J. F. (2021). Ecological validity and "ecological validity". *Perspectives on Psychological Science*, 16(2), 466-471.
- Kingstone, A. (2009). Taking a real look at social attention. *Current opinion in neurobiology*, 19(1), 52-56.
- Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, 99(3), 317-340.
- Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science*, 12(5), 176-180.
- Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., & Feldman, R. (2017). Brain-to-brain synchrony during naturalistic social interactions. *Scientific reports*, 7(1), 17060.
- Kleberg, J. L., Högström, J., Nord, M., Bölte, S., Serlachius, E., & Falck-Ytter, T. (2017). Autistic traits and symptoms of social anxiety are differentially related to attention to others' eyes in social anxiety disorder. *Journal of autism and developmental disorders*, 47(12), 3814-3821.
- Kleberg, J. L., Högström, J., Sundström, K., Frick, A., & Serlachius, E. (2021). Delayed gaze shifts away from others' eyes in children and adolescents with social anxiety disorder. *Journal of Affective Disorders*, 278, 280-287.
- Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological bulletin*, 100(1), 78.
- Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, 59(9), 809-816.
- Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and brain sciences*, 19(1), 1-17.
- Kompatsiari, K., Ciardo, F., Tikhonoff, V., Metta, G., & Wykowska, A. (2018). On the role of eye contact in gaze cueing. *Scientific reports*, 8(1), 1-10.
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, 93(3), 480-490.

- Kruglanski, A. W. (1975). The two meanings of external invalidity. *Human Relations*, 28(7), 653-659.
- Kuhn, T. S. (2012). *The structure of scientific revolutions*: University of Chicago press.
- Laidlaw, K. E., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, 108(14), 5548-5553.
- Langton, S. R. (2000). The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology: Section A*, 53(3), 825-845.
- Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in cognitive sciences*, 4(2), 50-59.
- Lappi, O. (2015). Eye tracking in the wild: the good, the bad and the ugly. *Journal of Eye Movement Research*, 8(5).
- Larison, B., Harrigan, R. J., Thomassen, H. A., Rubenstein, D. I., Chan-Golston, A. M., Li, E., & Smith, T. B. (2015). How the zebra got its stripes: a problem with too many solutions. *Royal Society open science*, 2(1), 140452.
- Latané, B. (1981). The psychology of social impact. *American psychologist*, 36(4), 343.
- Laursen, B., & Collins, W. A. (2004). Parent-child communication during adolescence. *The Routledge handbook of family communication*, 2, 333-348.
- Laursen, B., & Collins, W. A. (2009). Parent-child relationships during adolescence.
- Laursen, B., & Hafen, C. A. (2010). Future directions in the study of close relationships: Conflict is bad (except when it's not). *Social Development*, 19(4), 858-872.
- Leppänen, J. M., & Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. *Nature Reviews Neuroscience*, 10(1), 37-47.
- Levine, M. H., & Sutton-Smith, B. (1973). Effects of age, sex, and task on visual behavior during dyadic interaction. *Developmental psychology*, 9(3), 400.
- Levinson, S. C. (2016). Turn-taking in human communication—origins and implications for language processing. *Trends in cognitive sciences*, 20(1), 6-14.
- Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130302.
- Lewkowicz, D. J. (2001). The Concept of Ecological Validity: What Are Its Limitations and Is It Bad to Be Invalid? *Infancy*, 2(4), 437-450.
- Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of vision*, 13(4), 6-6.
- Macdonald, R. G., & Tatler, B. W. (2018). Gaze in a real-world social interaction: A dual eye-tracking study. *Quarterly Journal of Experimental Psychology*, 71(10), 2162-2173.
- Mastrotheodoros, S., Van der Graaff, J., Deković, M., Meeus, W. H., & Branje, S. (2020). Parent-adolescent conflict across adolescence: trajectories of informant discrepancies and associations with personality types. *Journal of Youth and Adolescence*, 49(1), 119-135.
- Matlin, M. (1989). *Cognition*. New York: Rinehart and Winston.
- Mavratzakis, A., Herbert, C., & Walla, P. (2016). Emotional facial expressions evoke faster orienting responses, but weaker emotional responses at neural and behavioural levels compared to scenes: A simultaneous EEG and facial EMG study. *Neuroimage*, 124, 931-946.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748.
- Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of vision*, 14(7), 6-6.

- Mills, W. (1899). The nature of animal intelligence and the methods of investigating it. *Psychological review*, 6(3), 262.
- Mirenda, P. L., Donnellan, A. M., & Yoder, D. E. (1983). Gaze behavior: A new look at an old problem. *Journal of autism and developmental disorders*, 13(4), 397-409.
- Moed, A., Dix, T., Anderson, E. R., & Greene, S. M. (2017). Expressing negative emotions to children: Mothers' aversion sensitivity and children's adjustment. *Journal of Family Psychology*, 31(2), 224.
- Moed, A., Gershoff, E. T., Eisenberg, N., Hofer, C., Losoya, S., Spinrad, T. L., & Liew, J. (2015). Parent-adolescent conflict as sequences of reciprocal negative emotion: Links with conflict resolution and adolescents' behavior problems. *Journal of Youth and Adolescence*, 44(8), 1607-1622.
- Möglich, M., & Hölldobler, B. (1974). Social carrying behavior and division of labor during nest moving in ants. *Psyche*, 81(2), 219-236.
- Mook, D. G. (1983). In defense of external invalidity. *American psychologist*, 38(4), 379.
- Moukheiber, A., Rautureau, G., Perez-Diaz, F., Soussignan, R., Dubal, S., Jouvent, R., & Pelissolo, A. (2010). Gaze avoidance in social phobia: objective measure and correlates. *Behaviour research and therapy*, 48(2), 147-151.
- Myllyneva, A., Ranta, K., & Hietanen, J. K. (2015). Psychophysiological responses to eye contact in adolescents with social anxiety disorder. *Biological Psychology*, 109, 151-158.
- Nasiopoulos, E., Risko, E. F., & Kingstone, A. (2015). Social attention, social presence, and the dual function of gaze. In *The many faces of social attention* (pp. 129-155): Springer.
- Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: is it normal? Is it automatic? Is it social? *Development and psychopathology*, 20(1), 79-97.
- Neath, K. N., & Itier, R. J. (2014). Facial expression discrimination varies with presentation time but not with fixation on features: a backward masking study using eye-tracking. *Cognition & emotion*, 28(1), 115-131.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman and Company.
- Neisser, U. (1991). A case of misplaced nostalgia. *American psychologist*, 46(1), 34-36.
- Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, 50(1), 213-227.
- Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T. C., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, 1-21.
- Nyström, M., Andersson, R., Holmqvist, K., & Van De Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, 45(1), 272-288.
- Onland-Moret, N. C., Buizer-Voskamp, J. E., Albers, M. E., Brouwer, R. M., Buimer, E. E., Hessels, R. S., . . . Mandl, R. C. (2020). The YOUth study: rationale, Design, and study procedures. *Developmental cognitive neuroscience*, 100868.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist*, 17(11), 776.
- Osborne-Crowley, K. (2020). Social Cognition in the Real World: Reconnecting the Study of Social Cognition With Social Reality. *Review of General Psychology*, 1-15.
- Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in human neuroscience*, 9, 660.

- Parsons, T. D., Carlew, A. R., Magtoto, J., & Stonecipher, K. (2017). The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. *Neuropsychological rehabilitation*, 27(5), 777-807.
- Patterson, M. L. (1982). A sequential functional model of nonverbal exchange. *Psychological review*, 89(3), 231.
- Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends in cognitive sciences*, 18(5), 242-250.
- Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of autism and developmental disorders*, 32(4), 249-261.
- Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological science*, 24(7), 1216-1225.
- Peterson, M. F., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of vision*, 16(7), 12-12.
- Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., & Schilbach, L. (2011). A non-verbal turing test: differentiating mind from machine in gaze-based social interaction. *PLoS One*, 6(11), e27591.
- Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience & Biobehavioral Reviews*, 37(10), 2516-2528.
- Puce, A., & Bertenthal, B. I. (2015). *The Many Faces of Social Attention: Behavioral and Neural Measures* (A. Puce & B. I. Bertenthal Eds.). Switzerland: Springer.
- Reindl, V., Gerloff, C., Scharke, W., & Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *Neuroimage*, 178, 493-502.
- Rice, K., Moriuchi, J. M., Jones, W., & Klin, A. (2012). Parsing heterogeneity in autism spectrum disorders: visual scanning of dynamic social scenes in school-aged children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(3), 238-248.
- Richter, J. N., Hochner, B., & Kuba, M. J. (2016). Pull or push? Octopuses solve a puzzle problem. *PLoS One*, 11(3), e0152048.
- Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Attention Perception Psychophysics*, 73(2), 291-296. doi:10.3758/s13414-010-0042-1
- Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Front Hum Neurosci*, 6, 143. doi:10.3389/fnhum.2012.00143
- Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. *Current Directions in Psychological Science*, 25(1), 70-74.
- Rogers, S. D., Kadar, E. E., & Costall, A. (2005). Gaze patterns in the visual control of straight-road driving and braking as a function of speed and expertise. *Ecological Psychology*, 17(1), 19-38.
- Rogers, S. L., Speelman, C. P., Guidetti, O., & Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific reports*, 8(1), 1-9.
- Rossano, F., Brown, P., & Levinson, S. C. (2009). Gaze, questioning and culture. *Conversation analysis: Comparative perspectives*, 27, 187-249.
- Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2017). Beyond differences in means: robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*, 46(2), 1738-1748.

- Rubo, M., & Gamer, M. (2018). *Virtual reality as a proxy for real-life social attention?* Paper presented at the Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications.
- Rubo, M., Huestegge, L., & Gamer, M. (2020). Social anxiety modulates visual exploration in real life—but not in the laboratory. *British Journal of Psychology*, 111(2), 233-245.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7-55): Elsevier.
- Schilbach, L. (2015). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, 3, 130-135.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and brain sciences*, 36(4), 393-414.
- Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, 2(4), 419-436.
- Schofield, T. J., Parke, R. D., Castaneda, E. K., & Coltrane, S. (2008). Patterns of gaze between parents and children in European American and Mexican American families. *Journal of Nonverbal Behavior*, 32(3), 171-186.
- Scott, S., Briskman, J., Woolgar, M., Humayun, S., & O'Connor, T. G. (2011). Attachment in adolescence: Overlap with parenting and unique prediction of behavioural adjustment. *Journal of Child Psychology and Psychiatry*, 52(10), 1052-1062.
- Senju, A., Hasegawa, T., & Tojo, Y. (2005). Does perceived direct gaze boost detection in adults and children with and without autism? The stare-in-the-crowd effect revisited. *Visual Cognition*, 12(8), 1474-1496.
- Senju, A., & Johnson, M. H. (2009a). Atypical eye contact in autism: models, mechanisms and development. *Neuroscience & Biobehavioral Reviews*, 33(8), 1204-1214.
- Senju, A., & Johnson, M. H. (2009b). The eye contact effect: mechanisms and development. *Trends in cognitive sciences*, 13(3), 127-134.
- Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114(2), 727-741.
- Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-Life Neuroscience: An Ecological Approach to Brain and Behavior Research. *Perspectives on Psychological Science*, 1745691619856350.
- Silverstein, C. H., & Stang, D. J. (1976). Seating position and interaction in triads: A field study. *Sociometry*.
- Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, 5(4), 644-649.
- Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123-1128.
- Smetana, J. G. (2011). Adolescents' social reasoning and relationships with parents: Conflicts and coordinations within and across domains. *Adolescent vulnerabilities and opportunities: Constructivist and developmental perspectives*, 139-158.
- Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006). Cognitive ethology and exploring attention in real-world scenes. *Brain Research*, 1080(1), 101-119.
- Smith, P. W., Feinberg, R. A., & Burns, D. J. (1998). An examination of classical conditioning principles in an ecologically valid advertising context. *Journal of Marketing Theory and Practice*, 6(1), 63-72.
- Smokowski, P. R., Bacallao, M. L., Cotter, K. L., & Evans, C. B. (2015). The effects of positive and negative parenting practices on adolescent mental health outcomes in a multicultural sample of rural youth. *Child Psychiatry & Human Development*, 46(3), 333-345.

- Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. *Trends in cognitive sciences*, 23(8), 699-714.
- Speer, L. L., Cook, A. E., McMahon, W. M., & Clark, E. (2007). Face processing in children with autism: Effects of stimulus contents and type. *Autism*, 11(3), 265-277.
- Sprengelmeyer, R., Rausch, M., Eysel, U. T., & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1409), 1927-1931.
- Steinberg, L. (2001). We know some things: Parent-adolescent relationships in retrospect and prospect. *Journal of research on adolescence*, 11(1), 1-19.
- Stephenson, L. J., Edwards, S. G., & Bayliss, A. P. (2021). From Gaze Perception to Social Cognition: The Shared-Attention System. *Perspectives on Psychological Science*, 1745691620953773.
- Sterling, L., Dawson, G., Webb, S., Murias, M., Munson, J., Panagiotides, H., & Aylward, E. (2008). The role of face familiarity in eye tracking of faces by individuals with autism spectrum disorders. *Journal of autism and developmental disorders*, 38(9), 1666-1675.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Yoon, K.-E. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587-10592.
- Stoffregen, T. A. (2003). Affordances as properties of the animal-environment system. *Ecological Psychology*, 15(2), 115-134.
- Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of graduate medical education*, 4(3), 279-282.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, 26(2), 212-215.
- Team, J. (2018). JASP (version 0.9)[computer software]. In.
- Ten Bosch, L., Oostdijk, N., & Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, 47(1-2), 80-86.
- Terburg, D., Hooiveld, N., Aarts, H., Kenemans, J. L., & van Honk, J. (2011). Eye tracking unconscious face-to-face confrontations: Dominance motives prolong gaze to masked angry faces. *Psychological science*, 22(3), 314-319.
- Thomas, S. A., Wilson, T., Jain, A., Deros, D. E., Um, M., Hurwitz, J., . . . Dunn, E. J. (2017). Toward Developing Laboratory-Based Parent-Adolescent Conflict Discussion Tasks that Consistently Elicit Adolescent Conflict-Related Stress Responses: Support from Physiology and Observed Behavior. *Journal of child and family studies*, 26(12), 3288-3302.
- Thorndike, E. (1899). A reply to "The nature of animal intelligence and the methods of investigating it". *Psychological review*, 6(4), 412-420.
- Thorndike, E. (2017). *Animal intelligence: Experimental studies*: Routledge.
- Timney, B., & London, H. (1973). Body language concomitants of persuasiveness and persuasibility in dyadic interaction. *International Journal of Group Tensions*.
- Tolman, E. C., & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological review*, 42(1), 43.
- Tucker, C. J., McHale, S. M., & Crouter, A. C. (2003). Conflict resolution: Links with adolescents' family relationships and individual well-being. *Journal of Family Issues*, 24(6), 715-736.

- Turing, A. M. (1950). Computing machinery and intelligence. In *Parsing the turing test* (pp. 23-65): Springer.
- Valenti, S. S., & Gold, J. M. (1991). Social affordances and interaction I: Introduction. *Ecological Psychology*, 3(2), 77-98.
- Valtakari, N. V., Hooge, I. T., Benjamins, J. S., & Keizer, A. (2019). An eye-tracking approach to Autonomous sensory meridian response (ASMR): The physiology and nature of tingles in relation to the pupil. *PLoS One*, 14(12).
- Valtakari, N. V., Hooge, I. T., Viktorsson, C., Nyström, P., Falck-Ytter, T., & Hessels, R. S. (2021). Eye tracking in human interaction: Possibilities and limitations. *Behavior Research Methods*, 1-17.
- Van Hooff, J. (1967). The facial displays of the catarrhine monkeys and apes.
- Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisualspeech perception. *Perception & Psychophysics*, 60(6), 926-940.
- Võ, M. L.-H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of vision*, 12(13), 3-3.
- Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception*, 6(3), 313-326.
- Wieser, M. J., Pauli, P., Alpers, G. W., & Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. *Journal of anxiety disorders*, 23(1), 93-103.
- Wilson, B. A. (1993). Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Applied and Preventive Psychology*, 2(4), 209-215.
- Winograd, E. (1988). Continuities between ecological and laboratory approaches to memory. *Remembering reconsidered: Ecological and traditional approaches to the study of memory*, 11-20.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: a professional framework for multimodality research*. Paper presented at the 5th International Conference on Language Resources and Evaluation (LREC 2006).
- Woody, M. L., Price, R. B., Amole, M., Hutchinson, E., Benoit Allen, K., & Silk, J. S. (2020). Using mobile eye-tracking technology to examine adolescent daughters' attention to maternal affect during a conflict discussion. *Developmental Psychobiology*.
- Wu, D. W.-L., Bischof, W. F., & Kingstone, A. (2013). Looking while eating: The importance of social context to social attention. *Scientific reports*, 3, 2356.
- Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and vision* (pp. 171-211): Springer.
- Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, 26(9), 1235-1240.
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149(3681), 269-274.
- Zhang, X., Dalmasso, M., Castelli, L., Fiorese, A., Lan, Y., Sun, B., . . . Galfano, G. (2020). Social attention across borders: A cross-cultural investigation of gaze cueing elicited by same-and other-ethnicity faces. *British Journal of Psychology*.
- Zuckerman, M., Miserandino, M., & Bernieri, F. (1983). Civil inattention exists—in elevators. *Personality and Social Psychology Bulletin*, 9(4), 578-586.