

Mobile Eye Tracking

EDITED BY

Elisabeth Zima
Anja Stukenbrock



John Benjamins Publishing Company

Mobile Eye Tracking

Pragmatics & Beyond New Series (P&BNS)

ISSN 0922-842X

Pragmatics & Beyond New Series is a continuation of *Pragmatics & Beyond* and its Companion Series. The New Series offers a selection of high quality work covering the full richness of Pragmatics as an interdisciplinary field, within language sciences.

For an overview of all books published in this series, please see
benjamins.com/catalog/pbns

Editor

Miriam A. Locher
University of Basel

Associate Editor

Anita Fetzer
University of Augsburg

Founding Editors

Jacob L. Mey
University of Southern
Denmark

Herman Parret
Belgian National Science
Foundation, Universities of
Louvain and Antwerp

Jef Verschueren
Belgian National Science
Foundation,
University of Antwerp

Editorial Board

Fabienne H. Baider
University of Cyprus

Piotr Cap
University of Lodz

Pilar Garcés-Conejos
Blitvich
University of North Carolina at
Charlotte

Elly Ifantidou
National and Kapodistrian
University of Athens

Andreas H. Jucker
University of Zurich

Kuniyoshi Kataoka
Aichi University

Daniela Landert
Ruprecht-Karls-Universität
Heidelberg

Nuria Lorenzo-Dus
Swansea University

Sophia Marmaridou
University of Athens

Srikant Sarangi
Aalborg University

Marina Sbisà
University of Trieste

Elizabeth Stokoe
The London School of
Economics & Political Science

Sanna-Kaisa Tanskanen
University of Helsinki

Chaoqun Xie
Zhejiang International Studies
University

Yunxia Zhu
The University of Queensland

Volume 351

Mobile Eye Tracking. New avenues for the study of gaze in social interaction
Edited by Elisabeth Zima and Anja Stukenbrock

Mobile Eye Tracking

New avenues for the study of gaze
in social interaction

Edited by

Elisabeth Zima

University of Freiburg

Anja Stukenbrock

University of Heidelberg

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/pbns.351

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2025001890 (PRINT) / 2025001891 (E-BOOK)

ISBN 978 90 272 1993 0 (HB)
ISBN 978 90 272 4492 5 (E-BOOK)

© 2025 – John Benjamins B.V.

This e-book is Open Access under a CC BY 4.0 license.
<https://creativecommons.org/licenses/by/4.0/>

This license permits reuse, distribution and reproduction in any medium, provided that the original author(s) and source are credited.

This work may contain content reproduced under license from third parties. Permission to reproduce this third-party content must be obtained from these third parties directly.

Permission for any reuse beyond the scope of this license must be obtained from John Benjamins Publishing Company, rights@benjamins.nl

John Benjamins Publishing Company · <https://benjamins.com>

Table of contents

CHAPTER 1. Introduction: Mobile eye tracking for the study of gaze in social interaction	1
<i>Anja Stukenbrock & Elisabeth Zima</i>	
PART 1. Methodological considerations on the use of mobile eye tracking to study gaze in social interaction	
CHAPTER 2. Why research on gaze in social interaction needs mobile eye tracking	24
<i>Elisabeth Zima, Peter Auer & Christoph Rühlemann</i>	
CHAPTER 3. The influence of the specificities of gaze behavior on emerging and ensuing interaction: A contribution to the discussion of the use of eye-tracking recordings for EMCA analysis	67
<i>Gitte Rasmussen & Elisabeth Dalby Kristiansen</i>	
CHAPTER 4. Mobile eye-tracking and mixed-methods approaches to interaction analysis	100
<i>Bert Oben, Clarissa de Vries & Geert Brône</i>	
PART 2. Exploring interactional phenomena with mobile eye tracking	
Stationary settings	
CHAPTER 5. On the relationship between gaze and the German recipient token <i>hm_hm</i>	132
<i>Johanna Masuch</i>	
CHAPTER 6. Gaze aversion as a marker of disalignment in interactions	165
<i>Maximilian Krug</i>	
CHAPTER 7. Pupil size indicates planning effort at turn transitions in natural conversation	188
<i>Mathias Barthel & Christoph Rühlemann</i>	

Mobile settings

CHAPTER 8. Laughter and gaze among talkers on a walk	208
<i>Peter Auer & Barbara Laner</i>	
CHAPTER 9. When the establishment of joint attention becomes problematic: How participants manage divergent and competing foci of attention	243
<i>Anja Stukenbrock & Angeliki Balantani</i>	
CHAPTER 10. Joint attention without language? On intersubjectivity and the joint experience of nature	277
<i>Kerstin Botsch, Peter Auer, Barbara Laner & Martin Pfeiffer</i>	
Appendix A	311
Appendix B	314
Index	315

Introduction

Mobile eye tracking for the study of gaze in social interaction

Anja Stukenbrock & Elisabeth Zima

University of Heidelberg | University of Freiburg

Over decades, pragmatics has expanded into a broad, multi-faceted research field that nowadays encompasses enormously diversified research traditions and methods (Jucker et al. 2018) — the broadest consensus of those quite heterogeneous approaches being that the object of study is language use. Pragmatics is grounded in an understanding of language as social action; it explores how language use contributes to the pursuit of communicative goals, the constitution of interpersonal relations (Locher & Graham 2010), the achievement of joint attention and mutual understanding (Clark 1996) in the myriad of contexts that constitute human sociality. It investigates how verbal actions are designed, how they draw on and reflexively bring forth the context within which they occur, including the historical reconstruction of language practices and the socio-cultural arenas of their use (Jucker & Taavitsainen 2010).

While already pioneering researchers such as Gregory Bateson, Margret Mead, Ray Birdwhistell, and Erving Goffman, among others, were keenly aware of the embodied, situated nature of language use and paid close attention to the temporal details of nonverbal behaviour, the institutional canonization of pragmatics initially took a different route. Following the seminal works of Karl Bühler, Emile Benveniste, John Austin, John Searle, Herbert Paul Grice etc., theories, concepts and topics such as deixis, speech acts, maxims of conversation, implicature, presupposition, relevance, etc., became canonical in pragmatics and were for a long time approached with predominantly theoretical interest. It was only with the upsurge of empirical research that the intrinsically embodied nature of human communication in its primordial habitat in face-to-face interaction was beginning to be reappraised. The present volume is inscribed in this empirical research tradition that, based on the systematic collection of audio- and/or videorecorded data, examines human verbal action and interaction as multimodal phenomena in copresent embodied configurations (Goffman 1963; Kendon 1967, 1990). Most notably, it aligns with the research tradition of embodied interaction analysis (Streeck,

Goodwin & LeBaron 2011) and its insistence on temporality and sequentiality (Deppermann & Streeck 2018) as fundamental organizing principles of social interaction. In light of the intrinsically multimodal nature of human interaction, Deppermann (2015: 323f.) argues that modern pragmatic approaches need to take into account the following four core dimensions: embodiment, temporality, sociality, and epistemicity. The widening of the field of pragmatics to incorporate these four dimensions leads to an understanding of the disciplines “as the study of the verbal-embodied actions of sociohistoric subjects in space and time”¹ (ibid: 327).

The chapters of this edited volume subscribe to this view of pragmatics in studying the fine-grained details of how human beings cooperate and achieve intersubjectivity in interaction. While the authors all show that and how participants use the full array of linguistic and embodied resources in the course of this process, special emphasis is given to the role of gaze. Most importantly, the studies in this volume converge methodologically in the use of mobile eye tracking as a state-of-the-art technology to explore the role of gaze in naturally occurring social interaction.

Recent years have seen the publication of a steadily growing number of studies that draw on mobile eye tracking to investigate gaze in interaction (Auer 2018, 2021a, b, Auer et al. 2024, Balantani 2022, Balantani & Lázaro 2021, Holler & Kendrick 2015, Kendrick & Holler 2017, Kristiansen & Rasmussen 2021, Krug 2020, 2022, Oben 2018, Oben & Brône 2016, Pfeiffer & Weiß 2022, Rühlemann 2022, Stukenbrock 2014a, 2018a, b, 2020, Stukenbrock & Dao 2019, Weiß 2019, 2020, Zima 2020, Zima et al. 2019). They are complimented by a couple of recent monographs and technical introductions to eye tracking research (Holmqvist et al. 2011, Duchowski 2017, Attardo & Pickering 2022) and, most notably, the hitherto only collective volume on eye tracking in interaction analysis by Brône & Oben (2018). In sum, they testify to the growing interest in the technology and its wide range of applications. The present volume adds to the body of research on eye tracking in linguistic and interaction research by focussing on the applicability of mobile eye tracking to analyse gaze “in the wild” (Stukenbrock 2018a), that is, in authentic and spontaneous interaction. It provides a collection of studies specifically dedicated to the methodological and analytical challenges and benefits of using mobile eye tracking data for multimodal interaction analysis.

1. The German original citation (and its wider context) is: “Ich plädiere dafür, vier Bestimmungsstücke ins Zentrum der Auffassung von ‚Pragmatik‘ zu rücken, die traditionell nicht als zentrale Aspekte von ‚Pragmatik‘ gesehen wurden: Leiblichkeit, Zeitlichkeit, Sozialität und Epistemizität. Zusammengenommen führen sie zu einem Verständnis von Pragmatik als der Wissenschaft vom sprachlich-leiblichen Handeln von soziohistorischen Subjekten in Raum und Zeit.” (Deppermann 2015: 327).

The growing interest in eye tracking and gaze in interaction is embedded in the more general development towards a growing awareness of the embodied nature of human communication, which is reflected in the rapidly increasing number of studies in Multimodal Conversation Analysis (Mondada 2019). Driven by the insight that verbally encoded information is not delivered independently of other semiotic modes, but as an integral part of multimodal units or Gestalts (Enfield 2009, Mondada 2014, Holler & Levinson 2019), many recent studies pay attention to gaze, either in passing or as an explicit research topic, building on and expanding the vast amount of research that illustrates the core role of gaze for, e.g., regulating turn-taking (for a research overview, see Degutyte & Astell 2021), negotiating participation (e.g., Rossano 2012a), coordinating joint action, and establishing joint attention (e.g., Stukenbrock 2015, 2018a, b, 2023). The following section briefly reviews the most pertinent findings in these research areas before zooming in on the unique potential of mobile eye tracking for the study of gaze in social interaction.

1. Gaze in social interaction: A multifunctional resource

Research on the role of gaze for interaction management dates back to early pioneering work by Kendon (1967), Duncan (1972), Argyle & Cook (1976), and Goodwin (1980, 1981, 1984). They presented the first systematic accounts of the role of speakers' and hearers' gaze for turn-taking and the negotiation of participation (Goffman 1981, 1986). Most importantly, they revealed that the gaze behaviour of speakers and recipients differs significantly. Whereas recipients tend to gaze at the current speaker, speakers usually shift gaze to and away from the recipient(s). Most notably, these gaze shifts are not random but tightly linked to interactional tasks and cognitive constraints.

In a similar vein, studying video recordings of dyadic interactions, Kendon (1967) and Duncan (1972) observed that speakers usually avert their gaze at the beginning of a turn, but shift it back to their interlocutor at its end. Both authors interpret this gaze pattern as serving a regulatory function, indicating the wish to take or allocate the turn, respectively, but Beattie (1978, 1979) provides a more cognitive explanation, according to which gaze aversion at the beginning of a turn is due to and mirrors the increased cognitive effort involved in planning the utterance. Complementary to that, gaze at co-participants towards the end of the turn is taken to be indicative of the fact that cognitive resources become available for monitoring and processing recipient reactions. The empirical observations of Kendon (1967) and Duncan (1972) are, however, not entirely compatible with Goodwin's (1980) observations, at least as far as speaker gaze at the begin-

ning of turns is concerned. He argues that speakers need to secure addressees' attention and active listening and deploy specific practices such as restarts and pauses to achieve mutual gaze with non-gazing addressees. However, as Rossano (2012a) argues, this seeming contradiction between Kendon's and Goodwin's findings results primarily from the fact that the authors study different social actions and activities, which involve different gaze patterns.

For example, in story-telling activities (Goodwin 1980), where, as an interactional rule, narrators are granted to hold the floor until the activity is brought to an end (Sacks 1992), the speaker's gaze towards a recipient does not select him or her as next speaker, but exerts a monitoring function instead (cf. also Sweetser & Stec 2016, Zima 2020) or, in multi-party constellations, is used to co-address recipients (Auer 2021a, Zima 2018). Also, question-answer sequences rely on specific gaze patterns. For example, Rossano (2012a, b), Stivers & Rossano (2010), and Stivers et al. (2009) show that recipients who are looked at by speakers during the first pair part of a question-answer sequence not only deliver the second pair part more frequently but also more quickly. Thus, gaze seems to be used to mobilise the response. Even more fundamentally, Auer (2021b) has shown that questions addressed to more than one interlocutor by use of a second person plural pronoun are overwhelmingly answered by the recipient who has been looked at last by the speaker. These results confirm previous studies on speaker gaze as a means to select next speakers (Kendon 1967, Jehoul et al. 2017, Streeck 2014; Auer 2018, 2021a, b, Weiß 2018, 2019, 2020, Zima 2018).

While the function of gaze to allocate turns is well supported by empirical evidence, some other functions are more strongly contested. For instance, Goodwin & Goodwin (1986) famously proposed that during word searches, speakers avert their gaze to signal an ongoing search for a missing word, while a gaze shift towards the interlocutor is argued to invite him/her to help with the word search. However, in a recent eye tracking study, Auer & Zima (2021) challenge this dichotomous description of gaze patterns and their functions, arguing for a statistically weaker link between gaze shifts and interlocutors' attempts to help overcome word search issues. In a similar vein, feedback behaviour has been argued to be tightly linked to mutual gaze (Bavelas et al. 2002). However, these findings have recently been challenged and refined by studies that make use of mobile eye tracking technology (Vranjes et al. 2018, Zima 2020, Masuch, this volume).

In addition to that, gaze is known to play a crucial role in the establishment of joint attention (Tomasello et al. 2007). However, the micro-temporal details of how participants mutually coordinate their gaze practices and calibrate them to the situated use of attention-directing cues such as demonstratives, gestures, object manipulation, body posture and movement are only slowly coming into focus. In its primordial understanding, joint attention concerns two or more par-

ticipants focussing their visual attention on the same object and being aware of that fact; i.e., joint attention must be mutually known in order to become part of the participants' shared common ground (Clark 1996, 2021). Therefore, it requires participants to mutually coordinate action and perception in order to achieve intersubjectivity on what they see and what it means to them in the local context.

In an early study, M. Goodwin (1980) found that processes of mutual monitoring in the course of evaluative object descriptions contribute to how speakers design their emerging actions. As the evaluated objects were part of the speakers' narratives and, therefore, materially absent, the focus of the study was, however, on response monitoring rather than on mutual gaze monitoring. An increasing number of studies, however, followed this early work and explored the interpersonal coordination of the participants' gaze to objects of joint attention in participants' shared surroundings. These studies cover a range of disciplines.

In child development research, the focus has been on the development of the prerequisites for the interactional accomplishment of joint attention as mutually known and socially consequential, such as the emergence of gaze following (Flom et al. 2007), the age-relatedness of infants' capacities for joint attention and cooperative engagement (Scaife & Bruner 1975, Carpenter et al. 1998, Tomasello et al. 2007), as well as the onset of infants' use of first-person experiences to make sense of the visual experiences of others (Brooks & Meltzoff 2014). In the same vein, conversation analytic work on the multimodal complexity of reference and joint attention has shown that the participants' coordinated gaze practices contribute to, and index the situated, dynamic accomplishment of shared orientation and understanding in a range of social activities (De Stefani 2014, 2021, Eriksson, 2009, Hindmarsh and Heath, 2000, Goodwin, 2003, Mondada, 2014, Stukenbrock, 2009, 2014a, 2015).

Gaze has thus long been known to play a crucial and manifold role for interaction management. However, since most studies on the role of gaze in social interaction rely exclusively on video recordings from an observer's perspective, they do not allow us to zoom in on the fine-grained details of participants' gaze behaviour. In particular, they do not permit robust observations on the exact location and duration of the participants' gazes, the trajectories of their gaze shifts, and, most notably, on the temporally fine-tuned interaction between speaker and addressee gaze. This is due to the fact that they are often based on researchers' extrapolations from the participants' head direction rather than on analytically reliable, i.e., videographic access to eye gaze. This means that even if observations in video-based studies seem plausible, they nevertheless risk to remain tentative. It is only with the advent of mobile eye tracking that researchers can now begin to reveal the minute details and subtleties of gaze practices that are interactionally relevant but escape the less fine-grained lens of the observer's video camera.

2. The advent of mobile eye tracking: A turning point in Conversation Analysis?

The first methods to track human gaze behaviour were developed already at the end of the nineteenth century. Drawing on mirrors and indirect observation, in 1897, French ophthalmologist Louis Emile Javal has first observed that eye movements during reading involved the succession of saccades and fixations (but see Wade and Tatler 2009 who argue that the merits should be given to one of his lab members, M. Lamare, instead; see also Plużyczka 2018:102 and Attardo & Pickering 2022:10–11). Not long after that, in 1908, Edmund Huey invented a mechanical apparatus, which became known as the first eye tracker. The device he had developed was technically very complex and relied on a highly invasive way to track people's eye movements, as it required the mechanical components of the tracker to have physical contact with the cornea. This procedure was not only painful for the participants but also massively restricted their freedom of movement. The device was thus by no means suitable for widespread use in gaze research. Although in the following decades, several non-invasive eye trackers have been developed (for an overview on the history of eye tracking, see Holmqvist et al. 2011, and Plużyczka 2018), it was not until the 1970s that eye tracking gained general acceptance, particularly in psychology, in cognitive sciences and the neurosciences but also in linguistic reading research, where the use of head-mounted eye trackers and screen-based eye tracking became well-established research methods.

With the development of mobile eye tracking glasses around the turn of the last century, which allow participants to move their heads and bodies freely and which can therefore also be used in mobile settings such as walking, cooking, playing music together, shopping and even doing sports, the possibilities to use eye tracking have increased dramatically. Mobile eye tracking thus opens up completely new prospects for the study of gaze in social interaction (see, e.g., Holler & Kendrick 2015, Brône & Oben 2015, Pfeiffer & Weiß 2022, Stukenbrock 2018a, 2018b, 2020, Auer 2018, 2021a, 2021b, Weiß 2018, 2020), as it enables researchers to reconstruct the interaction from a first-person perspective rather than from the observer's perspective as recorded by external cameras (Stukenbrock 2018a, 2018b, Zima et al., this volume).

Technically, eye tracking glasses rely on one or more cameras; these are usually integrated into the glasses' frame and directed towards the pupils. To determine the position and its movements, eye tracking glasses make use of the pupil-corneal reflection technique. More specifically, an infrared beam is emitted from the glasses and projected onto the cornea, where it forms a pattern that gets reflected by the cornea. This reflection is then recorded by the pupil cameras and serves as input to an external software to calculate the position and movement of the pupils.

In addition to that, a high-resolution scene camera is built into the nose bridge of the glasses. It faces outwards towards the surroundings and its recording is taken to represent the field of vision of the wearer of the eye tracking glasses. However, as Stukenbrock (2018b) has argued, this recording only approximates the field of vision of the wearer of the glasses. The actual field of vision is significantly larger at the edges, and the peripheral vision also extends over a larger section of the environment than the recording of the scene camera suggests (Rossano 2012b). Crucially, these restrictions must be considered when analysing eye tracking data (see also Stukenbrock 2018b for details on the difference between eye tracking recordings and the subjects' emic perspective).

For analysis, information from the pupil camera(s) and the recording of the scene camera is overlaid, resulting in a video in which a tracking cursor indicates the foveal vision of the person wearing the glasses within the field of vision delivered by the integrated scene camera. Some manufacturers, such as Tobii, also provide lenses for vision correction as well as sun protection devices. Finally, the speech of the participants is recorded by an integrated microphone. When interactants are recorded while walking, the audio recordings from the glasses deliver good enough data for verbal transcription. In stationary settings, most researchers use additional audio recording devices (see the chapters in this volume by Zima et al., Oben et al., Masuch, Krug, and Barthel & Rühlemann).

As technical development progresses rapidly, glasses are becoming more and more comfortable to wear and easier to use. In fact, there are several mobile eye tracking glasses from various manufacturers on the market (e.g., Tobii, Ergonneers, Pupils Lab, and others), which are worn like ordinary glasses, and some of them do not differ much in appearance from everyday glasses. At the same time, their measurements become more precise and also the possibilities for automated data evaluation and linking with other measuring devices, for example, for the synchronous recording of psychophysiological parameters such as skin resistance, are constantly increasing (see also Brône & Oben 2018 with reference to the development of eye tracking).

However, these technological benefits are not unanimously considered irrefutable arguments in favour of the use of eye tracking for interaction research. Most notably, in Conversation Analysis, eye tracking is met with more or less overt scepticism for various reasons. First, some consider it a too invasive method that jeopardises the ecological validity of the recordings. This argument is grounded in the assumption that interactants may not be able to (fully) block out the glasses. Thus, critics hold the view that studies based on eye tracking data violate the naturalistic approach to data collection, which is a fundamental tenet of Conversation Analysis and part of the goal to reconstruct the "endogenous organization of social activities in their ordinary settings" (Mondada 2013: 33).

The proponents of mobile eye tracking, in turn, argue that it is not only a rather non-invasive method of collecting data in naturally occurring social interaction, since the newest generations of these tracking glasses resemble normal glasses, but eye tracking is indeed indispensable to provide precise eye gaze data, and consequently, to conduct robust gaze analysis. While in Conversation Analytical research, the state of the art has always been to record and analyse gaze from a third-person perspective, proponents of eye tracking have begun to question this procedure as the sole way of understanding how gaze systematically contributes to the orderly organisation of social interaction. This scepticism arises from the observation that head direction and visual focus, while routinely being equated in analysis undertaken from the observer's perspective, as a matter of fact do not always coincide, but instead may very well diverge (for a detailed discussion, see Zima et al., this volume). Mobile eye tracking glasses, in contrast, record the visual focus of participants (their foveal vision) automatically and with high accuracy.

However, the fact that eye tracking data provide very detailed and accurate information on participants' visual foci, is not considered an advantage *per se*. Notably, the detailed information that researchers have on the participants' gaze behaviour may not be mutually known to the participants themselves and may therefore not be interactionally consequential. In other words, eye tracking is considered to bear the risk that researchers confuse the technically provided, high-resolution eye tracking data with the emic perspective of the participants (see also Rasmussen & Kristiansen, this volume), who, in contrast to the researcher, may have neither access to, nor knowledge of, the co-participant's gaze at the analytically focused-on moment.

The use of mobile eye tracking is hence by no means uncontroversial in Conversation Analysis. However, the discussion is less based on empirical arguments than on intuitions and assumptions about the impact on the data and the (dis)advantages of eye tracking compared to gaze analysis from an observer's perspective. Part 1 of our volume meets the pressing research desideratum to put the discussion on robust empirical grounds by focussing on methodological issues. The chapters in the first Part discuss the advantages and drawbacks of using mobile eye tracking to study gaze in social interaction. They provide a careful and data-based evaluation of the gains of mobile eye tracking while equally paying attention to potential pitfalls and the necessity to avoid faulty interpretation of eye tracking data. This includes the plea that in order to make sense of mobile eye tracking in research on social interaction, it is crucial that researchers be aware of the phenomenological difference between the technological output of eye tracking glasses, and human vision and perception as a socially situated, contextually embedded, and practical accomplishment (Goodwin 1994, 1996; Goodwin & Goodwin 1986; Nishizaka 2017). In fact, it is the use of both video and mobile eye

tracking recordings in mutually informative ways and with a critical eye on the affordances and constraints of either one that furthers scientific advancement on the technological as well as on the methodological plane.

In line with this overall goal, Part 2 presents empirical case studies on various topics and interactional phenomena, the understanding of which greatly benefits from the use of eye tracking. It is divided into two sub-sections, which complement one another. While the first is concerned with the role of gaze in stationary conversational settings, the studies in the second part analyse gaze practices in mobile settings, in which interlocutors engage with each other while being on the move together. In the following, we briefly present the individual chapters.

3. The chapters of this volume

The first chapter of our volume by **Elisabeth Zima, Peter Auer, and Christoph Rühlemann** sets the stage and zooms in on the reasons for “why research on gaze in social interaction needs mobile eye tracking”. Starting from the observation that the vast majority of studies on gaze are based on data recordings from an observer’s perspective, the authors first discuss whether and to what extent this perspective is ‘natural’ in providing access to the embodied cues that interlocutors make use of in interactional sense-making. They show that the perspectives of the interlocutors and the perspective(s) of the camera(s) systematically diverge in recordings of different F-formations (Kendon 1990). More specifically, the authors argue that the ‘standard procedure’ of video recording does not allow for reliable reconstruction of when interactants are looking at or away from each other. This argument is supported by three intercoder reliability studies conducted in ELAN (Wittenburg et al. 2006) that compare the transcription of mutual gaze in triadic interactions from an observer’s perspective with those undertaken on the basis of mobile eye tracking data. These studies show unsatisfactory reliability values for gaze coding from an interaction external perspective. This is due not only to the fact that annotators do not agree on which interlocutors are looking at each other, but also to a significant divergence in the transcribed length of these ‘seen’ gaze contacts. The design of the intercoder reliability studies on mutual gaze transcription in ‘standard video recordings’ is replicated for eye tracking data, which yield consistently high reliability scores. The authors conclude that the very fact that eye tracking minimises the need to infer gaze targets from ambiguous bodily cues makes it the preferred method for accurately reconstructing mutual gaze in interactional data.

This chapter on the advantages of using eye tracking to study gaze in social interaction is followed by a second contribution with a methodological focus. In

their chapter on “The influence of the specificities of gaze behaviour on emerging and ensuing interaction – A contribution to the discussion of the use of eye-tracking recordings for EMCA analysis” **Gitte Rasmussen** and **Elisabeth Dalby Kristiansen** draw on both video recordings and mobile eye tracking to analyse interactions in a retail store. The authors argue that in reconstructing the details of these interactions between salespeople, customers, and products in the store, both external and internal perspectives are informative and complement each other. However, both video recording and eye tracking come with advantages and disadvantages. More specifically, the authors hold that the use of cameras in mobile interactive settings is particularly complex. One static camera is insufficient to capture people on the move. Rather, it requires the use of a moving camera and continual decisions on camera positioning. This may result in blind spots, i.e., relevant details of the interaction may go unnoticed as they are not captured in the recording. In contrast, eye tracking recordings capture participants’ visual fields, addressing some practical challenges of mobile cameras, but still involve significant limitations, particularly in terms of the actual engagement with the physical surroundings. On the contrary, the combination of video and eye tracking recordings not only enables detailed analyses of how people navigate and negotiate social interactions in retail settings but also provides access to facets of the interaction that are otherwise hard to detect, such as the ‘avoidance of embodied interaction’.

The first part of this volume with a focus on methodological issues is rounded off by a chapter on “Mobile eye-tracking and mixed-methods approaches to interaction analysis” authored by **Bert Oben**, **Clarissa de Vries**, and **Geert Brône**. It contributes to the discussion of which methods are best suited to study the role of gaze in social interaction by demonstrating the efficacy of combining quantitative and qualitative methods to uncover general gaze patterns and nuanced factors that shape interaction dynamics. To that aim, they draw on two case studies.

The first case study delves into the nuanced negotiation of ironic meaning, showcasing how a blend of quantitative and qualitative analysis elucidates the role of eye gaze in irony. While the quantitative analysis reveals that speakers shift gaze towards recipients more during ironic utterances than during non-ironic utterances, the qualitative close-reading of the ironic sequences allows to formulate hypotheses on when these gaze shifts happen and which functions they fulfil. In turn, these hypotheses can then be tested quantitatively. This mutual informing of both approaches is conceptualised by the authors as ‘a feedback loop’ between quantitative and qualitative analysis.

The second study explores the contentious topic of interactional synchronisation, focussing on the making and breaking of eye contact in task-based dyadic interactions. The authors utilise cross-recurrence quantification techniques to

assess whether and when conversational partners align their gaze behaviour with each other. They discover that while breaking off mutual gaze involves a leader-follower pattern, interlocutors typically start to look at each other at exactly the same time, i.e., without one interactant reacting to the other's gaze shift. Zooming in on one context in which interlocutors typically gaze away from each other, i.e., 'looking away while thinking', the authors suggest that this perfect synchrony is due to implicit knowledge of speech pauses and turn management conventions.

The second part of the volume complements the methodological discussion by providing empirical studies on specific interactional phenomena. This section is divided into two parts. The first three chapters analyse the role of gaze in stationary conversational settings with seated participants. By contrast, the chapters in the subsequent part investigate gaze practices in mobile interaction with participants walking in the woods or moving together through a museum.

Part 2 of our volume is launched by **Johanna Masuch's** study "On the relationship between gaze and the German recipient token *hm_hm*". The author revisits the claim that the feedback behaviour of the recipients is causally related to the gaze behaviour of speakers. More specifically, Masuch departs from the pioneering study by Bavelas, Coates, and Johnson (2002), who argued that feedback tokens are systematically elicited by speakers who initiate brief periods of mutual gaze – the so-called *gaze windows* – to seek feedback from their recipients. Recipients react to this establishment of mutual gaze by giving feedback within this gaze window, which subsequently gets dissolved by the speaker. Masuch's study tests the generalisability of this claim by analysing a corpus of 536 instances of the German continuer (Schegloff 1982) *hm_hm* taken from triadic interactions. The findings of her corpus study suggest that while speaker gaze may prompt *hm_hm*-responses, they often occur outside periods of mutual gaze, disconfirming the gaze window hypothesis. Masuch provides a new and intriguing explanation for why *hm_hm* are more often than not uttered without mutual eye contact between speaker and recipient. Her analysis of the fine-grained timing of gaze shifts, the establishment of mutual gaze and its resolution, respectively, in relation to the speaker's turn and the utterance of *hm_hm* suggests that recipients strategically align their responses with the Feedback Relevance Space (FRS) at the boundary between Intonation Phrases. This, Masuch argues, is indicative of recipients' prioritising sequential appropriateness over responding within mutual gaze with speakers.

The establishment and dissolution of mutual gaze between speakers and recipients is also of central concern to **Maximilian Krug's** chapter on "Gaze Aversion as a Marker of Disalignment in Interaction". The focus is on disalignment (Stivers 2010) as a phenomenon which has not received much attention in spite of its consequentiality for social interaction. Participants' trajectories of action may

diverge so as to potentially disrupt the progressivity of the interaction, resulting in breaks of contiguity, or interactional impasses, which are indexed by extended pauses or silence. In his exploratory study, Krug, focusing on specific gaze practices associated with disalignment, aims to understand how gaze is used to manage transitions between aligned and disaligned states. He finds that disaligning participants avoid visually addressing other participants by orienting their gaze to less relevant areas until they realign with the course of the interaction. According to the author, this gaze pattern thus signals self-involvement (Goffman 1963) and nonavailability. By contrast, when participants construe themselves as active participants who wish to rejoin the interaction, they work to re-establish mutual gaze with their coparticipants. Interestingly, Krug finds that longer pauses are not treated as indicating disalignment as long as they are accompanied by mutual gaze with the co-participant.

Finally, in the last chapter of this Section, **Mathias Barthel** and **Christoph Rühlemann** break new ground by using pupillometry data to show that “pupil size indicates planning effort at turn transitions in natural conversation”. Drawing on English conversational data from the FreMIC corpus (Rühlemann & Ptak 2023) and pupillometric data provided by the Ergoneers Dikablis Glasses 3, the authors investigate the cognitive demands of speech planning and test the hypothesis that speech planning effort at turn beginnings is mirrored in increased pupil size. More specifically, they focus on question-answer sequences and compare pupil dilations of the answering to those of the not-answering participants. By demonstrating that pupil size increases in answerers significantly more than in non-answerers, they not only show that speech planning is a major contributor to increased cognitive effort during turn transitions, but they also provide evidence that this interaction-related cognitive effort becomes visible in the pupillometric data of the participants. In doing so, their study is the first to demonstrate the value of pupillometry for the study of gaze in spontaneous conversational data.

The chapter by **Peter Auer** and **Barbara Laner** on “Laughter and gaze among talkers on a walk” is the first of the volume’s section on mobile interaction. It investigates a particularly complex configuration, i.e., dyads of participants walking side-by-side through the Black Forest National Park while talking together. Since walkers normally gaze straight ahead on the ground or at the surroundings, gaze shifts to the co-participant as well as mutual gaze are very rare. Therefore, the authors ask when and why those ‘marked’ gaze shifts to the other *do* occur and to that end focus on speaker-gaze at the co-participant at candidate laughables, the most frequent pattern in their data. Based on 270 occurrences of candidate laughables in recordings of 12 dyads on a hike, they identify three patterns. The most frequent one consists of speaker-gaze at the co-participant in turns in which paraverbal keying of the laughable by laughing particles was missing and

the laughable had to be inferred solely on semantic grounds. In the second most frequent pattern, speaker-gaze and keying by the speaker's laughter co-occurred. These types have in common that speaker-gaze serves a monitoring as well as a response (i.e., laughter) mobilising function. By contrast, the lack of both paraverbal keying by laughter and of semantic evidence coincided with the absence of speaker-gaze to the addressee. Nevertheless, addressees could be observed to laugh in response to utterances that could not be identified as containing a laughable, so that the laughable was established by the addressee. Furthermore, the authors argue that the two main functions of gaze – regulatory and monitoring – often combine. In order to establish their specific role in a particular sequence, a fine-grained analysis of the relative timing of gaze, laughing particles, turn design, and addressee response is essential. To that end, high-precision eye gaze data is needed to give us more thorough insight into when exactly participants gaze at each other and why.

The chapter by **Anja Stukenbrock** and **Angeliki Balantani** explores the role of gaze for joint attention, focussing on moments during interaction “when the establishment of joint attention becomes problematic”. Their specific concern is “how participants manage divergent and competing foci of attention”. To that aim, they analyse data from a visit to a toy museum in which dyads of friends walk around the museum looking at the exhibits. Specifically, the chapter focusses on moments during these visits where participants are not attending to the same object and invite the co-participant to share attention on a phenomenon that they find interesting. Therefore, the points of departure for these sequences are divergent foci of attention and one participant's initiating action to establish joint attention. First, the authors reveal that there are various ways in which competing foci of attention sequentially unfold from states of divergent attention. Second, the authors show that participants may resolve the problem sequentially by sharing attention on the two competing objects successively, or, alternatively, by attending to one entity only while abandoning the other. However, participants may not always cooperatively resolve the problem of competing foci of attention. Instead, they may keep focussing their attention on their own object of interest without responding to the summons of their interlocutor at all. In all these cases, participants use different response mobilising practices, such as deictics, perceptual directives, response cries, noticings, assessments, and questions that are accompanied by pointing gestures or object manipulations. Based on the observation that the practices to mobilise recipient response are not equally successful in achieving joint attention, Stukenbrock and Balantani propose a continuum of more to less response mobilising practices as an object for further studies in research on joint attention.

Finally, in their chapter on “Joint attention without language? On intersubjectivity and the joint experience of nature” **Kerstin Botsch, Martin Pfeiffer, Barbara Laner, and Peter Auer** investigate the establishment of joint attention by bodily means only. Based on twelve mobile eye tracking recordings of dyads hiking through the Black Forest National Park together (cf. also Auer & Laner, this volume), the authors focus on noticings and challenge the assumption that joint attention and intersubjectivity are sequentially achieved by an assemblage of both verbal and embodied resources. To explain how the participants mutually know that they are looking at the same entity, the authors draw on the phenomenological sociology of Alfred Schütz (1953) and the two idealisations he proposes: first, that the perspectives of ego and alter are interchangeable, and, second, that in everyday, common sense thinking, the participants’ system of relevances is congruent. The “interchangeability of perspectives” and the “congruency of perspectives” constitute what Schütz terms “the *general thesis of reciprocal perspectives*” (ibid.: 8), emphasis in the original). The empirical challenge vis-à-vis those two “typifying constructs” (ibid.) consists in showing whether and how the Schützian common sense thinking as the social foundation for drawing inferences is put to work by participants in social interaction without talk. The concern of Botsch et al. is to show the great extent to which the participants’ joint experience of nature relies on practices of mutual monitoring and embodied displays of visual orientation. Crucially, these practices comprise gaze following (Stukenbrock 2020) by the second walker after the first walker has slowed down and stopped while looking (and gesturing) at an object. The authors argue that this is taken by the second participant as indexing a noteworthy object and ascribing meaning to it. The authors further argue that the embodied practices of the first walker “are understood as ‘symptoms of the other’s thoughts’ (Schütz 1953: 12)” by the second walker and thus enable him or her, even in the absence of talk, on the basis of the congruence of relevances, to infer the significance of the looked-at object. By contrast, verbal utterances in third position are taken as retrospective empirical evidence for the successful achievement of joint attention and shared understanding.

In sum, the chapters in this volume offer a rich and manifold overview of how future research on the temporal complexities of verbal and embodied interaction may benefit from the application of mobile eye tracking as a cutting-edge technology to provide robust analytic results to long-standing research questions in Conversation Analysis, Interactional Linguistics, and related domains involving gaze as a most fugitive interactional micro-phenomenon. While mobile eye tracking has been around for a while (see above), it is not as widely used for the study of naturally occurring social interaction as we would expect, the reason being, first, that the technology is challenging to use, and, second, that it is met with scepticism and resistance, for various reasons. The studies presented here have the great

merit, as does the edited volume as a whole, of taking the reservations seriously and meeting them with self-critical methodological reflection, involving qualitative, quantitative, and mixed-method approaches and, as a result, bringing to the fore highly detailed empirical analyses that attest to the enormous potential of mobile eye tracking for future studies on joint and mutual attention, self-, other- and mutual monitoring, participation, alignment, and affiliation in everyday and institutional settings, face-to-face as well as technologically mediated interaction, stationary and mobile activities and participation frameworks of varying sizes and configurations.

Acknowledgements

Authors are listed in alphabetical order. They both contributed equally to this chapter. We acknowledge generous support by the Open Access Publication Fund of the University of Freiburg. We are also grateful to all the anonymous reviewers of all the chapters of this volume and series editors Miriam Locher and Anita Fetzer for their unwavering support of our book project and their guidance throughout the publication process.

References

- Argyle, Michael and Cook, Mark. 1976. *Gaze and Mutual Gaze*. London: Cambridge University Press.
- Attardo, Salvatore and Pickering, Lucy. 2022. *Eye Tracking in Linguistics*. London: Bloomsbury Academic.
-  Auer, Peter. 2018. “Gaze, Addressee Selection and Turn-Taking in Three-Party Interaction.” In *Eye-Tracking in Interaction: Studies on the Role of Eye Gaze in Dialogue*, ed. by Geert Brône and Bert Oben. 197–231. Amsterdam, Philadelphia: John Benjamins.
-  Auer, Peter. 2021a. “Turn-allocation and gaze: A multimodal revision of the “current-speaker-selects-next” rule of the turn-taking system of conversational analysis.” *Discourse Studies* 23(2), 117–140.
-  Auer, Peter. 2021b. “Gaze Selects the Next Speaker in Answers to Questions Pronominally Addressed to More Than One Co-participant.” *Interactional Linguistics* 1(2): 154–182.
- Auer, Peter & Zima, Elisabeth. 2021. “On word searches, gaze, and co-participation.” *Gesprächsforschung — Online — Zeitschrift zur verbalen Interaktion* 22, 390–425.
-  Auer, Peter, Laner, Barbara, Pfeiffer, Martin, Botsch, Kerstin. 2024. “A multimodal investigation of the format ‘perception imperative + exclamative’ based on mobile eye-tracking data.” In *New Perspectives in Interactional Linguistic Research*. ed. by Margret Selting and Dagmar Barth-Weingarten. 242–271. Amsterdam: Benjamins.
- Auer, Peter and Laner, Barbara. This volume. “Laughter and gaze among talkers on a walk.”

- Balantani, Angeliki. 2022. "Transitioning between activities in joint projects: The case of German "so"." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 23, 343–370.
-  Balantani, Angeliki and Lázaro, Stefani. 2021. "Joint attention and reference construction: The role of pointing and "so"." *Language and Communication* 79(6), 33–52.
- Barthel, Mathias and Rühlemann, Christoph. This volume. "Pupil size indicates planning effort at turn transitions in natural conversation".
-  Bavelas, Janet, Coates, Linda and Johnson, Trudy. 2002. "Listener responses as a collaborative process: The role of gaze." *Journal of Communication* 52, 566–580.
-  Beattie, Geoffrey W. 1978. "Floor apportionment and gaze in conversational dyads." *British Journal of Social & Clinical Psychology* 17(1), 7–15.
-  Beattie, Geoffrey W. 1979. "Planning units in spontaneous speech: Some evidence from hesitations in speech and speaker gaze direction in conversation." *Linguistics* 17, 61–78.
-  Brône, Geert and Oben, Bert. 2015. "InSight Interaction: A multimodal and multifocal dialogue corpus." *Language Resources and Evaluation* 49(1), 195–214.
-  Brône, Geert and Oben, Bert Eds. 2018. *Eye-tracking in Interaction: Studies on the role of eye gaze in dialogue*. Amsterdam: John Benjamins.
-  Brooks, Rechele and Meltzoff, Andrew N. 2014. "Gaze following: A mechanism for building social connections between infants and adults." In *Mechanisms of social connection: From brain to group*. ed. by Mario Mikulincer and Phillip R. Shaver. 167–183. American Psychological Association.
-  Carpenter, Malinda, Katherine Nagell, Michael Tomasello, George Butterworth, and Chris Moore. 1998. "Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age." *Monographs of the Society for Research in Child Development* 63(4): 1–166.
-  Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.
- Clark, Herbert H. 2021. "Common ground." In *International Encyclopedia of Linguistic Anthropology*, ed. by James Stanlaw. New York: Wiley.
-  Degutyte, Ziedune and Astell, Arlene. 2021. "The Role of Eye Gaze in Regulating Turn Taking in Conversations: A Systematized Review of Methods and Findings." *Frontiers in Psychology* 12: 616471.
-  Deppermann, Arnulf. 2015. "Pragmatik revisited". In *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*. ed. by Ludwig M. Eichinger. 323–352. Berlin/Boston: de Gruyter.
-  Deppermann, Arnulf, Streeck, Jürgen. (Eds.). 2018. *Time in Embodied Interaction*. Amsterdam, Philadelphia: John Benjamins.
- De Stefani, Elwys. 2014. "Rearranging (in) space: On mobility and its relevance for the study of face-to-face interaction". In *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, ed. by Peter Auer, Martin Hilpert, Anja Stukenbrock and Benedikt Szmrecsanyi. 434–463. Berlin, Boston: De Gruyter.
-  De Stefani, Elwys. 2021. "If-clauses, their grammatical consequents, and their embodied consequence: making people turn around in guided tours." *Frontiers in Communication* 6:661165.
-  Duchowski, Andrew T. 2017. *Eye Tracking Methodology: Theory and Practice*. London: Springer International Publishing.

- doi Duncan, Starkey. 1972. "Some signals and rules for taking speaking turns in conversations." *Journal of Personality and Social Psychology* 23, 283–292.
- doi Enfield, Nick. 2009. *The anatomy of meaning. Speech, gesture, and composite utterances*. Cambridge: Cambridge University Press.
- doi Eriksson, Mats. 2009. "Referring as interaction: On the interplay between linguistic and bodily practices." *Journal of Pragmatics* 41(2), 240–262.
- Flom, R., Lee, K., Muir, D. (Eds.). 2007. *Gaze-Following. Its Development and Significance*. Lawrence Erlbaum, Mahwah, NJ.
- Goffman, Erving. 1963. *Behaviour in Public Places: Notes on the Social Organization of Gatherings*. New York: Free Press.
- Goffman, Erving. 1981. "Footings." In: *Forms of talk*. ed. by Erving Goffman. 124–159. Philadelphia: University of Pennsylvania Press.
- Goffman, Erving. 1986. *Frame Analysis. An Essay on the Organization of Experience*. Boston: Northeastern University Press.
- doi Goodwin, Charles. 1980. "Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning." *Sociological Inquiry* 50 (3–4), 272–302.
- Goodwin, Charles. 1981. *Conversational organization: Interaction between speakers and hearers*. Academic Press.
- Goodwin, Charles. 1984. "Notes on story structure and the organization of participation." In *Structures of social action: Studies in Conversation Analysis*, ed. by J. Maxwell Atkinson and John Heritage. 225–246. London: Cambridge University Press.
- doi Goodwin, Charles. 1986. "Between and within: Alternative sequential treatments of continuers and assessments" *Human Studies* 9(2–3), 205–217.
- doi Goodwin, Marjorie. 1980. "Processes of Mutual Monitoring Implicated in the Production of Description Sequences." *Sociological Inquiry* 50(3–4), 303–317.
- Goodwin, Charles. 2003. "Pointing as situated practice." In *Pointing: Where Language, Culture, and Cognition Meet*. ed. by Satoro Kita. 217–241. Lawrence Erlbaum, Mahwah, NJ.
- doi Goodwin, Charles. 1994. "Professional Vision." *American Anthropologist*. 96(3), 606–633.
- doi Goodwin, Charles. 1996. "Transparent Vision." In: *Interaction and Grammar*. ed. by Elinor Ochs, Emanuel A. Schegloff und Sandra A. Thompson. 370–404. Cambridge: Cambridge University Press.
- Goodwin, Marjorie and Goodwin, Charles. 1986. "Gesture and coparticipation in the activity of searching for a word." *Semiotica* 62, 51–75.
- doi Hindmarsh, Jon and Heath, Christian. 2000. "Embodied reference: A study of deixis in workplace interaction." *Journal of Pragmatics* 32(12), 1855–1878.
- doi Holler, Judith and Kendrick, Kobin. 2015. "Unaddressed participants' gaze in multi-person interaction: Optimizing reciprocity." *Frontiers in Psychology* 6: 98.
- doi Holler, Judith and Levinson, Stephen C. 2019. "Multimodal Language Processing in Human Communication." *Trends in Cognitive Science* 23(8), 639–652.
- Holmqvist, Kenneth, Nyström, Marcus, Andersson, Richard, Dewhurst, Richard, Halszka, Jarodzka, and Joost van de Weijer (eds). 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.

- Jehoul, Annelies, Brône, Geert and Feyaerts, Kurt. 2017. "Gaze patterns and fillers. Empirical data on the difference between Dutch 'euh' and 'euhm.'" *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016)*, 43–50.
- doi Jucker, Andreas H., Schneider, Klaus P., Bubnitz, Woldfram. Eds. 2018. *Methods in Pragmatics*. De Gruyter Mouton, Berlin, Boston.
- doi Jucker, Andreas H., Taavitsainen, Irma (Eds.). 2010. *Historical Pragmatics*. Berlin, Boston: De Gruyter Mouton, Berlin, Boston.
- doi Kendon, Adam. 1967. "Some functions of gaze direction in social interaction." *Acta Psychologica* 26: 22–63.
- Kendon, Adam. 1990. "Movement coordination in social interaction: some examples described." In *Conducting interaction. Patterns of behavior in focused encounters*. ed. by Adam Kendon. 91–115. Cambridge: Cambridge University Press.
- doi Kendrick, Robin H. and Holler, Judith. 2017. "Gaze Direction Signals Response Preference in Conversation." *Research on Language and Social Interaction* 50(1), 12–32.
- doi Kristiansen, Elisabeth Dalby and Rasmussen, Gitte. 2021. "Eye-tracking Recordings as Data in EMCA Studies: Exploring Possibilities and Limitations. Social Interaction." *Video-Based Studies of Human Sociality* 4(4).
- doi Krug, Maximilian. 2020. "Erzählen inszenieren. Ein Theatermonolog als multimodale Leistung des Interaktionsensembles auf der Probebühne." *Linguistik online* 104 (4), 59–81.
- doi Krug, Maximilian. 2022. *Gleichzeitigkeit in der Interaktion. Strukturelle (In)Kompatibilität bei Multiaktivitäten in Theaterproben*. Berlin: De Gruyter.
- Krug, Maximilian. This volume. "Gaze aversion as a marker of disalignment in interactions."
- doi Locher, Miriam A., Graham, Sage L. Eds. 2010. *Interpersonal Pragmatics*. Berlin, Boston: Mouton de Gruyter.
- Masuch, Johanna. This volume. "In the relationship between gaze and the German recipient token hm_hm."
- doi Mondada, Lorenza. 2013. "Conversation analysis: Talk and bodily resources for the organization of social interaction." In: *Body – Language – Communication*. Volume 1. ed. by Cornelia Müller, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeill and Sedinha Tessendorf, 218–226. Berlin, Boston: De Gruyter Mouton.
- doi Mondada, Lorenza. 2014. "Bodies in action: multimodal analysis of walking and talking." *Language and Dialogue* 4 (3): 357–403.
- doi Mondada, Lorenza. 2019. "Contemporary Issues in Conversation Analysis: Embodiment and Materiality, Multimodality and Multisensoriality in Social Interaction." *Journal of Pragmatics* 145: 47–62.
- doi Nishizaka, Aug. 2017. "The Perceived Body and Embodied Vision in Interaction." *Mind, Culture, and Activity* 24(2), 110–128.
- doi Oben, Bert. 2018. "Gaze as a predictor for lexical and gestural alignment." In *Eye-Tracking in interaction. Studies on the role of eye gaze in dialogue*. ed by Geert Brône and Bert Oben. 233–264. Amsterdam/Philadelphia: John Benjamins.
- doi Oben, Bert, Brône, Geert. 2016. "Explaining interactive alignment: A multimodal and multifactorial account." *Journal of Pragmatics*, 104: 32–51.
- Oben, Bert, Clarissa de Vries, and Geert Brône. This volume. "Mobile eye-tracking and mixed-methods approaches to interaction analysis."

-  Pfeiffer, Martin and Weiss, Clarissa. 2022. "Reenactments during tellings: Using gaze for initiating reenactments, switching roles, and representing events." *Journal of Pragmatics* 189, 92–113.
-  Płużyczka, Monika. 2018. "The First Hundred Years: a History of Eye Tracking as a Research Method." *Applied Linguistics Papers*. 25/4, 101–116.
- Rasmussen, Gitte and Kristiansen, Elisabeth Dalby. This volume. "The influence of the specificities of gaze behavior on emerging and ensuing interaction".
- Rossano, Federico. 2012a. Gaze behaviour in face-to-face conversation [Doctoral dissertation]. Radboud University Nijmegen.
-  Rossano, Federico. 2012b. "Gaze in Conversation." *The Handbook of Conversation Analysis*. ed. by Jack Sidnell and Tanya Stivers: 308–329. Oxford, U.K.: Wiley Blackwell.
-  Rühlemann, Christoph. 2022. "How is emotional resonance achieved in storytellings of sadness/distress?" *Frontiers in Psychology Volume 13* – 2022.
-  Rühlemann, Christoph, and Alexander Ptak. 2023. "Reaching beneath the Tip of the Iceberg: A Guide to the Freiburg Multimodal Interaction Corpus." *Open Linguistics* 9 (1).
- Sacks, Harvey. 1992. *Lectures on conversation*. Vol. 1 & 2 Oxford: Basil Blackwell.
-  Scaife, Michael and Bruner, Jerome S. 1975. "The capacity for joint visual attention in the infant." *Nature* 253(5489), 265–666.
- Schegloff, Emanuel A. 1982. "Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences." In *Analyzing Discourse: Text and Talk*, ed. by Deborah Tannen, 71–93. Georgetown University Press.
-  Schütz, Alfred. 1953. "Common-sense and Scientific Interpretation of Human Action." *Philosophy and Phenomenological Research* 14 (1): 1–38.
- Selting, Margret, Auer, Peter, Barth-Weingarten, Dagmar, Bergmann, Jörg, Bergmann, Pia, Birkner, Karin, Couper-Kuhlen, Elizabeth, Deppermann, Arnulf, Gilles, Peter, Günthner, Susanne, Hartung, Martin, Kern, Frederike, Mertzluft, Christine, Meyer, Christian, Morek, Miriam, Oberzaucher, Frank, Peters, Jörg, Quasthoff, Uta, Schutte, Wilfried, Stukenbrock, Anja, and Uhmann, Susanne. 2009. "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–402.
-  Stivers, Tanya. 2010. "An overview of the question–response system in American English conversation." *Journal of Pragmatics* 42(10), 2772–2781.
-  Stivers, Tanya and Federico Rossano. 2010. "Mobilising response." *Research on Language and Social Interaction* 43: 3–31.
-  Stivers, Tanya, Enfield, Nick., Brown, Penelope, Englert, Christona, Hayashi, Makoto, Heinemann, Trine, Hoymann, Gertie, Rossano, Federico, De Ruiter, Jan P., Yoon, Kyong-Eun and Levinson, Steven C. 2009. "Universals and cultural variation in turn-taking in conversation." *Proceedings of the National Academy of Sciences* 106(26), 10587–10592.
- Streeck, Jürgen. 2014. "Mutual gaze and recognition: Revisiting Kendon's 'Gaze direction on two-person interaction'" In *From Gesture in Conversation to Gesture as Visible Utterance: Essays in Honor of Adam Kendon*. ed. by Mandana Seyfeddinipur and Marianne Gullberg. 35–55. Amsterdam: Benjamins.
- Streeck, Jürgen, Goodwin, Charles, LeBaron, Curtis. Eds. (2011). *Embodied Interaction: Language and Body in the Material World*. Cambridge: Cambridge University Press.

-  Stukenbrock, Anja. 2009. "Referenz durch Zeigen. Zur Theorie der Deixis." *Deutsche Sprache* 37, 289–316.
-  Stukenbrock, Anja. 2014a. "Take the words out of my mouth: Verbal instructions as embodied practices." *Journal of Pragmatics* 65, 80–102.
-  Stukenbrock, Anja. 2014b. "Pointing to an 'empty' space: Deixis am Phantasma in face-to-face interaction." *Journal of Pragmatics* 74, 70–93.
-  Stukenbrock, Anja. 2015. *Deixis in der face-to-face-Interaktion*. Berlin, Boston: De Gruyter.
-  Stukenbrock, Anja. 2018a. "Mobile dual eye-tracking in face-to-face interaction. The case of deixis and joint attention." In *Eye Tracking in Interaction. Studies on the Role of Eye Gaze in Dialogue*. ed. by Geert Brône and Bert Oben. 265–301. Amsterdam: Benjamins.
- Stukenbrock, Anja. 2018b. "Blickpraktiken von Sprechern und Hörern bei der Lokaldeixis: Mobile Eye Tracking-Analysen." *Gesprächsforschung. Online-Zeitschrift zur verbalen Interaktion*, 132–168.
-  Stukenbrock, Anja. 2020. "Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention." *Frontiers in Psychology* 11, 1779–1779.
-  Stukenbrock, Anja. 2023. "Temporality and the cooperative infrastructure of human communication: Noticings to delay and to accelerate onward movement in mobile interaction." *Language & Communication* 92, 33–54.
- Stukenbrock, Anja and Balantani, Angeliki. This volume. "When the establishment of joint attention becomes problematic: How participants manage divergent and competing foci of attention."
-  Stukenbrock, Anja and Dao, Anh Nhi. 2019. "Joint Attention in Passing: What Dual Mobile Eye Tracking Reveals About Gaze in Coordinating Embodied Activities at a Market: Social Encounters in Time and Space". In *Embodied Activities in Face-to-face and Mediated Settings*, ed. by Elisabeth Reber and Cornelia Gerhardt. 177–213. Palgrave Macmillan.
-  Sweetser, Eve and Stec, Kashmiri. 2016. "Maintaining multiple viewpoints with gaze." In: *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools across Languages and Modalities*. ed. by Barbara Dancygier, Wie-lun Lu and Arie Verhagen. 237–258. Berlin: de Gruyter Mouton.
-  Tomasello Michael, Hare, Brian, Lehmann Hagen, and Call, Josep. 2007. "Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis." *Journal of Human Evolution* 52(3), 314–20.
-  Vranjes, Jelena, Brône, Geert and Feyaerts, Kurt. 2018. "Dual feedback in interpreter-mediated interactions: On the role of gaze in the production of listener responses." *Journal of Pragmatics* 134, 15–30.
-  Wade, Nicholas and Tatler, Benjamin. 2009. "Did Javal measure eye movements during reading?" *Journal of Eye Movement Research* 5, 5–7.
-  Weiß, Clarissa. 2018. "When gaze-selected next speakers do not take the turn." *Journal of Pragmatics* 133, 28–44.
- Weiß, Clarissa. 2019. "Blickverhalten des nicht-blickselegierten Sprechers während Korrekturen und Elaborierungen." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 20, 1–28.

- Weiß. 2020. *Blick und Turn-Taking in Face-to-Face-Interaktionen: Multimodale Interaktionsanalysen triadischer Gesprächssituationen mit Hilfe von Eye-Tracking*. Verlag für Gesprächsforschung. <http://verlag-gespraechsforschung.de/2020/weiss.html>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & H. Sloetjes. 2006. “ELAN: a Professional Framework for Multimodality Research.” *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 1556–1559.
- Zima, Elisabeth. 2018. “Multimodale Mittel der Rederechtsaushandlung im gemeinsamen Erzählen.” [Multimodal means to negotiate turn-taking in joint storytellings] *Gesprächsforschung – Online – Zeitschrift zur verbalen Interaktion* (18), 241–273.
-  Zima, Elisabeth. 2020. “Gaze and recipient feedback in triadic storytelling activities.” *Discourse Processes* 57 (9), 725–748.
-  Zima, Elisabeth, Weiß, Clarissa and Geert Brône. 2019. “Gaze and overlap resolution in triadic interactions.” *Journal of Pragmatics* 140, 49–69.
- Zima, Elisabeth, Auer, Peter and Rühlemann, Christoph. This volume. “Why research on gaze in social interaction needs mobile eye tracking.”

PART 1

**Methodological considerations on the use
of mobile eye tracking to study gaze
in social interaction**

Why research on gaze in social interaction needs mobile eye tracking

Elisabeth Zima, Peter Auer & Christoph Rühlemann
University of Freiburg

This chapter challenges the prevailing practice in ethnomethodologically inspired interaction research (EMCA) of recording and analyzing gaze in social interactions from an observer's perspective. Contrary to the assumption that this perspective is 'natural', we demonstrate systematic divergence between analysts' and participants' viewpoints and argue that 'the standard procedure' of video recording does not allow for a reliable reconstruction of when interactants look to or away from each other in a considerable number of cases. Three intercoder reliability studies, comparing the transcription of mutual gaze in triadic interactions from an observer's perspective with eye tracking data, support this argument. They reveal the inherent limitations of gaze coding from an observer's perspective, while showing that gaze transcription based on eye tracking data, which captures the participants' perspective, is much less error-prone. It minimizes the need to infer gaze targets from ambiguous bodily cues and thus emerges as the preferred method for accurately reconstructing mutual gaze as part of interactional sense-making.

Keywords: mobile eye tracking, video recording from a bystander's perspective, mutual gaze, intercoder reliability

1. Introduction

The multimodal turn in conversation analysis has led to the almost complete abandonment of monomodal audio recording in ethnomethodologically inspired interaction research (EMCA) and its replacement by multimodal video recording. Starting with the early pioneering work of Charles Goodwin and Christian Heath in the 1980s, which was based on analogue video recording, the advent of digital, small and unobtrusive camcorders has ushered in a new phase of video-based interaction analysis. Over time, a way of using the video camera to record 'naturally occurring' interactions has become established in EMCA, which we will refer

to hereafter as the ‘standard procedure’ (see Goodwin 1993, Mondada 2008, Heath, Hindmarsh & Luff 2010, among others). It is grounded in the assumption that the most ‘natural’ perspective from which to record an interaction is that of a (silent, non-participating) observer or, in Goffman’s terminology, a ‘bystander’ (Goffman 1979). Although alternative recording procedures have sometimes been discussed and used (such as multiple cameras, head-mounted cameras, 360° cameras; see discussion in the next section), the vast majority of multimodal studies in EMCA follow this standard recording procedure.

In this chapter we argue that the standard procedure of filming the scene from an observer’s perspective is by no means a straightforward or even ‘natural’ practice; in many cases it will be shown to fail to live up to the EMCA postulate that recordings should enable the analyst to reconstruct the participants’ methods of making sense of the interaction, understood as an endogenously organized, sequentially emergent social encounter. This failure is due to the fact that the analyst’s (i.e. a bystander’s) perspective and that of the participants (and hence their resources for constructing or reconstructing this encounter as meaningful) systematically diverge. This chapter provides evidence for this claim by presenting the results of two intercoder reliability studies of gaze transcription from a bystander’s perspective. The results are contrasted with a reliability study of gaze transcription based on eye tracking data, which captures the participants’ perspective.

The chapter is structured as follows: After some preliminary remarks on the methodology and epistemology of EMCA, we give a detailed account of why the interactants’ perspective and the camera’s perspective do not coincide in different types of F-formations (different interactional spaces and participant constellations, Kendon 1972, 1990). We then present the design of our intercoder reliability studies and discuss their results. In order to make the studies comparable, we focus on one particular constellation, namely static, triadic interactions, and restrict the transcribers’ task to the identification of mutual gaze phases. We will show that these cannot be reliably reconstructed in recordings that reflect a bystander’s perspective only. In the final part we look at some spatial constellations in which gaze transcription based on the ‘observer camera’ is particularly unreliable. Overall, our results strongly suggest that gaze coding from an observer’s perspective is – unlike gaze transcription based on eye tracking data – an inherently unreliable method.¹

1. We do not discuss the question of how an emerging interaction may be influenced by the very fact of its being recorded. The question is an empirical one, as also argued by Heath, Hindmarsh & Luff (2010: 47–49), Laurier & Philo (2006) and Mondada (2012b). Suffice it to say that standard video recordings as well as head-mounted recording glasses can become features of the interaction that participants actively orient to and thus have an impact on the interaction itself.

2. Epistemological and methodological questions of video recording in EMCA

The epistemological and methodological issues raised by the shift from audio to video data in EMCA have been discussed by various EMCA practitioners (including key figures in multimodal EMCA research such as Heath, Hindmarsh & Luff 2010 or Mondada 2008). Our reading of these texts reveals two potentially conflicting views on the status of video recordings as data. One can be traced back to Sacks' view of recordings as 'naturalistic' data that provide mundane evidence for claims about the interactional organization of sense-making in everyday verbal interaction, and is widely followed in conversation analysis. The other is rooted in the ethnographic tradition and is pursued in various fields of qualitative sociology, but only very rarely in EMCA.

For Sacks, the usability of (in his case, audio) recordings as everyday evidence rests on the simple fact that they allow otherwise fleeting verbal interactions to be replayed and presented to others as social 'facts' (see also Bergmann 1985 for a discussion). A well-known passage in Sacks' lectures in which he comments on this status of (audio) recordings as mundane evidence within the framework of what he calls "primitive" sociology is the following (from the introduction to the autumn 1967 lectures, cf. Sacks 1992: 622–623):

When I started to do research in sociology I had this particular aim: I figured that sociology couldn't be an actual science unless it was able to handle the details of actual events; handle them formally, and in the first instance be informative about them in the ways that primitive sciences tend to be informative. That is to say, you could tell your mother something you found out, and she could go and see that it was so. [...] So I started to play around with tape recorded conversations, for the single virtue that I could replay them; and then I could type them out somewhat, and study them extendedly, who knew how long it might take. And that was a good enough record of what happened, to some extent. Other things, to be sure, happened. But at least that happened. [...] I could get my hands on it, and I could study it again and again. And also, consequentially, others could look at what I had studied [...].

Showing the recorded document is, in this view, equivalent to the everyday practice of demonstrating to others that certain things "happened", as Sacks puts it in the above quote (cf. Lynch & Boden 1994 on this point). Of course, the recording is not a complete record of the interaction, but it is "good enough" to provide evidence of selected organizational features of the interaction. Lay people can see it as evidence that a certain kind of social action occurred. This is possible because the social world is already recognizably ordered. The recognition of this orderliness is based on the ability of ordinary members to see and understand social

facts in their prereflective, “natural attitude” (Schutz 1972, cf. Mondada 2013:34). This use of recorded data then presupposes that the document (the recording) preserves the structure of the original in such a way that its organizational features are recognizable when this “natural attitude” is taken. Ideally, the recording shows the features that are “constantly exhibited and interpreted by participants in order to coordinate their conduct” (Mondada 20012a: 54), so that they are also available to the analyst.

It stands to reason that a video-recorded documentation of the observables, which includes visual in addition to acoustic information, increases its usefulness in providing such “naturalistic” evidence, as it is able to capture many of those “[o]ther things, [...], [that] happened” which Sacks alludes to. From this perspective, the use of videotaped data is an extension of Sacks’ “naturalistic” program of documenting the social interaction such that it can be replayed as evidence. Sacks’ program is based on the belief that “usually, sufficiently adequate traces of what was originally seen and heard in the interaction are preserved in the recordings to permit a reasonable effort at reconstruction”, as Luckmann (2012: 31) remarks.

However, the multimodal extension of the “naturalistic” approach to video recording also raises a number of methodological issues that are not relevant to the same extent and in the same way for audio recording. The main difference is one of modality itself: video recordings allow us to see (as well as hear) social actions; they presuppose that the ways in which the co-participants saw these facts were (reasonably) identical to the ways in which the video document invites us to see the replayed event. This, however, is impossible in the standard practice of video recording, because the ‘vision’ of the camera is not the vision of the participants.

Vision as a sensory modality is always perspectival: nothing can be seen without being seen from a particular vantage point and from a particular angle. They fundamentally determine what is seen.² The vantage points of the co-participants in an interactive episode are, by definition, not those of bystanders. Therefore, the recorded observable as it appears to the analyst in front of the screen differs from its perception and apprehension by those who participated in the scene.

In early EMCA work on videotaped interaction (see e.g. Goodwin 1981, 1994), the main strategy for minimizing this difference was a technical one. It was hoped that the totality of the observable could be documented by placing wide-angle camera(s) in an appropriate position (see also Laurier & Philo 2006, Mondada 2012). However, this does not overcome the problem of the systematic divergence

2. This is not to say, of course, that hearing is not spatial at all: among other things, it allows us to locate the direction from which sound is coming. But this spatial distortion seems negligible compared to the way in which the perspective of the video camera differs from that of the participants.

between the (multiple) perspectives of the co-participants and that of the observing camera.

More recently, some EMCA video analysts have used multiple video recordings to capture (for example) mobile interaction (e.g. Mondada 2019, Deppermann, Laurier & Mondada 2018, Peräkylä & Ruusuvuori 2012), or experimented with 360° recording devices (see Laurier & Boelt Back 2023) and stereoscopic omnidirectional cameras (McIlvenny 2019). Note, however, that capturing the “totality” of the event is not identical to being able to take the perspective of the participants, as required in Sacks’ “primitive sociology”. On the contrary, increasing the number of recording devices or using innovative technology clearly produces data that go beyond the resources available to the participants. This is not unique to technologically innovative recording practices. It is a general feature of technical recordings that they “outdo what participants could have known about the situation” (Hirschauer 2006: 420, italics omitted). In his plea for new recording techniques that allow analysts to immerse themselves in the recorded scene *ex post*, McIlvenny (2019) argues that even the standard practice of placing the video camera in a “neutral position in the scene” is far from reflecting a “natural” position. Rather, it is a reflex of “a naively realist stance that relies on the cinematic genres of truth and subjectivity that have become naturalised in the past 130 years of moving images” (no page in original).

The second tradition of video recording in qualitative sociology emerged outside EMCA, where the availability of cheap and easily transportable digital video recording equipment also led to substantial methodological changes, the most important one in this context being the establishment of videography (cf. Knoblauch 2012). Videography sees itself as a continuation and refinement of traditional ethnography and participant observation. The video camera is understood as a technological extension or complement to the ethnographer’s eyes. It is a way of overcoming some of the shortcomings of classical ethnography, such as the forgetfulness of the human observer who relies on note-taking in a complex social field (cf. Knoblauch 2012: 71, Hirschauer 2006). Typically, ethnographers who use a camera do not withdraw from the scene, but instead use a roving camera that they direct at the focal scenes or events. There is no claim to capture the social event in its entirety. Each observation is acknowledged to be partial and perspectival, foregrounding some events and neglecting others, and it is based on the ethnographer’s preconception of, and even participation in, the scene. The camera’s vision is explicitly based on that of the human participant-observer. Video ethnographers such as Mohn (2012) emphasise the difference between video as documentation and video as a way of seeing (and therefore analyzing). As such, videography is much closer to traditional filmmaking than to the EMCA analyst trying to capture ‘everything’ from a supposedly neutral perspective (that of the passive camera).

In recent years, the Sacks-inspired and the ethnography-inspired approach seem to have converged to some extent. In EMCA, this has undoubtedly been driven by the emerging interest in mobile interaction. When recording mobile participants, it is often difficult, if not impossible, to capture the totality of the event, and it is therefore necessary to acknowledge the perspectival and selective nature of any camera recording. Indeed, the ‘participation’ of the camera can be analysed in the same way as the actions of the filmed participants. In a sophisticated study, Mondada (2019) uses a separate camera to film the camera-person filming the scene of a mobile group of people on a guided tour of a public garden. She shows how the camera-person anticipates (sometimes successfully, sometimes not) the projected movements of the participants in space and follows them with the camera, just as an ordinary participant would do (see also Heath, Hindmarsh & Luff 2010: 39–40, Laurier & Philo 2006, Pehkonen, Rauniomaa & Siitonen 2021). She also acknowledges that the camera-person’s decisions about how to adapt to the scene have an impact on the quality and completeness of the recording, but also on the recorded interaction itself. Mondada concludes that the moving cameras “both preserve the relevant features of interaction in a naturalistic perspective and configure them, shaping and arranging the view of the recorded action, giving it a particular orderliness and meaning” (Mondada 2012b: 42).

The discussion can be summarised as follows: First, Sacks’ program of a “primitive sociology” is no longer feasible, or at least no longer adequate, at this stage in the development of EMCA studies, in which video analysis predominates. A recording does not per se provide ‘naturalistic’ proof that the social events in the original situation took place and were meaningful to the participants in a specific way. Rather, every decision made by the analyst in the recording process affects the way in which the features of the original scene will be available to analysts (as well as lay people) on replay. But unlike some versions of video-ethnography, EMCA studies seek to maximise the convergence between the participants’ perspective and the perspective on the interaction provided by the recording. Secondly, any recording leaves some aspects of the original scene ‘out of the picture’ and unavailable for analysis. At the same time, every recording preserves information that was not available to the participants — although the amount and nature of this surplus of information varies greatly depending on the recording technology used. In particular, the recording may provide access to (meaningful) behaviour of an individual that is not available to the co-participant. It is therefore the task of the analyst to show that a given observable, which is accessible for analysis via the recording, was not only individually, but also socially and interactionally meaningful for the participants.³

3. This use of ‘surplus’ information is already common practice in EMCA work. For example, when analyzing the beginning or end of an interactional encounter, the recording should cap-

3. EMCA methodology and epistemology and the study of human gaze: Video recording versus eye tracking

We are now ready to apply the conclusions drawn at the end of Section 2 to the analysis of human gaze. In this section, we will ask in more detail whether and in which contexts standard video recording allows to reconstruct the participants' perspective, and whether the (additional) use of eye tracking as a recording technique can guarantee a better convergence between the participants' and the analysts' perspective.

As gaze has been the subject of numerous multimodal studies within and outside the EMCA framework, it is not necessary to justify the interest in this aspect of multimodal interaction in detail here (see Rossano 2013 for a summary). It is undisputed that gaze monitoring and gaze following are central to human sociability and that human interaction is based on the reciprocal analysis of interactants' gaze targets. A case in point is joint attention to objects or events in the environment, which is arguably one of the fundamental features of human sociability (Tomasello 2008) and depends crucially on gaze tracking (Stukenbrock 2018, 2020, 2021, Stukenbrock & Dao 2019, Stukenbrock & Balantani, in this volume; Auer et al. 2024, Botsch et al. in this volume). Most importantly, as Tomasello et al. (2007) have argued, the human eye owes one of its defining features, the large and white sclera, to the advantage it provides for gaze following and thus social cooperation. This is known as the cooperative eye hypothesis. Whereas other primates rely on body movements (especially head movements) to monitor where others (animals and humans) are looking (if this is relevant to them at all), the contrast between the human sclera and iris allows for much finer monitoring of gaze direction.

In addition to the white sclera, another defining feature of the human eye is its relatively large range of vision, which extends to approximately 100° to the left and

ture the behaviour of the potential participants before and after the social encounter (see e.g. Mondada 2009), as the transition from individual to interpersonally coordinated, socially relevant behaviour is the actual object of analysis. Similarly, in cases of mediated interaction (e.g. over the telephone), where the co-participants are not (fully) visible to each other, it may be useful to record their 'off-record' behaviour even though it is not accessible to the co-participants, since the transition from individual behaviour to social interaction and the impact of the off-record activities on the on-record exchange may be of analytical interest (Mondada 2008). In a similar vein, when investigating the establishment of joint attention, it is necessary to document the behaviour of individuals (especially their gaze) before they share their perception of the object or event with other individuals (cf. Stukenbrock 2015, 2020, 2021, and Botsch et al., this volume), as it is precisely this transition from personal to interpersonal orientation in the world that is of analytical interest.

right of the visual center line. However, the extreme 10° of the peripheral visual field (the monocular vision area) is mainly limited to the perception of motion. The focus detection area (foveal vision) extends only about 20° to either side of the center line (Zhisheng et al. 2019). This foveal vision is used to monitor the eye movements of other participants independently of head movements.⁴

Eye tracking with mobile eye tracking glasses captures participants' foveal vision very well. It differs from standard video recording in two ways. First, the glasses contain a scene camera which is installed in the bridge of the glasses. It records the scene from the perspective of the person wearing the glasses, similar to other types of head-mounted cameras (e.g. GoPro). Because the camera is mounted close to the eyes, it is almost perfectly aligned with the subject's perspective. The recording angle varies; in the recorders we used (SMI and Tobii Pro Glasses 2) it is relatively narrow, but wider than the foveal vision zone of the human eye.⁵

The second feature that distinguishes eye tracking glasses from other recording devices (including head-mounted ones) is the infrared cameras that record the movement of the wearer's pupils.⁶ Based on this information, an algorithm generates a marker ('cursor') that is superimposed on the scene camera image. For analysis, the position of the cursor can be interpreted as a very close approximation of gaze direction.

Together, these two features allow us to analyse not only participants' gaze, but also their monitoring of each other's gaze. More specifically, this holds for mutual gaze, which is the focus of our reliability studies below (Section 4). Under normal circumstances (short to medium distance between interactants, reasonably good lighting conditions), humans are able to monitor the gaze direction of co-participants quite well and can estimate whether they are being looked at or not. Eye tracking recordings provide researchers the possibility to reconstruct this information. Joint attention to an object or event, in contrast, is more difficult to analyse. While the tracker provides precise information about the tracked person's gaze, the co-participant's identification of the object being looked at by that person is a complex process that involves, in addition to the analysis of gaze angle,

4. Video-based gaze transcription in EMCA work is regularly based on head movements alone. As we will show below, gaze and head movements often occur together, but must not be equated. Therefore, we disagree with Dalby Kristiansen and Rasmussen's very general claim (2021: 6; but see Kristiansen & Rasmussen, this volume, for a more nuanced view) that gaze is only available as a resource to co-participants when it is accompanied by a convergent head movement and body orientation.

5. Eye movements without head movements occur in an area of maximally 60° to both sides of the center line.

6. For a technical description, cf. Holmqvist et al. (2011: 95–108).

aspects of the physical scene (e.g., the perceptual salience of objects and events in it), the verbal context, and common sense assumptions about the salience of objects and events (see the discussion in Stukenbrock 2015 and Stukenbrock & Balantani, in this volume).

We now move on to substantiate our claim that standard video recording from a bystander's perspective more often than not does not allow to reconstruct the participants' perspectives and discuss a small selection of F(acing)-formations (Kendon 1973), neglecting (for the sake of simplicity) mobile settings to which the standard method of recording is not usually applicable anyway. In particular, we focus on the dyadic and triadic constellations in Figure 1. We discuss the way in which standard video recording from the position of an observer/bystander allows, or fails to allow, the relevant features of the recorded scene to be preserved. The answer depends not only on the position of the camera, but also on the particular constellation of the participants being recorded. In each case, a schematic representation is given, and one or two examples (chosen more or less randomly from the EMCA literature) are reproduced to illustrate the type of F-formation. We have included the authors' description/transcription of the participants' gazes. Of course, by providing examples from recent EMCA publications, we do not intend to criticize the work of the authors or question the validity of their analyses. Rather, our goal is to illustrate the wide range of common recording practices in EMCA, and to discuss the possibilities and limitations of providing access to participants' gaze direction and targets. Our main point is that the video camera often 'see' less than the participants, a problem that can be avoided by using eye tracking equipment. But we also want to point out that the observer-camera can sometimes 'see' more than the participants.

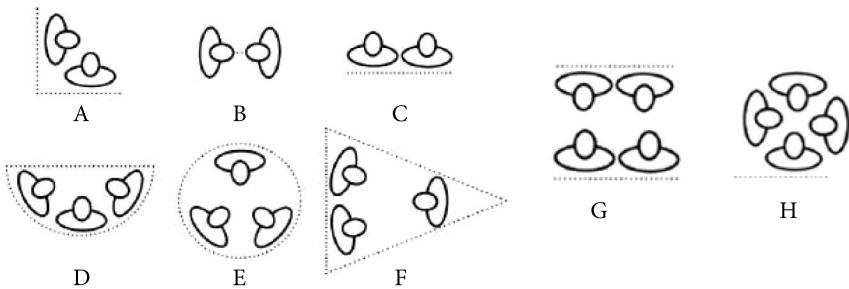


Figure 1. Some basic types of F(acing)-formations: L-shaped (A), vis-à-vis (B), side-by-side (C), semi-circular (D), circular (E), triangular (F), quadrangular/parallel (G), quadrangular/circular (A-F from Tong et al. 2016; G and H are added as they represent frequently video-recorded constellations.)

We begin with dyadic F-formations. The two most radically opposed types of F-formations for dyadic encounters are the vis-à-vis and the side-by-side constellation.

Vis-à-vis

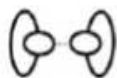


Figure 2. From Stoenica & Fiedler (2022: 259): Speakers are described as being in a state of mutual gaze



Figure 3. From Laurier (2008: 172): The woman on the sofa is described as looking at her friend. The gaze of her interlocutor, i.e. the woman in the chair, is not visible and not transcribed

Vis-à-vis constellations provide optimal spatial conditions for co-participants to monitor each other's gaze, and in particular to see whether the co-participant is looking at them or looking away. Even small deviations from the visual axis between them are easy to detect for the participants. The constellation is less suited for establishing joint attention to objects in the environment, especially those

behind one of the participants. Unless the objects are in the participants' o-space (Kendon 1990) in front of them, massive head and body movements are required to follow the gaze of the other participant.

Figure 2 is from a recording in which the camera is positioned roughly orthogonally to the axis connecting the participants' faces. Compared to the participants' perspective, the camera's observer perspective is not optimal for capturing participants looking at and away from each other's facial region, especially when no head movement is involved. Small eye movements toward and away from the other participant's face may go unnoticed.⁷ The camera position 'over the shoulder' (Figure 3) improves the situation with regard to the person on whom the camera is focused (although in this example the camera distance is too large to make a significant difference), but it completely neglects the other participant and does not allow for a sequential, interactional analysis of the gaze patterns. Although the camera in many ways fails to capture the interactionally relevant visual information available to the participants, it may also 'see' more than the participants when one of them is looking at an object that is not in the other's field of vision.

Side-by-side



Figure 4. From Dressel & Kalkhoff (2019, no page numbering): Both speakers are described as not looking at each other

In the first still (Figure 4), the camera is positioned vis-à-vis the two men sitting on the sofa. They are talking to the camera, which in this case seems to be more than a depersonalised observer (rather a silent listener). In the second case (Figure 5), the camera observes two people manipulating objects in front of them. It is positioned at an angle to the right of the person in front.

7. It should be added the authors of this study actually used two cameras, recording the interaction from two roughly opposing positions, which substantially reduces the risk of losing information. In their paper however, only this still is reproduced.



Figure 5. From Tuncer & Haddington (2019: 441): The authors use a red arrow to indicate that the man in the front is looking at the object in the woman's hand. The woman is described as gazing at the same object

In the side-by-side constellation, the participants' vision is optimised for the space in front of them, either for objects in close proximity (such as the table in Figure 5, in a work situation) or for objects located at a distance. This constellation is ideal for handling objects or scanning the scene in front of the participants, but not for fine-tuned monitoring of the other person's face. The gaze of one participant on objects in the space in front of them can only be monitored by the other participant if they turn their heads towards the gaze's face (as the male participant does in Fig. 5; see also, e.g., Stukenbrock 2015, 2020, and Stukenbrock & Balantani, in this volume, on museum exhibits). In the absence of such head movements, the camera's perspective provides information that is not available to the participants. The more it is positioned in front of the participants, the more it can capture their face and eye movements, thereby gathering information beyond the participants' visual grasp. Objects or events in the distance, in turn, can in this case be seen by participants, but not by the camera.

L-shaped



Figure 6. Example of an L-shaped arrangement (taken from own data)

Finally, L-shaped configurations (see Figure 6) are a compromise between the vis-à-vis and the side-by-side constellation. The more a particular arrangement approximates one or the other formation, the more it shares its advantages and disadvantages (for an analysis of gaze directions in this formation see Pekarek Doehler, Polak-Yitzhaki, Li, Stoenica, Havlik & Keevallik 2022: 224).

We now turn to some of the participant constellations frequently video-recorded when more than two participants are involved.

Semi-circular



Figure 7. From Stivers (2021: 10): Even though the person in the middle wears dark glasses, he and the woman to the right are described as being in a state of mutual gaze. The man on the left side is described as gazing toward the woman on the right side

The semicircular constellation opens the interaction space to one side. The camera takes the position of a fourth participant who closes the open space. This constellation has obvious advantages over the circular constellation (see below) and seems to be popular in EMCA research. We therefore used it as one of the settings we examined in our reliability study (Section 4).

The faces of all participants can be seen from the front or from the side in the recording. As in the side-by-side setup, the camera has an optimal view of the participant sitting opposite the camera; it can monitor his/her eye movements even better than either of the other two participants, who face each other but have to turn their heads to look at the middle person's face. In contrast, the gaze of the peripheral participants is sometimes difficult to reconstruct from the recording, especially when they turn away from the camera towards the central participant. Gaze shifts of one of the peripheral participants between the middle participant and the one sitting opposite are also often not accessible for analysis

(unless they can be inferred on the basis of head movements). For example, it is unclear whether the woman in Figure 7 is looking at the middle person or at the man sitting opposite her. In addition, it is often difficult to tell whether one of the peripheral participants is looking directly at the person opposite or away from that person (very similar to the problems encountered in the video recording of dyadic vis-à-vis constellations described above). Note that in all these cases the co-participants have access to information that is not available in the recording.

Triangular

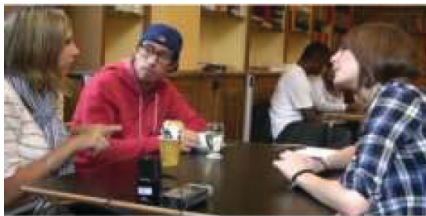
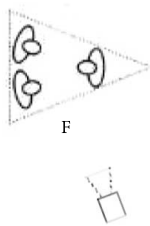


Figure 8. From Stoenica & Fiedler (2022: 256): Two opposing cameras; in the ‘standard recording method’, only the one on the left would be used. The woman in blue is described as being in a state of mutual gaze with the woman in white. The gaze of the man in the red sweater is not described

The triangular constellation is a variant of the circular constellation, which opens up the circular interaction space by grouping two participants in a quasi-side-by-side arrangement, facing the third participant. This allows the camera to ‘enter’ the interactional space as a (passive) observer. For the two participants in the side-by-side constellation, the comments made above about dyadic side-by-side constellations apply. These two participants have only limited access to each other’s eye movements when they look at the third participant. The camera records more information than is available to them. Only eye-movements accompanied by massive head movements (as in Figure 8) can be seen in peripheral vision by the participant looking at the person sitting or standing opposite to her/

him. If one of the participants positioned side-by-side and the participant opposite her/him are engaged in a verbal exchange, the observations made above about dyadic face-to-face constellations apply. This constellation will also be part of the reliability studies reported below.

Circular



The circular constellation is rarely videotaped (but see Oloff 2022, e.g. p.338). The reason for this is obvious: as in the case of dyadic constellations filmed over the shoulder of one of the participants, the focus is on the person sitting opposite the camera. The faces of the other participants are invisible and their gaze cannot be analyzed.

Quadrangular

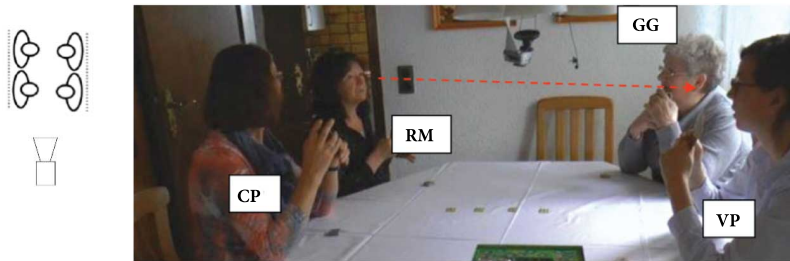


Figure 9. From Gubina (2022:307). As indicated by the arrow, participant RM is claimed to be looking at participant VP. The other participants’ gaze is not described

Quadrangular

The last arrangements we want to discuss include four participants. Various types of constellations are used. One often chosen type is similar to the triangular constellation, but combines two side-by-side arrangements. It is a common seating arrangement around a rectangular table and is therefore frequently video-recorded (Fig. 9). The standard method is to place a camera at the small end of the table, with the two rows of participants sitting at the long ends. From the recommended position of the camera (at the level of the participants’ heads), there

is a good chance that the bodies of the participants in front will block the camera's view of those in the back (which is not a problem for the participants themselves). The gaze of the two participants in the back is easier to reconstruct from the recording than that of the two in front, as they often turn their heads away from the camera to look at the other two participants sitting further away from the camera. For the pairs sitting opposite each other, the same limitations apply as for the vis-à-vis constellation. For participants sitting side by side, the restrictions on side by side recording apply.

Another variant are circular constellations with four participants, for instance four people sitting at the four sides of a quadrangular table. Here, the obvious problem is that the camera looks at the back of one (sometimes two) persons, which makes it impossible to reconstruct their gaze. Examples can be found in Gubina (2021, 289–290).

To sum up this discussion of different F-formations and camera perspectives, the perspective of the observing camera often does not coincide with the perspectives of the co-participants, but systematically diverges from them. This raises concerns about the reliability and consistency of analyses following the EMCA tradition of transcribing gaze direction and mutual gaze in such video data. To test this, we conducted three intercoder reliability studies. Two of these (Study 1a and b) are concerned with the standard EMCA procedure for recording and transcribing gaze from a bystander's perspective. The third study (here Study 2) replicates the design of the first two studies and applies it to eye tracking recordings.

4. Testing the reliability of gaze transcription in standard EMCA data versus eye tracking data

4.1 Study design

To test the possibilities and limitations of gaze analysis from an observer's perspective, we first conducted two related experimental intercoder reliability studies. We used eight video clips of triadic conversations that had been recorded using the 'standard procedure', i.e. from a bystander's perspective (see Figures 11–14) with only one camera. We asked experienced gaze researchers⁸ to individually transcribe mutual gaze phases in these clips in ELAN (Wittenburg et al. 2006) and then calculated intercoder reliability between transcriptions.

8. They were all researchers (from two universities) who have extensive expertise in the field of interactional gaze research, including but not limited to experience with mobile eye tracking data.

We chose to focus on mutual gaze because the information of being in a state of mutual gaze with another participant is available to the participants themselves. In other words, when any participant in an interaction looks at another participant and sees that that person is looking back at them, they both know that they are gazing at each other. The study was therefore designed to address the question of whether the standard recording procedure allows to reconstruct this part of the participants' shared knowledge. We have obtained written informed consent from all participants of the conversational recordings to record their interactions and to publish transcripts and screen shots. Given this informed consent, the study did not require approval from the Ethics Committee of the University of Freiburg.

In the first study (Study 1a), the clips were muted and the data were transcribed by nine analysts. In a follow-up study (Study 1b), we provided five of the transcribers from Study 1a with the unmuted video clips. We asked them to watch – and listen – to the clips again, and to correct their transcriptions whenever they considered it necessary.⁹

The eight clips used for these two studies were taken from four interactions between well-acquainted German L1 speakers who were engaged in free conversation in one of the interactants' homes. The clips, labelled "Triad A" to "Triad H", were between 30 seconds and two minutes long and were randomly selected from the conversations, which each lasted approximately 45 minutes. Figures 10–13 show screenshots from each conversation. All interactions were filmed by a camcorder positioned at a distance of one to two meters from the interlocutors. The perspective of the camera thus corresponds to that of a bystander in close proximity to the interactants. In addition, an audio recorder was placed on the dinner/coffee table.

The data set for the reliability studies consisted of two clips from each conversation. The interlocutors were all seated, but their seating positions varied. In Triad A/Triad H (see Figure 10), the three interactants were sitting on chairs around a dining table, with the participant in the center facing the camcorder and the other two facing each other (semi-circular F-formation, see above, Section 3). This means that the face and eyes of the participant in the center position were clearly visible on the recordings, while the faces and eyes of the interlocutors to the right and left were usually only visible in profile. In Triad B/Triad F

9. Note that the mutual gaze annotations in this study are therefore not independent ratings and the comparability to the results of Study 1a (no sound) is limited. To get comparable data, we would have needed to provide the empty ELAN-files with the unmuted data to nine new coders. However, it was not possible to recruit that many expert transcribers. Notwithstanding its limited scope and explanatory power, we opted for carrying out and reporting the results of this study as well. Our aim was to test whether full access to the recordings (video and sound) would lead to a significant increase in intercoder reliability.

(Figure 11), the three participants were also sitting in a semi-circular arrangement, but closer to each other on a corner sofa, almost forming an isocleses triangle (L-shaped F-formation). Again, the participant in the central position was facing the camera (at least when the head was not turned to the side, as in Figure 11), while the other two participants were filmed from the side (the speaker on the right more so than the one on the left).



Figure 10. Screenshot from Triad A/Triad H



Figure 11. Screenshot from Triad B/Triad F

In Triad C/Triad D and Triad E/Triad G, two interactants sat next to each other on one side of a table and on the same side of a corner sofa, respectively.



Figure 12. Screenshot from Triad C/Triad D



Figure 13. Screenshot from Triad E/Triad G

The person on the right (see the label *rechts* ('right') above the interactant's head) sat opposite the interactants in Triad C/D (Figure 3, triangular F-shape.) The person on the right in Triad E/Triad G sat on the other side of the sofa. This triad is similar to Triad B/Triad F, but the person in the center position sat a little closer to the person on the left, in a side-by-side arrangement with her. This type of F-formation is between semi-circular and triangular.

Mutual gaze was transcribed using the ELAN annotation tool. ELAN is widely used in interaction research and all our transcribers were highly familiar with it. The ELAN clips contained three pre-defined tiers labeled *Blickkontakt links_Mitte* 'mutual gaze Center_Left', *Blickkontakt Mitte_rechts* 'mutual gaze Center_Right',

and *Blickkontakt links_rechts* ‘mutual gaze Left_Right’ to facilitate annotation.¹⁰ To minimise the risk of erroneous annotations due to a confusion of the spatial position of the interactants (left- and right-hand side), we further inserted the labels *links* ‘left’, *Mitte* ‘center’, and *rechts* ‘right’ above the interactants’ heads into the video clips. We asked all transcribers to insert unlabeled, i.e. empty, annotations on the respective tiers for each period of time during which they believed that two interlocutors were gazing at each other. For example, if an annotator detected a mutual gaze phase between the participant on the left and the participant on the right, (s)he would add an empty annotation (see Figure 14) on the tier mutual gaze ‘Left_Right’ that stretches over the entire time span of the ‘seen’ gaze contact.

All transcriptions were done individually by all study participants on their own devices. We then combined all annotation tiers of all transcribers into one ELAN file for each triad. Figure 14 from Study 1a is a screenshot of one of the ELAN files. It shows all nine annotation tiers and all annotations (from all annotators) for “mutual gaze_Center_Left” in the first 27 seconds of the clip labelled “Triad A”.

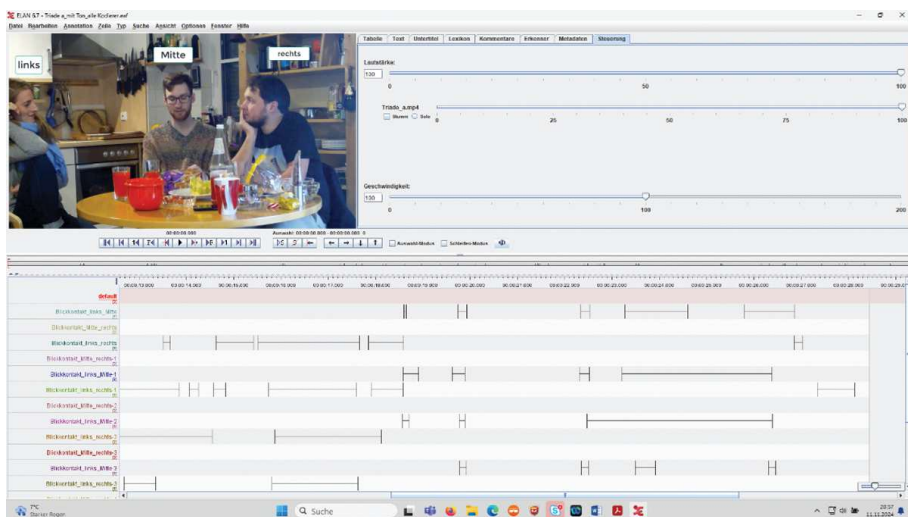


Figure 14. Screenshot from ELAN containing all annotations from all annotators

To calculate intercoder reliability, we exported our data into R, where we first structured them as a data frame and added temporal variables (start time, end time, duration). We further marked gaze annotations as ratings labeled ‘m’, filled in time gaps, and marked missing gaze annotations as ‘x’ (for no gaze contact).

10. All our transcribers were German L1 or L2-speakers.

We then duplicated the rows by ‘duration’ to obtain one trial every millisecond, resulting in, e.g., 10,000 trials every 10 seconds. To match the size of one frame in ELAN (30 ms) and hence one annotatable window, we summarised the ratings in 30 milliseconds windows. This means that if a single positive annotation “m” occurs in the window, the window as a whole is set to that annotation and only a single annotation is retained. We finally spread the data frame from long to wide format so that all ratings were in the same row and would allow for the calculation of interrater agreement on x- and m-annotations between the participants. Interrater reliability was calculated using Light’s Kappa.

Study 2 replicated the design of Study 1a (without sound), but was conducted with video clips that included eye tracking information. Interactants were again sitting around a coffee table in a semi-circular formation. They were all wearing mobile eye tracking glasses (SMI, Figure 15, or Tobii Pro Glasses 2, Figures 16 and 17).

In addition to the three eye tracking clips, the transcribers were also given the bystander clip¹¹ (bottom right, Figures 16–18). Again, we included labels (left, center, right) and lines to indicate which eye tracking video corresponded to which participant and spatial position. The transcription of reciprocal gaze in eye tracking data is a cognitively demanding task, as it requires paying close attention to the reciprocal relationship between gaze cursors in two or more recordings. The labels and connecting lines were intended to facilitate annotation and reduce the risk of accidentally placing an annotation on the wrong tier.

It might be assumed that annotators would reach perfect agreement when they can rely on eye tracking. However, this assumption is not justified. Apart from errors due to confusion of ELAN tiers, which can reduce reliability, it must be remembered that transcription based on eye tracking data still requires transcribers to make decisions about what counts as mutual gaze (see the discussion in Zima 2020).¹² By way of illustration, Figure 18 shows four screenshots of the same two participants’ eye tracking data from the same interaction. Are they in mutual gaze?

11. Combining the individual perspectives of the eye trackers with an external perspective is common practice in eye tracking research (for static settings). It makes it easier to assign the individual recordings and perspectives to the participants in the interaction. As mentioned by a reviewer of this contribution, this external perspective may be an additional source of information used by the study participants to transcribe mutual gaze. Space does not allow us to reflect on this issue in depth but as the results for Study 1a (and b) and Study 2 differ strongly from each other, it seems obvious that it is not used as the primary source of information.

12. Another reason are involuntary micro-motions of eye gaze (fixation tremor), that is, minute pupillary movements that go undetected in human observers but are captured by eyetrackers. Cf. Bowers et al. (2019). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6750810/>

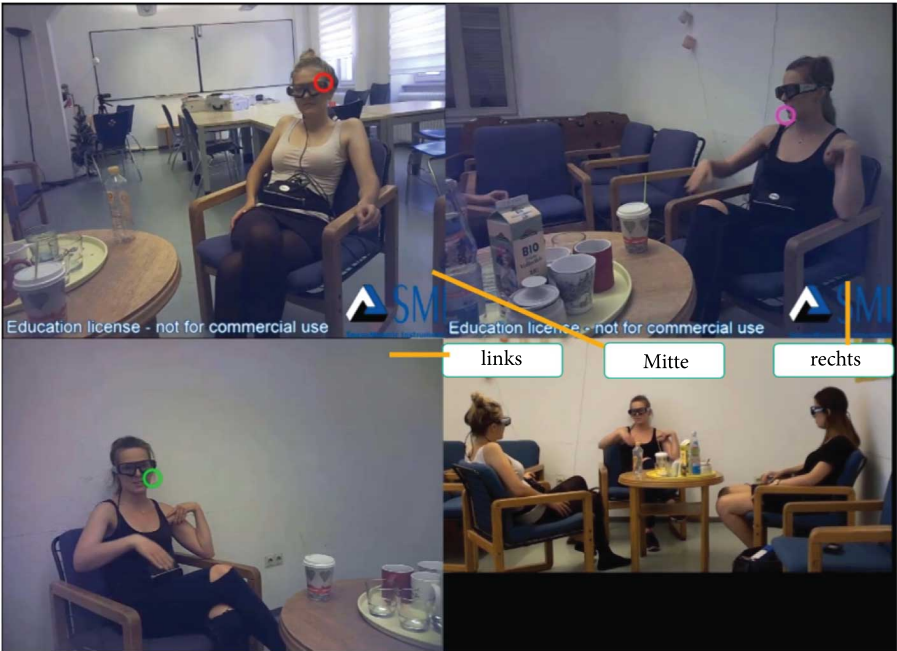


Figure 15. Study 2 – Triad A/E



Figure 16. Study 2 – Triad B/F



Figure 17. Study 2 – Triad C/D

While the two interactants in (a) are looking at each other's faces (the cursors are in the region of the eyes) and can therefore safely be said to be in a state of mutual gaze, (b)-(d) deviate from this prototypical case: In (b) and (d), only one participant's gaze cursor is in the other's facial region, and the other in a position slightly below her partner's chin in (b); in (d) it even deviates significantly to the right. In (c), both cursors are positioned outside the other's facial region. We did not give our transcribers instructions on how to proceed in these cases. Similar problems may arise in the case of very brief interruptions of mutual gaze by one participant (in the range of <50 ms), which could be transcribed as a continuous state of mutual gaze or as looking away. The intercoder reliability study on eye tracking data allows us to assess the degree of divergence between annotators due to these and similar standard problems of working with eye tracking data.

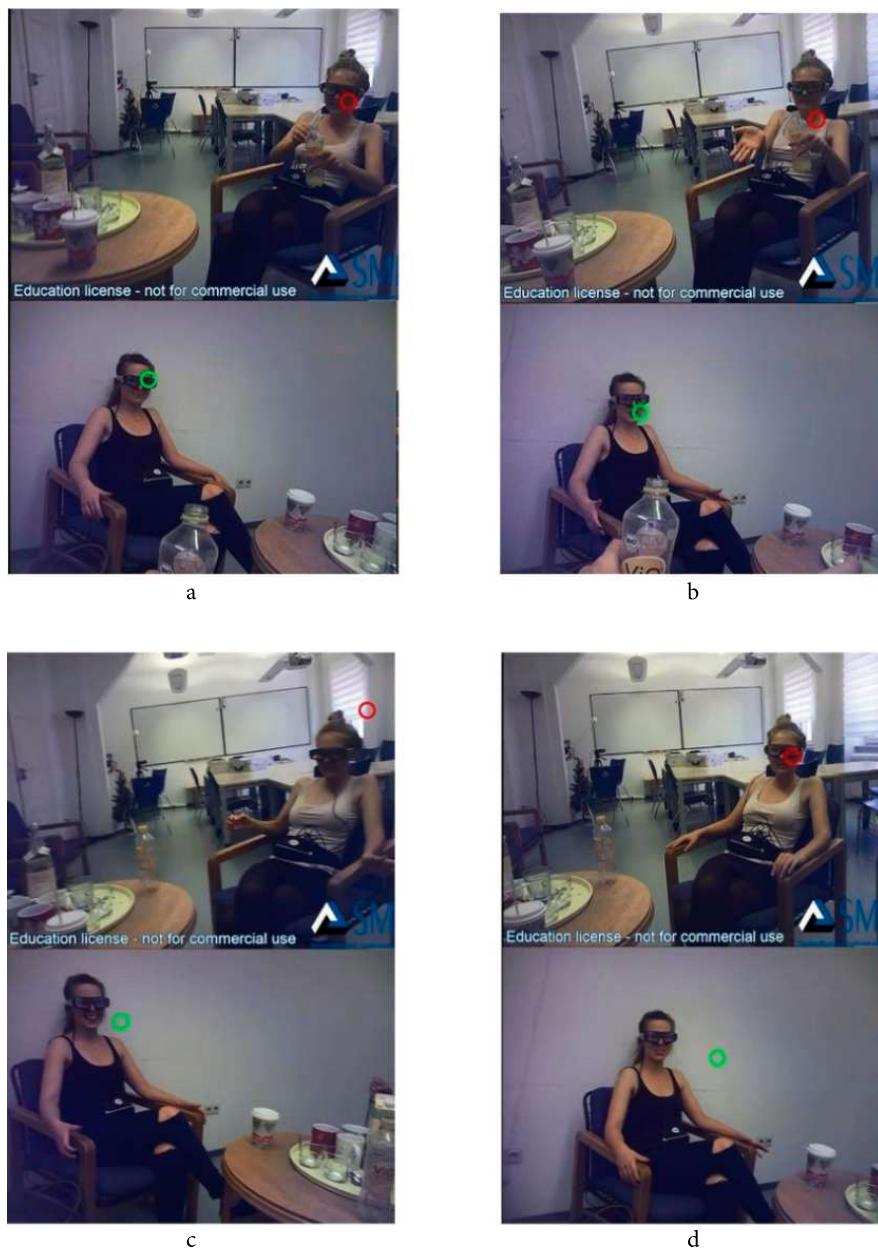


Figure 18. Illustrative gaze constellations taken from Zima (2020:7)

4.2 Results

Study 1a (no sound): Mutual gaze transcription from an observer's perspective in muted clips

Table 1 lists the Light's Kappa values for all interlocutor pairings in Triads A to H. Overall, intercoder agreement is moderate (Landis & Koch 1977) to low (McHugh 2012) with only very few pairings in selected clips reaching good agreement rates. The median Light Kappa value is 0.5491. The highest Kappa value is 0.8158 for the Center_Left pairing in Triad D. It is the only Kappa value above 0.8 (almost perfect agreement according to Landis & Koch 1997, strong agreement according to McHugh 2012). The lowest Kappa value is 0.16 (Triad B, Center_Right).

Table 1. Light's Kappa values for Study 1a (observer's perspective, no sound). NA indicates that the tiers for this particular pair of interlocutors contained no annotations, i.e. no transcriber saw any gaze contact

Video clip	Interlocutor pairing	Light's Kappa
Triad A	Left_Right	0.72941011
Triad B	Left_Right	0.21314475
Triad C	Left_Right	0.79514232
Triad D	Left_Right	0.73514523
Triad E	Left_Right	0.54722943
Triad F	Left_Right	0.44974103
Triad G	Left_Right	0.46237529
Triad H	Left_Right	0.67137633
Triad A	Center_left	0.70231077
Triad B	Center_left	NA
Triad C	Center_left	0.54910575
Triad D	Center_left	0.81579784
Triad E	Center_left	NA
Triad F	Center_left	0.25347729
Triad G	Center_left	0.55385514
Triad H	Center_left	0.66615359
Triad A	Center_right	NA
Triad B	Center_right	0.15825336
Triad C	Center_right	0.36423599
Triad D	Center_right	0.26487456
Triad C	Center_right	0.5507244
Triad F	Center_right	0.36749341
Triad G	Center_right	0.48101625
Triad H	Center_right	0.56972021

The boxplots in Figure 19 show the distributions of the Kappa values obtained from the interrater-reliability analyses of mutual gaze for interlocutor pairings. Agreement rates are highest for mutual gaze in the Center_Left pairing (mean Light's Kappa is 0.5901, one outlier) and lowest for the Center_Right pairing (0.3938). The mean of the Light's Kappa value for the Left-Right pairing is 0.5754.

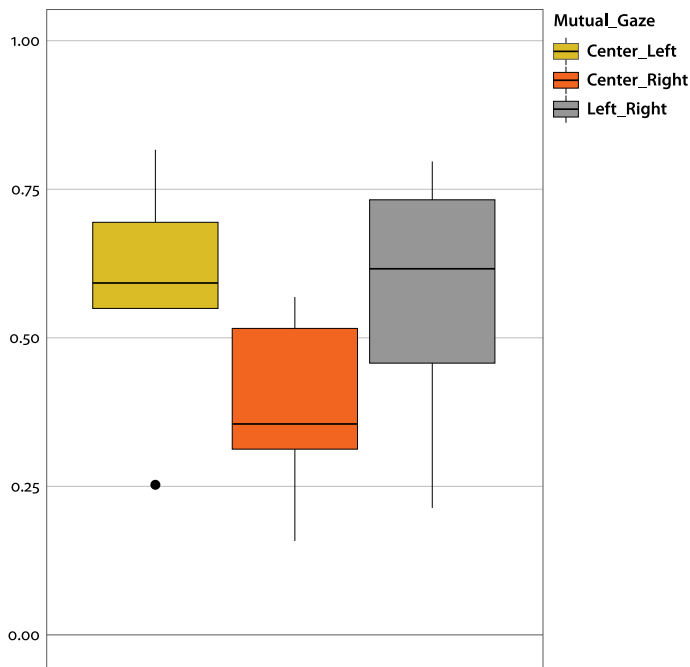


Figure 19. Box plots comparing interlocutor pairings according to spatial position in Study 1a

The following annotation density plots allow a deeper understanding of the causes of these unsatisfactory kappa values. Due to space constraints, we limit the discussion to three video clips from three different recordings (Figure 20–22). The discussion zooms in on the problematic data stretches highlighted by black rectangles. The annotation density plots show annotations by the same transcriber in the same color. Annotations for interlocutor pairings according to spatial position are grouped together vertically. Tiers without annotations are not included in the plots.

The density plot for Triad B (Figure 20) reveals that transcribers systematically differ in whether they see frequently interrupted gaze contacts or longer, uninterrupted reciprocal gaze phases; see e.g. first annotation line (pink) compared to fourth (light blue). More fundamentally, transcribers often disagree about *which* interactants are looking at each other. While the fourth (light blue annotations)

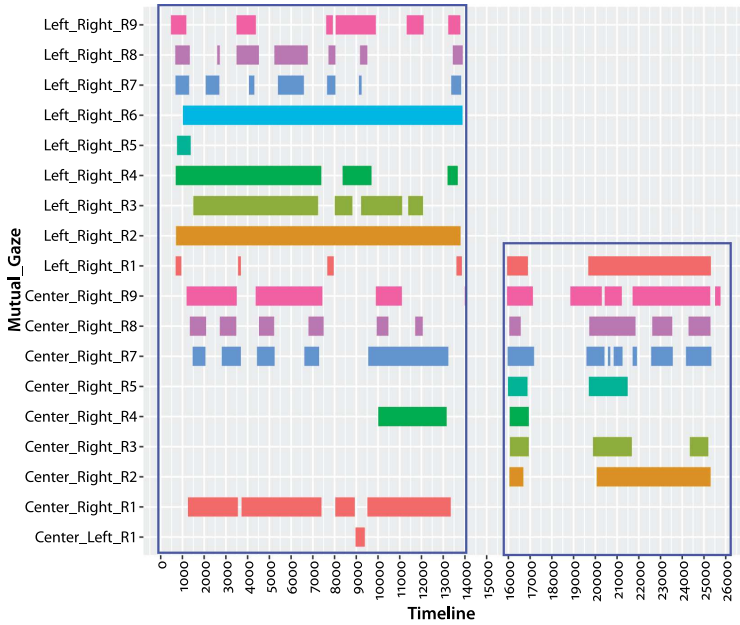


Figure 20. Annotation density plot for Triad B (study 1a)

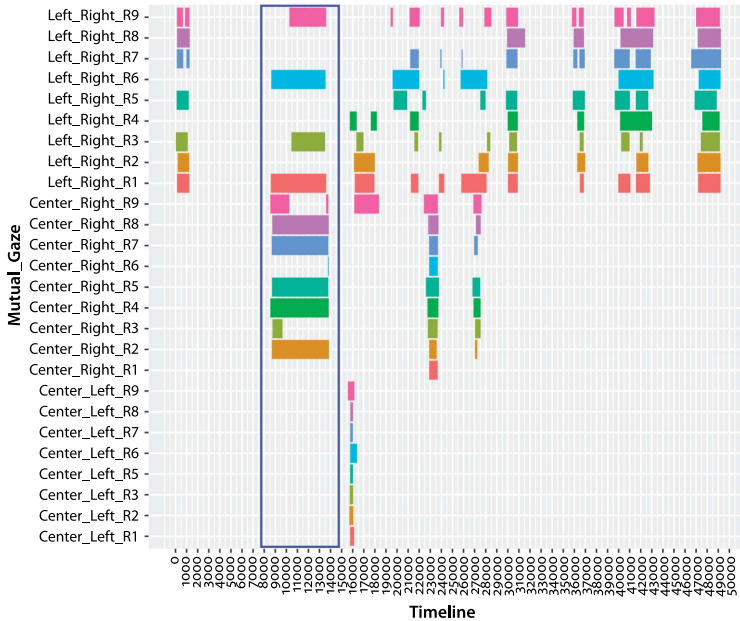


Figure 21. Annotation density plot for Triad G (study 1a)

and eighth (brown annotations) transcriber see gaze contact between the participants on the left and right in the phase between 1,000 and 14,000 milliseconds, the seventh (reddish annotations) sees the participant in the center position and the one on the right looking at each other for almost exactly the same stretch of time. This is not the result of a random confusion of tiers in ELAN or of left and right; other transcribers also see gaze contacts of the Center-Right pairing (see pink, mauve, navy blue and green annotations). From the beginning until almost half-way into the video clip, the analysts do not reach a satisfactory agreement on which interactants are looking at each other.

The same applies to the interval between 8,000 and 12,000 milliseconds in Triad G (Figure 21). Again, we see overlapping annotations on the Left-Right and Center-Right tiers. The rest of the data set contains mostly annotations of very short gaze contacts, which overlap quite well at the beginning and end of the clip (see the tiers for the Left-Right interlocutor pair). However, the middle of the clip shows only moderate to low agreement, with many gaze contacts identified by some, but not all transcribers.

Triad A (Figure 22), on the other hand, shows a much higher level of agreement. Although the annotators disagree to some extent about their length, they do not attribute mutual gaze phases to different speaker pairings. This is reflected in a better kappa value (around 0.6).

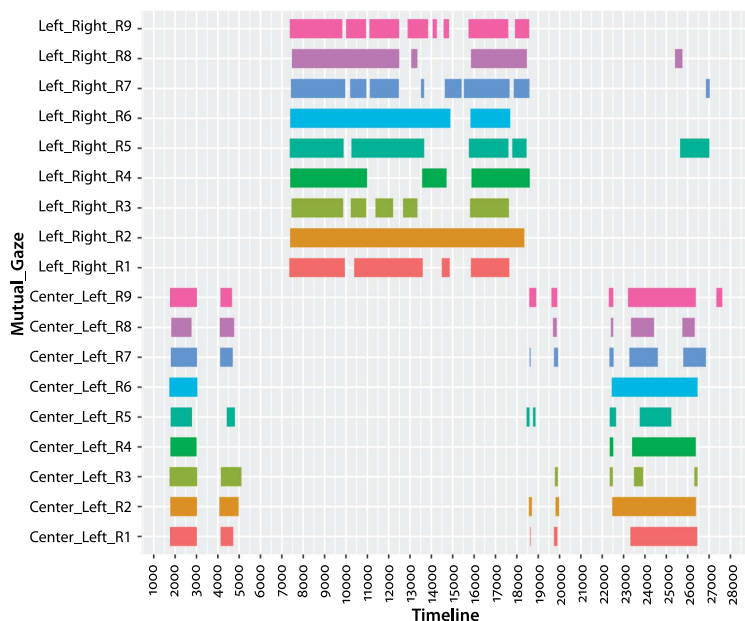


Figure 22. Annotation density plot for Triad A (Study 1a)

We can zoom in even further and look at some phases of the interaction for which the transcription of gaze was particularly unreliable. The first example is from Triad B (01:480 – 02:870). The utterance in line 02 of the transcript is approximately 1.4 seconds long. During this time interval, one transcriber saw no gaze contact at all, four transcribers saw mutual gaze between the participants sitting on the right and on the left, and four other transcribers saw gaze contact between the participants in the center position and on the right. No one annotated a gaze shift between the interactants during this sequence.

Extract 1.

((R, the person sitting on the right, tells a story from a seminar in pedagogy, where she was asked to give a didactic commentary.))

01 R: und dann hamma ANgefangen, (0.7)
and then we started
→ 02 #1 i*ch wa#2r #3grad am SCH#4REI#5ben#6,
I was writing
r: *typing gesture ----->
03 (1.0) (0.1) auf EINmal (1.0) word ZU,
all of a sudden: shut down word,
---->*



#1: 01:480



#2: 01:720



#3: 02:000



#4: 02:246



#5: 02:540



#6: 02:780

While R (right) tells her story, her two recipients remain in the same bodily position; both appear to be looking at her. With one exception, the annotators all saw (or perhaps inferred) that the two recipients were looking at the teller (even

in the muted condition, the teller is easily identified by her visible articulation movements and gestures). However, there was disagreement as to whom the teller was looking at. She gestures while speaking (right hand: iconic typing gesture), but she does not move her head. From the observer's position (with her face turned more than 90° away from the camera), it is extremely difficult to detect her eye movements, which may explain the high level of disagreement. The two recipients themselves, of course, cannot be assumed to have any difficulty following the teller's eye movements.

Similar problems arise in the following extract, taken from Triad G. In the bold-faced segment, which lasts a bit more than 5 seconds (08:580–13:860), five transcribers saw mutual gaze between the participants in the center and on the right side throughout the entire stretch; two saw gaze contact between the person at the center and the one on the right; one saw mutual gaze between the interlocutor at the center and the woman on the right from 08:750 to 09:650 and between right and left from 10:460 to 13:510. Finally, one transcriber saw gaze contact between center and right from 08:550 to 10:270 and from 13:595 to 13:805, as well as gaze contact between right and left from 10:270 to 13:660.

Extract 2.

((Three friends talk about their plans for New Year's Eve.))

01 R: aber wart ihr schon mal **SILvester** #1**auf_m schönberg?**
but have you ever been on the X-mountain on New Year's Eve?

02 M: (0.44)+ #2(0.4) **auf_m [SCH#3ÖNberg]?**
on the X-mountain?

03 L: [ne: :] **a#4be::r #5KANN**
man ma machen o#6der,

no:: : but you can do that, right?

04 R: is C00#6L.





#2: 09:603



#3: 10:605



#4: 11:669



#5: 12:669



The interactional space is structured similarly to the last example, but in this case all three participants are involved as speakers. R (right) is the speaker who starts the sequence with a polar question addressed to both co-participants (cf. the German second person plural pronoun ‘you PL.’). After a pause, M (center position) responds first with a repair initiation — he seems not to know the location. In overlap (as is typical for responses to *ihr*-questions, cf. Auer 2021), L (left) also responds and negates having been there in the past, adding ‘that this is something one could do’; she understands the question as a suggestion for a joint project for New Year’s Eve. R concludes the sequence with a positive evaluation. So R is the participant to whom both answers to the question are addressed. The sequence is between him and the other two, not between L and M.

The critical segment starts towards the end of R’s question and includes both answers as well as the evaluation that closes the sequence.

All transcribers agree that there is no mutual gaze between L and M; as they are sitting side by side, mutual gaze would require them to turn their heads, which is not the case in this extract. Apparently, this was taken as evidence for a lack of mutual gaze. Transcription difficulties once more concern possible mutual gazes between L(left) and R(right) vs. M(iddle; Centre) and R(ight). Again, one of the reasons for these difficulties is the lack of major head movements (with the exception of #6, where R slightly raises his chin). Also, both L and R are wearing glasses (but see also Figure 6–9), and L’s face is in the shadow. In addition, L and R can only be seen in profile, which limits the transcribers’ access to their eye movements (although L/R and M/R can, of course, see their gaze).

These examples thus show that the bystander’s perspective does not allow to reliably detect gaze directions of interactants that are filmed from the side, particularly if they can look at both interlocutors without moving their heads. This holds for both the Center_Right and the Center_Left pairings. For the same reason, the Center_Left constellation turns out to be the easiest to transcribe: In all video clips, the two participants at the center and the left-hand side have to turn their heads quite strongly to be able to gaze at each other.

However, one might assume that to have access to the verbal part of the interaction, i.e. to what is being talked about and who acts as speakers and recipients (see the co-addressed *ihr*-question in the previous example), should make it easier to annotate mutual gaze. Therefore, we expected intercoder reliability to increase for Study 1b, where we provided the transcribers with the unmuted video clips.

Study 1b (observer's perspective, with sound)

Table 2 lists Light's Kappa values for the unmuted condition and Figure 24 shows the box plot with the distribution of the median Kappa values across interlocutor pairings in space.

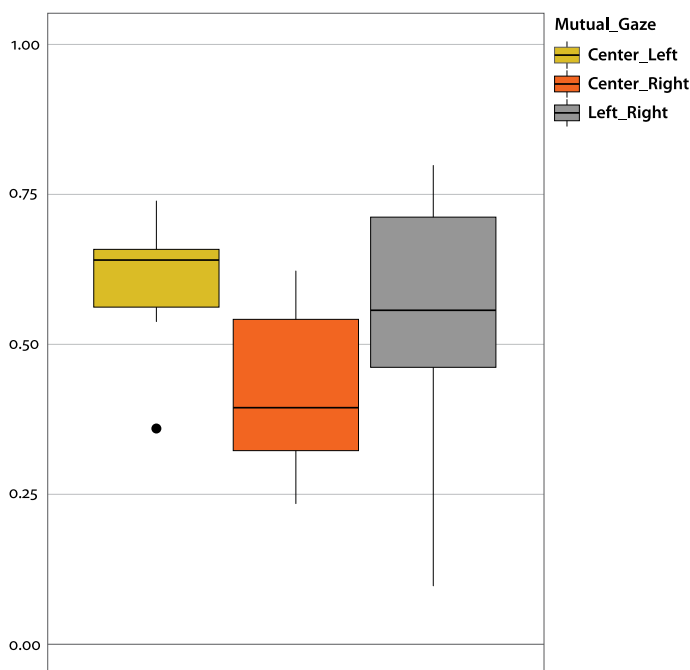


Figure 23. Box plots comparing interlocutor pairings according to spatial position in Study 1b

In the unmuted condition, the transcribers had a chance to correct their transcriptions based on the muted video clips. The video clip with the best agreement rates is now Triad C (Left-Right pair). Light's kappa is 0.7983, whereas the best performer from the first study (Triad D, Center_Left) now has a slightly lower kappa of 0.7379. The lowest kappa value is 0.0979 (Triad B, Left_Right), indicating virtually no overlap between the transcribers. The median of all kappa values is 0.5349, which is slightly lower than for the muted clips (0.5491). In terms

of interlocutor pairings, we see a pattern similar to that of the first study (see Figure 23). The Center_Right pairing turns out to be much more difficult to annotate than the other two.

Table 2. Light’s Kappa values for Study 1b (*observer’s perspective*, with sound). NA indicates that the tiers for this particular pair of interlocutors contained no annotations, i.e. no transcriber saw a gaze contact

Video clip	Interlocutor pairing	Light’s Kappa
Triad A	Left_Right	0.77128992
Triad B	Left_Right	0.09794655
Triad C	Left_Right	0.79830255
Triad D	Left_Right	0.69128738
Triad E	Left_Right	0.46035289
Triad F	Left_Right	0.48632148
Triad G	Left_Right	0.46367831
Triad H	Left_Right	0.62883591
Triad A	Center_Left	0.65282446
Triad B	Center_Left	NA
Triad C	Center_Left	0.53943484
Triad D	Center_Left	0.73789272
Triad E	Center_Left	NA
Triad F	Center_Left	0.35909172
Triad G	Center_Left	0.65753781
Triad H	Center_Left	0.62891836
Triad A	Center_Right	NA
Triad B	Center_Right	0.28683699
Triad C	Center_Right	0.39559416
Triad D	Center_Right	0.36162252
Triad E	Center_Right	0.47698268
Triad F	Center_Right	0.6197282
Triad G	Center_Right	0.23489498
Triad H	Center_Right	0.60033791

These low agreement rates are an unexpected result; we hypothesised that full access to the interactional record would make it easier to annotate mutual gaze.

This is clearly not the case. The study design does not allow for a thorough analysis of the reasons for this unexpected result. However, it suggests that access to the full recording including the audio may not help.

Study 2: Transcribing (mutual) gaze on the basis of eye tracking data

To test the reliability of gaze transcription on the basis of eye tracking data, we replicated the study design on muted video clips based on eye tracking data (supplemented by an external camera) and asked five of the transcribers from Study 1a to transcribe mutual gaze in six clips from three interactions. Table 3 lists the Light's Kappa values for all clips and interlocutor pairings.

Table 3. Light's Kappa values for Study 2 (eye tracking plus observer's perspective, no sound). NA indicates that the tiers for this particular pair of interlocutors contained no annotations, i.e. no transcriber saw a gaze contact

Video clip	Interlocutor pairing	Light's Kappa
Triad A	Left_Right	0.90098906
Triad B	Left_Right	0.88475207
Triad C	Left_Right	0.83002585
Triad D	Left_Right	0.7968849
Triad E	Left_Right	NA
Triad F	Left_Right	0.85896614
Triad A	Center_Left	0.88556465
Triad B	Center_Left	0.91222993
Triad C	Center_Left	0.68771256
Triad D	Center_Left	0.80410501
Triad E	Center_Left	0.95625479
Triad F	Center_Left	0.84231881
Triad A	Center_Right	0.78141959
Triad B	Center_Right	NA
Triad C	Center_Right	0.48137192
Triad D	Center_Right	0.85969991
Triad E	Center_Right	0.91360203
Triad F	Center_Right	0.86700502

Agreement between transcribers is high throughout the whole data-set, except for one outlier (Triad C, Center_Right) and two values that are in the mid-range (Triad C, Center_Left and Triad F_Center_Right).¹³ The highest Kappa value is 0.9563, indicating almost absolute agreement, and no less than twelve Kappa values are >0.8 (indicating perfect agreement according to Landis & Koch 1977). The box plot in Figure 24 compares interlocutor pairings according to spatial position. There are almost no differences between the medians of the kappa values for the three pairings, i.e. the spatial constellation does not play a role. This is of course to be expected, as the three eye tracking cameras do not privilege one perspective, as the camera in the observer’s position does.

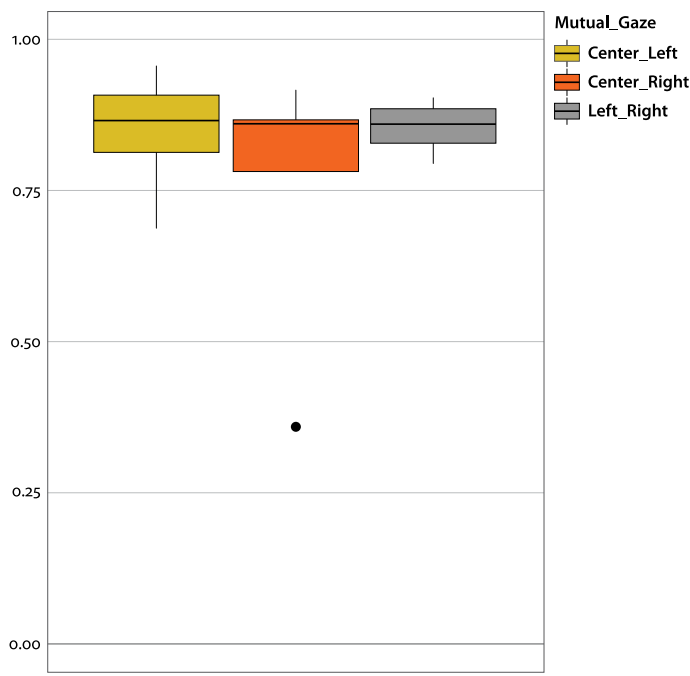


Figure 24. Box plots comparing interlocutor pairings according to spatial position in Study 2

The annotation density plots for all six clips in this study (Figure 25) also illustrate the remarkable degree of agreement between transcribers.

13. Triad C is a clip of an outdoor recording with difficult light conditions and an unstable cursor produced by the tracker worn by one participant. Not surprisingly, rather poor quality of the eye tracking leads to significantly more uncertainties among the transcribers.



Figure 25. Annotation density plots for all video clips from Study 2

The sole sources of disagreement are very short gaze contacts, which were not noticed or transcribed by all annotators, and small interruptions of mutual gaze. Space does not allow us to go into details of what happens in these moments in the data (such as missing gaze cursors, very short gaze aversions lasting only one or two frames of 30 msec). These minor disagreements confirm that also eye tracking requires decisions to be taken by the transcribers and is not an automatic process. However, despite these potential sources of disagreement, the transcribers' decisions are almost perfectly in agreement.

Finally, Figure 26 contrasts the three studies with each other. While the box plots for the muted and unmuted standard EMCA data look very similar and show unsatisfactory intercoder agreement, the reliability of the transcription based on eye tracking data by far exceeds that of recordings following the standard procedure.

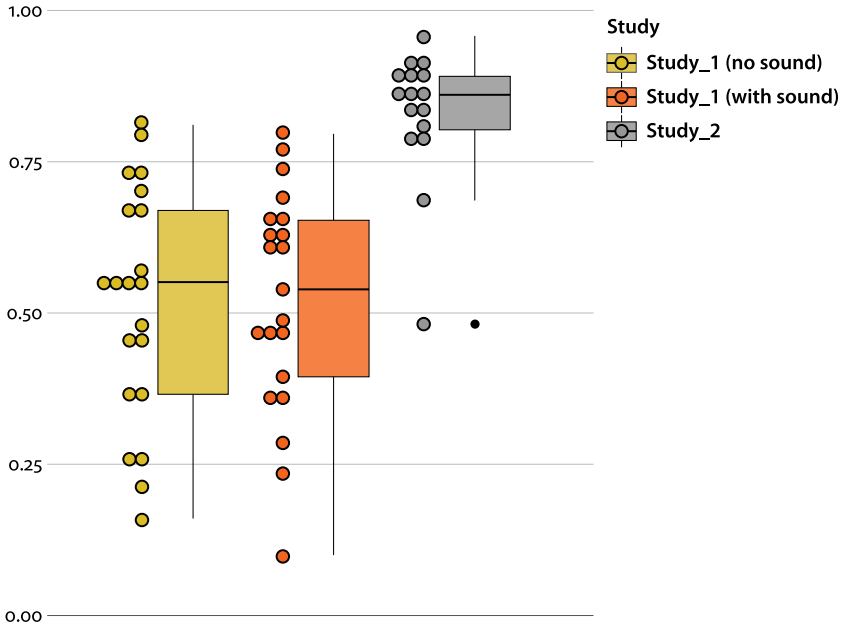


Figure 26. Boxplots for kappa values in all three studies

5. Conclusions

In this chapter, we have argued that the ‘standard’ procedure of EMCA to record and analyse gaze from a bystander’s (observer’s) perspective is often problematic. We have shown that for various F-formations and camera positions, the external camera’s perspective systematically diverges from that of the participants and can therefore make it difficult, if not impossible, to reconstruct their gaze direction and targets. This argument is supported by the results of a reliability study on interactional data that were recorded from a bystander’s perspective. Study participants who were asked to transcribe mutual gaze in the data were not able to do so in a reliable way. This suggests that the ‘standard’ recording procedure does not deliver well on the promise to provide “mundane proof” of what “has happened” in the interaction, particularly when changes of gaze direction are not accompanied by changes of head direction (see Schmitz (2020: 103) for a critique of the equation of head movement and gaze direction in EMCA). Furthermore, transcription of gaze on the basis of video recordings alone turns out to be especially problematic when two interactants are in a participant’s field of vision at the same time, i.e. if this participant can gaze at them without moving the head. Our tests show that disentangling gaze targets in these spatial constellations is particularly problematic.

As demonstrated by our second study, eye tracking resolves these issues by its very accurate measurements of gaze direction from the perspective of the participants. Particularly in the case of mutual gaze, eye tracking provides us with the visual information also available to the interactants themselves and hence allows to reconstruct the participants' *shared* knowledge.

For mutual gaze, we believe to have provided a range of arguments that call for a cautious evaluation of gaze analysis based on video recordings from a bystander's perspective. At the same time, we hope to have shown that eye tracking is very much in line with the EMCA postulate to take the participants' perspective and reconstruct their understanding of the situation from this perspective. As far as gaze at objects in the surroundings is concerned, eye tracking may, however, provide the analyst with a surplus of information that is not necessarily shared among interactants (which also holds for video data from the observer's perspective). Here, the analytical task of reconstructing what participants perceive from the data recorded by the eye tracking devices they are wearing is much more difficult.

References

-  Auer, Peter. 2021. "Gaze Selects the Next Speaker in Answers to Questions Pronominally Addressed to More than One Co-Participant." *Interactional Linguistics* 1 (2): 154–182.
-  Auer, Peter, Laner, Barbara, Pfeiffer, Martin, and Kerstin Botsch (2024). "Noticing and Assessing Nature: A Multimodal Investigation of the Format 'Perception Imperative + Exclamative' based on Mobile Eye-tracking Data." In *New Perspectives in Interactional Linguistic research*, ed. by Margaret Selting, and Dagmar Barth-Weingarten, 245–277. Amsterdam: Benjamins.
- Bergmann, Jörg R. 1985. "Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit: Aufzeichnungen als Daten der interpretativen Soziologie." In *Entzauberte Wissenschaft: Zur Relativität und Geltung soziologischer Forschung*, ed. by Wolfgang Bonß, and Heinz Hartmann, 299–320. Göttingen: Schwartz.
-  Bowers, N.R., Boehm, A.E., Roorda, A. 2019. "The effects of fixational tremor on the retinal image." *Journal of Vision* 19(11):8.
-  Dalby Kristiansen, Elisabeth, and Gitte Rasmussen. 2021. "Eyetracking Recordings as Data in EMCA Studies: Exploring Possibilities and Limitations." *Social Interaction. Video-Based Studies of Human Sociality* 4 (4).
- Deppermann, Arnulf, and Lorenza Mondada. 2018. "Overtaking as an Interactional Achievement: Video Analyses of Participants' Practices in Traffic." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 19: 1–131. www.gespraechsforschung-ozs.de
-  Goffman, Erving. 1979. "Footing." *Semiotica* 25: 1–29.
- Goodwin, Charles. 1981. *Conversational Organization: Interaction between Speakers and Hearers*. London: Academic Press.

- Goodwin, Charles. 1993. "Recording Human Interaction in Natural Settings." *Pragmatics* 3 (2): 181–209.
-  Goodwin, Charles. 1994. "Professional Vision." *American Anthropologist* 96 (3): 606–633.
- Gubina, Alexandra. 2021. "Availability, Grammar, and Action Formation: On Simple and Modal Interrogative Request Formats in Spoken German." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 19: 1–131. www.gespraechsforschung-ozs.de 22, 272–303.
- Gubina, Alexandra. 2022. *Grammatik des Handelns in der sozialen Interaktion*. Göttingen: Verlag für Gesprächsforschung.
-  Heath, Christian, Hindmarsh, Jon, and Paul Luff (eds). 2010. *Video in Qualitative Research. Analysing Social Interaction in Everyday Life*. London: Sage.
- Hirschauer, Stefan. 2006. "Putting Things into Words. Ethnographic Description and the Silence of the Social." *Human Studies* 29 (4): 413–441.
- Holmqvist, Kenneth, Nyström, Marcus, Andersson, Richard, Dewhurst, Richard, Halszka, Jarodzka, and Joost van de Weijer (eds). 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.
- Kendon, Adam. 1973. "The Role of Visible Behaviour in the Organization of Social Interaction." In *Social Communication and Movement: Studies of Interaction and Expression in Man and Chimpanzee*, ed. by Mario von Cranach, and Ian Vine, 29–74. New York: Academic Press.
- Kendon, Adam. 1990. "Movement Coordination in Social Interaction: Some Examples Described." In *Conducting Interaction. Patterns of Behaviour in Focused Encounters*, ed. by Adam Kendon, 91–115. Cambridge: Cambridge University Press.
- Knoblauch, Hubert. 2012. "Videography. Focused Ethnography and Video Analysis." In *Video Analysis: Methodology and Methods*, ed. by Hubert Knoblauch, Bernt Schnettler, Jürgen Raab, and Hans-Georg Soeffner, 69–84. Frankfurt am Main: Peter Lang.
- Rasmussen, Gitte and Kristiansen Dalby, Elisabeth. This volume. "The Influence of the Specificities of Gaze Behaviour on Emerging and Ensuing Interaction. A Contribution to the Discussion of the Use of Eye-Tracking Recordings for EMCA Analysis."
-  Landis, J. Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1): 159–174.
-  Laurier, Eric. 2008. "Drinking up Endings: Conversational Resources of the Café." *Language and Communication* 28 (2): 165–181.
- Laurier, Eric, and Chris Philo. 2006. "Natural Problems of Naturalistic Video Data." In *Video Analysis: Methodology and Methods*, ed. by Hubert Knoblauch, Bernt Schnettler, Jürgen Raab, and Hans-Georg Soeffner, 183–193. Oxford: Peter Lang.
-  Laurier, Eric, and Tobias Boelt Back. 2023. "Recurrent Problems and Recent Experiments in Transcribing Video: Live Transcribing in Data Sessions and Depicting Perspective." In *Ethnomethodological Conversation Analysis in Motion: Emerging Methods and Technologies*, ed. by Pentti Haddington, Tiina Eilittä, Antti Kamunen, Laura Kohonen-Aho, Tuire Oittinen, Lira Rautiainen, and Anna Vatanen, 245–263. London: Routledge.
- Luckmann, Thomas. 2012. "Some Remarks on Scores in Multimodal Sequential Analysis." In *Video Analysis: Methodology and Methods*, ed. by Hubert Knoblauch, Bernt Schnettler, Jürgen Raab, and Hans-Georg Soeffner, 29–34. Frankfurt am Main: Peter Lang.

- doi** Lynch, Michael, and David Bogen. 1994. "Harvey Sacks's Primitive Natural Science." *Theory, Culture & Society* 11 (4): 65–110.
- McHugh, Mary L. 2012. "Interrater Reliability: The Kappa Statistic." *Biochem Med* 22 (3): 276–82.
- doi** McIlvenny, Paul. 2019. "Inhabiting Spatial Video and Audio Data: Towards a Scenographic Turn in the Analysis of Social Interaction." *Social Interaction. Video-Based Studies of Human Sociality* 2 (1).
- Mohn, Elisabeth. 2012. "Permanent Work on Gazes. Video Ethnography as an Alternative Methodology." In *Video Analysis: Methodology and Methods*, ed. by Hubert Knoblauch, Bernt Schnettler, Jürgen Raab, and Hans-Georg Soeffner, 173–182. Frankfurt am Main: Peter Lang.
- doi** Mondada, Lorenza. 2008. "Using Video for a Sequential and Multimodal Analysis of Social Interaction: Videotaping Institutional Telephone Calls." *Forum Qualitative Sozialforschung / Forum Qualitative Social Research* 9 (3).
- doi** Mondada, Lorenza. 2009. "Emergent Focused Interactions in Public Places: A Systematic Analysis of the Multimodal Achievement of a Common Interactional Space." *Journal of Pragmatics* 41 (10): 1977–1997.
- Mondada, Lorenza. 2012a. "Video Recording as the Reflexive Preservation and Configuration of Phenomenal Features for Analysis." In *Video Analysis Methodology and Methods*, ed. by Hubert Knoblauch, Bernt Schnettler, Jürgen Raab, and Hans-Georg Soeffner, 51–68. Frankfurt am Main: Peter Lang.
- doi** Mondada, Lorenza. 2012b. "The Conversation Analytic Approach to Data Collection." In *The Handbook of Conversation Analysis*, ed. by Jack Sidnell, and Tanya Stivers, 32–56. Oxford: Wiley-Blackwell.
- doi** Mondada, Lorenza. 2019. "Practices for Showing, Looking, and Videorecording: The Interactional Establishment of a Common Focus of Attention." In *Embodied Activities in Face-to-Face and Mediated Settings: Social Encounters in Time and Space*, ed. by Elisabeth Reber, and Cornelia Gerhardt, 63–104. Cham: Palgrave Macmillan.
- doi** Oloff, Florence. 2022. "The Particle *jako* ('Like') in Spoken Czech: From Expressing Comparison to Mobilizing Affiliative Responses." *Frontiers in Psychology* 12 (662115).
- doi** Pehkonen, Samu, Rauniomaa, Mirka, and Pauliina Siitonen. 2021. "Participating Researcher or Researching Participant? On Possible Positions of the Researcher in the Collection (and Analysis) of Mobile Video Data." *Social Interaction. Video-Based Studies of Human Sociality* 4 (2).
- doi** Pekarek Doehler, Simona, Polak-Yitzhaki, Hilla, Li, Xiaoting, Stoenica, Ioana Maria, Havlík, Martin, and Leelo Keevallik. 2022. "Multimodal Assemblies for Prefacing a Dispreferred Response: A Cross-Linguistic Analysis." *Frontiers in Psychology* 12 (689275).
- doi** Peräkylä, Anssi, and Johanna E. Ruusuvaori. 2012. "Facial Expression and Interactional Regulation of Emotion." In *Emotion in Interaction*, ed. by Anssi Peräkylä, and Marija L. Sorjonen, 64–91. Oxford: Oxford University Press.
- Rossano, Federico. 2013. "Gaze in Conversation." In *The Handbook of Conversation Analysis*, ed. by Jack Sidnell, and Tanya Stivers, 308–329. Oxford: Wiley-Blackwell.
- Sacks, Harvey. [1967] 1992. *Lectures on Conversation*. (2 Volumes). Oxford: Basil Blackwell.

- doi Schmitz, H. Walter. 2020. "Wenn der Hörer sichtbar wird. Ch. Goodwins „The Interactive construction of a Sentence in Natural Conversation“ und die ethnomethodologische Konversationsanalyse." In: *Sine ira et studio: Disziplinübergreifende Annäherungen an die zwischenmenschliche Kommunikation*, ed. by Robin Kurilla, Karin Kolb-Albers, Hannes Krämer, and Karola Pitsch, 97–115. Berlin: Springer.
- Schutz, Alfred. 1972. "The Problem of Social Reality." In *Collected Papers. Volume 1*, ed. by Maurice Natanson. 277–293. The Hague: Martinus Nijhoff.
- doi Stivers, Tanya. 2021. "Is Conversation Built for Two? The Partitioning of Social Interaction." *Research on Language and Social Interaction* 54 (1): 1–19.
- doi Stoenica, Ioana-Maria, and Sophia Fiedler. 2022. "Multimodal Practice for Mobilizing Response: The Case of Turn-final *tu vois* 'You See' in French Talk-in-Interaction." *Frontiers in Psychology* 12: 249–268.
- doi Stukenbrock, Anja. 2015. *Deixis in der face-to-face-Interaktion*. Berlin, Boston: de Gruyter.
- doi Stukenbrock, Anja. 2018. "Forward-Looking: Where Do We Go With Multimodal Projections?" In *Modalities and Temporalities: Convergences and Divergences of Bodily Resources in Interaction*, ed. by Arnulf Deppermann and Jürgen Streeck, 31–68. Amsterdam: Benjamins.
- doi Stukenbrock, Anja. 2020. "Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention." *Frontiers in Psychology*, 11.
- doi Stukenbrock, Anja. 2021. "Multimodal Gestalts and Their Change Over Time: Is Routinization Also Grammaticalization?" *Frontiers in Communication*, 6 (6622406).
- doi Stukenbrock, Anja, and Anh-Nhi Dao. 2019. "Joint Attention in Passing: What Dual Mobile Eye Tracking Reveals About Gaze in Coordinating Embodied Activities at a Market." In *Embodied Activities in Face-to-face and Mediated Settings: Social Encounters in Time and Space*, ed. by Elisabeth Reber, and Cornelia Gerhardt, 177–213. Cham: Palgrave Macmillan.
- Stukenbrock, Anja and Balantani, Angeliki. This volume. "When the Establishment of Joint Attention Becomes Problematic: How Participants Manage Divergent and Competing Foci of Attention."
- doi Tomasello, Michael. 2008. *Origins of Human Communication*. Cambridge: MIT Press.
- doi Tomasello, Michael, Hare, Brian, Lehmann, Hagen, and Josep Call. 2007. "Reliance on Head versus Eyes in the Gaze Following of Great Apes and Human Infants: The Cooperative Eye Hypothesis." *Journal of Human Evolution* 52 (3): 314–320.
- Tuncer, Sylvaine, and Pentti Haddington. 2019. "Looking at and Seeing Objects: Instructed Vision and Collaboration in the Laboratory." *Gesprächsforschung — Online-Zeitschrift zur verbalen Interaktion* 20: 435–360. www.gespraechsforschung-ozs.de
- Wittenburg, Peter, Brugman, Hennie, Russel, Albert, Klassmann, Axel, and Han Sloetjes. 2006. "ELAN: a Professional Framework for Multimodality Research." In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, 1556–1559. Genua: European Language Resources Association (ELRA).
- doi Zhisheng, Wang, Nagai, Yukari, Zhu, Dan, Liu, Jiahui, and Ninyu Zou. 2019. "Based on Creative Thinking to Museum Lighting Design Influences to Visitors Emotional Response Levels Theory Research." *IOP Conference Series: Materials Science and Engineering* 573: 1–7.
- doi Zima, Elisabeth. 2020. "Gaze and Recipient Feedback in Triadic Storytelling Activities." *Discourse Processes* 57 (9): 725–748.

The influence of the specificities of gaze behavior on emerging and ensuing interaction

A contribution to the discussion of the use
of eye-tracking recordings for EMCA analysis

Gitte Rasmussen & Elisabeth Dalby Kristiansen
University of Southern Denmark

The integration of new technologies in Ethnomethodology and Conversation Analysis necessitates thorough discussion. This chapter explores the combination of recordings from a mobile eye-tracking device with recordings from an external mobile video camera, which may reveal intricate details of human activities. Focusing on customers' actions and interactions with salespersons, the chapter demonstrates how even brief observations made by customers, as captured by the eye trackers, are significant in understanding their subsequent actions when navigating amongst one another. In addition, it illustrates how customers' initiations and responses to salespersons' initiations of talk are to be understood in the context of the specificities of their prior observations. The main point emphasized is that eye-tracking recordings, along with video recordings from external cameras, capture essential behavioral nuances, leading to re-specifications of aspects of social action and interaction.

Keywords: eye-tracking data, video recordings, search activities, observations, conversation initiation, embodied actions

1. Introduction

Ethnomethodological Conversation Analytic (EMCA) studies aim to discover and describe participants' sense-making methods in and as participation in social interaction (Garfinkel & Sacks 1970, Mondada 2011). Audio and video recordings have proven to be invaluable resources for achieving this aim: EMCA analysis describes

members' unfolding understanding in interaction as it is publicly demonstrated in and through their actions which are witnessable and observable for co-present others, including analysts. EMCA analysts' repeated scrutiny of sequences of interaction, which is an essential part of the EMCA analytic practice, is made possible by the availability of video and audio recordings (Sacks 1984).

How best to use audio and video technologies to gain access to participants' perspectives and orientations has been and is still discussed within the field, from methodological discussions of video recordings (e.g., Heath et al. 2010, Knoblauch et al. 2006) and the technologies used to record and analyze video data (McIlvenny 2019) to studies of participants' use of video recordings as an object of analysis (Goodwin 1994) and researchers' practices for doing video recording (Mondada 2019a). These discussions continuously challenge existing paradigms and methods and contribute to the methodological development of the field, opening new avenues of EMCA research by, e.g., extending the analytic focus to multimodal and multisensorial interaction (e.g., Mondada 2019b, Deppermann 2013b), or exploring the possibilities of increasingly light and mobile cameras for studies of mobile interaction (LaBonte et al. 2021, McIlvenny et al. 2014).

The emergence of mobile wearable eye-tracking devices likewise provides new possibilities for data collection and analysis (Brône & Oben 2018, Stukenbrock & Dao 2019): Video recordings from the eye-tracking systems' built-in camera mounted on glasses, overlaid with visualizations of the wearer's eye movements and fixations, e.g., cursors in the form of red circles and lines, provide detailed information about the wearer's eye movements and fixations. However, as discussed in this volume, including eye-tracking data in EMCA analysis poses several methodological and analytic challenges.

In a paper (Kristiansen & Rasmussen 2021), we discussed how eye-tracking systems, originally developed for experimental studies (Wade & Tatler 2005), may be utilized for conducting EMCA analysis of multimodal interaction without losing sight of EMCA research interests and without distorting the data and methods necessary and relevant for this research framework (Garfinkel & Wieder 1992). We demonstrated that eye-tracking data may relevantly be used *if* they are fitted into an EMCA analysis of a local and social context, using video data from an external camera.

In this chapter, we will continue the discussion. It is based on a methodological and analytic concern with avoiding focus on gaze behavior at the cost of the ecology of the organization of actions which gaze behavior is part of, and avoiding the uncritical use of information obtained through 'objective,' 'scientific' methods to achieve insight into the organization of social interaction as oriented to by the participants.

We will show how cameras mounted on glasses capture details in individuals' visual field and how overlaid visualizations of fixations may provide information about individuals' observations which are relevant to the organization of their subsequent actions and interactions. Moreover, we will show how this kind of information may contribute to understanding that conduct and enable us to respecify aspects of social actions and maybe even to describe social phenomena previously unavailable to EMCA analysis. Specifically, we will show how what individuals may have a glimpse of, may notice, or observe, in short, perceive,¹ in their surroundings influences subsequent conduct.

Although this chapter is a contribution to a discussion of how recordings that track eye movements (in short eye-tracking recordings) may be used in EMCA analysis, we maintain that interactions should be recorded with an external video camera and that eye-tracking recordings should be analyzed in combination with these external recordings to preserve the core EMCA research interests and methodologies. As the analyses exhibit, they are derived from a combination of external video recordings and eye-tracking recordings, with one exception. This Example (5) illustrates the limitations of analyzing eye-tracking recordings without external video recordings that reveal the wearer's embodied conduct in the physical world.

1.1 Research on pre-activities and pre-sequences

This chapter's EMCA analytic interest concerns the relationship between participants' actions and interactions and their perceptions in the surroundings that preface them. Previous EMCA research on prefaces and pre-beginnings of interaction is extensive.

Studies by for instance De Stefani and Mondada (2010, 2018), Kendon (1973), Mondada and Schmitt (2010), and Müller and Bohle (2007) demonstrate how prospective conversationalists arrange and continuously rearrange interactional spaces through embodied conduct. They show how conversations may be opened step-by-step as co-participants approach one another, achieve a first mutual eye contact, and initiate a first turn at talk (see also Kendon & Ferber 1973, Oloff 2010, Rasmussen 2023a, Schmitt & Deppermann 2007).

In a further study, Mondada (2009) shows how the first turn of an encounter is finely designed with respect to the co-participants' walking bodies (see also Broth & Mondada 2013). In addition, De Stefani (2014) and Stukenbrock (2018)

1. Throughout this chapter, we will use 'perception' and 'perceive' to denote what participants may observe or notice in the physical environment.

show how couples in markets and supermarkets coordinate embodied conduct and initiate talk in relation to product displays which, as it were, spawn topics of talk.

Deppermann and colleagues (Deppermann et al. 2009) show how walking into sight works to attract attention, and Mondada (2007) shows how pointing may serve the same purpose, whereas Day and Rasmussen (2019) show how the manipulation of objects is associated with entitlements to speak. Deppermann and colleagues (2009) furthermore show how the manipulation of a folder in a meeting works to attract attention.

Schegloff (1996) shows how audible inbreath may be used to indicate speaker-ship and preface Turn Construction Units' (TCU) beginnings, as may discourse particles (Schegloff 1987), and Deppermann (2013a) makes the point that any turn-at-talk at any moment in interaction is prepared, and needs being prepared, multimodally, before being produced and designed for a recipient to respond to it (see also Lindstrøm 2006).

Finally, EMCA studies have demonstrated how prospective conversationalists and participants in ongoing conversations coordinate gaze behavior before initiating a TCU (see for example Goodwin 1980, 1981, Rossano 2012, Rossano et al. 2009). Auer (2018, 2021) and Stukenbrock & Dao (2019) add important insights into CA research in gaze behavior by describing the details of how mutual gaze is accomplished, by analyzing the co-participants' eye movements and fixations as recorded by an eye-tracker device.

This study contributes to the field, as it describes instances of how embodied action in interaction, conversation initiating turns and responses to them, and pre-segments to TCUs are related to participants' perceptions in the physical environment.

2. Data collection

The data analyzed in this chapter is part of a corpus of recordings collected for the Velux-funded research project RESEMINA (The Digital (Re)semiotization of Buying and Selling Interactions). The corpus consists of approx. 30 hours of video and eye-tracking recordings of online shopping and shopping in brick-and-mortar shops and supermarkets in Denmark, collected in 2018/2019.

The data were collected with the written informed consent of the participants prior to the recordings, including consent to publish transcripts and images of the recorded interactions. The data are managed and stored in accordance with Dan-

ish law and EU regulations as sanctioned and monitored by the Data Protection Office of the University of Southern Denmark.²

The chapter analyzes and discusses data from brick-and-mortar shops and supermarkets. All participating customers were approached outside the shops and asked to participate in the research project. They thus had an errand in the shops independently of the research project. Shop personnel were informed about the recording and agreed to participate before any customers were approached.

To make the eye-tracking recording, the participating customer wore a pair of Tobii Pro Glasses 2 eye-tracking glasses while shopping. The eye-tracking glasses were calibrated immediately before entering the shop. Other customers and members of staff whom the participating customer might get in contact with were not equipped with eye-tracking glasses. Such a design (i.e., using more than one pair of glasses) would necessitate assumptions about who that customer would come to interact with, or it would impose on the customer whom to interact with, whereas we designed the study to sustain as much naturalness of the setting as possible. Because of our choice to use only one pair of eye-tracking glasses, if a pair of customers agreed to participate, it was only possible to equip one of them with eye-tracking glasses (see Kristiansen & Rasmussen (2021) for a discussion of the selection process).

As noted in the introduction, due to our EMCA focus on co-participants' observable (Garfinkel 1967), situated, embodied and multimodal actions and interactions, we recorded the shopping experience using an external mobile video camera. The researcher operating the camera moved with the wearer through the setting. We manually synchronized these video recordings with the eye-tracking recordings as part of the analytic process using Movavi Video Editor. To illustrate the behavior of analytical interest, we will present stills from the external video camera alongside corresponding stills from the eye-tracking recordings. The eye-tracking system records the wearer's eye movements and calculates gaze location over time (see the introduction to this volume), among other things, while simultaneously capturing the environment in front of the wearer of the portable eye-trackers through its built-in camera. It identifies periods during which the eye remains still, known as 'fixations' (Carter & Luke 2020). These fixations are represented as a cursor shaped as a red circle overlaid on the video recorded by the eye-tracking camera. However, in some cases, the quality of the video data acquired by the eye-tracking camera lacks the necessary quality for the equipment to accurately compute eye movements and fixations. Therefore, recordings

2. In Denmark, approval from centralized ethical committees is not required for the use of video data as applied in the present study.

obtained through the eye-tracking system may lack indications of eye movements and fixations. Also, due to the synchronization of recordings from the external camera and the eye-tracking recordings, it is not always possible to include frames that capture the wearer's brief fixations at specific moments. We use such recordings and frames as evidence to demonstrate our analytical points as well.

Stills from the external video recordings are indicated by Figure (X), while corresponding stills from eye-tracking recordings are indicated by Figure (X⊙). The stills from eye-tracking recordings are pinned to the stills from the external video recordings. This is done to prevent that eye-tracking material is treated as analytically independent from the wearer's embodied engagement with their multimodal environment as captured on the external video recording. Each still from the eye-tracking recordings is placed so that it does not cover important details in the still from the external video recordings. In examples where the co-participants also engage in talk in interaction, the stills are incorporated into transcripts of the talk and embodied conduct (see transcription conventions in the Appendix of this chapter). In this case, the stills are referenced as # (X) for external video recordings and # (X⊙) for eye-tracking recordings.

How the different types of recordings are used will be pointed out and discussed as part of the EMCA analyses presented in the sections below.

3. Customers' perceptions and their relation to subsequent embodied conduct

Actions are carried out in relation to someone and/or something that reveals itself as the action unfolds, and how the actions make sense may be worked out based on various assumptions about that relation. When, e.g., a customer walks up to a shelf with product displays and picks up a product for further examination, co-present others may make assumptions that this product was aimed for at the outset. An alternative understanding may be, of course, that the customer moved toward the shelf and the product caught their attention and interest, and so they it picked up. The difference may be unimportant for understanding that and how the customer walks up to the shelf and examines the product, but it may in some cases be crucial if we want to understand the social significance of the customer's navigation in the environment to get to the shelf and the product.

In Example (1) (below), customer, A, who is wearing eye-tracking glasses, enters a shop to browse. At one point she stops, facing a shelf with product displays.

Example 1.³

#1



#1 ⊙ ↓



As is visible from the external video recording (#1), she stops more specifically in front of a table next to a group of other customers who are seated at a coffee table (to the left in #1). Figure 1 also shows that she gazes in the direction of the shelf behind the table, i.e., her attention is drawn towards something in this area. The external video recording does not allow for describing exactly what her attention is drawn to. To obtain this information, it is necessary to consult the eye-tracking recording (#1⊙) which reveals that she focuses on (fixates) a yellow bowl placed next to the other customers. As it turns out, her noticing of the bowl on this shelf is consequential for understanding her subsequent embodied behavior. Also, this behavior is analyzed through a combination of the two data types:

3. This example is used in Rasmussen et al. (2024)

#2.



#2⊙

#3



#4



↓#3⊙

#4⊙



A walks around the table (#2 and #3) as she focuses on the bowl, as indicated by the fixation (#3⊙). She then bends forward (#4) to pick it up (#4⊙).⁴ In other words, she navigates for a specific purpose: to get closer to just this yellow bowl. As is visible to everyone in the area, a path runs between the table and a lounge section to the bowl (See above, #1⊙), though a dining chair has been moved slightly away from the table, partly blocking the path. A, however, chooses a path around the table to it. Her action may well be responsive to the location of other

4. Although it is not important for the point made in this analysis, notice that the still from the eye-tracking recording does not exhibit the customer's fixation. Based on a combination of the recordings, it is likely, though, that she gazes at the bowl in this moment.

customers occupying the lounge section. According to Kendon (1990), interactional spaces are formed as co-participants and co-located others take the co-participants' personal spaces into account. The personal space includes personal belongings and a space in front of and behind the back of the participants.

Walking around the table is thus a practical solution to the organization of conduct in an interplay with the physical environment and co-present others. As made describable by the information about customer A's gaze conduct in the eye-tracking recording in addition to the information about her bodily movement in the shop that is accessible in the external video recordings, she takes a *detour* to get around obstacles, that is, the seated customers in front of her, and *avoids* making the customers move to reestablish a personal space and interrupting their activity as she works her way toward the bowl.

Example 2⁵ below serves as material for another analysis of an action of avoidance which is made possible by adding eye-tracking recordings to external video recordings. In this example, the action serves to give another customer time to get out of the way.

Customer, A, wearing eye-tracking glasses, walks along an aisle in front of a man in a blue shirt, customer B (#5). B obstructs the camera's view of A as it follows her. Thus, only her right leg (#6 and 7) and the left wheel of her trolley (#5) are visible. However, the stills from the eye-tracking recording reveal how A moves in space (#5○-8○). They also reveal A's gaze direction through cursors. According to the cursors, A gazes in the direction of biscuits on shelves to her right behind yet another customer (C) in a red shirt (#5○) who has walked towards the main aisle which A and B are walking along and has stopped (#6) while attending to a small piece of paper in his hand. Although this is captured on the external video recordings, it is much clearer in the eye-tracking recording. Finally, the stills from the eye-tracking recording show how A's gaze wanders from the biscuits to C (#6○), back to the biscuits (#7○), and then back to C (#8○). The sequence of her gaze behavior is significant to understanding her subsequent embodied action: She walks past C (#7-8, #7○-8○).

5. This example was used in Rasmussen & Kristiansen (2022).

Example 2.

#5



↓#5⊙



#7



↓#7⊙



#6



↓#6⊙



#8



#8⊙

In the context of her gaze behavior, captured and described with the use of the cursors on the eye-tracker recordings, A's walking past C is understandable as a response to C occupying just this space in front of the perceived biscuits for doing his business. C's position (and occupation) is an obstacle to A using the possibility to walk to the biscuits, because it would interfere with C's space and interfere with his activity. Thus, she avoids entering his space, at least for the moment. Her following conduct aligns with this. Importantly, however, the analysis and understanding of this action is achieved *as it happens*, not in hindsight. This was also the case in Example 1 when understanding A's detour to reach the yellow bowl.

In Example 2, A proceeds past C, but after a few steps she stops and leans onto her right foot (#9). This is visible only on the external video recording, whereas the eye-tracking recording reveals how she gazes in the direction of other biscuits (#9⊙). This lasts for 0.3 seconds. Next, A leaves the trolley and walks back to the biscuits that she had noticed before passing C (#10, #10⊙).

#9



#9⊙

#10



#⊙10

Walking a few steps further, stopping, and leaning on to her right foot serves a specific purpose: A gives C time to finish his business and avoids, for another moment, working her way around him, interrupting his activity.

In conclusion, a combination of the data available in terms of external video recordings and eye-tracking recordings with registered fixations allows for an analysis of how co-participants' embodied conduct indicates avoidance, that is, an analysis and understanding of it as it unfolds rather than retrospectively.

In the remainder of the chapter, we will show how not only fixations but also what the eye-tracking recordings do not register as fixations (for reasons which we will point out as we unfold our analysis) add new details to the information provided by external video recordings about what influences the design of turns of talk.

4. Customers' perceptions and their relation to sequence initiations and responses

As shown in Stukenbrock & Dao (2019), and Auer (2018, 2021) for seated configurations, understanding customers' specific gaze behavior before sequence initiation and responses to sequence initiating turns may add new insight into the design of such turns and the action they accomplish. Our research and findings add to this body of research. De Stefani (2014) shows how product displays may serve as topics in multimodal interaction (see also Rasmussen 2023b). We begin our analyses by briefly illustrating how the specificities of the customers' perceptions of the surroundings, e.g., the features of multimodally designed products (Kress 2010), influence the design of such topic-initiating turns. We then proceed to analyze how customers' perceptions may as a matter of fact influence the design of responding turns to sequence initiations as well. For this purpose, we focus on customers' responses to salespersons' offers to help (Section 4.1).

In Example 3, a pair of shoppers are moving through a bookshop, browsing shelves with product displays. In the process, they notice and comment on various products (see also Stukenbrock & Balantani, Chapter 9, this volume). At the beginning of the extract below, A, wearing eye-tracking glasses, and B are moving along shelves with product displays (#11). B is walking in front of A who blocks B from view in the video recording. A gazes at a small red book with '*adressebog*' ('address book') written in gold letters displayed on the top shelf to her left (#11 ☉). B, standing in front of A, is looking at products on the lower shelf to her right with her back to the shelf displaying the book, without visual access to either the book or to A's gaze at this moment. While B is still blocked from the view of the external video camera, this information is provided by the eye-tracking recording. B (with the backpack) is visible in #12 ☉.

Example 3.

#11



#11 ☉

#11◉

A gazes at the book for 0.5 seconds before initiating a sequence:

1 A: *ej (.) jeg ku virkelig godt tænke mig en adressebog=*
 'oh I would really like an address book'
 >small red book
 #110

2 A: *=ku det ik være [sødt*
 'wouldn't that be sweet'
 >B
 #12◉

#12



#12◉

3 B: [kontorma=
 'office files'

4 B: *=hvaffor noget*
 'what'
 B straightens up

5 B: [at ha en adressebog
 'to have an address book'
 B turns towards the shelves on the other side
 #13

#13



#13◉

6 A: [at ha en adressebog
 'to have an address book'
 > the address book
 #130

7 A: *prøv at se*
 'look'

A points at the address book
 >the address book
 #14, #14◉

#14



#14◉

Of specific interest here is that while the external video recording shows that A gazes toward the shelves to her left, the eye-tracking recording reveals that A fixates on the address book for 0.5 seconds (#11◉) before telling B that she would really like an address book (line 1).

Specifically, she uses the indefinite article ('an address book', line 1) to announce the new topic, gain joint attention towards it, and elicit talk about it (lines 5–7). Extending her turn, A asks B whether she agrees that having an address book would be 'sweet' (line 2). The book is small, its color red, and it is covered in leather with gold lettering (see #11◉), all of which make it stand out against the surrounding stationery (Rasmussen et al. 2024) and draw attention to further inspection. Moreover, it appears to prompt an evaluation that suggests cuteness, intimacy, and a sense of nostalgia, i.e. 'sweet'. A incorporates this assessment into her topic and sequence initiating turn, which builds on and references the specific multimodal features of what she specifically looked at for 0.5 seconds. Access to these features and her gaze upon them is obtained through eye-tracking recordings, as illustrated.

4.1 Search activities and their relation to recruitment sequences

Browsing and searching activities may preface shop encounter openings and constitute the environments in which help is requested or offered (Kendrick and Drew 2016). As demonstrated by the analyses below regarding the sequential organization of the encounter, openings by salespersons may in fact intervene in the customers' search activity. Consequently (some of) the particulars of that activity shape the customers' accepting response to the salespersons' offer.

Example 4 (below) serves as an illustration of how the utilization of eye-tracking recordings in addition to the external video recordings furnishes us with

information that enables the aforementioned analyses and findings. A mother, B, wearing eye-tracking glasses, is searching for a book for her son. The video recordings show how she walks along shelves with books at a rather quick speed, going from one to the other across sections with specific genres (#15- #17). In addition, the eye-tracking recordings show how she lets her gaze wander over the books without focusing on anything specific (#15○-#17○). Eventually, she stops at some shelves (#18-#19), but her gaze keeps wandering across the book displays (#18○-#19○). Due to the synchronization of the recordings from the external video and the eye-tracking recordings, it was not possible to include frames that display B's brief fixations at this moment. The activity comes off as a search in the kids' book section, but as an unsystematic search in relation to topic and genre:

Example 4.

#15



#15○

#16



#16○

#17



#17○

#18



#18○

#19



#19○

The transcript begins as the salesperson, A, who comes into view later and is initially only visible on the eye-tracking recording (#22○), intervenes in the search activity. At this moment, B bends forward (#20):

#20



#20○

- 1 A *har du brug for lidt hjælp*
'do you need a little help'
2 B *Øh øh øh he he h.jh.a*
'eh eh yes he'
Δturns around tw A
#21

#21



#21○

B><A

#22

#22

#22



#22○

3 A [he he]
 'he he'
 smiles
 *>customer-->
 takes a step tw customer

4 B [he he] Δøh øh øh det er det lige nu der er jeg fuldstændig=
 'he he eh he eh well right now I am totally'
 Δ>to her right
 =lost fordi min søn han viste mig en bog .hhh Δhan ønskede sig
 'lost because my son showed me a book that he wanted'
 Δ*turns to her left-->

In line 2, A offers help: 'do you need a little help'. As visible in the external video recording (#21), B responds by turning around. As revealed exclusively, and beyond any reasonable doubt, by the eye-tracking recording with no cursors, she turns around to face A (#22○). Turning around, she accepts the offer (line 2). Note that the still of the eye-tracking recording (#21○) represents one frame out of a sweeping motion. The sweeping motion is continuous and therefore, as opposed to what the still may suggest, the frame does not represent a moment at which B's eyes stabilize. Note also that the researcher operating the camera steps backward as B turns around, which leads to B being only partially visible in #22.

Of analytic interest is especially how B's acceptance of help is postponed by a series of 'eh' and a laugh particle. They indicate that B may not be able to deliver the expected answer. Whereas research in conversation analysis has revealed how pre-segments may relate to previous turns at talk (see e.g., Schegloff & Lerner 2009), these pre-segments of B's turn aggregate and hint at the specificities of the multimodal search activity in which B is engaged when the salesperson intervenes, and the encounter commences. The search leads the customer to stand facing shelves while her gaze is wandering across the displays, jumping from one section and genre to the next. B's behavior is understandable as a search basically without knowing what she is looking for. Thus, the customer not only aligns with the salesperson's action as she accepts her offer to help but simultaneously relates her action to the details of her search activity so far.

In contrast to Example 4 above, Example 5 (below) shows how a salesperson offers help *after* the customer's local search has ended and she has left the search area. This influences how the offer and the acceptance of help are made. The specificities of what the customer was looking for before leaving the space influence the construction of the customer's acceptance as well. What she specifically looked at and searched for is revealed by the eye-tracking data.

In the example, the customer, B, wearing eye-tracking glasses, is looking at teapots. She picks up a specific small one and then puts it down to grasp one in a box. She turns the box around as illustrated in the external video recording (#23) and the eye-tracking recording (#23○). In addition, the eye-tracking recording reveals how her gaze wanders over the text at its bottom, and it shows that the

text is in English and Chinese. The text addresses the materiality of the teapot, its price, and its size. B then turns the packaging around several times again (#24, #24⊙, #25⊙, #26⊙) while her gaze keeps wandering across pictures and texts on it as depicted by the eye-tracking recording. She seems, in other words, to search for something specific.

Example 5.

#23



#23⊙

#24



#24⊙

#25



#25⊙

#26



#26⊙

B then puts the package down, turns around, and walks away. The salesperson (A) passes by. The transcript exhibits how A initiates an offer to help as she turns around towards B from across the room (line 1):

- 1 A *Δdu skal altså bare sige til hvis du vil ha hjælp ik os*
'you just have to say if you want help okay'
- 2 B *Δja*
'yes'
Δturns around tw A
#27⊙
#270



- 3 A *du er velkommen*
'you are welcome'
- 4 B *øhm je- jeg ville bare spørge HVOR MANGE*
'ehm I just want to ask how many'
ØH HVA HEDDER DET HV- HVOR HVOR MANGE MI- MILLILITER=
'eh what is it called ho-how how many milliliters'
- 5 B B approaches A
- 6 A *=der ku være I Δden he:r*
'can be in this one'
Δreaches for teapot
- 7 A reaches for teapot
uhh
'uhh'
- 8 B withdraws her hand
- 9 A *[det ved jeg faktisk ik=*
'I actually do not know'
- 10 C [approaches A and B from a close distance
- 11 A *=jeg tror bare Δen kop*
'I think just a cup'
Δlifts cup up
#28^o
#28^{o6}



In line 1 A's offer to help interrupts B in walking away. A speaks in a slightly raised voice considering the distance between them (see #27^o) and constructs the offer with an emphasis on the fact that B has the possibility to recruit help ('just/'*altså*') (line 1). B acknowledges the offer (line 2), which A emphasizes for a second time 'you are welcome' (line 3). As was the case with the eye-tracking recordings in Examples 1 (#4^o) and 4 (#18^o and 19^o), the still from this eye-tracking recording (#27^o) is not overlaid with a cursor. Whereas the lack of cursors in (#4^o) and (#18^o and #19^o) are results of the synchronization, in #27^o the quality of

6. Unfortunately, there is no external video recording corresponding to the eye-tracking recording at this moment because there were some technical issues with the external video camera.

the recording from the video camera in the eye-tracker equipment is not good enough for the fixations to be calculated. Still, as in Examples 1 and 4, we take it to be more than likely that B gazes in the direction of A in line 4 as A gazes in the direction of B.

Through her offer (line 3), A pursues a request for help from B as an alternative to B walking away.

B responds by indicating, loudly, just what kind of help she needs (line 4): She needs to know how much the tea pot can contain. As in Example 4 above, B's request is informed by some of the particulars of her prior action: she uses the term 'milliliters' which is used in technical specifications that are usually found in texts on the packages. As a matter of fact, B searches for the term (line 4) which may indicate that she may not use the term on an everyday basis.

Also, of importance is if and how customers', in this case, B's, responses, influenced by the specificities of preceding search activities, influence the assistance provided by the salespersons. As it turns out, A does not know the answer to B's question (lines 7 and 9). Instead, she makes a guess (line 11). Rather than aligning with B's request that relates to the specificities of B's prior search activity, A indicates the number of cups, that is, one cup, that the teapot may contain. The indication of size in terms of number of cups is not equivalent to the indication of size in terms of milliliters, and 'cup' is not a unit of measurement in Danish as it is in English. A accompanies her talk by picking up the teapot. This is illustrated in the eye-tracking picture, #28○.

While A is delivering her answer (lines 9–11 above), however, another salesperson, C, approaches (line 10), which B, by the way, seems to orient towards. #28○ exhibits how she gazes between A and C, registered as a fixation on the rod upon which the shelves are mounted in the eye-tracker recording.

Only C's foot is visible in #28○. Due to technical challenges, the external video camera stopped recording, as evidenced in the still (#27○) from the eye-tracking recording at line 2 in the transcription, which reveals how the external video camera operator examines the camera. This has consequences for the information available to the analyst compared to what is available to the participants. External video recordings are inherently limited by the camera's sensor, which determines its angle of view, and it may of course not capture what specifically a participant focuses on. Similarly, the angle of view of the eye-tracking camera is also restricted (see also Auer & Laner, Chapter 8, and Botsch et al., Chapter 10, this volume) and, importantly, it does not capture the wearer's embodied conduct and engagement with the physical world. As shown in the examples above, having access to the participants' embodied conduct is crucial for describing the ecology of physically co-present face-to-face interaction.

In this example, only information from the eye-tracking recording is available. This is limited to A's hand and arm, C's foot, and B's visual field and a fixation. Information that might have been available from a recording from an external video camera must be inferred from these details, e.g., that A is standing to B's right, facing the shelf, that C is moving toward A and B, and that C is approaching from the counter which is right next to the shelf. While the video camera was operational, it captured how C moved past B to this area and greeted B during B's initial inspection of the products (not in the transcript).

Of analytic interest is how C, who has been in the vicinity and within hearing distance (Goffman 1981) of A and B's exchange of talk (lines 1–9), stops next to the shelf as A's turn comes to completion. After the pause following A's answer, B reaches for a package on the shelf (line 14 below). Simultaneously, B accepts A's answer ('okay', line 13 below), but she does so with some delay (line 12):

Example 5. (continued)

12 Ps (0.7)

13 B [*>okay<*
'okay'

14 C [reaches for package, [grasps package
#29◉
#29◉



15 A [hva siger du
'what do you say'
>C

16 C *det må stå på pakken*
'it must be on the package'

17 A *nå ja ja det ku da godt være*
'oh yes yes it could well be'

18 B *hm*
'hm'

The pause (line 12) indicates that B approves A's answer while not fully accepting it, that is, the basis on which A makes her guess (by gazing at the teapot and holding/weighing it in her hand). C responds to this pause by reaching for the package, which indicates the initiation of a search for the requested technical

specifications, i.e., a search strategy that remedies A's strategy. Notice that A requests C's help (line 15) as C is moving her hand toward the package. Notice also that the request assumes that C has heard the prior exchange ('what do you say'). C responds that 'it must be on the package' (line 16), which provides an answer to A's question while at the same time accounting for C's reaching for the package.

Salesperson C delivers the requested help, originally requested from A, by aligning her action to the details of the customer's, B's, talk, which in turn was based on the details of the customer's previous search activity. Consequently, the salesperson, C, helps tie back to B's original search. A confirms the relevance of C's suggested search strategy (line 17), and it receives a confirmation from B (line 18).

Despite the disalignment of A's attempt to deliver help, B approves and does not initiate repair of it just like she approves and accepts C's (aligned) help. In fact, none of the customers in our data interfere with the salespersons' search methods once they have accepted the offer to help. That is, the sequential organization of the salesperson's offering (in this example as in our data material in general) and the customer's accepting help constitute a pivot around which the role of the customer as the searcher is changed into an information provider who enables the salesperson to do their job, that is, deliver help fitted to the customer's request which is touched off by the details of the customer's previous, so far unsuccessful, searching activities.

Our final example, Example 6, is meant to support this analytic point. Again, based on a combination of different types of recordings and eye-tracking data, it illustrates how the customer, B, provides information when requested by the salesperson, A, as part of A's search method. The information can, as a matter of fact, be inferred from salient features of the local environment in which B conducts his search, and thus from his search activity, which the salesperson intervenes in (line 1) and is accepted to participate in.

In this example, customer B, who wears the eye-tracking glasses, engages in searching amongst first red and then pink items. The transcription begins when B is standing in front of a shelf displaying pink items that are by convention childish and girlish. At this point, A offers her help (line 1). B accepts her offer through the phrase 'a small purse' (line 5). In response, A engages in unpacking this information in interaction with the customer (line 7 and line 14 (Example 6 (continued))).

Example 6.

1. A *Åer der ïnoget du leder efter?*
 'is there something you are looking for'
 Åstands a little away from B, slightly to B's left, bends her head
 on her left shoulder, >B's face
 #30, #300

#30



#30⊙

2. Ps (0.2)

3. B *en lille Δpung*
'a small purse'

Δturns slightly towards A, >A

#31

#31



#31⊙

4. Ps (0.5)

5. A *en lille Δpung*
'a small purse'6. B *ja h.=*
'yes'
steps back from A7. A *=til børn eller voksne*
'for children or adults'8. B *[Δbørn*
'children'

Δ>the shelf with pink products

#32, #320

#32



#32⊙

9. Ps (0.3)

10. A *ja-*
'yes'

[turns to the right, lifts right arm pointing to the right

11. B [turns around almost simultaneously

12. Ps (0.5)

13. A (*denne vej*)
(‘this way’)

As is visible on the external video recording, A approaches B, orienting toward the framework of observation that B has established by standing in front of the shelves: A stops to B’s left and leans slightly toward him, bending her head toward her left shoulder (#30). In that way, A works to enter B’s peripheral field and omits to stand in his way before offering her help ‘is there something you’re looking for’ (line 1).

B’s answer ‘a small purse’ (line 3) agrees to A’s offer of assistance, outlining the kind of help that he needs and aggregating the specificities of what B was doing when A intervened.

The search is put on hold while he answers. As visible only on the external video recording, B’s bodily position (legs and torso) remains directed toward the shelves while his head turns slightly toward A. Our access to information on B’s bodily position and posture, which A also has access to, permits an understanding of how B not only accepts A’s offer but lets her into his framework of observation. The eye-tracking recording reveals (as does the external video recording) that B gazes at A. A confirms B’s request (line 5). B affirms and takes a step back (line 6), which indicates that he is making room for A to step toward him and the shelves.

A asks B whether he is looking for a purse for a child or an adult (line 7). This question does not take into consideration the specificities of how B has arranged himself in relation to the pink items, nor to his turn at talk, i.e., his acceptance, which is tailored to him standing in just that way in just that place – in front of the ‘pink’, ‘childish’ items – ‘a small purse’ (line 3). Nevertheless, B provides the

requested information. He tells her that it is for a child, which he underlines by redirecting his gaze toward the pink items (line 8, #32, #32○). Notice that his gaze direction and bodily behavior are visible on the external video recordings. However, his focus on the pink items is only available on the eye-tracking recording, indicated by the cursor.

Instead of entering the search space arranged by B, A directs B to other shelves, thereby remedying his search strategy. Standing in front of these shelves, she asks him about the gender of the recipient of the purse:

Example 6. (continued)

14. A *er det en dreng eller pige*
 ‘is it a boy or a girl’
 puts her left hand on the shelf
 >B

15. B *Δdet er en pige=*
 ‘it is a girl’
 Δ>row with pink purses
 #33, #33◉

#33



#33○

16. A [*Δde:t en pige ja*
 ‘it’s a girl yes’
 Δflips through a row of purses

17. B [*reaches for a pink purse*

A’s request for information (line 14) is not responsive to what seems to be significant details of the pre-activities, which are also indicated through B’s responsive actions to her offer (he stepped back and gazed at ‘pink’, ‘childish’ items) and re-instantiated through his subsequent responses to her requests for information. Once again, he delivers the requested information, that it is for a girl (line 15), while gazing toward pink purses on the shelf in front of him as indicated by the cursor showing a fixation on a pink purse. This is illustrated in the still from the eye-tracking recording (#33○). Finally, he picks a pink purse for the girl (line 17). Information about B’s taking the pink purse is only available on the eye-tracking

recording, as indicated by the cursor, due to the position of the external camera in relation to the participants and the physical surroundings at this moment (#33).

Of specific analytic interest here is how B does not deliver all the information in his first answer to A's first offer to help ('is there something you are looking for', line 1) in terms of for example 'a small pink purse for a little girl'. Instead, he provides information that is assumed to be relevant and enough for A to assist him. "Relevant" and "enough" information is provided by B drawing on talk and the spatial and material specificities of the interactional space which they specifically arranged, as captured by the external video camera and the eye-tracking camera in combination when A offered her help in response to B's search activity (#30, #30○). That information, assumed by B to be "relevant" and "enough", is, however, subsequently unpacked through sequences of talk, but only *on demand*, so to speak. In that way, B leaves it to A to develop a search strategy, as A makes relevant what kind of information she needs at specific moments in time during the process. The roles of the co-participants thus pivot around the acceptance of help. In this example, the salesperson basically resets the entire search process.

5. Discussion

As described in this chapter, actions in interaction may be responsive to participants' specific perceptions of specific items and their features or actions and events. They may influence customers' packaging of topic-initiating turns, and they may influence how customers package their responses to salespersons' offers to help search for (finding) products. Whereas the salespersons' offers are based on customers' demonstrably doing looking for something, the customers' responses, which work to request help, aggregate some of the specificities of "looking for something". In contrast, other specificities may still be exhibited, e.g., frameworks of observation indicated through bodily positioning. As exemplified in this chapter, the salespersons may align with customers' responses and their details pertaining to the specificities of their prior search activities. Conversely, they may disalign and remedy them and, in certain instances, even initiate a complete reset of the entire search process. Furthermore, our analyses show how the salespersons' responses in any of these forms serve as a pivot around which the customers' activity shifts from being an individual search to becoming an effort that they assist in.

Finally, the chapter's analyses demonstrate how embodied actions not only provide the context for subsequent embodied actions and/or talk but may themselves be responsive to particular perceptions of specific items and their features and surroundings.

These findings were achieved through analyses of data obtained using two types of technology, one well-known, i.e., ‘external’ mobile video cameras, and one comparably new, i.e., mobile eye-tracking glasses. The chapter has shown and discussed how these technologies and the recordings and data they generate may be used to conduct EMCA analysis of multimodal interaction among participants engaged in everyday activities:

The recordings from the external video camera provide access to the participants’ embodied interaction and conduct in time and space (see also e.g., De Stefani and Mondada 2014). However, the use of mobile external cameras necessitates ongoing local decisions on how the camera should be positioned, as the direction in which co-participants (and researchers) move develops. And no matter where the camera is positioned, there will be blind spots in terms of space on the other side of the co-participant (from the camera’s perspective), and at longer distances, that are not visible on the recordings.

As illustrated in this chapter, the camera mounted on the eye-tracking glasses, on the other hand, captures what is in front of the co-participant, that is, in their visual field. In that way, the recordings from the eye-tracking glasses’ built-in camera solve (some of) the practical problems resulting from using an external mobile camera (e.g., Example 2, 3, 4, and 6 (#33, #33○)). In that way, the eye-tracker camera serves as a camera on par with a traditional external video camera. However, as amply demonstrated in Example 5, there are limitations to what they can show both in relation to the physical surroundings and other individuals and to the wearer’s engagement with both. This conduct is a crucial premise for carrying out EMCA research in multimodal interaction.

In our data, only one participant is wearing eye-tracking glasses. Intuitively, it might seem like a possible solution to the problem to equip a co-participant (or maybe even more) with a second pair of eye-tracking glasses. However, although this might provide additional information, it would not provide information on the second wearer’s positioning in the space in relation to the environment, unless the first wearer’s eye-tracking glasses capture this, for instance because the first wearer gazes in the direction of the second wearer at some distance. This also goes for what the second wearer’s glasses may capture. Furthermore, equipping more than one individual with eye-tracking glasses assumes that these individuals will come to be close enough to one another for interaction of interest to EMCA to emerge. In our research design, we strove to interfere as little as possible with the naturalness of the situation. For that reason, we did not impose any restrictions on whom to interact with by, e.g., equipping specific others (customers or salespersons) with eye-tracking glasses.

Eye-tracking recordings aim to provide information about what the wearer fixates on and when. This information *may* be relevant for analysis. By using this

type of information, we have described how what customers specifically fixated on and engaged with through their fixations (and other embodied conduct) influenced their turn initiation in ensuing interaction (Example 3) and how customers responded to turn initiations in such interactions by others who may have observed them “looking for something” or searching (Examples 4, 5, and 6). We also used eye-tracking recordings to describe details of customer responses as these interactions emerged and unfolded (Examples 5 and 6). As noted, however, fixations indicated on the eye-tracking recordings are calibrations that register some fixations, and what is categorized as fixations depends on the settings of the eye-tracking software (see also Krug, Chapter 6: 216–217, this volume). In this chapter, we have used information on fixations to describe situations where participants have responded to what was fixated on in the recording as indicated by the cursor.

However, as demonstrated in this chapter as well as in other works (e.g. Rasmussen 2023b; Rasmussen and Kristiansen 2022), we do not use fixations to exclude the possibility that participants may see, notice, and respond to items or co-present others in their visual field, which the participants did not fixate on as displayed by the cursor in the eye-tracking software, either because they fixated on something else (see e.g. Example 1),⁷ because their fixations were not exhibited in specific frames resulting from synchronization (see e.g. Example 4, #18○ and #19○), or because the quality of the data obtained by the eye-tracking camera was not adequate to compute their fixations (see e.g. Example 5, #27○). In conclusion, the type of data obtained from the eye-tracking equipment is not crucial for determining whether they are utilized for analysis or not; what holds significance is how they correlate with the external video recordings and what they reveal.

It is of utmost importance for understanding our work and the way we use our data to keep in mind that whatever the type of data, data combination, and generated information, we seek to describe what the wearer of the eye-tracking glasses or co-present others demonstrably respond to, rely on, and draw upon in interaction. As shown in this chapter, the combination of external video recordings and eye-tracking recordings (with overlaid cursors indicating fixations) permits, for instance, analyses and descriptions of how customers respond to specificities of prefacing observations and noticings of, or glances at, something in their visual field in embodied actions and how they refer to (aggregates of) them in turn initiations and responses, which leads to a respecification of our understanding of these actions. The information which this combination of recordings provides may even give access to phenomena that have not been accessible until now, e.g.,

7. In this example the fact that the customer fixated on the yellow bowl and not on the co-present other customers was used to make the point that she made a detour – thus orienting to both the yellow bowl and the other customers.

‘intended’, ‘avoiding’ embodied actions, or ‘withholding of information’ in order to provide it on demand. It paves the way for the possibility of describing these actions *as they occur* rather than retrospectively and allows for descriptions of how co-present others may deal with them interactionally.

References

-  Auer, Peter. 2018. “Gaze, addressee selection and turn-taking in three-party interaction.” In *Eye-tracking in Interaction: Studies on the role of eye gaze in dialogue*, ed. by Geert Brône, and Bert Oben, 197–232. Amsterdam/Philadelphia: John Benjamins Publishing Company.
-  Auer, Peter. 2021. “Turn-allocation and gaze: A multimodal revision of the “current-speaker-selects-next” rule of the turn-taking system of conversation analysis.” *Discourse Studies* 23 (2): 117–140.
-  Brône, Geert and, Bert Oben (eds). 2018. *Eye-tracking in Interaction: Studies on the role of eye gaze in dialogue*. John Benjamins Publishing Company.
-  Broth, Mathias, and Lorenza Mondada. 2013. “Walking away: The embodied achievement of activity closings in mobile interaction.” *Journal of Pragmatics* 47 (1): 41–58.
-  Carter, Benjamin T., and Steven G. Luke. 2020. “Best practices in eye tracking research.” *International Journal of Psychophysiology* 155: 49–62.
- Day, Dennis, and Gitte Rasmussen. 2019. “Interactional Consequences of Object Possession in Institutional Practices.” In *Objects, Bodies and Work Practice*, ed. by Dennis Day, and Johs. Wagner, 87–112. Blue Ridge Summit, PA: Multilingual Matters.
-  Deppermann, Arnulf. 2013a. “Turn-design at turn-beginnings: Multimodal resources to deal with tasks of turn-construction in German.” *Journal of Pragmatics* 46: 91–121.
-  Deppermann, Arnulf. 2013b. “Conversation Analytic Studies of Multimodal Interaction.” *Journal of Pragmatics* 46 (1): 1–172.
-  Deppermann, Arnulf, Schmitt, Reinhold, and Mondada, Lorenza. 2009. “Agenda and emergence: Contingent and planned activities in a meeting.” *Journal of Pragmatics*, 42(6): 1700–1718.
-  De Stefani, Elwys. 2014. “Establishing joint orientation towards commercial objects in a self-service store.” In *Interacting with Objects: Language, materiality, and social activity*, ed. by Maurice Nevile, Pentti Haddington, Trine Heinemann, and Mirka Rauniomaa, 271–294. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- De Stefani, Elwys, and Lorenza Mondada. 2010. „Die Eröffnung sozialer Begegnungen im öffentlichen Raum: Die emergente Koordination räumlicher, visueller und verbaler Handlungsweisen.“ In *Situationseröffnungen Zur multimodalen Herstellung fokussierter Interaktion*, ed. by Lorenza Mondada, and Reinhold, Schmitt R., 103–170. Tübingen: narr verlag.
-  De Stefani, Elwys, and Lorenza Mondada. 2014. “Reorganizing mobile formations: When “guided” participants initiate reorientations in guided tours.” *Space and Culture* 17 (2): 157–175.

-  De Stefani, Elwys, and Lorenza Mondada. 2018. "Encounters in Public Space: How Acquainted Versus Unacquainted Persons Establish Social and Spatial Arrangements." *Research on Language and Social Interaction* 51 (3): 248–270.
- Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.
- Garfinkel, Harold, and Harvey Sacks. 1970. "On formal structures of practical actions." In *Ethnomethodological studies of work* ed. by Harold Garfinkel, 160–193. Routledge & Kegan Paul.
- Garfinkel, Harold, and D. Lawrence Wieder. 1992. "Two incommensurable asymmetrically alternate technologies of social analysis." In *Text in Context: Contributions to Ethnomethodology*, ed. by Graham Watson, and Robert Morris Seiler, 175–206. Sage Publications.
- Goffman, Erving. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
-  Goodwin, Charles. 1980. "Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning." *Sociological Inquiry* 50 (3–4): 272–302.
- Goodwin, Charles. 1981. *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.
-  Goodwin, Charles. 1994. "Professional Vision." *American Anthropologist* 96 (3): 606–633.
- Heath Christian, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research: Analysing Social Interaction in Everyday Life*. London: Sage.
- Kendon, Adam. 1973. "The role of visible behaviour in the organisation of social interaction." In *Social communication and movement*, ed. by Marion Carnach, and Ian Vine, 29–74. London: Academic Press.
- Kendon, Adam. 1990. *Conducting interaction: Patterns of behaviour in focused encounters*. CUP Archive.
- Kendon, Adam, and Andrew Ferber. 1973. "A description of some human greetings." In *Comparative ecology and behaviour of primates*, ed. by Richard P. Michael, and John Crook, 591–668. London/New York.
-  Kendrick, Kobin H., and Paul Drew. 2016. "Recruitment: Offers, requests, and the organization of assistance in interaction." *Research on Language and Social Interaction* 49: 1–19.
-  Knoblauch, Hubert, Bernt Schnettler, Jürgen Raab, and Hans-Georg Soeffner (eds.). 2006. *Video Analysis: Methodology and Methods. Qualitative Audiovisual Data Analysis in Sociology*. Peter Lang.
- Kress, Günther R. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge.
-  Kristiansen, Elisabeth Dalby, and Gitte Rasmussen. 2021. "Eye-tracking Recordings as Data in EMCA Studies: Exploring Possibilities and Limitations." *Social Interaction. Video-Based Studies of Human Sociality* 4 (4).
-  LaBonte, Andrew, Jon Hindmarsh, and Dirk vom Lehn. 2021. "Data collection at height: Embodied competence, multisensoriality, and video-based research in an extreme context of work." *Social Interaction. Video-Based Studies of Human Sociality* 4 (3).
-  Lindström, Jan. 2006. "Grammar in the service of interaction: Exploring turn organization in Swedish." *Research on Language and Social Interaction* 39 (1): 81–117.

-  McIlvenny, Paul. 2019. "Inhabiting Spatial Video and Audio Data: Towards a Scenographic Turn in the Analysis of Social Interaction." *Social Interaction. Video-based Studies of Human Sociality*, 2 (1).
-  McIlvenny, Paul, Mathias Broth, and Pentti Haddington. 2014. "Moving Together: Mobile Formations in Action." *Space and Culture* 17 (2): 104–106.
-  Mondada, Lorenza. 2007. "Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers." *Discourse studies* 9 (2), 194–225.
-  Mondada, Lorenza. 2009. "Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space." *Journal of Pragmatics* 41 (10): 1977–1997.
-  Mondada, Lorenza. 2011. "Understanding as an embodied, situated, and sequential achievement in interaction." *Journal of Pragmatics* 43 (2): 542–552.
-  Mondada, Lorenza. 2019a. "Practices for Showing, Looking, and Videorecording: The Interactional Establishment of a Common Focus of Attention." In *Embodied Activities in Face-to-face and Mediated Settings: Social Encounters in Time and Space*, ed. by Elisabeth Reber, and Cornelia Gerhardt, 63–104. Springer International Publishing.
-  Mondada, Lorenza. 2019b. "Rethinking bodies and objects in social interaction: A multimodal and multisensorial approach to tasting." In *Discussing new materialism: Methodological implications for the study of materialities*, ed. by Ulrike Tikvah Kissmann, Joost van Loon, 109–134. Wiesbaden: Springer.
- Mondada, Lorenza, and Reinhold Schmitt (eds.). 2010. *Situationseröffnungen. Zur multimodalen Herstellung fokussierter Interaktion*. Tübingen: Narr Francke Attempto Verlag.
- Müller, Cornelia, and Ulrike Bohle. 2007. "Das Fundament fokussierter Interaktion. Zur Vorbereitung und Herstellung von Interaktionsräumen durch körperliche Koordination." In *Koordination. Analysen zur multimodalen Interaktion*, ed. by Reinhold Schmitt, 129–165. Tübingen: Narr Francke Attempto Verlag.
- Oloff, Florence. 2010. "Ankommen und Hinzukommen: Zur Struktur der Ankunft von Gästen." In *Situationseröffnungen Zur multimodalen Herstellung fokussierter Interaktion*, ed. by Lorenza Mondada, and Reinhold Schmitt, 171–228. Tübingen: Narr Francke Attempto Verlag.
-  Rasmussen, Gitte. 2023a. "Analysing interaction involving wheelchairs: configuring interactional spaces to engage in activity and to initiate conversation." *Journal of Interactional Research in Communication Disorders*, 14 (2): 328–355.
-  Rasmussen, Gitte. 2023b. "Multimodal engagement and interaction and the appearance of contemporary non-traditional retail shopping." In *Multimodality and Social Interaction in Online and Offline Shopping*, ed. by Gitte Rasmussen, and Theo van Leeuwen, 104–127. New York: Routledge.
-  Rasmussen, Gitte, and Elisabeth Dalby Kristiansen. 2022. "The sociality of minimizing involvement in Danish self-service shops: Danish customers' multi-modal practices of being, getting and staying out of the way." *Discourse & Communication* 16: 200–232.
-  Rasmussen, Gitte, Elisabeth Dalby Kristiansen, and Søren Vigild Poulsen. 2024. E-pub ahead of print. "The World of Daily Life: Doing a search for (e-)shopping purposes." *Pragmatics & Society*.

- Rossano, Frederico. 2012. Gaze behaviour in face-to-face interaction, PhD thesis, Radboud University, Nijmegen: The Netherlands.
-  Rossano, Frederico, Paul Brown, and Stephen Levinson. 2009. "Gaze, questioning and culture." In *Conversation analysis: Comparative perspectives*, ed. by Jack Sidnell, 187–249. Cambridge: Cambridge University Press.
- Sacks, Harvey. 1984. "Notes on methodology." In *Structures of social action: Studies in conversation analysis*, ed. by Jim M. Atkinson, and John Heritage, 21–27. Cambridge, Paris: Cambridge University Press.
-  Schegloff, Emanuel A. 1987. "Recycled turn beginnings." In *Talk and Social Organization*, ed. by Graham Button, and John R.E. Lee, 70–85. Clevedon: Multilingual Matters.
- Schegloff, Emanuel A. 1996. "Turn organization: one intersection of grammar and interaction." In *Interaction and grammar*, ed. by Elenor Ochs, Emanuel Schegloff, and Sandra Thompson (Eds.), 52–133. Cambridge University Press.
-  Schegloff, Emanuel A., and Gene Lerner. 2009. "Beginning to respond: Well-Prefaced Responses to Wh-Questions." *Journal of Research in Language and Social Interaction* 42 (2): 91–115.
- Schmitt, Reinhold, and Arnulf Deppermann. 2007. "Monitoring und Koordination als Voraussetzung der multimodalen Konstitution von Interaktionsräumen." In *Koordination. Analysen zur multimodalen Interaktion*, ed. by Reinhold Schmitt, 95–128. Tübingen: Narr Francke Attempto Verlag.
-  Stukenbrock, Anja. 2018. "Mobile dual eye-tracking in face-to-face interaction: The case of deixis and joint attention." In *Eye-tracking in Interaction: Studies on the role of eye gaze in dialogue*, ed. by Geert Brône, and Bert Oben, 265–300. Amsterdam/Philadelphia: John Benjamins Publishing Company.
-  Stukenbrock, Anja, and Anh Nhi Dao. 2019. "Joint attention in passing: What dual mobile eye-tracking reveals about gaze in coordinating embodied activities at a market." In *Embodied Activities in Face-to-face and Mediated Settings*, ed. by Elisabeth Reber, and Cornelia Gerhardt, 177–213. Cham: Palgrave MacMillan.
-  Wade, Nicholas, and Benjamin W. Tatler. 2005. *The Moving Tablet of the Eye: The Origins of Modern Eye Movement Research*. Oxford University Press, USA.

Appendix. Transcription conventions

-	cut off
:	prolongation of sound
>okay<	higher tempo
Ps	Pause
(0.2)	Pause, measured in seconds
(.)	Micropause, less than 0.2 seconds
.hhh	Hearable in-breath
=	latched talk
[	overlapping talk

>	gaze at
><	Mutual gaze
↑	High pitch
Δ	Co-occurring
*	beginning of action and conduct that continues across more turns and turn constructional units (TCU) as indicated through -->
-->*	indicates termination of continued action and conduct across more turns and TCUs
MANY	capital letters indicate increased volume

Mobile eye-tracking and mixed-methods approaches to interaction analysis

Bert Oben, Clarissa de Vries & Geert Brône
University of Leuven

The integration of mobile eye-tracking technology in linguistic research has catalyzed a surge of investigations across diverse linguistic subdisciplines. This chapter advocates for a mixed-methods approach in analyzing eye gaze behaviour during face-to-face interactions. Through two case studies, examining eye gaze in interactional irony and gaze synchronization, we demonstrate how this approach can help enhance our understanding of conversational eye gaze behaviour.

Keywords: eye gaze, interaction, mixed-methods, irony, synchronization

1. Introduction

The introduction of mobile eye-tracking technology to the study of language in interaction has, over the last decade, led to a growing body of research in several subdisciplines of linguistics, including Interactional Linguistics, Conversation Analysis, Cognitive Linguistics and psycholinguistics. Each of these disciplines comes with its own research questions, methodological traditions and toolkits, which has led to a fragmented picture (see Brône & Oben 2018 for a first overview of the field). One of the ways in which studies may differ, is the use of a qualitative, quantitative or mixed-methods approach to the data collected by mobile eye-tracking systems. Whereas studies in Conversation Analysis tend to use the eye-tracking data as a resource for fine-grained qualitative analyses on interactional data, psycholinguists have mostly resorted to task-based interactions and/or experimental designs to provide quantitative results on gaze distribution patterns. Only few studies have opted for a balanced mix-methods approach in which quantitative and qualitative analyses feed into each other (Kendrick & Holler 2017, Auer 2021, Zima et al. 2019, Auer & Zima 2021) (see Brône & Oben 2023 for a methodological overview).

In this chapter, we provide an argument in favour of mixed-methods approaches, using examples from various studies on different topics, and highlighting different levels of analysis. In a first case study, we zoom in on the cognitively and interactionally complex phenomenon of irony. The main aim of this study is to show that a back and forth between quantitative and qualitative analysis may provide valuable insights into the level(s) at which the role of eye gaze for the negotiation of ironic meaning can best be studied. A second study deals with the much-debated question of synchronisation in interaction, zooming in on the synchronisation of gaze behaviour across participants. Similar to the first study, we illustrate how quantitative approaches can provide insights into general patterns, whereas a qualitative micro-analysis can help to uncover additional factors that may support or clarify the results.

2. Defining and refining units of analysis: A case study on irony in interaction

As a first example of how quantitative and qualitative approaches inform each other, we zoom in on a set of studies that explored the role of eye gaze in the construction of irony in face-to-face interaction. Given the complexity of this phenomenon, involving the negotiation of intention and stance between speakers and their addressees, it constitutes a challenge to any multimodal-pragmatic approach. More specifically, the studies to be discussed here may serve to show how the results of a quantitative analysis may call for a fine-grained qualitative (re)consideration, which then again feeds into a recalibrated quantitative approach. As such, we aim for our case studies to illustrate the benefits of a continuous feedback loop between qualitative and quantitative analysis.

The starting point for a first exploratory quantitative analysis (reported in Brône & Oben 2022), was the observation that theories in (cognitive) pragmatics tend to describe interactional irony in terms of participants' setting up a pretence layer in discourse (Clark 1996, 2016, Coulson 2005, Brône 2008, Tobin 2016, Barden 2017, a.o.). According to this view, when speakers produce an ironic utterance, they pretend "to be an injudicious person speaking to an uninitiated audience" (Clark & Gerrig 1984), as is the case in a simple example such as (1), taken from a dataset of spontaneous triadic interactions in Dutch (Brône & Oben 2015). This specific sequence takes place at the start of the recording, when the experimenters have just left the three participants (Jesse, Emma, and Sophie) alone. After a brief silence during the conversational opening, Emma suggests a topic that they can talk about, viz. their adventures during their exchange semesters abroad (*Erasmus stories*, line 1), and adds that this in fact is the only thing that

unites them (line 3), followed by turn-final laughter. With this utterance Emma sets up an ironic pretence, possibly building on the difficulty in getting the conversation started, in which the assumed speaker Emma¹ seriously claims that she and her assumed co-participants Sophie' and Jesse' have only few things in common (which is not the case in reality). Importantly, setting up and managing the pretence is argued to be a joint action between the participants, as addressees need to be able to see through (and navigate between) the discursive layers that are being set up. In Example (1), Sophie first acknowledges Emma's playful utterance (*yeah*) and then continues on the pretence layer by ironically stating that it is a sad thing that they have so few things in common, but that it did bring them closer together.

(1) Pilot_2 — 00:00:16.173 — 00:00:26.216

01. Emma oké we zullen het over onze erasmusverhalen hebben [zeker hè,
okay let's talk about our Erasmus stories, okay]
02. Sophie [ja.
yeah]
03. Emma das het enige da we gemeenschappelijk hebben,
((lacht))=
that's the only thing we have in common
((laughs))
04. Sophie =ja [da's](.) eigenlijk wel echt zielig ma ok-
yeah that is actually really sad but ok-
05. Emma [dus;]
so
06. (0.3)
07. Sophie het heeft ons wel dichter [bij elkaar gebracht;
it has brought us tighter together though]
08. Emma [da's waar;
that's true]

Given the wealth of literature on the layered nature of irony in interaction (Clark 1996, 2016, Coulson 2005, Brône 2008, Tobin 2016, Barnden 2017, a.o.), it is somewhat surprising that only few studies addressed the question how participants use various semiotic resources to communicate their ironic intention and understanding (e.g. de Vries et al. 2021, Gironzetti et al. 2016, Gonzalez-Fuente et al. 2015, Tabacaru & Lemmens 2014). One would expect, for instance, that the complexity of navigating through different layers of action would be reflected in the use of resources that are used for monitoring and providing feedback, such as eye gaze and head movements. Specifically for eye gaze, a few studies did suggest that participants display particular gaze behaviour, such as increased attention to mouth and eyes of both speaker and recipients in occurrences of irony and humour (Gironzetti et al. 2016), as well as an increase in the amount of gaze shifts

1. We use the labels Emma', Jesse' etc. to indicate the counterparts of the real participants on the pretence layer (or in the pretence space).

by speakers during the production of ironic utterances (Gonzalez-Fuente et al. 2015). These initial probes were indicative of a higher degree of engagement with co-participants, but a more systematic approach is needed to get a better grasp on the correlation between gaze distributions and irony production/reception in interaction. One way of providing such an account is through the use of mobile eye-tracking data. The use of eye-tracking as a method, allows for a more accurate and reliable estimation of both the gaze target as well as the precise timing of shifts in the focus of attention, which are extremely difficult to annotate with ‘external’ camera’s only (see Brône & Oben 2018, for an elaborate discussion, or Zima et al. this volume). More specifically, mobile eye-tracking, as opposed to remote eye-tracking in which participants look at each other on computer screens (as in the study by Gironzetti et al. above), enables researchers to study gaze behaviour in a non-mediated and face-to-face setting.



Figure 1. Recording set-up for the triadic interactions in the Insight Interaction Corpus. Numbers indicate the camera perspectives of the respective participants, as shown on the bottom right image of the external camera perspective

In our analytical approach, we started from the above-mentioned observation reported in Gonzalez-Fuente et al. (2015) that speakers produce more gaze shifts in ironic vs. non-ironic utterances, and expanded the perspective to all participants, using a dataset of 5 spontaneous triadic conversations (approx. 100 minutes of data) from the Insight Interaction Corpus (Brône & Oben 2015). Figure 1 shows the interactional set-up for this data set, which includes an external camera perspective (Sony HDRFX1000E, 25 frames per second, 720 × 576 pixels) and

three participant perspectives generated by mobile eye-tracking systems (Pupil Pro Eye-Tracking Glasses and Tobii Pro Glasses 2), which record the participants' visual field as well as their gaze behaviour. All participants in this corpus gave informed consent to participate in the study and that images as well as transcripts could be produced on the basis of the video recordings. The study was approved by the local ethics committee with case number G-2021-3303.

Our procedure to annotate ironic utterances was based on Gibbs' (2000) classification of irony. For a more detailed report on this, see Brône & Oben (2022). Next, we calculated the average amount of gaze shifts per second for both speakers and their addressees, using conversational turns as the basic unit of analysis. A comparison between the ironic utterances in the dataset and a random control set of non-ironic utterances from the same interactions produced the following initial results:

- Speakers produce significantly more gaze shifts during the production of ironic utterances in comparison to non-ironic utterances. This result holds for both shifts between co-participants as for shifts from and towards the background;
- Addressees also produce significantly more gaze shifts during ironic utterances compared to non-ironic utterances, and the effect was even stronger than was the case for speakers. The result again holds for different gaze shift types: to and from the background, between speaker and other addressee;
- There are more instances of mutual gaze between addressees in the ironic vs. non-ironic utterances, and these moments of mutual gaze tend to be longer in the case of ironic utterances, reflecting moments of grounding or reaction monitoring between the addressees.

The basic pattern reported in this study thus reveals an increased engagement of participants (both speakers and their addressees) with their co-participants, which may be linked either to the cognitive complexity of the staged communicative act that is irony (Clark 1996, *supra*) or to the delicate social dynamics involved. This social dimension is strongly present in irony, since the pretence is often argued to function as a stance act, with speakers distancing themselves from the staged utterance (see also the echoic mention theory, developed by Sperber & Wilson (1981), for a model that stresses the critical attitude expressed by the ironist) and/or playfully targeting a co-participant (ironic teasing) or external target. Walking such a social tightrope may require additional coordination between the participants, which then again may be reflected in participants' gaze behaviour.

When we apply the findings of this initial quantitative analysis to the introductory example in (1), we recognize the general patterns, but also more. The representation in (1') adds a gaze score for each of the participants below the tran-

scription lines (infra), in which gaze shifts are aligned with the speech sound of the corresponding transcription. The symbols in the score represent the gaze target at each point in time (e.g. the first score line below line 3 of the transcript represents Emma's gaze behaviour while she produces the ironic utterance, showing that she shifts her gaze from Sophie (S) to Jesse (J) while saying "gemeenschappelijk" (*in common*)). In the ironic utterances in lines 3 and 4, the ironist shifts gaze multiple times (including gaze shifts to the background (BG)), while the addressees tend to monitor both the current speaker and the co-addressee. It seems that participants are aware of the potentially face-threatening situation, in case the utterance would be interpreted literally, and thus closely monitor the others. What the relatively crude quantitative analysis did not reveal, however, is the relevance of temporal positioning and sequential organization. The example in (1') shows that both speakers and their addressees tend to shift their gaze most towards the end of the utterance, which again may be of particular relevance for response monitoring and/or feedback elicitation.

(1') Pilot_2 — 00:00:16.173 — 00:00:26.216

01. Emma	oké we zullen het over onze <u>erasmus</u> verhalen hebben [zeker hè, okay let's talk about our Erasmus stories, okay
02. Sophie	[ja. yeah
03. Emma	das het enige da we gemeensch <u>appelijk</u> hebben, ((lacht))= that's the only thing we have in common ((laughs))
Gaz_J	E-----S-----
Gaz_E	S-----J-----
Gaz_S	E-----
04. Sophie	=ja [da's](.) eigenlijk wel echt <u>zielig</u> ma ok- yeah that is actually really sad but ok-
05. Emma	[dus;] so
Gaz_J	S-----
Gaz_E	J-----S-----J-----
Gaz_S	E-BG-----E-----BG-----J----
06.	(0.3)
07. Sophie	het heeft ons wel <u>dichter</u> [bij elkaar gebracht; it has brought us tighter together though
08. Emma	[da's <u>waar</u> ; that's true

In a qualitative study that built on the first quantitative analysis of the eye-tracking data, Brône (2021)² dives deeper into the question how gaze relates to other semiotic resources, verbal as well as nonverbal, in the interactional process of meaning making in irony. More specifically, a multimodal micro-analysis of a set of representative examples from the same data set as Brône & Oben (2022) shows how the different temporalities and dependencies among the different semiotic resources are crucial in the negotiation of irony in interaction. For the analysis of gaze behaviour, this entails that participants' gaze behaviour may at the same time be affected by and itself affect (non)verbal behaviour by co-participants, leading up to a complex network of interrelated actions. For instance, gaze shifts by one participant may serve as a trigger for specific (non)verbal actions by other participants, as is the case when gaze is used by an ironist as an invitation for others to join in the tease. In other cases, there is a strong synchronisation of gaze behaviour, that is a strong temporal interdependency between the gaze of the different participants, which may be indicative of a sense of complicity between the participants (for a deeper exploration of gaze synchronisation, see Section 3 below). This is the case, for instance, in line 4 of Example (1'), where Emma and Sophie shift their gaze synchronously to establish mutual gaze. In sum, the results of the micro-analysis provide a qualification of the quantitative results summarized above. More specifically, the basic gaze patterns uncovered in the initial quantitative study are studied in relation to the interpersonal dynamics between the participants (including invitational cues by the ironist, where establishing eye contact with one or more addressees may be interpreted as an invitation to join the staged communicative act).

Building on the two initial quantitative and qualitative studies described above, De Vries et al. (2021) refine the findings by taking into account both the interplay of different bodily resources in the production of irony in interaction and the relevance of timing in relation to speakers' gaze behaviour, thus improving the granularity of analysis. For the former aspect, the study included laughter, body movement (including gestures, head movements and shoulder shrugs) and attempted to provide some first evidence of multimodal Gestalts (Mondada 2014a) for the construal of irony. For the latter dimension, the study started from the hypothesis that speakers would mainly be involved in visual grounding (reaction monitoring) towards the end of their utterances, at a point at which addressees should arguably be aware of the speaker's ironic intentions. This increased addressee orientation could be reflected in the amount of speaker gaze shifts at the end of ironic utterances.

2. Although this study was published shortly before the quantitative corpus analysis presented in Brône & Oben (2022), it explicitly builds on the latter. The somewhat confusing publication dates are merely the result of our choice of publication venues for these studies.

To test these hypotheses, De Vries et al. used the same corpus of triadic face-to-face interactions as in Brône & Oben (2022) and Brône (2021) but included more interactions (16 videos, amounting to 256 minutes of recordings). Importantly also, in the selection of ironic items and non-ironic control items, we narrowed down the scope to scalar evaluative utterances, i.e. utterances that can be situated on an evaluative scale. In the example in (1) above, both ironic utterances in lines 3 and 4 would be categorized as scalar, as they include stance acts that can be compared to other stances, and can be downgraded or enhanced (*the only thing we have in common* vs. *the many things we have in common; that is actually really sad* vs. *that is actually really cool*). Choosing this particular focus on scalar utterances was done to improve the comparability between the test items ($n=123$) and control items ($n=123$). Examples of such control items would be non-ironic scalar evaluative utterances such as ‘*that is very smart*’, or ‘*so it’s sort of cheap*’. By having similar utterance types for both the ironic and non-ironic utterances in the dataset, we can more confidently claim that any difference between the two are attributed to irony only, rather than other potentially relevant factors such as the stance that is being taken. And finally, the segmentation of the data was done on the basis of intonation units rather than turns-in-interactions (as was done in Brône & Oben 2022), which allowed us to take into account transition-relevance places (TRPs) and the importance of speaker gaze behaviour at those points. In order to capture the temporality of gaze behaviour within these segments, annotation of speaker gaze shifts was done both at the level of the entire segments and for the final 1000 ms of a segment.

The results of this more fine-grained quantitative analysis suggest that, on average, speakers recruit more bodily resources during the production of ironic utterances compared to non-ironic stance acts, which was most prominently manifested in the use of laughter, head movements, body repositionings and specific gaze shifts (infra). This finding suggests that speakers are actively engaged in communicating their ironic intent, using several embodied resources at their disposal. Apart from the observation that more embodied resources are used in the production of ironic utterances, the analysis also revealed particular clustering patterns. Using Kendall’s Tau correlation coefficient, correlations were calculated between the presence (a binary variable) of all of the resources under scrutiny in this chapter. Most notably, there is a correlation between the presence of laughter³ and specific gaze measures (gaze shifts, mutual gaze), which corroborates the claim made in the qualitative study that embodied behaviour in one semiotic

3. More specifically, in ironic segments where there is laughter, there are also more gaze shifts ($T=0.218$, $p=0.009$), more instances of mutual gaze ($T=0.177$, $p=0.040$) and speaker-to-listener gaze ($T=0.219$, $p=0.010$), as well as gaze aversions ($T=0.243$, $p=0.007$).

channel (e.g. gaze) may affect or be affected by other (non)verbal behaviour, thus leading to interrelations or multimodal patterning in the construal of ironic utterances. Second, by looking in more detail into the gaze distribution data, we found that the basic observation reported in the first exploratory study, viz. that speakers produce more gaze shifts in ironic compared to non-ironic utterances, needs to be refined. More specifically, the analysis at the level of intonation units revealed that speakers only produce more gaze shifts towards the end (i.e. during the last 1000 msec) of the ironic segment, compared to non-ironic segments. This difference holds for both gaze shifts to and from the background (gaze aversions) as well as shifts between the two co-participants. This result is in line with the above-mentioned claim that the creation of ironic meaning in face-to-face interaction is a highly interactive process, which requires a substantial amount of visual checking, most notably towards TRPs, when speakers expect their addressees to have parsed the segment as intended.

What the back and forth between quantitative and qualitative analysis for this case study has shown is that it can be a fruitful undertaking to start off from relatively course-grained distributional analyses that test basic claims emanating from the literature, and then work towards a more fine-grained picture through a feedback loop between qualitative and quantitative analyses. This way, the quantitative analyses provide corpus-based support for insights from qualitative studies, but also help to substantiate claims on recurrent multimodal patterns or Gestalts. Indeed, one could argue that if quantitative patterns emerge across different settings and sequential contexts, taking into account all contingencies that may affect such patterns and result in ‘noisy data’, this provides an extra argument in favour of their existence.

3. Mutual gaze during face-to-face interaction: A second case study

In the first case study, we demonstrated that eye gaze is temporally anchored, i.c. relative to transition relevance places. In a second exploration, we want to highlight another temporal dimension of eye gaze, viz. the synchronisation of one interlocutor’s gaze behaviour with gaze and speech of the other interlocutor. In the first pioneering empirical studies into gaze behaviour in interaction, Kendon (1967) and Argyle & Cook (1976) observe an asymmetry in gaze behaviour between speaker and addressees. They found that addressees look at their speaking partners more than the other way around, that mutual gaze nearly always occurs during turn transitions, that speakers typically briefly look away at the beginning of their turn or during hesitations and pauses, but that they do systematically look at their partner at the end of longer turns. Later work confirmed or added to the results from these pioneering studies (see e.g. Brône & Oben 2018)

by showing that verbal and non-verbal feedback markers sometimes do (Bavelas et al. 2002, De Stefani 2021) and in other times do not synchronise with mutual gaze (Zima 2020, Masuch, this volume), or by illustrating that turn transitions not always happen in mutual gaze (cf. Kendon 1967), but can involve gaze aversion by the incoming speaker as well (Oertel et al. 2012).

Focusing further on (dis)establishing eye contact, Zima, Weiss & Brône (2019) show how breaking mutual gaze, can help interlocutors ‘win’ the battle over the turn in cases of speech overlap. Vranjes, Brône & Feyaerts (2018) demonstrate how interpreters make and break eye contact with their interlocutors to manage turn-taking in more elaborate multi-unit turns, and Krug (this volume) shows how mutual eye gaze is relevant in overcoming conversational disalignment. These studies, along with more quantitatively-oriented approaches, such as the support vector machines approach in Jokinen et al. (2013), conclude that information on the gaze direction by all conversational partners is a good predictor for turn holding. Making and breaking eye contact has not only been related to turn taking, but also to other interactional phenomena such as establishing joint attention (e.g. Goodwin 1981, Mondada 2014b, Stukenbrock 2020), collaborative music making (Bischof et al. 2021, Vandemoortele et al. 2018) or marking shifts in viewpoint during reenactments (Pfeiffer & Weiss 2022, Sidnell 2006, Thompson & Suzuki 2014).

Not all research on gaze synchronisation, involves the analysis of face-to-face interactions. For example, one group of studies starts from a joint-attention paradigm in which participants are not looking at each other, but at a computer screen while playing a map, puzzle or matching game. Participants are reported to perform matching tasks (e.g. find a target object in a complex picture) faster if they have visual information on where their partner is looking at (Brennan et al. 2008, Frischen et al. 2007, Lachat et al. 2012, Neider et al. 2010, Richardson & Dale 2005) or perform complex tasks better when higher levels of gaze synchronisation are observed (Vrzakova et al. 2019). Also, participants synchronise their eye movements more as they interact longer with each other (Dale et al. 2011, Hadelich & Crocker 2006).

Summing up this brief literature overview, studies in the domain of cognitive science have demonstrated that there is a fine-grained systematic coupling between interlocutors’ gaze fixations: interlocutors tend to look at the same thing at the same time, or at least with a systematic time lag between the regions of interest involved. Studies in the domain of Conversation Analysis and interactional linguistics have shown how establishing eye contact is highly relevant in a wide range of conversational phenomena such as turn taking, overlap resolution, joint attention or viewpoint shifting. What is still lacking, however, is a focus on the temporal synchronisation of establishing and breaking this eye contact: to what extent is the gaze behaviour by one interlocutor temporally dependent on the gaze behaviour by the other? More specifically, our research questions are:

- i. Do interlocutors start and stop looking at each other at exactly the same time, or is there a systematic time lag between the onset and offset of making and breaking eye contact?
- ii. If we observe synchronisation of making and breaking eye contact, what interactional role would such a synchronisation fulfil?

3.1 Data and method

To answer our research questions, we use data from the Insight Interaction Corpus (Brône & Oben 2015), as we did in the first case study. From this corpus we selected the data from an animation description task (mainly because this task is highly comparable to the existing gaze synchronisation studies in a non-face-to-face setting, e.g. Richardson & Dale (2005) or Louwerse et al. (2012)),⁴ and from a brainstorm task (because this is a type of interaction that is less structured and affected by an experimental design). In the picture description task, 15 pairs of participants were shown a range of complex pictures (15 sets of pictures in total). Participants watched the animations simultaneously, yet separately (i.e. on different screens that were located in the same room; participants were not able to see each other's screen) and there were subtle differences between participants' animations. After the animations were presented, they disappeared and participants were asked to find the differences between the animations. Participants were free to discuss as long as they pleased until they had found all the differences (or gave up without finding them). In the brainstorm task participants were asked to come up with new features for a mobile phone that was branded specifically for women. Again, participants were free to take as much time as they wanted and be as specific or realistic as they wanted, to allow for a maximally unrestricted conversation. Figure 2 shows the set-up of the interaction, which included an external camera perspective (left) and two mobile eye-tracking systems (one for each participant, on the right).

To study synchronisation of making and breaking eye contact, we used cross-recurrence quantification analysis (CRQA). A substantial body of research (for overviews, see Fusaroli et al. 2014 or Xu et al. 2020) has used this CRQA technique to study phenomena of behaviour matching, including the synchronisation of eye gaze (Richardson & Dale 2005, Richardson et al. 2009, Dale et al. 2011, Vrzakova

4. In these studies (such as Richardson & Dale (2005) or Louwerse et al. (2012)), participants are assigned alternating roles as 'descriptor' and 'matcher'. The former describes a picture or video; the latter is asked to use the information in the description in a subsequent task (e.g. retelling the scenario in the video or redrawing the picture). During their interaction, both descriptor and matcher are looking at a computer screen that displays relevant objects (e.g. characters from the video or landmarks from the picture). Screen-based eye-trackers measure what both interlocutors are looking at while completing the task.



Figure 2. Recording set-up for the two-party interactions in the Insight Interaction Corpus

et al. 2019). CRQA is a type of correlation analysis that looks for a time lag at which the overlap between two time-series is maximal. This allows the analyst to check whether events typically occur simultaneously, with a given time lag, or completely unrelated from one another.

Without diving all too deep into the mathematics behind the quantification (see Coco & Dale 2014 or Wallot 2017 for useful tutorials), we provide a brief overview of our method. To allow for CRQA, the existing transcriptions and annotations were sampled (10 Hz) into categorical time-series: every 100 ms we polled the gaze annotation tiers and scored, per participant, whether there was gaze at the face of the co-participant (“1”) or not (“0”). Figure 3 shows a schematic overview of that procedure for the first thirty seconds (of conversation *bl33*). The blue line indicates the gaze behaviour of participant 1; the orange line that of participant 2; the shaded grey areas are the moments of mutual face fixations, i.e. moments of eye contact.

Because we are zooming in on the synchronisation of making and breaking eye contact, and not on the full duration of the resulting eye contact in between, the next step in our procedure was to turn the continuous annotation of eye gaze into event-based data. To mark the moments of making and breaking eye contact, only the first 500 ms of a fixation towards the face of an interlocutor, and the first 500 ms of a fixation away from the face of an interlocutor were retained and marked with the value “1”. Figure 4 provides a visual representation of the result of this event-based procedure for the phenomenon of making eye contact. The spikes (value “1”) correspond to the moments at which the participants shift their eye gaze towards their interlocutor’s face.

After sampling the data into event-base time series, we used the R-package developed by Coco & Dale (2014) to perform the CRQA. Figure 5 shows an example of such an analysis of synchronisation of eye gaze for two dyads in the animation description task.

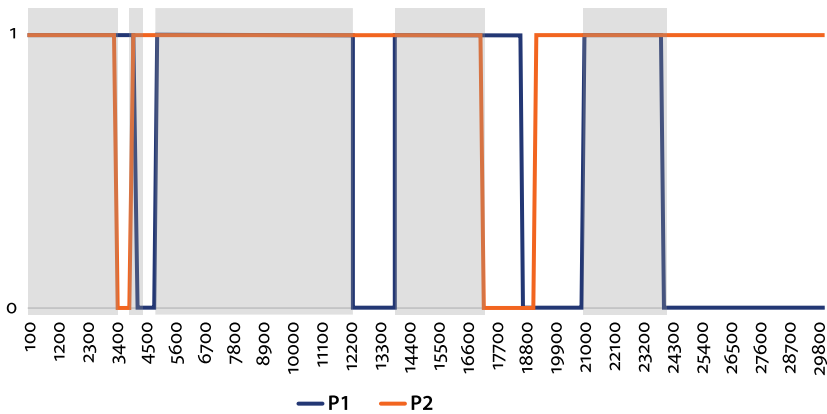


Figure 3. Example of the sampling method for eye gaze. “1” corresponds with a fixation at the face of the other participant; “0” with a fixation away from the face. This example shows the first thirty seconds from one conversation (bl33): x-axis shows time in milliseconds. ((P1 = participant1; P2 = participant2; shaded areas are moments of eye contact)

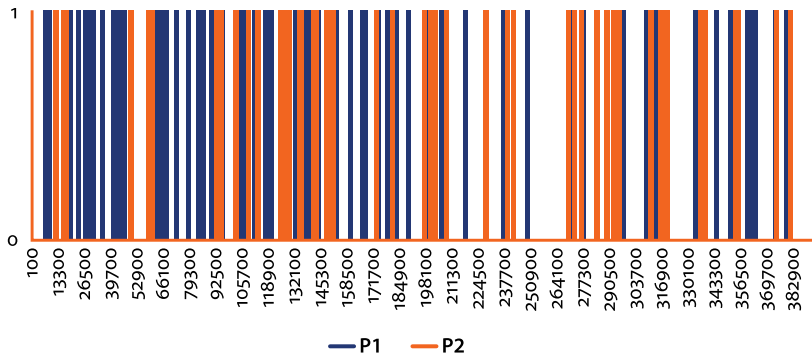


Figure 4. Visual representation of event-based annotation. The spikes indicate when each participant starts looking at the other participant’s face (value “1” on y-axis). This example shows the data from a full conversation (lw33): x-axis shows time in milliseconds. (P1 = participant1; P2 = participant2)

The Y-axis in Figure 5 indicates the recurrence rate; the X-axis represents a time scale in seconds with t_0 in the middle. The values on the Y-axis are less intuitive to interpret, because they are dependent on the frequency of the phenomenon (i.c. making eye contact) and on the sample rate. In general terms, the Y-axis shows the relative amount of cases for which the time series of both participants show the value “1” (i.e. the first 500 ms of gaze towards the face of the co-participant). What matters most for the interpretation of the plot, are the peaks (in blue), their position relative to t_0 , and the baseline (in orange, cf. *infra*).

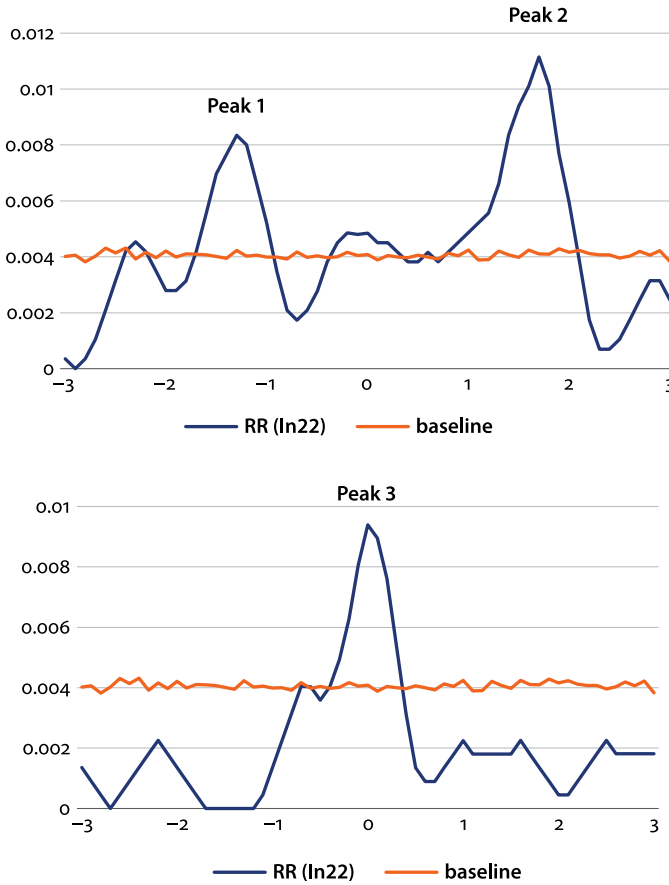


Figure 5. Plots of the recurrence rates from the CRQA analysis. The plot at the top (data from conversation ln22) a pattern with systematic time lags; the bottom plot (data from conversation ci21) a pattern of instantaneous synchronisation. (x-axis: time lag in seconds; y-axis: recurrence rate; RR=recurrence rate)

In the top panel of Figure 5, and based on the gaze behaviour in conversation *ln22*, we see an example of synchronisation of making eye contact with a given time lag. Peak 1 should be interpreted as participant 2 following participant 1,⁵ i.e. the most recurring pattern when participant 1 is the first to look at the face of par-

5. The participants in a dyad were labelled as “participant 1” and “participant 2” according to their seating arrangement: participant 1 sat on the left side of the room; participant 2 on the right side. This order dictates the directionality (who is following who?) of the recurrence plots: the location of the peaks (left, right or spot on) relative to t_0 is determined by the researchers’ choice of assigning a participant to the label “participant 1” or “participant 2”.

ticipant 2, is that participant 2 returns that look 1.3 seconds later. Similarly, peak 2 shows how also participant 1 follows participant 2, i.e. when participant 2 is the first to look at the face of participant 1, it typically takes this participant 1.7 seconds to look at his interlocutor's face. This pattern is different from the one in the bottom panel of Figure 5, based on the data from conversation *ci21*, where we observe a simultaneous synchronisation: the most recurring pattern in this conversation is that participants start looking at each other at exactly the same time, hence the position of peak 3 at t_0 .

In line with the common practise in research on gaze synchronisation (following among others Richardson & Dale (2005) and Louwerse et al. (2012)), and to rule out that our results were obtained by chance or factors unrelated to the interaction, we computed a baseline level of synchronisation. This is relevant because if both interlocutors would start looking at the face of their conversational partner at purely random points in time (i.e. completely regardless of what happens in the interaction), they would still be looking at each other's face at some points in time. These points in time would constitute the chance level of the synchronisation of making and breaking eye contact. We have simulated this random behaviour by reshuffling our event-based time-series data, i.e. every data point in our sampled data was randomly assigned a different position in the time series. Using this procedure we created 1.000 pairs of temporally randomised gaze data. On each of those pairs we then performed a CRQA. The average of those cross-recurrence analyses should be read as the chance level of synchronisation: only if the CRQA plot of the actual data is above the averaged baseline plot, the synchronisation is real and not due to chance alone. An example of this was already shown in Figure 5. The orange line in that plot represents the baseline obtained by our randomisation procedure, the blue line is the CRQA plot of the real data for the conversation under scrutiny.

3.2 Results

The interlocutors in our corpus synchronise their eye gaze during face-to-face conversation, or to be more precise, they synchronise the start of looking at each other's face. If one participant starts to look at the face of the other, the most recurrent pattern is that the other participant reciprocates that fixation on the face of the conversational partner simultaneously. We already saw an example of this type of synchronisation for one dyad in the bottom panel of Figure 5. When averaged across dyads and interaction type (i.e. animation description and brainstorm task), we find that gaze in face-to-face conversation is indeed strongly synchronised (see Figure 6 for the averaged CRQA plot). The peak of the bell curve is

exactly at t_0 , meaning that interactants typically (start) look(ing) at each other at exactly the same time, without any significant time lags.

To test for the significance of the difference between the gaze synchronisation in the real data (Figure 6, blue line) and the baseline data (orange line), we calculated a mixed effects model. As fixed effect we entered the variable *real-vs-base* (binomially indicating whether the data come from the shuffled baseline or the real interactions). As random effect we added *dyad* (categorically indicating a code for each dyad) to the model. The *recurrence rates* (i.e. the values on the vertical axes in the individual cross-recurrence plots) were the dependent variable. The mixed effects model confirms that the synchronisation we observe from a visual inspection of the plot in Figure 6 is not due to chance, i.e. the recurrence rates in the real data are significantly larger than those in the baseline data ($t = 24.12, p < 0.001$).

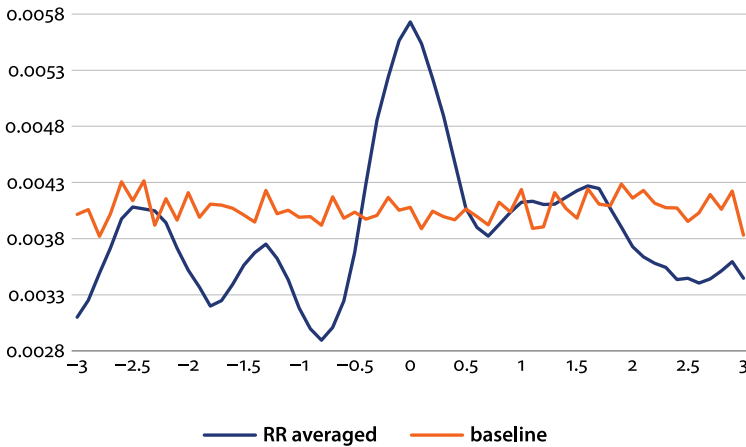


Figure 6. Recurrence rate plot of starting to look at the co-participant, averaged for all conversations in the data set. (x-axis: time lag in seconds; y-axis: recurrence rate; RR=recurrence rate)

Participants in our data not only synchronised their making of eye contact, they also synchronised their breaking of eye contact. To be more precise, participants synchronised the onsets of fixations away from the face of their conversational partners. In contrast to making eye contact, breaking eye contact appears to occur with an asymmetrical temporal pattern. The peak of the plot is flattened and skewed to the right. This pattern is indicative of a leader-follower pattern: participant 1 follows participant 2 rather than the other way around. More precisely, participant 2 is systematically the first one to gaze away from the face of participant 1, with the latter reciprocating this gaze aversion, typically between 100 and

700 ms later. Analogous to the statistical test for making eye contact, also for the synchronisation of breaking eye contact our mixed effects model showed that the recurrence rates in the real data are significantly larger than those in the baseline data ($t=18.49$, $p<0.001$).

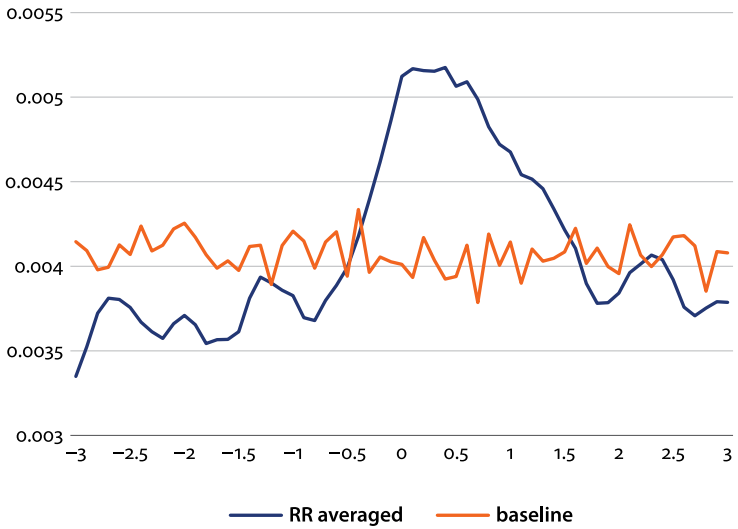


Figure 7. Recurrence rate plot of looking away from the co-participant, averaged for all conversations in the data set. (x-axis: time lag in seconds; y-axis: recurrence rate; RR=recurrence rate)

As a final step in our quantitative analysis, we checked whether the conversational task affected gaze synchronisation, i.e. we compared the synchronisation in the animation description task to that in the brainstorm task. Mixed effects models with *recurrence rates* as dependent variable, *dyad* as random factor and *task* as fixed factor revealed that, both for the making and the breaking of eye contact, the recurrence rates between the two tasks did not differ. The gaze synchronisation we observe, thus holds for both types of conversational task.

3.3 Discussion of the quantitative results

For making eye contact, the CRQA paints a picture of perfect synchronisation, i.e. the most recurrent pattern in the data is that of both participants starting to look at each other's face at exactly the same time. This perfect synchrony cannot arise if interlocutors wait to be looked at, before looking at their conversational partners themselves. Rather, interlocutors are able to mutually predict a gaze shift by their co-participant towards their own face. The mechanisms driving this synchroni-

sation, cannot be directly deduced from our corpus-based quantitative approach. One line of reasoning would be that the gaze synchronisation we observe is in line with behaviour matching in other bodily articulators and is comparable to synchronisation of body posture (Shockley et al. 2003), adaptors like face scratching or foot shaking (Chartrand & Bargh 1999), blinking (Nakano & Kitazawa 2010), head movements or combinations at multiple levels (cf. Louwerse et al. 2012). In this vein, the tight temporal coupling of making eye contact results from very mechanistic processes in which bodily synchrony is both a result of and a facilitator for fluent social interaction (cf. Fusaroli & Tylén 2016). Another line of reasoning could be found in the well-developed literature on the relationship between eye gaze and turn taking: the observed synchronisation between participants might be a mediated one, and thus explained as the simultaneous reaction to a gaze-external event (for example the projectability of a transition relevance place), rather than a gaze-internal dependency. In this sense, gaze synchronisation arises because participants jointly respond to a conversational event, rather than to the eye gaze behaviour of their conversational partner. This point will be further discussed in Section 3.4.

Next to the synchronisation of making eye contact, we also zoomed in on the synchronisation of breaking eye contact. The profile of the plot in Figure 7, which is skewed to the right and has a flat peak, requires some further discussion. A leader-follower pattern, with the peak of the curve not occurring at t_0 , would in itself be quite a straightforwardly explainable result: in such a pattern participants would react to each other's gaze behaviour. In other words, if one participant breaks the eye contact, the other participant will follow soon after. This mechanism would, however, result in an averaged plot with a double peak (as in the left panel of Figure 5). Our data show a systematicity that is harder to explain: participant 1 follows participant 2 in breaking eye contact, but not the other way around. Because both participants have equal roles in the conversations (unlike for example in dyads where one is the director and the other the matcher, or one is the interviewer and the other the interviewee), such an unbalanced skewness-to-the-right is unexpected. The only systematicity in how our participants were labelled as "participant 1" and "participant 2", resided in the seating arrangement: participant 1 sat on left side of the room; participant 2 on the right side. This issue is further discussed in Section 3.5.

Notwithstanding the unexpected one-sided skewness for synchronisation of breaking eye contact, we want to stress there was synchronisation, i.e. participants do not break eye contact at points in time that are unrelated to when their conversational partners break eye contact. Unlike for making eye contact, the synchronisation for breaking eye contact was not simultaneous: a gaze aversion by one interactant was reciprocated with a gaze aversion by the other within a 100

to 700 ms timespan. This time lag disfavors the interpretation of a simultaneous reaction to a gaze-external event. If the gaze aversion was a joint reaction to an observable (or projectable) conversational event, we would expect the gaze aversions to occur simultaneously. In our data this is not the case.

Because our quantitative approach is only informative of whether participants time-align their making and breaking of eye contact, and not of why they do so, a further qualitative exploration of the data is required.

3.4 Further explaining the observed synchronisation in qualitative observations

The quantitative analysis painted a picture of simultaneous synchronisation for eye contact making, and a leader-follower pattern with time lags ranging between 100 and 700 ms for eye contact breaking. To further investigate these occurrences of simultaneously establishing eye contact and delayed eye contact breaking, we used the search tools in ELAN (version 6.4) to:

- i. locate all cases in which both interlocutors start looking at each other within a 100 millisecond time range ($n = 111$)
- ii. locate all cases in the 100 to 700 ms range for interlocutors who stop looking at each other ($n = 314$)

One of the recurrent patterns we observed in the cases resulting from the ELAN search string described in (i) and (ii) was that of ‘looking away while thinking’ (cf. Auer & Zima 2021, Bavelas & Chovil 2018, Goodwin & Goodwin 1986, Heller 2021). More specifically, we found that pauses, hesitation markers and information structure might play a role in allowing interlocutors to time-align their making and breaking of eye-contact.

In Excerpt 1, S2 first asks whether she can start with describing her animation, to which S1 responds that this is ok (“ja” *yes*) in line 1 and S2 starts with the overall scene in which the animation takes place (i.e. the sea) in line 2. S2 displays different features of conversational disfluency, i.e. the use of two pauses and the use of “ja” (*yeah*) acting as a stalling device, and averts her gaze away from S1. Line 2 therefore constitutes a case of ‘looking away while thinking’. Even though the message is fairly simple (S2 expresses that her animation took place in the sea), S2 hesitates to use the label “sea”, possibly considering alternatives such as “under water”, “in a lake”, etc. During S2’s gaze aversion, S1 looks away as well. Both re-establish mutual gaze when S2 retrieves and utters the words “de zee” (*the sea*). This synchronised re-establishing of eye-contact occurs after a 350 ms pause by S2.

This excerpt does not only showcase an instance of the simultaneous re-establishing of eye contact, it also shows the more consecutive pattern for gazing

away from the interlocutor's face: in line 2, S1 averts her eye gaze simultaneously with the onset of her speech, which is followed by a gaze aversion by S1 509 ms later. This time lag of the gaze aversion by S1 lies within the 100 to 700 ms range we observed in the CRQA.

Excerpt 1. (conversation *bl23* in the animation description task)

```

1. S1: ja (0.4)
      yes (0.4)
      G1: S2-----
      G2: S1-----
2. S2: ik zag (0.5) ja (0.3) de zee
      I saw (0.5) yeah (0.3) the sea
      G1: S2-----away-----S2-----
      G2: away-----S1-----

```

In Excerpt 2, taken from the brainstorm task, we see a similar pattern. Both interlocutors are discussing whether a mobile phone designed for women should contain very stereotypical female features (e.g. colourful, shiny, containing make-up tools, etc.) or not. In the excerpt, S1 sets up her line of reasoning in line 1 (“het is niet omdat ge een vrouwelijke gsm wilt” *it's not because you want a female mobile phone*), but she starts hesitating in formulating the outcome of that set-up. This hesitation is apparent from S1 looking away, and the adverb “dan” (*then*) in between two pauses. S1 looks away from S2 right at the start of her first pause, followed by S2 (160 ms later) who also averts his eye gaze. The very short latency of 160 ms makes it unlikely that the gaze aversion by S2 is (only) a reaction to the gaze aversion by S1. The projectability of the pause might play a role here as well. Both participants re-establish eye-contact after the second pause and on the stressed word “roze” (*pink*) following that pause.

Excerpt 2. (conversation *paar3* in the brainstorm task)

```

1.S1: maar(0.5)het is niet omdat ge een vrouwelijke gsm wilt
      but(0.5)it's not because you want a female mobile phone
      G1: away-----S2-----
      G2: S1-----
2. S1: dat die (0.4) dan (0.3) roze moet zijn metglitters
      that it (0.4) then (0.3) has to be pink with glitter
      G1: S2-----away-----S2-----
      G2: S1-----away-----S1-----

```

In both excerpts, and in many of the examples labelled as ‘looking away while thinking’ we observe the same pattern. The current speaker is searching for words, looks away from the conversational partner, buys some conversational time by producing (filled) pauses and hesitations markers, and re-establishes eye-contact after a brief pause. The re-establishing of eye-contact occurs simultaneously for

both interactional partners; the gaze aversion is manifested with a leader-follower pattern in which the current speaker is the first to break eye contact, followed by the conversational partner a few hundred milliseconds later.

This gaze aversion while thinking is consistent with the findings of Jehoul et al. (2017) on filled pauses: these authors found that the non-nasal filler “*eh*” is used for word searches and often entails gaze aversion, and that the nasal filler “*eu*” is associated with more complex thinking processes and nearly always involves the speaker looking away from the addressee. The more elaborate hesitations we encounter in our data, occur at points in the discourse where a transition to a next speaker is highly unlikely. In Excerpt 1 for example, the speaker sets up a clause in which a direct object is required: the set-up “*ik zag*” (*I saw*) before the hesitation is not a complete clause and requires a further complementation with what was seen by the subject “*ik*” (*I*). In addition, the gaze aversion is a further element disfavouring a turn transition. Because the addressee is given signals that the current speaker wishes to keep the floor, and because within-speaker pauses longer than 500 ms are rare (see Ten Bosch et al. 2005 for pause durations in Dutch conversations), the exact moment at which the speaker re-engages in finishing the construction that was set up, is quite predictable. This projectability in turn taking is of course far from new (see the seminal work in Conversation Analysis by Sacks et al. (1974) or Auer (2005), or experimental work by De Ruiter et al. (2006)). However, what mobile eye-tracking data in the current analysis add, is providing evidence for the projectability of turn management within turns (rather than between them).

Both for a speaker and an addressee, knowing that the speaker intends to keep the turn, and that within-utterance pauses are typically at around 300 ms (Ten Bosch et al. 2005: 83), conversational partners can anticipate when the speaker will resume. If addressees were to wait until the onset of speech, the pattern of synchronised eye-contact we observe, would not be possible because they would be ‘too late’ in jointly establishing eye-contact with the speaker.

3.5 Functional quantification

The CRQA analyses allowed us to unearth the exact temporal relationship between interlocutors’ making and breaking of eye contact. The subsequent qualitative analysis highlighted one type of pattern that can explain the quantitative observations. In a final step, we want to further annotate the search results obtained in (i) and (ii) to arrive at a more complete picture of the interactional processes that might give rise to the synchronisation patterns provided by the CRQA.

As is apparent from Figure 8, nearly half of making/breaking eye contact is linked to our animation description task: interlocutors start and stop looking at

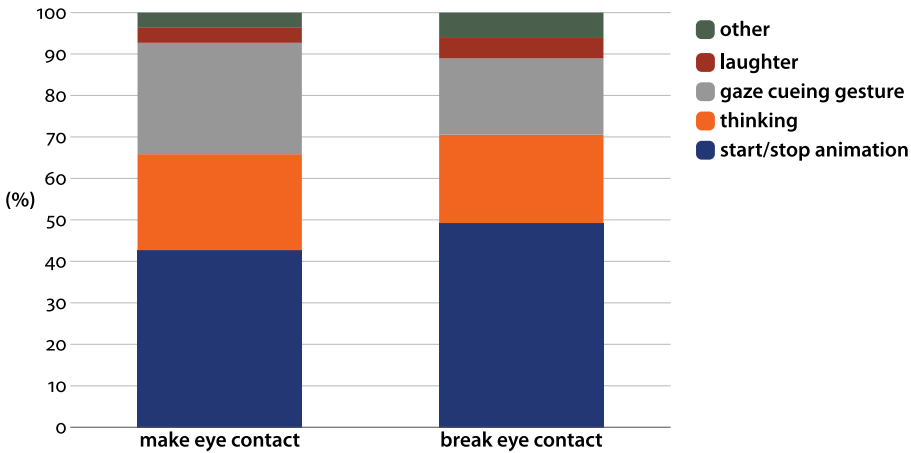


Figure 8. Distribution of functions of making ($n = 111$) or breaking eye contact ($n = 314$)

each other when resp. stopping and starting the next animation. This means that a large part of the observed synchronisation can be explained by the specificity of the task. This is not a trivial observation, given that most studies on behavioural synchronisation in the domain of cognitive science do not perform a functional corpus linguistic or qualitative conversational analytic analysis. That is especially relevant knowing that most of these studies also start from explicitly task based interactions, as was the case in our study. As a consequence, we advocate that researchers should take sufficient care in extrapolating observations of synchronisation in task-based interaction beyond this specific context.

Second, interlocutors time-align their eye contact when looking away while thinking. Both in the animation description task (when participants seem to visualize their animation while looking away) and in the brainstorm task (when participants appear to think of a next possible feature while looking away), gaze synchronisation frequently occurs during these moments of filled or unfilled pauses while thinking what to say next. A third frequent moment of gaze synchronisation arises during gaze-cued gestures. A large body of research (for an overview, see Frischen et al. 2007) has demonstrated that people’s gaze is drawn towards the gaze direction of co-present, or even very sketchily displayed, faces. In our data set, gaze cueing occurs when speakers look at their own gestures as a cue for the other participant to focus on that gesture as well. We observe that both breaking (gaze towards the gestures) and making (gaze back towards the face) eye contact, is accounted for by cases of gaze cueing. Two final observations to be made from Figure 8 are that gaze synchronisation (both making and breaking eye contact) sporadically occurs when interlocutors laugh, and that only a few cases of gaze synchronisation could not be attributed to any of the categories

described above. This is again not a trivial observation, because it emphasizes that the gaze synchronisation we observe in our data is not only, as one might expect, governed by the turn taking machinery (and the role of eye gaze therein). Participants in our data do not typically start or stop looking at each other because they or their conversational partners start or stop speaking. Rather, they synchronise their making and breaking of eye contact to perform the joint action of playing the video animations, to look at their own and each other's gestures or when jointly laughing.

As such, our more in-depth qualitative and functional corpus-based enrichment of the eye-tracking data has allowed us to put the synchronisation of making and breaking eye-contact (as observed through a purely quantitative analysis, void of any functional or contextual interpretation) into perspective.

4. Conclusion

Eye gaze behaviour is highly dynamic yet highly structured. Eye-tracking allows researchers to capture this behaviour in great detail. With this contribution we wanted to highlight how, next to qualitative approaches (e.g. in Conversation Analysis) and quantitative approaches (e.g. in experimental psychology), also a mixed-methods approach (i.e. a corpus-linguistics inspired approach) can yield fruitful insights into gaze behaviour during face-to-face interactions.

In a first case study on interactional irony, we started from a quantitative study that revealed how interlocutors (both speakers and addressees) tend to produce more gaze shifts, and more often (and for longer periods of time) establish eye contact during ironic utterances, compared to non-ironic ones. A qualitative follow-up study highlighted how eye gaze and head movements jointly operate when performing ironic utterances, and how ironists use eye gaze to invite co-participants to join in the ironic pretence. Based on these findings, a second quantitative study was conducted to dig into the multimodal clustering of non-verbal behaviour during ironic utterances, and to sharpen the temporal/sequential unfolding of that behaviour. This study revealed that ironists recruit more of the multimodal resources under scrutiny than non-ironists, and that gaze shifts and mutual eye gaze (already observed in the first quantitative study) appear to cluster with laughter into a multimodal ironic package. What this overview shows, is how the qualitative part of our study was used to nuance the findings of the earlier quantitative study. We feel that intertwining approaches in this case led to a better understanding of how eye gaze functions as a tool to display and monitor interactional irony.

In the second case study on gaze synchronisation, we used cross-recurrence quantification techniques to check whether interlocutors time-align their gaze behaviour relative to the gaze behaviour of their conversational partners. We found that this type of synchronisation indeed occurs, more specifically, interlocutors synchronise their making and breaking of eye contact. In a functional annotation, typical for corpus-linguistics approaches to interactional phenomena, we analysed all cases of simultaneous (i.e. within a 100 ms window) making or breaking of eye-contact. We observed that roughly half of these cases were due to the specificity of the task (i.e. starting and stopping the animations participants had to discuss), and that roughly one third of the cases occurred during gaze-cueing of hand gestures (i.e. simultaneously looking at the hands of one of the interlocutors) or during ‘looking away while thinking’. This final category was subjected to further close-reading, in an attempt to explain how interlocutors manage to truly synchronise their looking at each other. To achieve that level of synchronisation, interlocutors have to make a projection of when their conversational partner will be looking at them (rather than wait for that moment to occur, and only then respond to it). The close-reading allowed us to put forward the hypothesis that implicit knowledge of the duration of utterance-internal pauses, together with other turn-management conventions (i.e. speakers using hesitation markers in the middle of an unfinished clause want to maintain the conversational floor) allow for near-perfectly simultaneous cases of re-establishing eye contact.

Even though the insights from our cases studies are not necessarily groundbreaking, they do make a case for letting qualitative and quantitative techniques feed into each other. The fine-grained and abundant data that mobile eye-tracking devices deliver, can be used to look at the same data from multiple angles. Such a multiperspective approach can lead to new insights, nuance existing findings, or provide fertile ground for formulating hypotheses that can be further tested with any type of scientific design.

References

- Argyle, Michael and Cook, Mark. 1976. *Gaze and Eye contact*. London: Cambridge University Press.
-  Auer, Peter. 2005. “Projection in Interaction and Projection in Grammar.” *Text – Interdisciplinary Journal for the Study of Discourse* 25(1), 7–36.
-  Auer, Peter. 2021. “Turn-allocation and gaze: A multimodal revision of the “current-speaker-selects-next” rule of the turn-taking system of conversation analysis.” *Discourse Studies* 23(2), 117–140.
- Auer, Peter & Zima, Elisabeth. 2021. “On word searches, gaze, and co-participation.” *Gesprächsforschung – Online – Zeitschrift zur verbalen Interaktion* 22, 390–425.

- doi Barnden, John A. 2017. "Irony, pretence and fictively-elaborating hyperbole." In *Irony in language use and communication*, ed. by Aneqliki Athanasiadou and Herbert L. Colston, 145–178. Amsterdam: John Benjamins.
- doi Bavelas, Janet and Chovil, Nicole. 2018. Some pragmatic functions of conversational facial gestures. *Gesture* 17, 98–127.
- doi Bavelas, Janet, Coates, Linda and Johnson, Trudy. 2002. Listener responses as a collaborative process: The role of gaze. *Journal of Communication* 52, 566–580.
- doi Bishop, Laura, Cancino-Chacon, Carlos, and Goebl, Werer. 2021. "Beyond synchronization: Body gestures and gaze direction in duo performance." In *Together in Music: Participation, Co-Ordination, and Creativity in Ensembles*. ed. by Renee Timmers, Freya Bailes, and Helena Daffern, 182–188. Oxford: Oxford University.
- doi Brennan, Susan, Chen, Xin, Dickinson, Christopher A., Neider, Mark B., and Zelinsky, Gregory. 2008. "Coordinating cognition: The costs and benefits of shared gaze during collaborative search." *Cognition* 106, 1465–1477.
- doi Brône, Geert. 2008. "Hyper- and misunderstanding in interactional humour". *Journal of Pragmatics* 40(12), 2027–2061.
- Brône, Geert. 2021. "The multimodal negotiation of irony and humour in interaction: On the role of eye gaze in joint pretense". In *Figurative Thought and Language* (Vol. 11), ed. by Augusto Soares da Silva, 109–136. Amsterdam: John Benjamins Publishing Company.
- doi Brône, Geert and Oben, Bert. 2015. "InSight Interaction: A multimodal and multifocal dialogue corpus." *Language Resources and Evaluation* 49(1), 195–214.
- doi Brône, Geert and Oben, Bert (Eds.) 2018. *Eye-tracking in Interaction: Studies on the role of eye gaze in dialogue*. (Advances in Interaction Studies (AIS)). Amsterdam: John Benjamins.
- Brône, Geert and Oben, Bert. 2022. "Monitoring the pretence. Intersubjective grounding, gaze and irony." In *Cognitive Sociolinguistics Revisited*. ed. by Gitte Kristiansen, Karlien Franco, Stefano De Pascale, Laura Rosseel and Weiwei Zhang, 544–556. Berlin: De Gruyter Mouton.
- doi Brône, Geert & Oben, Bert. 2023. Mobile eye-tracking for multimodal interaction analysis. In *The Routledge Handbook of Experimental Linguistics*. ed. by Sandrine Zufferey and Pascal Gygas, 283–298. London: Routledge.
- doi Chartrand, Tanya L. and Bargh, John A. 1999. "The chameleon effect: The perception behaviour link and social interaction." *Journal of Personality and Social Psychology* 76(6), 893–910.
- doi Clark, Herbert H. 1996. *Using Language*. Cambridge University Press.
- doi Clark, Herbert H. 2016. "Depiction as a method of communication." *Psychological Review* 123–3, 324–347.
- doi Clark, Herbert H. & Gerrig, Richard. 1984. On the pretense theory of irony. *Journal of Experimental Psychology: General* 113, 121–126.
- doi Coco, Moreno I. and Dale, Rick. 2014. "Cross-recurrence quantification analysis of categorical and continuous time series: an R package." *Frontiers in Psychology* 5, 510.
- Coulson, Seana. 2005. "Sarcasm and the space-structuring model." In *The Literal and Nonliteral in Language and Thought*. ed. by Seana Coulson and Barbara Lewandowska-Tomaszczyk. 129–144. Frankfurt: Peter Lang.

- Dale, Rick, Kirkham, Natasha and Richardson, Daniel. 2011. "How two people become a tangram recognition system." *Proceedings of the European Conference on Computer-Supported Cooperative Work*. Berlin: Springer Verlag.
- doi de Ruiters, Jan, Mitterer, Holger and Enfield, Nick. 2006. "Predicting the end of a speaker's turn: A cognitive cornerstone of conversation." *Language* 82, 515–535.
- doi De Stefani, Elwys. 2021. "Embodied responses to questions-in-progress: silent nods as affirmative answers." *Discourse Processes* 58(4), 353–371.
- doi de Vries, Clarissa, Oben, Bert and Brône, Geert. 2021. "Exploring the role of the body in communicating ironic stance." *Languages and Modalities* 1(1), 65–80.
- ELAN (Version 6.4) [Computer software]. 2023. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>
- doi Frischen, Alexandra, Bayliss, Andrew P. and Tipper, Steven. 2007. "Gaze cueing of attention: Visual attention, social cognition, and individual differences." *Psychological Bulletin* 133(4), 694–724.
- doi Fusaroli, Riccardo and Tylén, Kristian. 2016. "Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance." *Cognitive Science* 40(1), 145–71.
- doi Fusaroli, Riccardo, Konvalinka, Ivana and Wallot, Sebastian. 2014. "Analyzing social interactions: The promises and challenges of using cross recurrence quantification analysis." *Springer Proceedings in Mathematics and Statistics* 103, 137–155.
- doi Gibbs, Raymond W. 2000. Irony in Talk Among Friends. *Metaphor and Symbol* 15(1). 5–27.
- Gironzetti, Elisa, Pickering, Lucy, Huang, Meichan, Zhang, Ying, Menjo, Shigehito and Attardo, Salvatore. 2016. "Smiling synchronicity and gaze patterns in dyadic humorous conversations." *Humour: International Journal of Humour Research* 29(2), 301–324.
- Goodwin, Charles. 1981. *Conversational Organization: Interaction Between Speakers and Hearers*. Cambridge, MA: Academic Press.
- doi González-Fuente, Santoago, Escandell-Vidal, Victoria and Prieto, Pilar. 2015. "Gestural codas pave the way to the understanding of verbal irony." *Journal of Pragmatics* 90, 26–47.
- Goodwin, Marjorie and Goodwin, Charles. 1986. "Gesture and coparticipation in the activity of searching for a word." *Semiotica* 62, 51–75.
- doi Hadelich, Kerstin and Crocker, Matthew. 2006. "Gaze alignment of interlocutors in conversational dialogues." *Proceedings of the 2006 Symposium on Eye Tracking Research and Applications*, 38.
- doi Heller, Vivien. 2021. Embodied displays of "Doing Thinking." Epistemic and interactive functions of thinking displays in children's argumentative activities. *Frontiers in Psychology* 12: 636671.
- Jehoul, Annelies, Brône, Geert and Feyaerts, Kurt. 2017. "Gaze patterns and fillers. Empirical data on the difference between Dutch 'euh' and 'euhm'." *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016)*, 43–50.
- doi Jokinen, Kristiina, Furukawa, Hirohisa, Nishida, Masafumi, Yamamoto, Seiichi. 2013. Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems* 3 (2), Article 12, 30 pages.
- doi Kendon, Adam. 1967. "Some functions of gaze-direction in social interaction." *Acta Psychologica* 26, 22–63.

- doi Kendrick, Robin H. and Holler, Judith. 2017. "Gaze Direction Signals Response Preference in Conversation." *Research on Language and Social Interaction* 50(1), 12–32.
- doi Lachat, Fanny, Conty, Laurence, Hugueville, Laurent and George, Nathalie. 2012. "Gaze cueing effect in a face-to-face situation." *Journal of Nonverbal Behaviour* 36, 177–190.
- doi Louwerse, Max, Dale, Rick, Bard, Ellen and Jeuniaux, Patrick. 2012. "Behaviour matching in multimodal communication is synchronized." *Cognitive Science* 36, 1404–1426.
- doi Mondada, Lorenza. 2014a. "The local constitution of multimodal resources for social interaction." *Journal of Pragmatics* 65, 137–156.
- doi Mondada, Lorenza. 2014b. "Pointing, talk and the bodies: reference and joint attention as embodied interactional achievements" In *From Gesture in Conversation to Visible Utterance in Action*, ed. by Mandana Seyfeddinipur and Marianne Gullberg, 95–124. Amsterdam: John Benjamins, 95–124.
- doi Nakano, Tamami and Kitazawa, Shigeru. 2010. "Eyeblink entrainment at breakpoints of speech." *Experimental Brain Research* 205, 577–581.
- doi Neider, Mark B., Chen, Xin, Dickinson, Christopher, Brennan, Susan and Zelinsky, Gregory. 2010. "Coordinating spatial referencing using shared gaze." *Psychonomic Bulletin & Review* 17, 718–724.
- doi Oertel, Chatarine, Wlodarczak Marcin, Edlund Jens, Wagner Petra and Gustafson Joakim. 2012. "Gaze patterns in turn-taking." *Proceedings of Interspeech*, 2247–2250.
- doi Pfeiffer, Martin & Clarissa Weiss. 2022. Reenactments during tellings: Using gaze for initiating reenactments, switching roles and representing events. *Journal of Pragmatics* 189. 92–113.
- doi Richardson, Daniel and Dale, Rick. 2005. "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension." *Cognitive Science* 29, 1045–1060.
- doi Richardson, Daniel, Dale, Rick and Tomlinson, John. 2009. "Conversation, gaze coordination, and beliefs about visual context." *Cognitive Science* 33, 1468–1482.
- doi Sacks, Harvey, Schegloff, Emanuel. and Jefferson, Gail. 1974. "A simplest systematics for the organization of turn-taking for conversation." *Language* 50, 696–735.
- Shockley, Kevin, Santana, Marie-Vee and Fowler, Carol. 2003. "Mutual interpersonal postural constraints are involved in cooperative conversation." *Journal of Experimental Psychology: Human Perception and Performance* 29, 326–332.
- doi Sidnell, Jack. 2006. Coordinating Gesture, Talk, and Gaze in Reenactments. *Research on Language and Social Interaction* 39(4). 377–409.
- Sperber, Dan and Wilson, Deirdre. 1981. "Irony and the use-mention distinction." In *Radical Pragmatics*. ed. by Peter Cole. 295–318. New York: Academic Press.
- doi Stukenbrock, Anja. 2020. "Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention." *Frontiers in Psychology* 11: 1779.
- doi Tabacaru, Sabina and Lemmens, Maarten. 2014. "Raised eyebrows as gestural triggers in humour: The case of sarcasm and hyper-understanding." *The European Journal of Humour Research* 2(2), 11–31.
- doi Ten Bosch, Louis, Oostdijk, Nelleke and Boves, Lou. 2005. "On temporal aspects of turn taking in conversational dialogues." *Speech Communication* 47(1), 80–86.
- doi Thompson, Sandra A. & Ryoko Suzuki. 2014. Reenactments in conversation: Gaze and reciprocity. *Discourse Studies* 16(6). 816–846.

-  Tobin, Vera. 2016. Performance, Irony, and Viewpoint in Language. In *Theatre, Performance and Cognition: Languages, Bodies and Ecologies*. ed. by Amy Cook and Ronda Blair. 54–67. London: Bloomsbury Methuen.
-  Vandemoortele, Sarah, Feyaerts, Kurt, Reybrouck, Mark, De Bièvre, Geert, Brône, Geert, De Baets, Thomas. 2018. “Gazing at the partner in musical trios: a mobile eye-tracking study.” *Journal of Eye Movement Research* 11 (2), 1–13.
-  Vrzakova, Hana, Amon, Mary J., Stewart, Angela and D’Mello, Sidney. 2019. “Dynamics of Visual Attention in Multiparty Collaborative Problem Solving using Multidimensional Recurrence Quantification Analysis.” *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ‘19)*, Paper 342, 1–14.
-  Wallot, Sebastian. 2017. “Recurrence quantification analysis of processes and products of discourse: A tutorial in R.” *Discourse Processes* 54(5–6), 382–405.
-  Xu, Tian, de Barbaro, Kaya, Abney, Drew and Cox, Ralf. 2020. “Finding Structure in Time: Visualizing and Analyzing Behavioural Time Series.” *Frontiers in Psychology* 11, 1457.
-  Zima, Elisabeth. 2020. Gaze and Recipient Feedback in Triadic Storytelling Activities. *Discourse Processes* 57(9). 725–748.
-  Zima, Elisabeth, Weiß, Clarissa and Geert Brône. 2019. “Gaze and overlap resolution in triadic interactions.” *Journal of Pragmatics* 140, 49–69.

PART 2

Exploring interactional phenomena with mobile eye tracking

Stationary settings

On the relationship between gaze and the German recipient token *hm_hm*

Johanna Masuch

University of Freiburg

This study examines the interplay between speaker gaze, the German recipient token “hm_hm,” and the Feedback Relevance Space (FRS) in triadic conversations. Our findings underscore the role of speaker gaze in eliciting “hm_hm” tokens: over three-quarters of them were produced by the recipient last looked at by the speaker. However, it challenges previous accounts of the relationship between gaze and recipient feedback (Bavelas et al. 2002) by showing that recipients prioritise the timely placement of “hm_hms” and orient to the FRS — typically located towards the end of the speaker’s intonation phrase — over giving feedback while in mutual gaze with the speaker. This is consistent with the function of “hm_hm” as a continuer. The findings of this study contribute to a refined understanding of the relationship between gaze, recipient feedback and turn-taking.¹

Keywords: recipient feedback, hm_hm, continuer, German triadic interactions, gaze window, Feedback Relevance Space

1. Introduction

While listening to a speaker produce a turn, recipients often respond to the talk by means of vocalisations such as *uh huh*, *yeah*, and *okay*, or non-verbally through gestures, smiles, or head nods. These forms of recipient behaviour provide information on how the talk has been understood and what kind of stance the recipient takes (Gardner 2013). As such, they are an integral part of every conversation and actively shape the course of the interaction (Tolins and Tree 2014: 152). This primordial role of feedback behaviour for human communication is reflected in

1. I would like to thank Peter Auer and Elisabeth Zima for helpful advice on earlier versions of this chapter. I am also grateful for the comments of two anonymous referees, which were of great help in improving my manuscript.

the plethora of studies dedicated to understanding the functional distinctions between various response tokens (Goodwin 1986, Clark & Krych 2004), their temporal placement relative to the speaker's talk (Duncan & Fiske 1977, Goodwin 1986), and the linguistic, prosodic, and embodied cues that invite them (Goodwin 1981, Koiso et al. 1998, Ward & Tsukahara 2000, Morency et al. 2009, Stivers & Rossano 2010, Bavelas et al. 2002). Among these cues, gaze seems to play a particularly prominent role. However, recent studies on the relationship between gaze and backchannel behaviour (Bavelas et al. 2002, De Kok & Heylen 2012, Rossano 2012a, Zima, 2020) do not draw a uniform picture, as authors reach partly contradictory conclusions.

In their seminal paper on dyadic storytelling, Bavelas, Coates, and Johnson (2002) proposed that feedback tokens are systematically elicited by speakers' gaze shifts and occur within brief periods of mutual gaze with the recipient, a phenomenon known as the 'gaze window' pattern. However, more recent studies have contested this concept. Rossano (2012b) argues that the notion of the 'gaze window' lacks precise definition, particularly with respect to its temporal boundaries, while Zima (2020) contests the generalisability of Bavelas et al.'s (2002) observations. In her study on triadic storytelling activities, she used mobile eye tracking glasses to record interactants' gaze behaviour and found that only about one third of feedback tokens were embedded in gaze windows. The main reason for this divergent finding is that speakers often resolve mutual gaze phases before recipients actually give feedback. The feedback tokens (verbal tokens and nods) are thus more often than not not embedded in a mutual gaze phase. Zima explains this observation with competing demands on speakers at the boundary of Turn Constructional Units (TCUs), the *natural home* of feedback tokens as in triadic interactions, they not only need to monitor their recipients' displays of (mis)understanding, but also manage turn-taking and plan their next utterance.

Although recent efforts have thus been made to clarify the relationship between gaze and feedback, its exact nature remains largely unclear. This chapter aims to contribute to the topic by zooming in on the details of the temporal relationship between speaker gaze shifts, the establishing of mutual gaze and the utterance of feedback tokens focussing on the German recipient token *hm_hm*. More specifically, we report the findings of a corpus study on 536 tokens of *hm_hm* extracted from appr. 5.5 hours of recorded triadic interactions during which all participants' gaze behaviour was tracked by mobile eye tracking glasses (SMI and Tobii Pro Glasses 2).

In line with Zima's findings, we will show that speaker gaze may indeed prompt *hm_hm* responses, but often they are placed outside periods of mutual gaze, disconfirming the gaze window hypothesis (Bavelas et al. 2002). However, we propose a different interpretation than Zima (2020) on why recipients do not

always immediately respond to gaze cues. Our analysis suggests that recipients strategically align their *hm_hm* responses with the Feedback Relevance Space (FRS) at the boundary between Intonation Phrases, prioritising sequential appropriateness over responding within mutual gaze with speakers.

The chapter is structured as follows: First, we provide an overview of previous research on the German recipient token *hm_hm* and its English counterpart, with a focus on their functions and temporal placement relative to the speaker's utterance. Second, we summarise the literature on the relationship between speaker gaze and recipient tokens. The corpus and methodology used to annotate and analyse gaze patterns are then presented in Section 3. Sections 4 and 5 present our corpus study's analysis, providing both quantitative and qualitative insights into the observed gaze patterns. Finally, Section 6 summarises the main findings of our study.

2. Previous research on the function of gaze and the placement of the recipient token *hm_hm*

2.1 The placement of *hm_hm* relative to the speaker's turn

Hm_hm and its equivalents in other languages are often referred to as *generic backchannels* (Goodwin 1986, Bavelas et al. 2000). They do not respond to the topic of the previous talk in a specific way, but rather signal general “understanding and continued attention to the speaker” (Tolins & Tree 2014: 154). Most importantly, by signalling this continued attention, recipients actively construe themselves as being in the recipient role and forgo the opportunity to take the turn or to initiate repair of some problem of hearing or understanding.

In line with its basic function, speakers react to *hm_hm* by continuing their turn in a systematic way, i.e., by adding new information, transitioning “from one subtask to another at the same level of the hierarchy” (Tolins & Tree 2014: 157). Schegloff (1982: 87) coined the term ‘continuer’ to describe this function. It determines when, within the speakers' turn, they can be uttered (cf. Tolins & Tree 2014: 154) as it is structurally relevant for recipients “to display their understanding of the current state of the talk” (Schegloff 1982: 81) in places where a possible completion of the turn is reached and the opportunity to take the turn opens up. In a similar vein, Goodwin (1986: 108) points out that *uh huh's* – the English equivalent of *hm_hm* – “do not occur just anywhere within the turn but rather at the boundaries of turn-constructional units”, i.e. at positions that show that one unit has been understood and that the next one is now anticipated. Often beginning within one turn constructional unit (TCU) and ending in the next, these

feedback tokens thus function as “bridges between units” (Goodwin 1986:208). Accordingly, speakers regularly treat these tokens as signals to continue by beginning the next turn in overlap with *uh huh*.

These findings on the placement of recipient tokens are used by models that predict possible placement points in order to adapt spoken dialogue systems to provide authentic recipient tokens (cf. Cathart et al. 2003:1 f.). Recipient tokens are expected to occur at transition relevance places (TRP) based on the idea that they will be most interpretable by the speaker if they occur at or before an utterance reaches *pragmatic completion* (Cathart et al. 2003:2, emphasis in original). However, evidence suggests that recipient tokens do not occur exclusively at TRPs (cf. Heldner et al. 2013:1). Rather, they are often uttered in-between turns or at IP boundaries that are not TRPs though (e.g., in storytelling activities, see also Howes and Eshghi 2021:335). These places have been referred to as *backchannel relevance spaces* (Heldner et al. 2013) or *feedback relevance spaces* (Howes & Eshghi 2017, 2021), analogous to the term *transition relevance place* (Sacks et al. 1974).

Although all TRPs are FRSs, the opposite does not hold (cf. Howes & Eshghi 2021:335). According to Heldner et al. (2013:2), there are more backchannel relevance spaces than there are vocal or visual recipient tokens actually uttered. Thus, interlocutors do not give feedback at every possible FRS. Heldner et al. found an average backchannel frequency of 14 backchannels per minute, with “on average 3.5 times more backchannel relevance spaces than actual backchannels” (Heldner et al. 2013:7f.).

In this chapter, we use the term *feedback relevance space* to refer to the position where recipient tokens occur, and we show that these are typically located around the end of an intonation phrase (specifically, Sections 3 and 4). Dittmann and Llewellyn’s work (1967) provides converging evidence that the intonation unit (rather than the TCU) is the most relevant unit in speech decoding, and thus a relevant unit for listeners to provide recipient tokens showing that listeners’ responses are “almost exclusively located at the ends of the speakers’ phonemic clauses rather than within them” (Dittmann & Llewellyn 1967:341).

Moreover, a few studies have focused on how speakers invite listeners to give feedback. Quasthoff (1981:301) demonstrated that short pauses, tag questions, and explanations often prompt recipients to produce recipient tokens. Additionally, it has been shown that prosodic and syntactic cues within the speaker’s talk also serve as invitations for recipients to provide feedback (Koiso et al. 1998, Bavelas et al. 2002, Tolins & Tree 2014). Most importantly, it has been argued that also gaze may be used as such a cue.

2.2 The function of gaze to mobilise recipient responses

In a pioneering paper published in 1967, Adam Kendon revealed differences in the gaze behaviour of interlocutors when acting as speakers or recipients. Accordingly, we tend to look at the interlocutor “more while [...] listening than while [...] speaking” (Kendon 1967: 37). Kendon was also the first to illustrate the role of the speakers’ gaze in dyadic interaction to mobilise a response from recipients: “during the course of a long utterance, p’s [i.e. the speaker’s] glances at q [i.e. the recipient] come at the points at which he receives an accompaniment signal [i.e. a recipient token] from him, and so may function not only as checks on q’s [the recipient’s] behaviour, but as signals to p [the recipient] that q [the speaker] wants confirmation that what he is saying is getting across” (Kendon 1967: 56). He called these recipient tokens “accompaniment signals” and defines them as “the short utterances that the listener produces as an accompaniment [...] when the speaker is speaking at length” (Kendon 1967: 43). They are said to help the speaker understand how the speech has been received by the recipient.

Another important study on the function of gaze in mobilising a response in dyadic interactions comes from Goodwin & Goodwin (1986), who examined the role of gaze in the activity of searching for a word. During word searches, recipients typically display engagement with the ongoing activity of searching for a word by gazing at the speaker and thereby signalling continued attention (Goodwin & Goodwin 1986: 55, 67). The authors further argue that when a speaker is unable to solve a word search alone (this solitary activity is marked by an averted gaze), (s)he shifts gaze to a recipient in order to solicit help with the word search, thus transforming the solitary activity into a cooperative one (Goodwin & Goodwin 1986: 64, 67, but for a more nuanced view, see Auer & Zima (2021) corpus study). However, recipients may not always be able to suggest a potentially appropriate word. In these cases, feedback tokens, such as nods or continuers may act as an appropriate response, as they signal to the speaker that there is “some sort of adequate comprehension of what the speaker is trying to say” (Goodwin & Goodwin 1986: 71).

In a similar vein, Stivers & Rossano (2010) argue that gaze is systematically used to mobilise response. They study first pair parts that are less normatively structured, such as announcements, noticings, and assessments, to examine the resources available to speakers to hold the recipients accountable for responding (Stivers & Rossano 2010: 9). Alongside intonation, morphosyntax, and epistemics, the speaker’s gaze towards the recipient is identified as an important characteristic of turn design that mobilises a response in such actions (Stivers & Rossano 2010: 8). The authors further demonstrate that the cumulative use of multiple resources increases the response relevance (Stivers & Rossano 2010: 9).

In another study on Italian question-answer sequences, Rossano (2012a) provides a more in-depth account and argues for the importance of mutual gaze for

mobilising a response: „The occurrence of mutual gaze [...] is an important predictor both of the occurrence of a response and of its occurring promptly” (Rossano 2012a: 153). He further claims that while response mobilisation is driven by the speaker’s gaze shift towards the recipient, the timing of the response (immediate or delayed) depends on whether the recipient is already looking at the speaker or not (Rossano 2012a: 153). However, in general, the power of gaze to mobilise response is limited: “speaker gaze can pressure for responses not at any point in time but rather when it occurs in specific sequential environments” (Rossano 2012a: 154; for a similar argument Auer & Zima 2021). Further evidence for the role of gaze to mobilise response comes from experimental research by De Kok and Heylen (2012) and work on interpreter-mediated interaction by Vranjes (2018).

After this overview of work on response mobilisation in general, the next section zooms in on the role of gaze to mobilise recipient feedback, starting with the most influential account by Bavelas, Coates, and Johnson (2002).

2.3 The gaze window hypothesis

The most explicit account on the relationship between gaze and feedback behaviour has been presented by Bavelas et al. (2002). Their study, in which they propose the *gaze window* pattern, has been highly influential in gaze research (Zima 2020: 3, Rossano 2012b). The authors follow Kendon (1967) in assuming a temporal, causal link between the speakers’ gaze behaviour and the production of recipient feedback, and examine the timing of the recipient’s response² (Bavelas et al. 2002: 571 f.). In their semi-experimental study on dyadic storytelling activities, they discover a specific gaze pattern, which they call *gaze window* (Bavelas et al. 2002: 569). This asymmetric pattern consists of three interdependent parts, which they describe as follows (Bavelas et al. 2002: 569 ff.):

1. Given that recipients typically gaze at speakers for long periods of time (Argyle & Cook 1976, Duncan & Fiske 1977, Goodwin 1981, Kendon 1967, Rossano 2012a, 2012b), a speaker looking at the recipient in order to seek a response from them most often results in a brief period of mutual gaze.
2. Within this period of mutual gaze, the recipient responds immediately.
3. Shortly after the response, the speaker closes the gaze window by averting his gaze and continues to hold the turn.

2. Bavelas et al. (2002) use the term “listener response” but in line with conversation analytic work that has shown that recipients are by no means passive listeners, we prefer the term “recipient”.

According to the authors, the production and timing of the recipient token constitute a collaborative process: the speaker's gaze provides an opportunity to give feedback, the recipient takes advantage of this opportunity and offers a recipient token, which in turn ends the mutual gaze phase (Bavelas et al. 2002: 572). Most importantly, the speaker's gaze at the recipient is taken to *solicit* that response (Bavelas et al. 2002: 578).

Although the gaze window concept has received a lot of attention in interactional linguistic research on gaze, it has also faced criticism, primarily regarding the methodological approach used (cf. Rossano 2012b, Zima 2020: 3). Rossano (2012b: 40) argues that the coding system does not provide evidence "that listeners are responding specifically because of the gaze", because neither the exact temporal relationship between the gaze and the response nor other communicative behaviours that might influence the solicitation of a response were considered. He further criticises the lack of sequential specifications such as actions performed in relation to the gaze window (Rossano 2012b: 40).

Also Zima (2020: 3) takes issue with the study design used in Bavelas et al.'s (2002) approach. She argues that the predetermined communicative roles of the interactants (storyteller or recipients), the very short duration of their elicited interactions, and the method used to annotate gaze, which relies on the use of multiple cameras and mirrors instead of direct measurement, may reduce the reliability of the annotations. Furthermore, she also criticised the lack of a specific set of temporal criteria that constitute gaze windows. In her study on triadic storytelling activities, she proposes setting the following time limits for the three phases of the gaze window pattern and uses mobile eyetracking to analyse participants' gaze behaviour:

- (1) Mutual gaze was established no longer than 1.5 seconds and no less than 250 ms prior to the onset of the feedback token.
- (2) Mutual gaze lasted at least 750 ms and was not ended by either party prior to 250 ms after the onset of the feedback token.
- (3) Mutual gaze was dissolved within 1.5 seconds after the feedback onset.

(Zima 2020: 5)³

Under these conditions, only one third of her data falls into the gaze window category. This refutes the quantitative dominance of the gaze window pattern and

3. The temporal boundary of 1.5 seconds is based on the average length of intonation phrases, as well as the average length of mutual gaze in her data. The time frame of 250 ms is meant to capture the minimum reaction time for recipient to respond to a gaze shift by giving feedback. It is based on psycholinguistic evidence, which shows that it takes at least 200 to 250 ms (under experimental conditions) to react to another person's actions (Zima 2020: 9, Marslen-Wilson 1985).

shows that shifting gaze to a recipient is not as tightly linked to feedback production as suggested by Bavelas et al. (2002). Instead, several different gaze patterns coexist. Zima demonstrates that mutual gaze frequently occurs within turns, but is often broken off shortly before feedback tokens are expressed (both verbal tokens and nods). As a result, these tokens are frequently not produced during mutual gaze.

Following this line of research, this chapter aims to explore the temporal relation between speakers' gaze behaviour and the production of recipient tokens in more detail. Based on the assumption that different functions of recipient tokens may be reflected in different placements (cf. Tolins & Tree 2014: 154, Goodwin 1986), we decided to focus on one specific recipient token, i.e. German *hm_hm*, and to provide an in-depth analysis of its temporal placement in relation to participants' gaze behaviour. More specifically, the aim of our study is to shed light on the reasons why recipients do not always react immediately to a gaze mobilisation by placing the token within a mutual gaze phase. We hypothesise that recipients orient towards the Feedback Relevance Space to give feedback, prioritising this 'timely' placement over its production while being in a state of mutual gaze with the speaker.

3. Corpus and methods

In line with our interest in the fine-grained interplay between gaze behaviour and verbal feedback, we used mobile eye-tracking glasses (SMI and Tobii Pro 2) to record participants' gaze behaviour during conversations. While video recordings from an external view only provide an approximate estimation of head and gaze direction, eye-tracking glasses measure gaze movements and fixations with high accuracy, which is essential for annotating mutual gaze (Zima, Auer & Rühlemann, this volume).

The study is based on eight informal German triadic interactions with a total duration of approximately 5.5 hours. Twelve male and twelve female students took part in the recordings. All participants gave informed written consent to the publication of transcripts and stills from the recordings. approval by the ethics committee at Freiburg University was not required. Some of them knew each other well, while other participants were unacquainted. For each recording, three participants were seated around a table in a triangular formation and asked to freely discuss topics of their choice. The environment was intentionally designed to be minimally stimulating, allowing participants to use their gaze for interactional purposes without being distracted by other tasks such as manipulating objects or observing their surroundings. All three participants wore mobile eye-tracking glasses throughout the entire interaction, while an external video camera recorded the interaction from an observer's perspective at a distance of a few

metres from the group. This recording and the three eye tracking recordings were then synchronised and arranged on a split screen using Adobe Premiere Pro CC (Adobe Inc., San José). Subsequently, the split-screen video and audio files were imported into ELAN (Wittenburg et al. 2006), where the conversations were transcribed according to GAT2-conventions (Selting et al. 2009, see Appendix A).

As a multifunctional signal in spoken German (cf. Ehlich, 2007), *hm(_hm)* is not only used as a recipient token. To keep the analysis consistent, we excluded all non-response tokens of *hm_hm* from our dataset, including the lapse terminator token occurring after long silences, gustatory *hms*, repair initiator *hms*, hesitation marker *hms* and *hms* as answers to questions immediately preceding the token (cf. Gardner 2001).⁴ After the identification of 536 instances of *hm_hms* used as a recipient token in the data, the gaze behaviour of all three participants was coded.⁵ Following the standard procedure for coding gaze fixations (see Jokinen et al. 2013: 12.9, Weiß 2020), a tracking cursor resting on a participant's facial region for at least three frames (40 ms each) was coded as a gaze fixation. A cursor directed to another place in the room was coded as gaze aversion.

These differences in gaze constellations are represented in the notation of the gaze behaviour above the verbal transcripts (cf. Figure 1.). Three triangularly arranged circles reflect the spatial arrangement of the participants and contain the first letter of their abbreviated pseudonyms. Different arrow shapes indicate different gaze patterns: whereas a *double arrow* ($\Rightarrow$) indicates that the participant is looking at another participant's facial area, a *single arrow* ($\rightarrow$) pointing to the environment represents gaze aversion. A *single arrow* ($\rightarrow$) pointing to another participant (but not toward the facial area) represents a gaze toward a participant. Mutual gaze is represented by a double-headed arrow ($\Leftrightarrow$). In the case of a gaze shift, the gaze arrow for the respective participant is missing from the transcript until a specific gaze target is reached. The curled brackets above the verbal transcript mark the relation to the verbal expressions during which the specific gaze constellation is observed. The *hm_hms* that are the focus of the sequential analysis are written in bold.

Figure 2 is a representative screenshot from our split-screen videos (last gaze constellation of Example 1). Dennis (seated on the left) and Zac (seated in the

4. These different uses of *hm* differ greatly in terms of prosody. The gustatory *hm* is elongated and does not occur in our data, as no food was allowed during the recordings. All other functions can be identified by their sequential context. Apart from the answer to a question, these other uses are exclusively monosyllabic.

5. This approach, which starts from the actually expressed recipient token in order to investigate gaze patterns in relation to it, has the disadvantage that the reversed cases were not included, i.e. the cases where the speaker gaze fails to mobilise a recipient token.

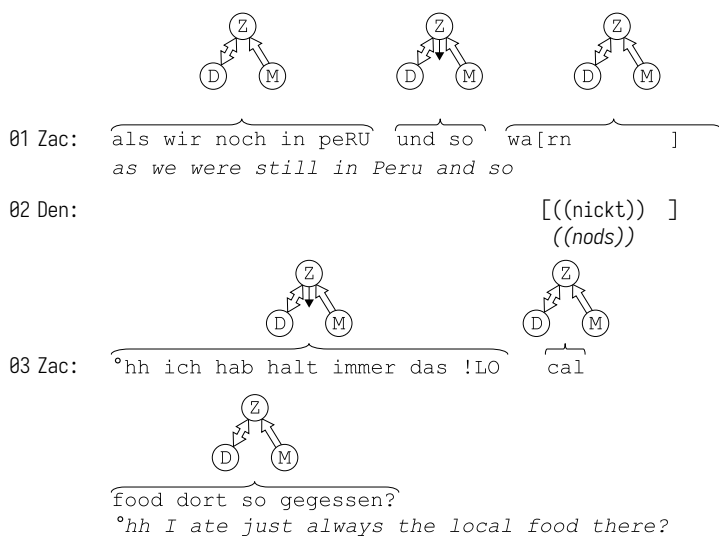


Figure 1. Exemplary transcript (cf. Example 1)

middle) gaze at each other, i.e. both participants' tracking cursors are on each other's faces. Max (on the right) looks at Zac. The split screen is arranged as follows: bottom left is Dennis' view, top left is Zacs' view (sitting in the middle), top right Max's view, and bottom right is the view from the external camera.



Figure 2. Screenshot from a split screen showing a moment of mutual gaze⁶

6. We have obtained written informed consent from all study participants to publish transcripts and stills from the recordings. Given this informed consent, approval by the ethics committee of the University of Freiburg was not required.

The starting point of our gaze analysis is always the articulatory onset of *hm_hm*. For each instance of *hm_hm*, we coded whether or not the speaker and recipient were looking at each other at the onset of the feedback token. When mutual gaze was present, we measured the time elapsed between the speaker's gaze shift to the recipient and the onset of *hm_hm*. This coding is based on the assumption that it is in fact the speaker's gaze shift to the recipient which mobilises feedback rather than the establishment of mutual gaze *sensu stricto*.

In addition to that, the relation to the speaker's talk was taken into account by coding whether or not the gaze shift happened during the Intonation Phrase to which the recipient token responds or in the IP before that. This is based on the observation, discussed in more detail in Section 4, that the intonation phrase boundary is a more relevant unit for the placement of feedback tokens than the TCU, as feedback is regularly given at IP boundaries where no syntactic-pragmatic completion point, and thus no TCU boundary has been reached (Heldner et al. 2013, *supra*). This is illustrated in the following example, where the speaker Zac expresses a compound TCU consisting of an *if* and a *then*-clause (cf. Lerner 1996:240). Max utters a feedback token after the *if*-clause (line 01) in line 03, although this projects a continuation by a *then*-clause, which only starts after the feedback token at the end of line 02.

Example 1. Meat production

01 Zac:   
ja:: wenn man einmal geSEHN hat,
yes:: if you have seen once

02   
wie fleiSch produziert wird [in DEUTSCH] land so dann;
how meat is produced in Germany, so then

03 Max: [hm_hm,]
mm hm

04 Zac:   
WEIß man_s ja irgendwie so-
you somehow know it so

4. Results: The relationship between gaze and *hm_hm*

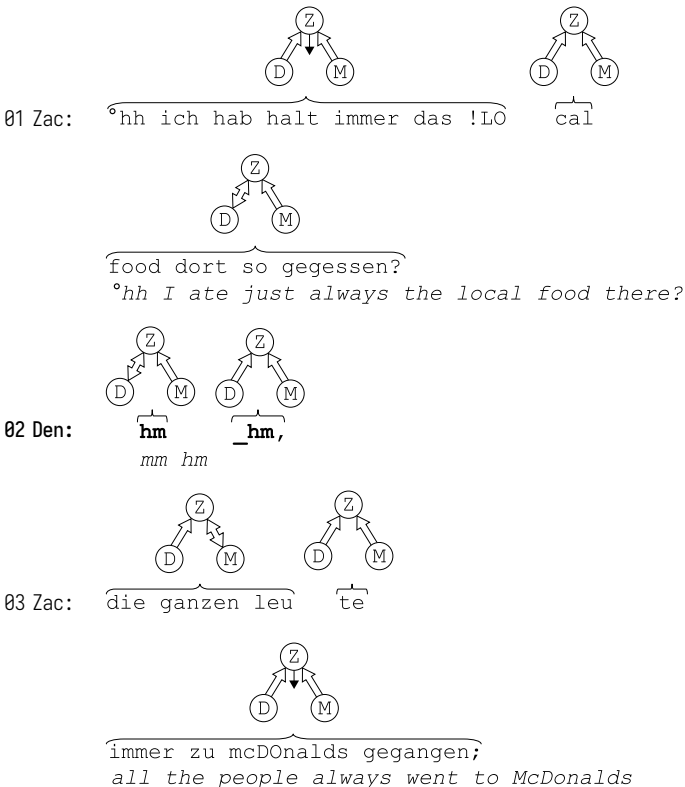
4.1 Description of attested patterns

We identified five different gaze patterns in our data set. They each describe a different temporal organisation and interrelatedness of the speaker's and recipient's gaze behaviour, the speaker's talk, and the utterance of *hm_hm*. Before going into the qualitative and quantitative details of these patterns, we briefly describe the patterns and give a short example for each of them.

Pattern 1 (mutual gaze established in current IP): The onset of the feedback token is embedded in a mutual gaze phase between the speaker and the feedback-giving recipient. The mutual gaze phase starts within the current IP.

This pattern is instantiated in Example 2, where the mutual gaze between the speaker and the recipient begins in line 01 with the word *food* and ends shortly after the first syllable of the recipient token with the gaze aversion of the speaker (line 02).

Example 2. Local food

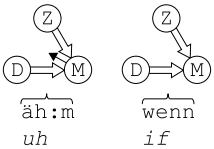


Pattern 2 (mutual gaze established in previous IP): The onset of the feedback token is embedded in a phase of mutual gaze between the speaker and the recipient giving feedback, but mutual gaze is established before the onset of the Intonation Phrase that the recipient token semantically and pragmatically relates to (which means that at least one opportunity to give feedback has passed).⁷

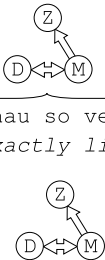
Example 3 is an instantiation of this pattern: here, the mutual gaze starts early in line 01 and continues until the beginning of the recipient token (line 03).

Example 3. 1,8 Earths

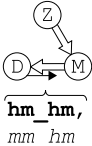
01 Max:



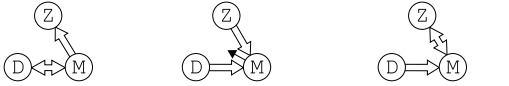
02



03 Den:



04 Max:

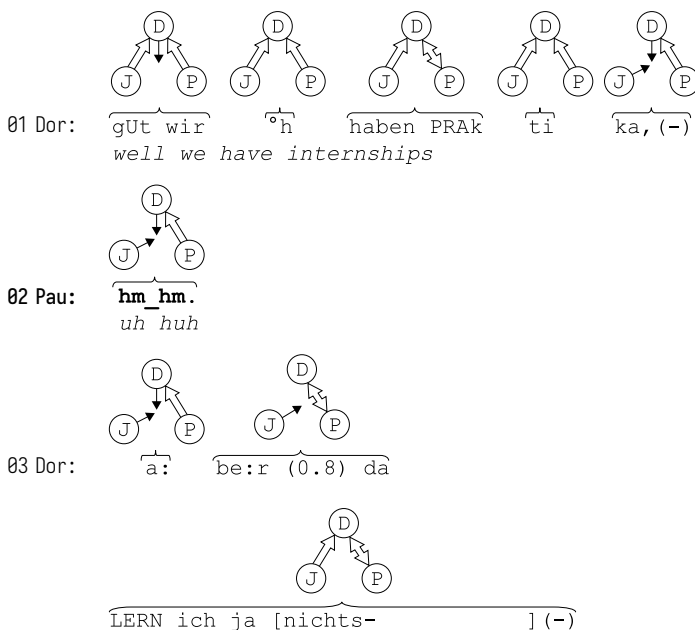


7. In these cases, the recipient does not react immediately to the mobilisation by the speaker's gaze. Instead, the maintained mutual gaze across IP boundaries seems to increase the pressure on the recipient to provide a recipient token.

Pattern 3 (dissolved mutual gaze): Mutual gaze is established within the current IP, but it is dissolved by one of the participants before the onset of *hm_hm*.

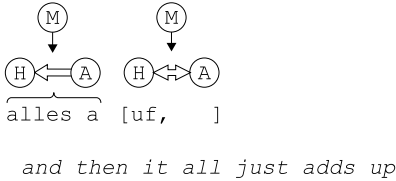
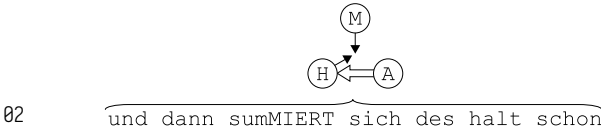
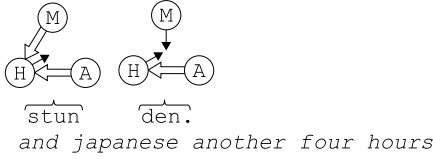
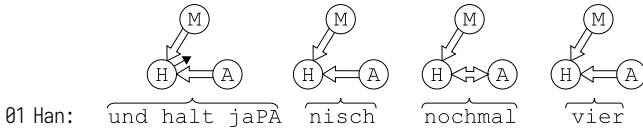
Example 4 illustrates this pattern: the short phase of mutual gaze between Doris and Paul ends before the end of the intonation phrase (line 01). Paul utters a recipient token in line 02, although the speaker is no longer looking at him.

Example 4. Internships

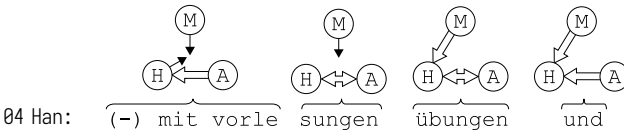


Pattern 4 (gaze shift only): The speaker initiates a gaze shift towards the recipient, but at the onset of the recipient token, mutual gaze has not yet been established. It is established during or shortly after the recipient token. This pattern is instantiated in Example 5. The speaker Hannah shifts her gaze towards Annika shortly before the end of the IP. The mutual gaze phase between Hannah and the recipient Annika is only established simultaneously with the onset of the recipient token in line 03. Arguably, Annika thus does not react to the establishment of mutual gaze but – at least potentially – to the gaze shift towards her.

Example 5. Japanese



03 Ani: [hm_hm,]
mm hh



Pattern 5 (no gaze): The Speaker and the recipient who utters *hm_hm* have not been engaged in mutual gaze prior to the feedback token.⁸

8. Unfortunately, there is not enough space here for a detailed analysis of these cases. These recipient tokens are not a reaction to the speaker's gaze mobilisation, but seem to attract the speaker's gaze to themselves. In more than half of these cases, the recipient becomes the next target of gaze addressing. This suggests that recipient tokens in this position have different functions in conversation than those placed within or after a mutual gaze phase.

Example 6 illustrates the pattern: Katrin, who utters *hm_hm* in line 02, is not looked at by the speaker, as Holli looks into the space between the two recipients throughout the entire storytelling sequence (lines 01–04).

Example 6. Children

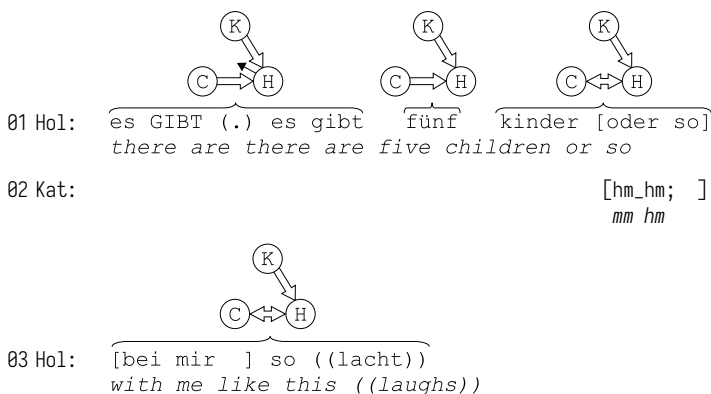


Figure 3 is a schematic representation of the different patterns (1 to 4), illustrating the temporal order of gaze shifts, the establishment of mutual gaze, and the production of *hm_hm* in relation to the speaker's IP.⁹ The arrow represents the gaze shift; the dashed lines enclosed between two vertical lines represent mutual gaze between speaker and recipient. The current IP defines the FRS that follows it, i.e. the space in which the recipient token that responds to it can be placed.

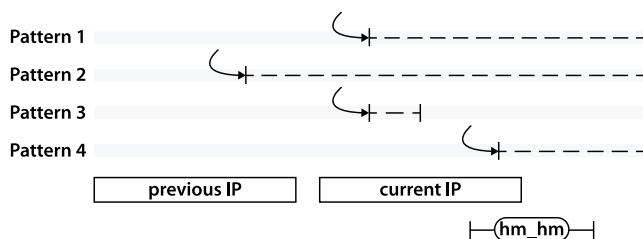


Figure 3. Schematic representation of the different gaze patterns resulting from the relationship between gaze shifts, the establishment of mutual gaze, the utterance of *hm_hm* and the speaker's turn production

9. As there was no eye contact in this pattern, pattern 5 is not included in the figure.

4.2 Quantitative distribution of gaze patterns

Table 1 gives a quantitative overview of the gaze patterns found in our data set of 536 instances of *hm_hm*. It shows the important role that gaze seems to play for the recipient token *hm_hm*: In more than half of the cases, the recipient token is uttered during mutual gaze (frequencies of patterns 1 and 2 added up). However, this result does not confirm the gaze-window hypothesis either (cf. Zima 2020), as only pattern 1 contains clear instantiations of the gaze-window pattern. Pattern 2 refers to the constellation where one FRS and thus an opportunity to utter *hm_hm* within a gaze window has not been used by the recipient. Thus, the 90 cases that instantiate pattern 2 include an instance of the gaze window pattern, but also an FRS in which the recipients do not behave as predicted by Bavelas et al. (2002). Most notably, about half of the *hm_hms* in our dataset are not produced during mutual gaze at all.

Table 1. Overview of gaze patterns

Gaze pattern	Total (<i>n</i> = 536)
Pattern 1: Mutual gaze, current IP	210 (39.2%)
Pattern 2: Mutual gaze, previous IP	90 (16.8%)
Pattern 3: Dissolved mutual gaze	104 (19.4%)
Pattern 4: Gaze shift only	28 (5.2%)
Pattern 5: No gaze	104 (19.4%)

To test the hypothesis that gaze can be used to elicit recipient feedback – even if it is not produced during mutual gaze (Zima 2020) – we looked more closely at patterns 3 to 5. Our aim was to reveal in how many cases the recipient token is delivered by the last looked-at recipient rather than by the third, not looked-at recipient. The results are shown in Table 2.

Table 2. Overview of recipient tokens with no mutual gaze at onset of the RT

Gaze pattern	<i>hm_hm</i> uttered by last-looked-at recipient	<i>hm_hm</i> uttered by recipient not last- looked-at	Total (<i>n</i> = 236)
Pattern 3: Dissolved mutual gaze	62	42	104
Pattern 4: Gaze shift only	6	22	28
Pattern 5: No mutual gaze	27	77	104

Table 2 provides us with additional information on the relationship between gaze and *hm_hm*. For example, as in pattern 3, in 62 cases (59.6%), the recipient giving feedback is the last looked-at recipient. If we combine the frequencies of patterns 1 and 2 and add the cases in which it is the last looked-at recipient who utters the recipient token from Table 2, we can infer the importance that gaze seems to have for mobilising *hm_hms*: 73.6% of all cases are produced by the last or currently looked-at recipient. In patterns 4 and 5, the picture is, however, inverse with the majority of *hm_hms* being produced by the not last-looked at recipient. This is not surprising in pattern 4, where the speaker initially looks away and already turns his gaze to the recipient, but does not yet look at them at the onset of the recipient token. Space does not allow us to discuss all five patterns and especially the deviant pattern 5 in detail (but see Masuch, in preparation). In the following, our focus is on the 73.6% of *hm_hms* that are clearly gaze-related. In the remainder of the chapter, we will show that while gaze seems to play an important role in mobilising feedback, it does not seem to be equally important for the temporal placement of the recipient token: Rather than trying to say *hm_hm* while being looked at by the speaker, recipients (in most cases, see Discussion) orient themselves to the Feedback Relevance Space around the end of an intonation phrase in order to actually utter *hm_hm*.

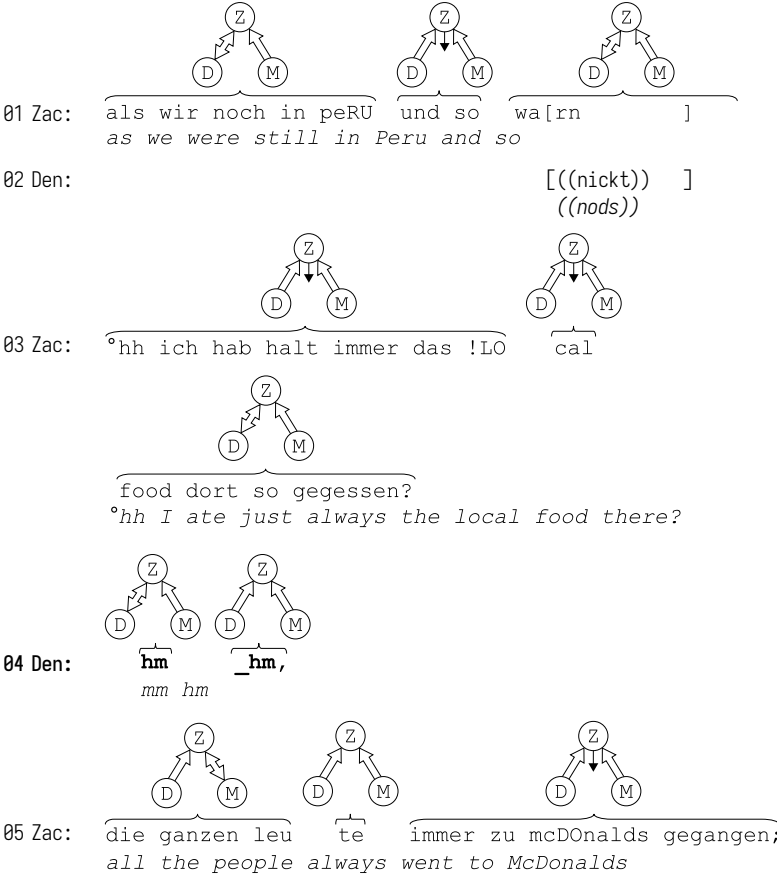
4.3 Analysis of the temporal placement of gaze-mobilised *hm_hms*

4.3.1 *Pattern 1: Timely placement of hm_hm as a reaction to the gaze mobilisation*

The first Example (7) to be discussed here instantiates the gaze-window pattern. The recipient reacts to the gaze mobilisation by immediately responding with the recipient token *hm_hm*, which is embedded in a mutual gaze phase. We argue that this is mainly due to the fact that mutual gaze is established shortly before the next FRS is reached, so that the production of *hm_hm* at the Feedback Relevant Space coincides with mutual gaze.

Dennis, Zac, and Max are sharing their experiences of travelling to South American countries and talking about local food. Earlier, Zac said that the Bolivian government had banned McDonald's because the company did not use local potatoes.

Example 7. Local food



Zac says that during his stay in Peru (line 01), he ate only *local food* (line 02), while the locals *always went to mcdonalds* (line 04). At the end of line 01, Zac begins a period of mutual gaze, which Dennis responds to by nodding. After the nod, the speaker averts his gaze and begins a new turn. An instance of the gaze-window pattern can be seen in lines 02–03. Zac begins his turn (topicalising the food he has eaten) with averted gaze, which according to Kendon (1967) is typical of the beginning of a turn, and then turns his gaze back to Dennis while uttering the second syllable of the word !LO!cal. This leads to a period of mutual gaze, during which Dennis, the looked-at recipient, utters the recipient token hm_hm, (line 03). The speaker then averts his gaze after the first syllable of the recipient token. Note, however, that this time, the mutual gaze is maintained for 1170 milliseconds before the recipient starts his recipient token. The speaker then continues to speak, wondering about the people going to mcdonalds (line 04).

Although there are several cases in which recipients respond with a recipient token after a short phase of mutual gaze with the speaker, the next examples will show that even if a gaze shift is understood by recipients as a request for feedback, recipients typically wait until the Feedback Relevance Space is reached to act upon this request. Note that these cases still instantiate Pattern 1, given that the *hm_hm* is realised during mutual gaze.

4.3.2 Pattern 1: More time passes between the beginning of mutual gaze and the utterances of *hm_hm*

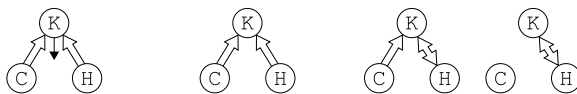
Excerpt 8 supports our argument that the timing of the response depends on whether or not a Feedback Relevance Space is reached. In this conversation, Holli talks about the complexity of learning Arabic. Quite some time elapses between the onset of mutual gaze phase between Katrin and Holli and the delivery of the recipient token at the first possible FRS (therefore the example is assigned to pattern 1 (see Table 1)).

Example 8. Tutoring student

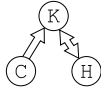
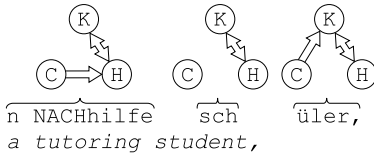
- 01 Hol: es funktioniert einfach so ANders,=
it works simply so differently
- 02 dass es viel SCHWIERiger is erstmal so:: überhaupt es auch
that it is more difficult firstly so:: to be able
DENken zu können,
to think like that at all
- 03 dass de dich so AUSdrückst;
that you express yourself like that
- 04 Kat: ja.
yes.
- 05 Hol: [also-]
[thus-]
- 06 Cla: [<pp>krass<]
[<p<rad>]



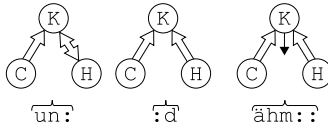
- 07 Kat: { ich glaub } { die ZÄHln auch irgnwie (.) } { ANders; }
i think they also count differently;



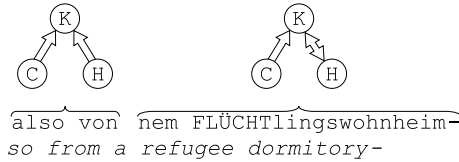
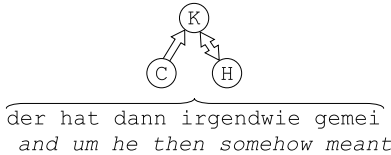
- 08 { oder_also weil } { ich hatte des } { ma ich ha } { tte mal }
or_also because I had that once I had once



09 Holli: hm hm?
hm hm?



10 Kat: un: :d ähm::



- 11 und äh: der hat dann AUCH gesagt dass; (---)
and uh: he than also said that; (---)
- 12 für ihn es irgendwie Anders is ähm:: (-) denn z
for him it is somehow different um:: (-) because z
- 13 also so: EINS, ZWEI und VIEle; oder so.
so like this: ONE, TWO and MANY; or like this.

In lines 01–03, Holli explains her difficulties in learning Arabic by saying that the structure of the language is so different that it is not easy to think like that at all (line 02) and to get used to it. With also (05) she announces a continuation of her talk, but Katrin takes over in line 06, naming another feature in which the Arabic language differs from the German language: *they also count differently*; (line 07). She marks her low epistemic status (especially compared to Holli, who is learning the language) with uncertainty markers such as *I believe* (line 07) and *somehow* (line 07 and 10) (cf. Heritage 2012; Delettres and Jallerat-Jabs 2018) and bases her statement on the testimony of an Arab tutoring student (lines 08–11): *He said that* (line 11) *it is somehow different* (line 12) to count. Finally, Katrin gives an example of a number line in counting: *so like this: ONE, TWO and MANY*;

or like this. (line 13). The relevant gaze pattern occurs in line 08, where Katrin establishes a mutual gaze phase with Holli after the words *I had that*. This mutual gaze is maintained until the beginning of the next intonation phrase in line 10. A total of 1974 milliseconds pass between the beginning of Katrin's gaze shift to Holli and the onset of the recipient token *hm_hm?* (line 09), which Holli places immediately after the end of Katrin's Intonation Phrase. Holli therefore does not react immediately to the gaze (as is predicted by the gaze window hypothesis). Instead, Holli waits until the end of the intonation phrase to utter *hm_hm*. This timing of the recipient token fits well with its function: On the one hand, it does not disturb Katrin's emerging utterance by overlapping with it. On the other hand, it signals understanding of the short explanation (lines 07–09) provided by Katrin, which is a prerequisite for her to be able to continue her turn.

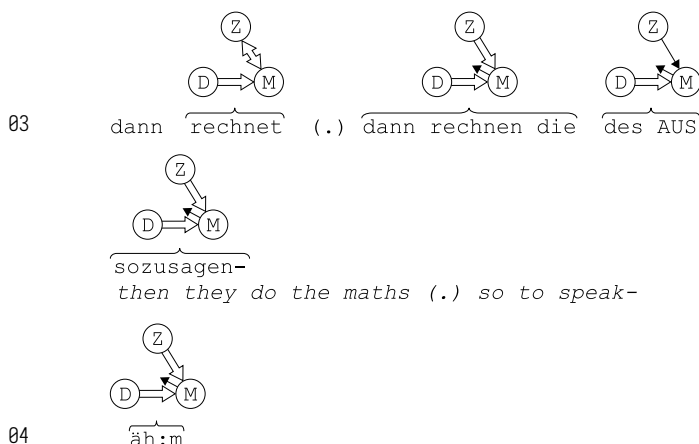
4.3.3 Pattern 2: The recipient utters *hm_hm* at the second FRS

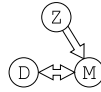
Example 9 is an exception to our previous argument. We will show that there are interactional reasons why a recipient sometimes does not immediately utter a feedback token at the first FRS but waits until the next FRS is reached. The excerpt is part of a conversation about an online test that is designed to measure the personal ecological footprint. Max, who tried out this test, received a result of “1.8 Earths”. After the participants first clarify which criteria are included in this test, Dennis asks what this “1.8 Earths” are all about.

Example 9. 1,8 earths

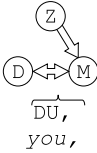
01 Den: und (.) was hat des dann mit den eins komma acht Erden auf
and (.) *what's the deal with the one point eight earths?*
sich,

02 Max: also und DANN genau;
so and *THEN* exactly

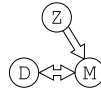




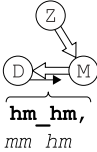
wenn jEder so (--) sich genau so verhalten würde wie
uh:m if everyone (--) behaved exactly like



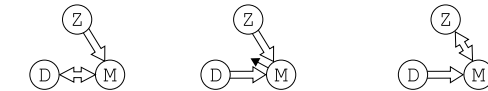
05



dann bräuchten wir EINS komma acht mal die erde? (-)
then we would need one point eight times the Earth?



06 Den:



07 Max:

damit's Ganze: ähm:: (-) wie sagt man?
so that the whole um:: (-) how do you say?

08 Zac:

((schnalzt)) Ökosystem?
((clicks)) ecosystem?

09 Max:

ja also dass es ähm: (-) dass es sich nicht AUFbraucht.
yeah so that it um: (-) that it doesn't use itself up.

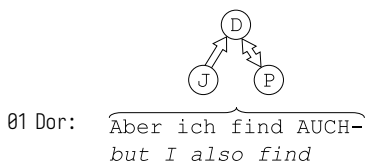
In response to Dennis' question in line 01, Max explains that to compute one's ecological footprint, different criteria are added up (line 03). The resulting value is how many earths we would need if everyone (-) behaved exactly like the person being tested (line 04). Accordingly, if everyone lived like Max, we would need *ONE point eight times the earth* (line 05) in order not to use up the Earth's natural resources (line 09). The speaker's word search in line 07, which he also makes explicit with the question *how do you say?* is answered by Zac who suggests *ecosystem?* (line 08) as a potentially fitting word. However, Max does not take up this suggestion, but begins with a paraphrase of the expression he is looking for, *that it does not use itself up* (line 09). Max utters the hesitation marker *font* (line 04) with averted gaze and then, after the conjunction, initiates mutual gaze with Den-

nis. This eye contact is not only maintained throughout the protasis, i.e. the if-clause, but also throughout the apodosis (more than nine seconds in total), which is introduced with the resumptive element *then* (Z. 05). The fact that Max is primarily addressing Dennis here is sequentially related to the previous question (line 01). He monitors whether Dennis understands the explanation he provides by looking at him. Only at the end of the Compound TCU, consisting of the protasis and the apodosis, does Dennis deliver the reception signal *hm_hm*. We argue that this ‘delay’ of the recipient token, which is only uttered at the second FRS, is primarily due to the content of the speaker’s response. The protasis in line 04 not only projects on a syntactic level that an apodosis will follow, but is above all an answer to the question of what the 1.8 Earths are all about. It is only in line 05 that the reference to the question posed at the beginning is established. Only at this point, a display of understanding is meaningful and relevant. This is mirrored in Max behaviour who leaves a short pause for Dennis to display understanding by delivering a recipient token. If the explanation at this point had not been sufficient for Dennis, he could have used this opportunity to initiate a repair.¹⁰

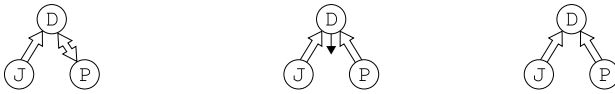
4.3.3 Pattern 3: The speaker dissolves mutual gaze, and the recipient waits until the FRS to utter *hm_hm*

Pattern 3 is similar to Example 8, which instantiates pattern 1, in that the recipient does not react to mutual gaze by uttering *hm_hm* immediately, but waits until the next FRS is reached. The difference is that at this FRS, the speaker no longer looks at the recipient. This is exemplified in Example 10: The recipient (Paul) waits until the end of the intonation phrase (and a short pause) is reached to utter *hm_hm*. At this point, the speaker (Doris) does not look at him anymore, but gazes away. The participants are talking about their experiences with their study subjects. They agree that their study programmes should be much more practical in order to prepare students adequately for work.

Example 10. Internships

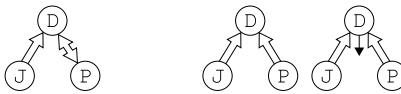


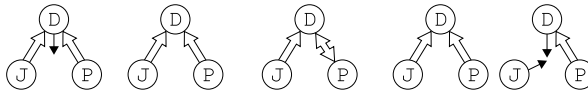
10. Here, the FRS overlaps with the TRP, but this is not always the case.

02 
 äh: so: (0.3) (0.8) ähm (0.9) dass wir nUr so
 uh so um that we only learn so


 theoREtisches [zEug lernen,]
 theoretical stuff

03 Pau: [((nickt))]
 ((nods))

04 Dor: 
 und überhAupt nichts PRAKti sches;
 and nothing at all practical

05 
 gUt wir °h haben PRAK ti ka, (-)
 well we have internships

06 Pau: 
 hm hm.
 uh huh

07 Dor: 
 a: be:r (0.8) da LERN ich ja [nichts-] (-)
 but there i learn nothing

08 Pau: [((Schulterzucken))]
 ((shrugs))

09 ja.
 yes

In this sequence, Doris complains that there is too little practical content taught in law school: she would only learn *theoretical stuff* (line 02) and *nothing at all practical* (line 04). The fact that Doris mainly gaze-addresses Paul (lines 01–07) while expressing her opinion can be explained by the distribution of epistemic status. Her opinion refers to the general course of studies, which Paul knows and

can judge as well, whereas Julian has no knowledge of what is taught at the Law Faculty. After Doris has evaluated the learning material in her study programme (lines 01–04), she admits that there are at least internships (line 05). At the beginning of the IP, Doris still looks away (cf. *well we*), but then she takes a breath, turns her gaze to Paul and establishes mutual gaze with him. However, she averts her gaze again for the last two syllables of the IP, so that at the end of the IP, Paul is no longer looked at. Nevertheless, he delivers the recipient token shortly after the end of the IP (line 06).

From the speaker's point of view, gaze aversion could be a sign of an ongoing planning activity that the speaker is engaged in (as argued in Zima 2020). A number of studies confirm this relationship between gaze aversion and planning activity (e.g., Kendon 1967), some arguing for the regulatory function of this gaze aversion, others attributing it to the cognitive load of the planning activity (e.g., Beattie 1981). However, while this speaker-centred explanation may be valid for speakers, it does not account for the motivation of recipients to wait for the next FRS instead of uttering *hm_hm* when being in mutual gaze with the speaker. We argue that the reasons lie in the need of gaze-addressed recipients to place feedback signals in a timely manner and to wait until a Feedback Relevance Space is reached. Apparently, recipients prioritise this timely placement over providing feedback while being looked at.

After this short overview of gaze patterns in which gaze may reasonably be claimed to be used as a cue to mobilise response, we will now turn to the question of where the Feedback Relevance Space is located in relation to the speaker's talk.

5. The placement of *hm_hm* in relation to gaze and the Feedback Relevance Space

In Section 2.1, we introduced the concept of the Feedback Relevance Space. In order to determine exactly where this space is located, and, more precisely, to define its temporal extent, this section will zoom in on the quantitative distribution of recipient tokens with respect to the end of the speaker's intonation phrases.

Before doing so, we will briefly present qualitative arguments for why the FRS does not simply correspond to the place where the IP ends, but constitutes a space around the end of the IP. Consider Example 11, where the FRS starts before the end of an IP. It is another example of pattern 3 (see Table 1), in which the speaker ends a phase of mutual gaze before the end of the IP. The recipient (Lina) waits until the end of the intonation phrase to place the recipient token *hm_hm*. At this point, the speaker does not look at her anymore, but instead gazes at the

other recipient, Tobias. In this extract, the interactants compare different streaming platforms such as Amazon Prime and Netflix.

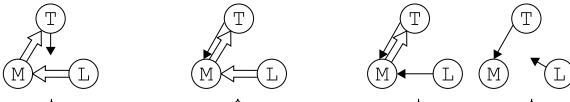
Example 11. Stream

01 Mar: °hh und dann LOHNT sich_s eigentlich auch fast, (1.9)
and then it is actually almost worth it

02 (-) des zu HOLN,
 (-) *to get it*

03 (-) wenn du VIEL schaust.
 (-) *if you watch a lot.*

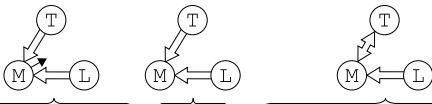
04 (-) weil ich mein es gibt ja auch die LEUde==
 (-) *because there also are the people*

05 
 =die streamen halt (--) nUr serien
they stream (--) just series


 über irgendwelche: (--)
on any (--)


 KI nox oder mo[vie:] (.) sachen?
kinox or movie things?

06 Lin: [hm_hm;]
mm hm;

07 Mar: 
 (--) und des is ja immer so die FRAge;
 (--) *and it it always the question*

08 (-) inWIEweit des halt immer-
 (-) *to what extent it always-*

09 (--) so WEIßt du halt zumindest==
 (--) *like that you know at least*

10 es isch irgendwie in_ner geWISsen weise;
it is somehow in a certain way

11 oder es is einfach leGAL was du halt machst,
or it is simply legal what you are doing,

Marcel argues in lines 01–03 that a Netflix subscription is worth the price (cf. line 01 LOHNT sich_s) for people who frequently watch shows (cf. line 03 *if you watch a lot*). The relevant sequence for the recipient token starts in line 05. Marcel starts to talk about people who only stream series illegally via websites like *kinox* or *movie things*, starting the first words of his relative clause *they stream just* while looking at Tobias. He then begins to look at Lina, establishing mutual gaze between him and her during the preposition *on* the indefinite determiner *any* and a short hesitation phase (cf. (–) line 05). At this moment, the turn has not yet reached a possible point of completion, as a projected nominal phrase is still missing. With the resolution of the hesitation phase, Marcel shifts his gaze back to Tobias and finishes his turn with the nominal phrase *kinox* or *movie things* (line 05). Overlapping with the last syllable of the word *movie*, at the now reached Feedback Relevance Space (= FRS), Lina utters the recipient token *hm_hm* (line 06), although the speaker no longer looks at her. Zima (2020:17 f.) explains this pattern by the fact that the speaker addresses both recipients during the turn but can only look at one recipient at the end of the turn. The last-looked-at recipient then chooses verbal feedback over visual signals because “articulated feedback reaches the narrator also without eye contact” (Zima 2020:18). This example shows not only that recipients wait until the FRS is reached, but also that this FRS begins before the actual completion of the IP. This point of recognition has been described by Jefferson (1984: 26) with the term *pre-completor onset*: “the talk can be seen to be ‘all over but for the last word(s)’. At such a point the ‘thing’ has been adequately said [and] the ‘place’ has been adequately arrived at.” In this case, only the projection of the nominal *things* remains open. This interpretation of the FRS as a space, beginning before syntactic completion, rather than being a place is supported by a quantitative analysis of the placement of *hm_hm* in relation to speaker’s gaze shifts and the Feedback Relevance Space, which is plotted in Figure 4. The end of the intonation phrase corresponds to the zero point on the x-axis. We measured the time between the onset of the recipient token and the end of the intonation phrase in milliseconds. If a recipient token starts before the end of the intonation phrase, it is part of the minus range (to the left of zero). If it starts after it, it is in the columns to the right of the zero point.

The analysis shows that for the majority of *hm_hms* in our data set, the articulatory onset falls within the time frame of 600 ms before and 600 ms after the end of an intonation phrase (= 455 instances, representing 85% of all *hm_hms*).¹¹ Our

11. However, we have to keep in mind that the picture is slightly distorted by the fact that on the one hand, *hm_hm* in our data have an average length of 440 milliseconds and on the other hand, there is sometimes a short pause between the end of the IP and the beginning of a recipient token.

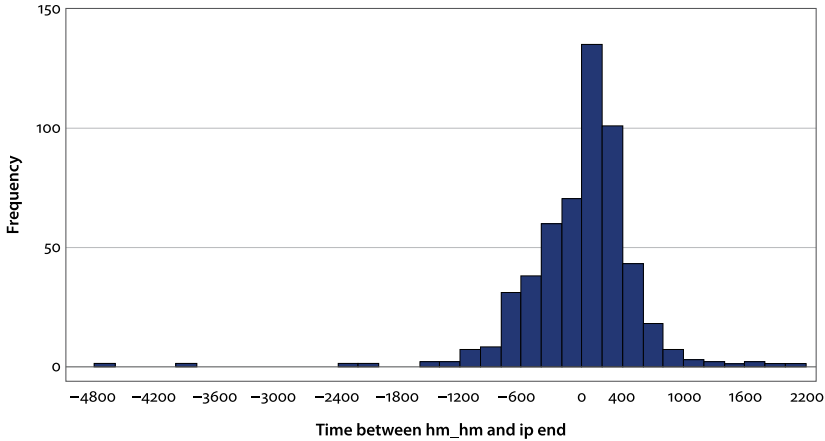


Figure 4. Distribution of *hm_hm* in relation to the speaker's IP end

analysis confirms that this slot, i.e., the Feedback Relevance Space corresponds to the time span within which feedback is acceptable or preferred and is indeed better conceptualised as a space rather than an exact place.

To prove our main claim that recipients interpret a gaze shift towards them within the IP as an attempt to mobilise feedback, but orient towards the IP boundary to give this feedback, we tested our data for a correlation between the temporal distance between the IP boundary and the utterance of *hm_hm*, and the onset of a speaker's gaze shift towards the recipient. For this calculation (see Figure 5), we excluded all *hm_hms* that are not directly related to the speaker's gaze and therefore cannot be seen as an immediate response to it (= patterns 2 and 5 as well as some cases from pattern 3, see Table 2). This includes all cases in which a mutual gaze is established in the intonation phrase preceding the one to which the recipient token reacts. This leaves us with 293 cases.

We hypothesised that the temporal distance between the IP boundary and the onset of the *hm_hm* depends on how much time elapses between the speaker's gaze shift towards the recipient and the end of the intonation phrase: If the gaze shift occurs shortly before the end of the IP, the time between the gaze shift and *hm_hm* will be shorter than in cases where the gaze shift occurs early in the IP because the recipient has to wait (longer) until the FRS is reached. To test this hypothesis, we determined the Spearman's rank correlation between the variables "time between gaze shift and *hm_hm*" and "time between gaze shift and IP end", as it is also suitable for non-normally distributed data and is less sensitive to outliers. The test showed that the two variables correlate with each other ($r=0.661$, $p=2.2e-16$, $N=293$). The correlation coefficient (0.66) indicates a significant positive correlation, that is, a higher value of the variable "time between gaze shift and IP end" is associated with a higher value of the variable "time between gaze shift

and *hm_hm*". The significance test yielded a p-value of $2.2e-16$, indicating high significance.

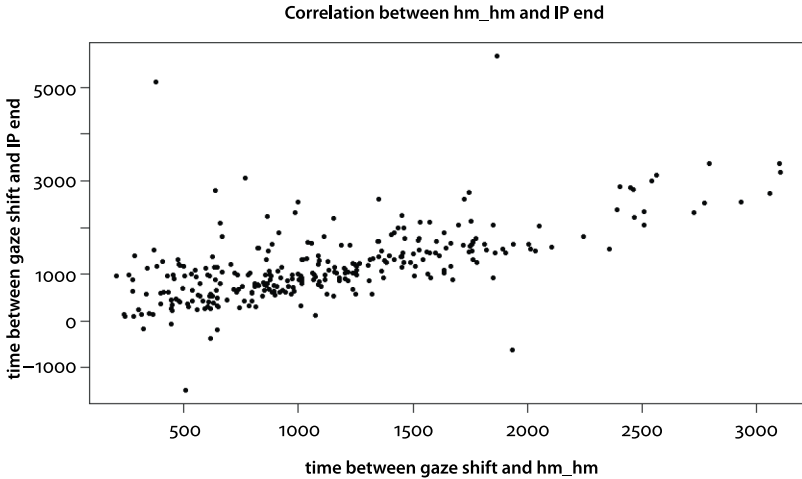


Figure 5. Scatterplot showing the linear relationship between the IP end and the placement of the *hm_hm*

The scatterplot in Figure 5 visualises the correlation between the IP end and the placement of the recipient token. It can thus be concluded that the time interval between the gaze shift and the *hm_hm* is greater if the time interval between the gaze shift and the IP end is also greater. The hypothesis that recipients orient to the FRS when they place a recipient token is thus confirmed, and this seems to be more important than the question of whether or not they are still engaged in mutual gaze with the speaker.

Although the previous analysis provides information about where the FRS is to be located, the quantitative evaluation does not explain why recipients choose a particular placement, i.e., whether they start giving feedback early at the recognition point or later after the start of the next IP.¹² Although we have proposed the recognition point as a possible starting point of the left boundary of the FRS, it is difficult to determine it precisely. However, its variability could simply be related to the fact that recipients cannot predict when exactly the speaker's IP will end, and therefore they orient to an approximate end point. Furthermore, the possibility remains that in some cases the recipients are simply "late" with their feedback.

12. A chi-square test was performed to test the hypothesis of an interaction between the placement of the feedback token (turn-initial, within, turn-final, in a pause at the TRP) and the patterns described in Table 1. The result was not significant, $\chi^2(18, N=536) = 31.81, p = .023$. Further research is needed to clarify the question of why recipients choose a specific placement.

6. Conclusions

In this chapter, we investigated the relationship between speaker gaze, the German recipient token *hm_hm*, and the Feedback Relevance Space in triadic interactions. Our data confirm the importance of gaze for mobilising the recipient token *hm_hm*, with more than three quarters of the *hm_hms* in our data set being produced by the last looked-at recipient. However, the gaze-window pattern postulated by Bavelas et al. (2002) can only explain a minority of our cases, while coexisting gaze patterns show different temporal interrelationships between mutual gaze and the placement of the recipient token. In a next step, we showed that the Feedback Relevance Space (FRS) plays an important role for when a recipient token is delivered. Recipients mostly do not react immediately to the gaze shift of the speaker (and the establishment of mutual gaze), but rather orient to the Feedback Relevance Space to produce a recipient token. Based on the distribution in our data set, we argued that the FRS should be seen as a space located around the end of the speaker's intonation phrase. This placement between one unit and the next is well in line with the function of a continuer. Temporally "late" recipient tokens or a placement after a phase of mutual gaze has already ended can be explained by the fact that recipients wait until the FRS has been reached before giving feedback.

References

- Argyle, Michael und Mark Cook (1976): *Gaze and mutual gaze*. Cambridge [etc.]: Cambridge University Press.
- Auer, Peter and Elisabeth Zima. 2021. "On word searches, gaze, and coparticipation." *Gesprächsforschung — Online-Zeitschrift zur verbalen Interaktion* 22: 390–425.
-  Bavelas, Janet, Coates, Linda and Trudy Johnson. 2000. "Listeners as co-narrators." *Journal of Personality and Social Psychology* 79 (6): 941–952.
-  Bavelas, Janet, Coates, Linda and Trudy Johnson. 2002. "Listener responses as a collaborative process: The role of gaze." *Journal of Communication*, 52 (3): 566–580.
-  Beattie, Geoffrey W. 1981. "A further investigation of the cognitive interference hypothesis of gaze patterns during conversation." *British Journal of Social Psychology* 20 (4): 243–248.
- Cathart, Nicola, Carletta, Jean and Ewan Klein. 2003. "A Shallow Model of Backchannel Continuers in Spoken Dialogue." *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*.
-  Clark, Herbert and Meredyth A. Krych. 2004. "Speaking while monitoring addressees for understanding." *Journal of Memory and Language* 50 (1): 62–81.
-  De Kok, Iwan and Dirk Heylen. 2012. "Analyzing nonverbal listener responses using parallel recordings of multiple listeners." *Cognitive Processes*, 13 (2): 499–506.

-  Delettres, Cécile and Jallerat-Jabs, Britta. 2018. ““Ja Müssen Sie jetzt nix irgendwie aufschreiben?” – Eine Empirische Untersuchung zur Verwendung von irgendwie im gesprochenen Deutsch”. *Diskursive Verfestigungen: Schnittstellen zwischen Morphosyntax, Phraseologie und Pragmatik im Deutschen und im Sprachvergleich*, edited by Laurent Gautier, Pierre-Yves Modicom and Hélène Vinckel-Roisin, Berlin, Boston: De Gruyter: 229–242.
-  Dittmann, Allen T. and Lynn G. Llewellyn. 1967. “The phonemic clause as a unit of speech decoding.” *Journal of Personality and Social Psychology* 6 (3): 341–349.
- Duncan, Starkey and Donald W. Fiske. 1977. *Face-to-face interaction: Research, methods, and theory*. Lawrence Erlbaum Associates.
- Ehlich 2007. „Formen und Funktionen von hm: Eine phonologisch-pragmatische Analyse.“ *Sprache und sprachliches Handeln, Bd.2: Prozeduren des sprachlichen Handelns*. Berlin, New York: de Gruyter. 273–286.
-  Gardner, Rod. 2001. *When Listeners Talk: Response Tokens and Listener Stance*. Amsterdam: John Benjamins.
- Gardner, Rod. 2013. “Conversation analysis and recipient behaviour.” *The Encyclopedia of Applied Linguistics*, ed. by Carol A. Chapelle: 1086–1094. Oxford, U.K.: Wiley-Blackwell.
- Goodwin, Charles. 1981. *Conversational organization: Interaction between speakers and hearers*. Academic Press.
-  Goodwin, Charles. 1986. “Between and within: Alternative sequential treatments of continuers and assessments” *Human Studies* 9 (2–3): 205–217.
- Goodwin, Marjorie Harness and Charles Goodwin. 1986. “Gesture and coparticipation in the activity of searching for a word.” *Semiotica* 62 (1–2): 51–75.
- Heldner, Mattias, Hjalmarsson, Anna and Jens Edlund. 2013. “Backchannel relevance spaces.” In *Nordic Prosody: Proceedings of the XIth Conference Tartu 2012* ed. by E. L. Asu and P. Lippus.: 137–146. Frankfurt am Main: Peter Lang.
-  Heritage, John. 2012. Epistemics in Action: Action Formation and Territories of Knowledge, *Research on Language & Social Interaction*, 45:1, 1–29.
- Howes, Christine and Arash Eshghi. 2017. “Feedback relevance spaces: The organisation of increments in conversation. *Proceedings of the 12th International Conference on Computational Semantics*.
-  Howes, Christine and Arash Eshghi. 2021. “Feedback Relevance Spaces: Interactional Constraints on Processing Contexts in Dynamic Syntax.” *Journal of Logic, Language and Information* 30: 331–362.
- Jefferson, Gail. 1984. “Notes on some orderliness of overlap onset.” *Discourse analysis and natural rhetoric*. ed. by V. d’Urso und P. Leonardi: 11–38. Padua, Italy: Cleup Editore.
-  Jokinen, Kristiina; Furukawa, Hirohisa; Nishida, Masafumi and Yamamoto, Seiichi. 2013. “Gaze and turn-taking behaviour in casual conversational interactions.” *ACM Trans. Interact. Intell. Syst.* 3 (2), S. 1–30.
-  Kendon, Adam. 1967. “Some functions of gaze direction in social interaction.” *Acta Psychologica* 26: 22–63.
-  Koiso, Hanae, Horiuchi, Yasuo and Den Yasuharu. 1998. An Analysis of Turn- Taking and Backchannels Based on Prosodic and Syntactic Features in Japanese Map Task Dialogs. *Language and Speech* 23: 296–321.

- doi Lerner, Gene H. 1996. "On the "semi-permeable" character of grammatical units in conversation: conditional entry into the turn space of another speaker." *Interaction and Grammar* (Elinor Ochs, Emanuel A. Schegloff, Sandra A. Thompson, eds.), Cambridge, Cambridge University Press: 238–276.
- doi Marslen-Wilson, William. 1985. "Speech shadowing and speech comprehension." *Speech Comprehension*, 4 (1–3), 55–57.
- doi Morency, Louis-Philippe, de Kok, Iwan and Jonathan Gratch. 2009. "A probabilistic multimodal approach for predicting listener backchannels" *Autonomous Agents and Multi-Agent Systems*. 20: 70–84.
- Quasthoff, Uta. 1981. „Zuhöreraktivitäten beim konversationellen Erzählen.“ In *Dialogforschung* ed. by Peter Schröder and Hugo Steger: 287–313. Düsseldorf: Schwann.
- Rossano, Federico. 2012a. Gaze behaviour in face-to-face conversation [Doctoral dissertation]. Radboud University Nijmegen.
- doi Rossano, Federico. 2012b. "Gaze in Conversation." *The Handbook of Conversation Analysis* ed. By Sidnell, Jack and Tanya Stivers: 308–329. Oxford, U.K.: Wiley Blackwell.
- Sacks, Harvey, Emanuel A. Schegloff und Gail Jefferson (1974). "A Simplest Systematics for the Organization of Turn-Taking for Conversation". In: *Language* 50.4, S. 696–735.
- Schegloff, Emanuel A. 1982. "Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences." In *Analyzing Discourse: Text and Talk*, ed. by Deborah Tannen: 71–93. Georgetown University Press.
- Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., Couper-Kuhlen, E., Deppermann, A., Gilles, P., Gunthner, S., Hartung, M., Kern, F., Mertzlufft, C., Meyer, C., Morek, M., Oberzaucher, F., Peters, J., Quasthoff, U., Schutte, W., Stukenbrock, A., and Uhmann, S. 2009. „Gesprächsanalytisches Transkriptionssystem 2 (GAT 2).“ *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–402. <http://www.gespraechsforschung-online.de/heft2009/heft2009.html>
- doi Stivers, Tanya and Federico Rossano. 2010. "Mobilising response." *Research on Language and Social Interaction* 43: 3–31.
- doi Tolins, Jackson and Jean E. Tree. 2014. "Addressee backchannels steer narrative development." *Journal of Pragmatics* 70: 152–164.
- Vranjes, Jelena. (2018). "On the role of eye gaze in the coordination of interpreter-mediated interaction: an eye-tracking study". Dissertation. Katholieke Universiteit Leuven.
- doi Ward, Nigel and Wataru Tsukahara. 2000. "Prosodic features which cue backchannel responses in English and Japanese." *Journal of Pragmatics* 32: 1177–1207.
- Weiß. 2020. Blick und Turn-Taking in Face-to-Face-Interaktionen: Multimodale Interaktionsanalysen triadischer Gesprächssituationen mit Hilfe von Eye-Tracking [Doctoral dissertation]. Albert-Ludwigs-Universität Freiburg.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556–1559).
- doi Zima, Elisabeth. 2020. "Gaze and recipient feedback in triadic storytelling activities." *Discourse Processes* 57 (9): 725–748.

Gaze aversion as a marker of disalignment in interactions

Maximilian Krug
University of Duisburg-Essen

Social interaction requires participants to be aligned with each other. Interactional disalignment occurs when actions are inappropriate for a given situation or when actions are not followed up by interlocutors. This study examines gaze aversion as a visual practice in which participants display and maintain interactional impasses that result from disalignment. As the data suggest, the participants redirect their foveal attention to interactional less relevant areas of interest to avoid visually addressing other participants and show self-involvement as a state of unavailability. The data basis is two video recordings (40 min each) of triadic interactions in a laboratory. The participants wear mobile eye-tracking glasses (Tobii Pro Glasses 2). The data are in English.

Keywords: conversation analysis, disalignment, mobile eye tracking, multimodality, progressivity, self-involvement

1. Introduction

Aligning with each other is one of the basic requirements of social interaction (Stivers 2008). Alignment is accompanied by the mutual display of presence, attention, and shared understanding regarding an ongoing activity in terms of the participation framework (Goodwin and Goodwin 2004). Practices for establishing alignment include, but are not limited to, reception signals, such as continuers (Goodwin 1986) or nodding (Stivers 2010). However, these do not necessarily indicate congruence of opinion among participants. Instead, participants in an argument can have contrary opinions and still be aligned regarding the common activity of arguing (Morek 2016). In the interaction analysis research literature (e.g., Stivers, Mondada & Steensig 2011, Steensig 2019), alignment is often discussed together with affiliation, but it is defined as distinct from it. While affiliation can be accompanied by a display of congruent stance-taking and positioning,

alignment is considered to operate on the structural level of joint action. This study focuses primarily on the structural aspects of alignment to maintain progressivity.

Central to alignment is the idea of “situational appropriateness” (Heller 2016: 91), which refers to actions that are deemed acceptable and fitting by the participants in a particular situation. According to Stivers et al. (2011: 20), actions align when the interactants accept and share the preconditions and conditions of the proposed action or activity. In this context, alignment acts as a form of social calibration (cf. Stivers 2008: 36) in which the interactants indicate to each other the extent to which they approve of the current course of action. In this regard, Stokoe et al. (2020: 73) assume an alignment continuum “with responses to preceding actions being more or less productive.” This means that courses of action can also only be marginally aligned without endangering the progressivity of an interaction. However, if participants’ actions disalign too much, for example, if two local interactional projects compete, the progressivity of an interaction is at risk. Examples are breaks in contiguity (Stivers and Robinson 2006) and interactional impasses (Park 2010). Their main characteristic is the lack of relevant next turns, which may induce silences that lead to a break in progressivity (cf. Park 2010: 3297). According to Hoey (2018), such lapses can be filled with other activities (e.g., drinking), indicating that an ongoing activity is paused but will be resumed shortly. However, what happens if no such substitute activity is started and no pause but an actual break in progressivity is displayed is still largely understudied. Recent studies on disalignment in interaction have focused mainly on verbal (Stokoe et al. 2020) or gestural (Klatt & Krug, 2023) practices that display disalignment, while some studies have highlighted the relevance of gaze (e.g., Kendrick and Holler 2017, Pekarek Doehler et al. 2021, Robinson 2020). However, the details of the gaze practices that participants use to display disalignment in interactions remain unknown.

To address this research gap, this study uses eye tracking to provide a detailed analysis of participants’ gaze practices in situations in which interactants disalign with regard to the *in situ* joint activity, including their gaze targets and measurement of interactional timing. This level of precision is necessary to examine the subtle transitions from aligned to disaligned situations and the ways in which participants visually display disalignment in joint activities.

2. The role of gaze in disalignment sequences

Using their bodies as part of an “embodied participation framework” (Goodwin 2007), interactants display to each other that they understand themselves to be part of an ongoing activity and accept the participant status they occupy in it. Although gaze plays a central role in showing participation (Rossano 2013), most instances of gaze aversions do not indicate problems with the progressivity of an ongoing activity (Goodwin 1981, Weiß 2018). Thus, from an ethnomethodological perspective (Garfinkel 1967), practices are needed in which participants perform gaze aversions in such a way that they become visible to other members as changes in the participation framework – as disaligning with the current course of action.

The literature on gaze in disalignment situations is relatively limited. Kendon (1967) observed that in non-cooperative situations, mutual gaze between speaker and listener becomes less frequent. Recently, this observation was empirically supported by Kidwell (2006), who showed that gaze aversion could be understood as an act of resistance. In an eye-tracking study, Kendrick & Holler (2017) found that dispreferred responses are often produced with gaze aversion from recipients. The recipients also gaze away from the questioner when a dispreferred answer is imminent (Robinson 2020). These results were confirmed by Pekarek Doehler et al. (2021) in a cross-linguistic analysis. However, in their study, the sequences examined largely contained dyadic constellations. For this reason, although disaligned actions were visible through verbal and bodily conduct, the ongoing activity was rarely in danger because, in dyadic constellations, the individual participant is under more pressure to maintain progressivity. Thus, to investigate the role of gaze in situations in which the progressivity of interactions is arguably at risk, the present study examines triadic constellations (Kendrick & Holler 2017).

Accessing gaze by means of (mobile) eye tracking is a comparatively recent trend within the conversation analytic framework. Analogous to the focus of the first conversation analytic studies on the phenomenon of turn-taking (Sacks, Schegloff & Jefferson 1974), the majority of eye-tracking studies currently deal with the importance of gaze in turn-taking. These include analyses of next speaker selection (Auer 2018, Weiß 2018), gaze aversion for turn-holding (Brône et al. 2017) and overlap resolution (Zima, Weiß & Brône 2019), gaze during speaker hesitation (Weiß & Auer 2016, Krug 2023), and inquiries into the role of gaze as backchannel responses (Vranjes et al. 2018). Other studies have considered gaze in specific sequential environments, such as in question – answer sequences (Holler & Kendrick 2015, Kendrick & Holler 2017), correction or elaboration sequences (Weiß 2019), storytelling activities (Zima 2020), or word searches (Auer & Zima

2021). Furthermore, conversation analytic eye-tracking studies have investigated how interactants use gaze to negotiate joint attention. These include studies on local deixis (Stukenbrock 2018a, 2018c), attention in mobile settings (Stukenbrock & Dao 2019, Laner 2022), and gaze behaviour in the context of multimodal projections (Stukenbrock 2018b).

The aforementioned studies employ eye tracking for the purpose of fine-grained measurement and reconstruction of eye movements, which would not be possible with the usual conversation analytical methods used to study gaze from a bystander's perspective (cf. Goodwin 1980; for a discussion, see also Zima, Auer, Rühlemann, this volume). Eye-tracking glasses can obtain more specific information on both the timing of gaze movements (e.g., the exact moment when the gaze reaches its counterpart) and gaze targets (e.g., where a person looks in detail). This is a more accurate description with more robust conclusions about participants' visual orientation, but it also comes with some methodological challenges, which are briefly discussed in the following section.

3. Data and methodology

The data basis for this chapter is two video recordings of triadic interactions in a university laboratory. In each session (approximately 40 minutes), two of the participants wore mobile eye-tracking glasses (Tobii Pro Glasses 2). Six disalignment sequences were identified in the data. These six cases, along with a comparative seventh case, form a case collection analyzed sequentially according to multimodal conversation analysis (Mondada 2019). The data are presented in transcripts that follow GAT2 conventions (Couper-Kuhlen & Barth-Weingarten, 2011, see Appendix A). Mondada's (2018) conventions were used for multimodal annotations (see Appendix B). This study is exploratory and utilizes a relatively small dataset compared to other chapters in this volume (e.g., 30 hours of eye-tracking recordings in Rasmussen & Kristiansen, in this volume). Despite the limitations, the findings offer valuable initial insights into practices of gaze aversion as markers of disalignment in interactions. Future research with larger datasets is necessary to build upon these preliminary results.

Prior to the recording, the participants gave their informed consent. All data excerpts presented in this chapter were taken before the participants were given any tasks. Although authentic conversations can be collected in this way, the interactions can be considered elicited. This becomes evident when the participants endure long periods of silence without doing anything despite breaks in the progressivity of the interaction.

The data are in English, but the participants are not native speakers. At the time of recording, all participants studied English at the C1 and C2 levels. They knew each other but were not friends. The names of the six participants are pseudonyms according to the color of their sweaters (red, white, and black). As only two mobile eye-tracking glasses were available at the time, in both sessions, the participants labeled “Red” did not wear eye-tracking glasses (for an empirical explanation of why the analysis of gaze behavior from a bystander perspective can be problematic, see Zima, Auer, and Rühlemann, in this volume). The participants sat around an oblong table. The distance between them was more or less equal. Due to the table’s rectangular shape, two participants sat opposite each other, while the third participant took a more sideward position. To compensate for this imbalance, the participant in the sideways position and one of the persons sitting opposite this participant both wore eye-tracking glasses.

The participants in the study wore Tobii Pro Glasses 2 as a mobile eye-tracking device. The eye tracker’s sampling rate was 50 Hz and recorded a maximum of 50 data points per second. The eye tracker also had a scene camera mounted at the center of the eyeglass bridge, which recorded videos at 1920×1080 resolution at 25 frames per second. This made it possible to record the direction of the wearer’s head or gaze and was used as an approximation of the wearer’s field of vision. The technical design of the eye-tracking glasses results in two possible applications, which can be combined but should be separated methodologically: the use of the eye-tracking glasses as a head camera and as a measuring device for eye movements. Conversely, data from the scene camera could help researchers take the perspective of the wearer of the eye-tracking glasses and follow the interaction from their point of view. Eye-tracking glasses can be applied as a measuring device. As with any measuring device, what it can (and cannot) measure should be defined beforehand. Even if it may seem counterintuitive at first, it becomes clear in the following that, contrary to what one may assume, eye-tracking glasses do not measure where a person is looking: “First, let us make clear that we cannot know where a human is looking. Even when a participant says she looks at a point, the center of the fovea can be slightly misaligned” (Holmqvist, Nyström & Mulvey 2012: 48). Instead, 50 data points per second were recorded, which were plotted on a two-dimensional surface (in this case, the image of the scene camera). Thus, it was not the gaze that was measured but rather the position of the pupil, which was correlated with the scene camera image. As the human gaze does not represent a sum of data points, most eye-tracking manufacturers offer software solutions for visualizing the data points with which a so-called eye cursor can be generated in the form of a colored circular ring that is placed on the image of the scene camera. To obtain usable gaze data for the analysis, the data must be interpreted as standardized as possible. For this reason, concepts have been devel-

oped in psychology in which eye tracking is often used under controlled laboratory conditions (e.g., Frank et al. 2013, Kurzhals et al. 2017, Rayner 2009). One of these concepts, which will be applied below, is the area of interest (AOI) – that is, an area where a stimulus is placed that is of specific interest to researchers. When a certain threshold of data points from a participant reaches this area, the eye movements that led to this area and took place within the area are examined (Holmqvist et al. 2011: 187).

In the present case, gaze movements within disalignment situations are investigated. Therefore, of interest only are the gaze movements to and within the following AOIs: the participants, the desk, and the walls. As the measurement accuracy of the eye-tracking glasses decreases, the further an AOI is away from the calibration point of the eye tracker (cf. Holmqvist et al. 2011: 128), the sizes of the AOIs must be adjusted accordingly. The eye trackers in the present data were calibrated to a distance of 1 m. This corresponds to the distance between the person on the long side of the table (participant Black) and the participants on the short sides (participants White and Red). However, the distance between the participants on the short sides (participants White and Red, who do not wear eye-tracking glasses) is 160 cm. This means that glances by White toward the AOI Red have greater inaccuracy than those by White toward the AOI Black. Consequently, for White, the AOI Red must be 60% larger than the AOI Black.

An AOI is considered hit (the so-called AOI hit) as soon as at least one fixation has fallen within the AOI. According to Duchowski (2007), fixation can be conceptualized as comparatively static eye movements of a 150–600 ms duration with which humans can obtain information about their environment: “Fixations are eye movements that stabilize the retina over a stationary object of interest. [...] Miniature eye movements that effectively characterize fixations may be considered noise present in the control system (possibly distinct from the smooth pursuit circuit) attempting to hold the gaze steady. This noise appears as a random fluctuation about the area of fixation, typically no larger than a 5° visual angle” (Duchowski 2007: 46).

In the transcripts, the entry and exit times of the AOI hit are annotated. However, saccades, which are rapid eye movements between two fixations, often at a duration of 30–80 ms (cf. Holmqvist et al. 2011: 23), within the AOIs are not annotated separately. Thus, the eye movements between the AOIs are mainly annotated but not the eye movements themselves. On the one hand, this allows for the appropriate handling of the measurement inaccuracies of the eye tracker; on the other hand, this results in more readable transcripts while maintaining fine-grained accuracy in the description of eye movements for the present research question.

In eye-tracking studies, there is always the question of how large the boundaries of AOIs are drawn. In this study, AOI hits are counted even if the eye-

tracking cursor is 3° away from the actual object. Under optimal conditions (cf. Tobii Pro 2007), the accuracy (deviation of the measured to the “real” gaze target) (cf. Holmqvist, Nyström & Mulvey 2012: 45–46) of Tobii Pro Glasses 2 is between 0.3° and 1.1° . For large viewing angles (more than 15° from the calibration point), this deviation increases to an average of 3.05° . This means that AOIs are not hit if the viewing angle exceeds a threshold of more than 15° with respect to the calibration point. As the precision (constancy of deviation) is only a 0.05° – 0.62° deviation according to the manufacturer’s test (Tobii Pro 2007: 6) (i.e., there is only a small scatter of data points), extending the AOI by about 3° is a viable method for dealing with these technical difficulties. Accordingly, AOIs are still considered hit, even if the deviation of the gaze point from the gaze target is less than 3° .

Therefore, large AOIs allow us to deal with both the technical limitations of eye trackers and the anatomical limitations of the human eye in such a way that microanalyses of interactions are possible. The saccadic gaze changes between AOI hits are also recorded because of the traceability of the visual event, but they are analytically irrelevant due to the phenomenon of saccadic blindness (restriction of perceptual ability during saccadic gaze movements; cf. Geise 2011: 169–170). These are annotated with “...” and only provide information when one AOI is left and the next one is visually selected. If there is no AOI hit (e.g., when looking over or under the glasses), the annotation falls back on the eye-tracking glasses as the head camera. In such cases, “looks in direction x” is annotated to indicate that an AOI hit may be present but cannot be measured due to technical limitations.

4. Analysis

The following presents seven cases in which disalignment occurs at varying degrees. The cases are ordered so that the strongest forms of disalignment are presented first, and the weaker ones are presented later in the analysis. The focus is on gaze and its function in displaying and maintaining disalignments, but particular attention is given to pauses, as these are potential markers of disaligned actions. The analyses illustrate how participants treat such pauses as part of an ongoing action despite gaze aversion, for example, through facial gestures, such as raised eyebrows.

When participants’ actions in interactions disalign with respect to the common course of action, the progressivity of the interaction may be at risk (cf. Stivers & Robinson 2006). In the following paragraphs, two cases are presented in which disalignment becomes recognizable based on the participants’ orientation toward the progressivity of the interaction.

The first case shows that the participants negatively evaluate the breaks in progressivity. In the excerpt, Red, Black, and White sit at a table and talk about their hobbies. Both Black and White wear eye-tracking glasses. Black talks about how he likes to do archery with his friends. During the interaction, a disalignment phase (noted by the lack of uptake by the co-participants) is followed by an extended pause, which Black describes as an “awkward silence.”



Figure 1. Both black (left image) and White (right image) look at the AOI desk

Excerpt 1. Awkward silence (T₃₁)

```

001 BLA  *%we go where nobody's aROUND,* (-)*
    bla  *@RED-----*...*
    whi  %@BLA----->

002      *and we * can't* HARM anybody (-) apart from ourselves;*(-)*
    bla  *@Wall--*.....*@RED-----*...*
003      *and then (-) yeah (-) we go* (-)* to shoot some Arrows.=ya?
    bla  *@desk-----*...*@RED----->

004 RED  yeah it's COOL;
005 bla  robin HOOD style.
006 RED  *((laughs))*
    bla  *.....*
007 WHI  *%((laughs))
    bla  *@WHI----->
    whi  %@desk--->

008 RED  *%have*% you ever SPLIT an arrow (-) yet,
    bla  *...*@RED----->
    whi  %.....%@RED----->

009 BLA  *%n::o*% (-) NO % (-) % (--)% not% yet;
    bla  *...*@desk--->>
    whi  %.....%@BLA---%.....%@RED--%.....%@desk--->>

010      (8.5)#1
011 BLA  awkward SiLence.

```

As Black describes the circumstances of his hobby in more detail (001–003), his gaze alternates between Red, the wall, and the desk. White’s gaze is fixed on Black as the speaker throughout his turn. Black receives a positive evaluation from Red (“cool,” 004), which Black picks up on with the category “robin hood style” (005).

On a thematic and multimodal level, the participants are aligned in the first part of the sequence. This continues as Black's self-categorization receives a laugh from both Red (006) and White (007). Following up on this humorous sequence, Red asks whether Black has ever split an arrow (008). Black responds with a prolonged "no," which he repeats again after a short pause and specifies as pending after another pause (009). His gaze lingers on the desk without producing another situation-appropriate action or gaze-selecting one of the co-participants as the next speaker. The fact that the conversation has come to a standstill is evident not only in the absence of further turns by Red or White but also in White's gaze behaviour, which first alternates between the potential next speakers Red and Black and finally gazes on an interactionally less relevant area (the desk). As the interaction partners fail to produce a situation-appropriate action for more than 8 seconds (010), they are disaligned with regard to their common course of action. This can be observed in their interactional impasse, which Black finally refers to as an "awkward silence" (011).

This case shows that in disalignment situations, the progressivity of an interaction is disrupted. These disruptions are negatively evaluated by the participants and are usually avoided by them (cf. Stivers & Robinson 2006). Here, interruptions in progressivity are brought about by a gaze practice that consists of withdrawing the gaze from the co-participants and looking at interactionally less relevant areas (the wall, the desk). The illustrated gaze practice of averting one's gaze from the interaction partners during disalignments appears to be robust in the present data.

This also comes into play in the following excerpt. Unlike the previous example, not all participants disalign here. Instead, an utterance by one participant (White) is treated as situationally inappropriate, in which the participant briefly withdraws from the course of action.

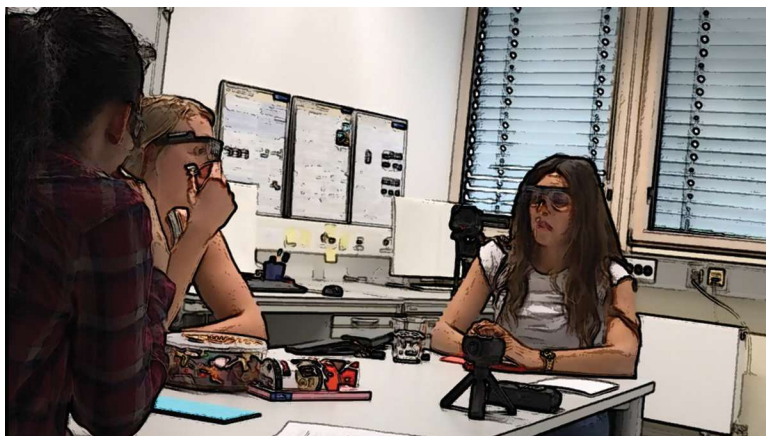


Figure 2. White grimaces while looking down at the AOI desk

Excerpt 2. Riverdale (T23)

```

001 RED  $*+they are doing a spinoff,=right?
        red  $@WHI1----->
        whi  *@RED----->
        gre  +@RED----->

002 BLA  what,
003 WHI  yeah- (-)
004      [I think] it's RIVerdale?
005 RED  [they  ]
006 WHI  were [they-  ]
007 RED  [no +no +]
        bla  +...+

008 BLA  +*no *river*dale+*(-) % #2is+ something% else
        bla  +@WHI-----+.....+@RED----->>
        whi  *...*@BLA-*.....*@desk----->
        whi                                     %grimaces-----%

009 WHI  $<<pp>0kay.>
        red  $@BLA----->>

010 RED  I'm not going* to WATCH it;
        whi  →*@RED----->

011 bla  me NEither;
012 RED  I think it's TORTure;
013 bla  yeah;

```

Participants Black and White (both wearing eye-tracking glasses) and Red are talking about a TV series that all three have seen. Red introduces the information that this series will get a spinoff (001), which Black answers with a marker of disbelief (“what,” 002). At this point, both Black and White look to Red, as she is the narrator. While Black’s gaze alternates between White and Red, Red’s gaze initially remains fixed on White, who is sitting opposite her. White epistemically aligns with Red’s information (003) and introduces a potential name for the spin-off (“Riverdale,” 004). As she begins to elaborate on this information (006), she is interrupted by Red (007), and Black rebukes her that Riverdale is “something else” (008). As a result of this rebuke of her utterance, which is disaligning with the current joint activity, White withdraws from the conversation. She accomplishes this in three steps. First, she directs her gaze, which up to this point has been regularly shifting between the co-participants and the desk, which is an area of little interactional relevance. Her gaze remains there until she later rejoins the interaction and indicates her participation by visually addressing Red (010). Second, in addi-

1. As Red is not wearing eye-tracking glasses in this case, these annotations are based only on a visual estimate of her eye movements (cf. Zima, Auer, and Rühlemann, in this volume).

tion to averting her gaze, White grimaces (#2), and third, whispers an acceptance of this rebuke (“okay,” 009). The other participants also treat White as temporarily no longer participating in the interaction. This is particularly visible in Red’s gaze behaviour. She pulls her gaze away from White and directs it to the remaining member, Black, when White displays multimodal non-availability.

As White disengages from the interaction after the disalignment, White exhibits the same gaze behaviour that Black and White showed in the previous case. Thus, the participant seems to visually display disalignment. By gazing at interactionally less relevant areas, White deals with the interactional problem that the resource gaze can never be completely unused, as every glance at something can always be potentially regarded as a gaze at something else, thus making it interactionally relevant. Therefore, looking at interactionally less relevant areas avoids visually addressing any of the other participants, which could be a marker of alignment.² This form of participation is what Goffman (1963: 69–73) calls self-involvement: “This kind of inward emigration from the gathering may be called ‘away.’” When participating in the form of self-involvement, the interactants produce so-called disengagement displays (Goodwin 1981). Interactants monitor, among other things, bodily resources, such as gestures, orientation of the upper body, and gaze, which provide them with information about the engagement display of the other participants (cf. Oben, de Vries, and Brône, this volume). With the help of such an engagement display as an orientation of one participant toward another within a social situation, the interactants indicate to each other what participant status they attribute to each other. Therefore, interactors with an active engagement display make themselves communicatively available to other participants, whereas a disengagement display (e.g., as a consequence of a disalignment) indicates unavailability.

When all participants are engaged in self-involvement, there is no more exchange or interaction. Such a state seems to be dispreferred by the participants in the data because it is always the initial *no-sayer* who produces the next action to evoke a resolution of self-involvement to reengage the disinvolved participant, thus reestablishing progressivity. This practice is illustrated in the following two sequences. In Excerpt 3, the *no-sayer* is the only one who does not display self-involvement, while in Excerpt 4, the *no-sayer* also turns away from the interactional events.

2. Certainly, as already shown by Goodwin (1980), mutual gaze is not necessarily an indicator of participation, but it also depends on speakership to a great extent.



Figure 3. White looks at Red, who turns her head away

Excerpt 3. Abroad (T₀₁)

```

001 RED  *but YOU went +abroad+ as well,
        whi  *@RED----->
        bla           +.....+@WHI---->

002 WHI  NO i didn't;

003      (0.5)

004 RED  no;

005      (0.2) % (0.3)#3 + (0.2) + (0.2) * (0.2)
        red      %turns head away----->
        bla      -->+.....+@Wall----->
        whi      -->*@BLA->

006 WHI  *but* %i% +would+ LIKE to;
        whi  *...*@RED----->
        red      %.%@towards WHI-->
        bla      +.....+@WHI-->

```

Looking at White, Red asks a question about whether White also did a semester abroad (001). By gaze-selecting White, Red disambiguates the potentially ambiguous pronominal address term “you.” The fact that Black does not feel addressed is shown by her gaze, which switches to White as the addressed participant (001). White negates the question (002), which is acknowledged by Red through repetition of the answer particle (004). When no further uptake happens – that is, when it becomes unclear in which direction the conversation will continue – the questioner Red orients herself away, and Black also turns away and fixes her gaze on the wall opposite her (005). Similar to the first example, in which there was an “awkward silence,” the participants consider it strange to simply say “no” without providing another account. This is evident in White’s behaviour, which takes responsibility for restoring the dysfunctional progression of interaction that has occurred. Therefore, the no-sayer White, who is the only one who does not indicate

self-involvement, turns from the questioner Red to the next potential speaker Black. When the latter indicates no willingness to interact, White resumes her turn (006). This earns her the reciprocity displays of Red and Black, who both end their self-involvement and turn their attention to White. Thus, the interactional impasse of the disaligned situation is overcome, the interactants align with each other again, and the conversation continues with verbal contributions by all three participants.



Figure 4. White (right image) looks at Black, who looks at the AOI desk (left image)

Excerpt 4. Glasses (T25)

```

018 RED  *and are you NERvous with* those* thingies on,=or?
        bla  *@RED-----*.....*@Wall----->
        whi  +@RED----->
019 BLA  +no +[not ] nervous they're just (.) really annoying;
        [mhmh;]
020 WHI  whi  +...+@BLA----->
021 RED  [((laughs))]
022 BLA  [cause you ] feel like you have this massive THING on your head,
023      and IT'S (-)
024 RED  but +you+ *don't* wear GLASSes+ in+ [xxx,]
        bla  *.....*@RED----->
        whi  +...+@RED-----+...+@BLA-->
025 BLA  [no ]
026      *i never*+ i NEVer+ wore glasses.
        bla  *.....*@desk----->
        whi  -->+.....+@desk----->
027      (2.5)
028 BLA  so i'm +not USED+ to #4it at all;
        whi  -->+.....+@BLA----->

```

In this excerpt, participants Red, White, and Black talk about the eye-tracking glasses they are wearing (018–23). Red (who is not wearing eye-tracking glasses) asks the thematically related question of whether White and Black normally wear glasses (024). Unlike in the previous case, his alternating gaze indicates that, this time, he is addressing both participants with the ambiguous address term “you.” Similar to Black in the previous excerpt, White’s gaze on Black indicates that she understands him to be the addressee of this question or rather that she expects

him to answer first (cf. Weiß 2019). Black negates the question (025) and follows it up with concretization (026). Simultaneously, with his last utterance, he lowers his gaze to the desk. Interestingly, White, who has been watching Black for a while, also lowers her gaze as soon as Black takes his eyes off Red. Therefore, White, like Black, indicates self-involvement, projecting a reluctance to answer the question Red verbally addresses to both of them. For more than 2 seconds, neither party provides an action that would maintain progressivity.³ This interactional disalignment is resolved only by Black, who offers another thematic contribution and draws an arc to the initial theme of “eye-tracking glasses.” White resolves her self-involvement and directs her gaze to Black. In the interaction that follows, the participants align again and continue their conversation.

Both sequences show that the initial no-sayer seems to be responsible for resolving the interactional impasse caused by a lack of uptake by the co-interactors. The sequences also show that self-involvement displayed through foveal attention to interactionally less relevant areas can be understood as displays of non-reciprocity by the co-participants. When several participants withdraw from the interaction in this way, there is a noticeable disruption in progressivity.

The above cases show situations in which participants’ actions are disaligned, either by treating an action as inadequate or by not giving an uptake of a previous turn, in which the progressivity of the ongoing interaction is interrupted. The following three cases represent the other side of the disalignment continuum. They demonstrate the principle of contiguity in interactions (Sacks 1987) – that is, that participants have a basic assumption that other participants will maintain progressivity. Excerpt 5 shows that, while conversational pauses can be indicative of disalignments, the gaze shows whether disalignment is real for the participants. Excerpt 6 illustrates in this context that other multimodal resources can also indicate continuing alignment. Finally, Excerpt 7 demonstrates a case that looks like a disalignment on the verbal level, but only with a focus on gaze as a resource does it become apparent that the participants are still aligned with each other.

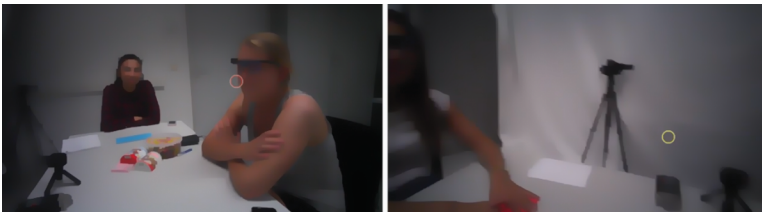


Figure 5. White looks at Black (left image), who looks at the AOI wall (right image)

3. As Red does not wear eye-tracking glasses, it is not clear from the data whether Red is looking at one of the two participants or fixating on a point between the two interactants.

Excerpt 5. Future plans (To3)

001 RED *%do you* wanna*%do it% HERE,
 bla *@desk--*.....*@RED----->
 whi %@RED-----%,....%@BLA-->>
 002 BLA +no. +
 bla +shakes head+
 003 RED [((laughs))]
 004 WHI [((laughs))]
 005 BLA *[((laughs))]* no WAY;
 bla *.....*@desk-->
 006 (0.7)
 007 BLA *no there* is a:: #5 f: i don't KNOW,
 bla *.....*@wall----->>
 008 it's not a university,
 009 but it's a: *westfälische* HOCH*Schule
 bla -->*.....*@RED----->>
 010 RED ah;
 011 BLA and they HAVE-

Red asks the other participants about their plans after graduating with a bachelor's degree and whether they want to do their master's at their current university (001). Based on White's visual orientation to Black, it can be stated that White understands Black as an addressed participant. Black denies the question and shakes her head energetically (002). The other participants react laughingly to this display of determination (003–004). As in the previous cases, a single “no” does not seem to be sufficient for the participants because Black, the no-sayer, follows it up with the affirmation “no way” (005). At the same time, she lowers her gaze to the desk as an interactionally less relevant area and remains silent for 0.7 seconds (006). In this way, Black indicates the end of the sequence (cf. Rossano 2013) and self-involvement. However, unlike in the previous cases, the other participants treat this pause as an activity that contributes to the ongoing interaction. This is especially evident in White's gaze behaviour of keeping her gaze fixed on Black the whole time, indicating that she considers a continuation of Black's turn likely and still treats her as the current speaker. Consequently, as White continues to look at her, Black continues her turn. When Black continues with a longer turn in which she presents her concrete plans (007–011) and visually addresses the questioner Red again, it becomes apparent that the short disruption of progressivity has been overcome by the participants.

A similar situation can be seen in the following case, in which a pause of comparable length is accompanied by a facial gesture.



Figure 6. White looks at Red, who raises both eyebrows (left image), while Black looks at the AOI wall (right image)

Excerpt 6. Vacation (T12)

```

001 WHI *do you have* um any plans* o (.)* for vaCation,
    whi *.....*@BLA-----*.....*@RED----->
002     are *you* GOing on +vacation? +* =or?*
    whi *...*@BLA-----*.....*
    bla                                     +shakes head+
005 RED *in SUMmer,
    whi *@RED----->
006 WHI mhm,
007 BLA +no;
    bla +@wall-->
008 RED no; % (0.2) %
    red %moves head up%
009 RED % (0.4) #6 %
    red %raises both eyebrows%
010 RED i %have% to write TERM <<laughing>[paper> ]
    bla %...*@RED----->
011 BLA [(laughes)]
012 RED THAT'S my vacation;
013 BLA ((laughes))

```

White asks her co-participants about their vacation plans (001–002). From her alternating looks, it is clear that her question addresses both Red and Black (cf. Auer 2021). While the question is still being asked, Black provides a negative answer by shaking her head (002). Red asks for a specification (“in summer,” 005), to which White replies (006). Black gives another verbal response before visually targeting the wall in front of her, providing a potential display of self-involvement (007). Red also responds negatively to the question, turning her head away and upward (008). Gaze aversion is not exclusively used to mark self-involvement, as this may also indicate that she is preparing the next utterance (cf. Kendon 1967). However, unlike Black, Red indicates continued interactional alignment with raised eyebrows (009). The questioner, White, keeps her gaze fixed on Red the entire time, indicating that she is inviting a continuation of Red’s turn (010).

Black then participates in Red's turn visually and through laughter (011), thus restoring the progressivity of the interaction. This case shows that, for White, the raised eyebrows are an indication of continued alignment between the participants. This is evidenced by the fact that her gaze does not wander to a less relevant field but continues to look at the person from whom she could expect the next interactional contribution.

The same practice is shown in the last case in this study, in which it is clear from the gaze behaviour that the participants' actions do not disalign, despite the obvious disagreement.

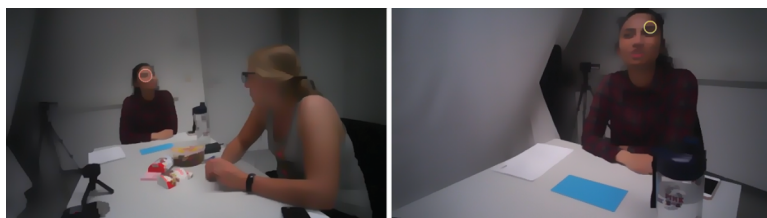


Figure 7. White (left image) and Black (right image) look at Red

Excerpt 7. Personal matter (T09)

```

001 RED  *+but (.) i don't KNOW;
      whi  *----->
      BLA  +@RED----->>

002      i don't (-) i don't FEEL- (-)

003 BLA  *COM*FORT*able,*
      whi  *...*@BLA*....*

004 RED  *yeah NO it's (-) comfortable,
      whi  *@RED----->>

005 WHI  it's not YOUR thing;

006 RED  no it's not THAT;
      (1.0)#7

008 RED  i MEAN [i would-]
009 BLA      [oKAY;  ]

```

Here, the participants are discussing an event in Red's life that is not elaborated on for privacy reasons. In her narrative, Red signals uncertainty about the evaluation of the event she is discussing (001–002). Black first offers a syntactically integrated word suggestion (003) (cf. Auer & Zima 2021), which Red rejects (004). White then provides a possible evaluation (005), which is rejected again by Red (006). Thus, although Red does not reach an agreement with either co-participant and thus treats both suggestions as inadequate in her ongoing narrative activity, neither participant lowers her gaze. Even in the 1-second pause that follows, the par-

ticipants remain oriented toward each other and indicate alignment with each other. On a purely verbal level, the progressivity of the interaction seems to have stalled. Red rejects a statement as inadequate for her current course of action. The 1-second pause that follows could be an indication of an interactional impasse and, thus, a disalignment. It is also fitting that Red, as the no-sayer, seems to be responsible for overcoming this interactional impasse. However, the gaze behaviour of the other participants makes it clear that they do not assume a disalignment in this situation. Rather, their continued visual orientation toward Red as the storyteller seems to show that there is no break in the interaction for them and that they are still aligned with the current course of action. Thus, against the background of the other cases, this case demonstrates the important role of gaze in the organization of alignment in interactions. Although previous studies (e.g., Goodwin 1980) have demonstrated that interactions can still occur in a structured manner even without mutual gaze, it seems that in situations that involve potential danger to the advancement of the interaction, mutual gaze between individuals is adequate to signify alignment.

5. Conclusion

In interactional situations, participants strive to maintain progressivity in interactions. Actions that disrupt this progressivity are dispreferred. One of the important prerequisites for progressivity is the alignment between participants. Alignment occurs when the interactants indicate to each other that they support the current course of action. This includes deontically asymmetric situations in which one participant contributes more than the others (e.g., storytelling activities) as well as conflict interactions (e.g., disputes). If actions do not contribute to an ongoing course of action, they are disaligning. One way of indicating disalignment is to avert one's gaze. However, as not every averted gaze indicates disalignment, it requires practice on the part of the participants to communicate availability.

In this study, actions are regarded as disaligning when they are considered inappropriate for a given situation or when no further uptake occurs – that is, when it becomes unclear in which direction the conversation will proceed. In disalignment, participants redirect their gaze, which until then has frequently shifted between the co-participants (cf. Oben, de Vries, Brône, this volume), to interactionally less relevant areas. Their foveal attention remains until they rejoin the interaction and indicate participation. Thus, by looking at interactionally less relevant areas, they avoid visually addressing other participants, which could be interpreted as a display of alignment. This form of participation is called self-involvement (Goffman 1963), and it indicates non-availability. By not addressing

the persons engaged in self-involvement visually and verbally, participants treat them as temporarily not participating in the conversation until they rejoin it on their own accord. There is a noticeable disruption of progressivity when several participants withdraw from an interaction in this way. In a self-involved situation, there is no exchange or interaction. Based on the data, it appears that the participants disprefer such a state because the person causing an interactional impasse always produces a next action that evokes a resolution of self-involvement.

Furthermore, the analysis shows that mutual gaze plays an important role in overcoming disalignment in interactions. Thus, participants understand pauses, which can indicate disalignments, as part of an ongoing utterance if they are accompanied by a mutual gaze. If no mutual eye contact is established, for example, if one of the participants turns away, then facial gestures, such as raised eyebrows, can indicate continued alignment. Thus, when considering alignment in social situations as one of the basic requirements for interactions (cf. Stivers 2008), gaze seems to be crucial in displaying orientation to the current course of action.

Compared with Pekarek Doehler et al.'s (2021) study, the sometimes long interruptions of the interaction flow are striking. In their study, disalignment is also used to indicate that a certain course of action is not understood as adequate. However, progressivity itself is rarely in danger. Participants also show no self-involvement by visually focusing on interactionally less relevant areas. One possible explanation for these differences could be the configuration of the participants. While Pekarek Doehler et al. (2021) examined dyadic participant settings, the present study considers triadic constellations, in which there is less responsibility for individual participants to maintain progressivity. In addition, the degree of familiarity may play a role. That is, it may be easier for participants to disengage from an interaction after disalignment if the other interactants are unknown; consequently, little social calibration needs to occur. Accordingly, self-involvement, as indicated by gaze aversion to interactionally less relevant areas, could be a practice for enduring the "awkward silence" of interrupted progressivity.

Although the results suggest a systemic pattern, the study is limited in two aspects. First, the use of only two pairs of eye-tracking glasses by three participants can be considered a technical limitation. This is not a deliberate decision in the study design but is due to the fact that no third pair of eye-tracking glasses was available at the time of recording. Thus, although the multimodal practices of the person without eye-tracking glasses are also annotated and analyzed, this limits the case selection and perspective on the situation when different granularity levels have to be used for the same resource for different participants. Second, it is difficult to empirically grasp disalignment. While the study is able to help determine the endpoint of the alignment continuum (Excerpt 1, "awkward silence"), in which the participants are maneuvered into an interactional state of rupture in the

progression of the interaction, the intermediate steps are more difficult to detect. As the participants are oriented toward maintaining progressivity, many actions contain at least traces of alignment (e.g., turn-taking and gaze practices). Therefore, it is difficult to find and systematize disalignment in its pure form. As disalignment also appears to be rare, a larger-scale study and larger case collection could help here.

Acknowledgments

I would like to thank the two anonymous reviewers and the editors, Anja Stukenbrock and Elisabeth Zima, for their many helpful comments in the earlier versions of this article.

References

-  Auer, Peter. 2018. "Gaze, Addressee Selection and Turn-Taking in Three-Party Interaction." In *Eye-Tracking in Interaction: Studies on the Role of Eye Gaze in Dialogue*, edited by Geert Brône and Bert Oben, 197–231. Amsterdam, Philadelphia: John Benjamins.
-  Auer, Peter. 2021. "Gaze Selects the Next Speaker in Answers to Questions Pronominally Addressed to More Than One Co-participant." *Interactional Linguistics* 1(2): 154–182.
- Auer, Peter, and Elisabeth Zima. 2021. "On Word Searches, Gaze, and Co-participation." *Gesprächsforschung – Online – Zeitschrift zur verbalen Interaktion* 22: 390–425.
-  Brône, Geert, Bert Oben, Annelies Jehoul, Jelena Vranjes, and Kurt Feytaerts. 2017. "Eye Gaze and Viewpoint in Multimodal Interaction Management." *Cognitive Linguistics* 28 (3): 449–483.
- Couper-Kuhlen, Elizabeth, and Dagmar Barth-Weingarten. 2011. "A System for Transcribing Talk-in-Interaction: GAT 2." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 12: 1–51.
- Duchowski, Andrew T. 2007. *Eye Tracking Methodology: Theory and Practice*. London: Springer.
- Zima, Elisabeth, Peter Auer and Christoph Rühlemann, this volume. "Why research in gaze in social interaction needs mobile eye tracking."
-  Frank, Stefan L., Irene Fernandez Monsalve, Robin L. Thompson, and Gabriella Vigliocco. 2013. "Reading Time Data for Evaluating Broad-Coverage Models of English Sentence Processing." *Behaviour Research Methods* 45 (4): 1182–1190.
- Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.
-  Geise, Stephanie. 2011. "Eyetracking in der Kommunikations- und Medienwissenschaft: Theorie, Methode und Kritische Reflexion." *Studies in Communication/Media* 2: 149–263.
- Goffman, Erving. 1963. *Behaviour in Public Places: Notes on the Social Organization of Gatherings*. New York: Free Press.
-  Goodwin, Charles. 1980. "Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn Beginning." *Sociological Inquiry* 50 (3–4): 272–302.

Goodwin, Charles. 1981. *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.

doi Goodwin, Charles. 1986. "Between and Within: Alternative Sequential Treatments of Continuers and Assessments." *Human Studies* 9: 205–217.

doi Goodwin, Charles. 2007. "Participation, Stance and Affect in the Organization of Activities." *Discourse & Society* 18 (1): 53–73.

Goodwin, Charles, and Marjorie H. Goodwin. 2004. "Participation." In *A Companion to Linguistic Anthropology*, edited by Alessandro Duranti, 222–244. Malden, Mass.: Blackwell.

Heller, Vivien. 2016. "'das_s VOLL verARsche hier': Aligment und Disalignment mit jugendsprachlichen Praktiken in der Unterrichtsinteraktion." In *Jugendsprache in Schule, Medien und Alltag*, edited by Carmen Spiegel and Daniel Gysin. 91–108. Frankfurt am Main: Peter Lang.

doi Hoey, Elliott M. 2018. "Drinking for Speaking: The Multimodal Organization of Drinking in Conversation." *Social Interaction. Video-Based Studies of Human Sociality* 1(1).

doi Holler, Judith, and Kobin H. Kendrick. 2015. "Unaddressed Participants' Gaze in Multi-Person Interaction: Optimizing Reciprocity." *Frontiers in Psychology* 6: 1–14.

Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.

doi Holmqvist, Kenneth, Marcus Nyström, and Fiona Mulvey. 2012. "Eye Tracker Data Quality: What It Is and How To Measure It." *Proceedings of the Symposium on Eye Tracking Research and Applications*, 45–52.

doi Kendon, Adam. 1967. "Some Functions of Gaze-Direction in Social Interaction." *Acta Psychol.* 26: 22–63.

doi Kendrick, Kobin H., and Holler, Judith. 2017. "Gaze Direction Signals Response Preference in Conversation." *Research on Language and Social Interaction* 50(1): 12–32.

doi Kidwell, Mardi. 2006. "'Calm Down!' The Role of Gaze in the Interactional Management of Hysteria by the Police." *Discourse Studies* 8: 745–770.

doi Klatt, Marie, and Maximilian Krug. 2023. "Von der Disalignierung zum Disengagement. Aushandlung von Partizipation in konfliktären Eltern-Kind-Interaktionen." *fokus:interaktion* 1: 29–66.

doi Krug, Maximilian. 2023. "Overcoming Blanking: Verbal and Visual Features of Prompting in Theatre Rehearsals." *Human Studies* 46: 221–246.

doi Kurzahls, Kuno, Michael Burch, Tanja Blascheck, Gennady Andrienko, Natalia Andrienko, and Daniel Weiskopf. 2017. "A Task-Based View on the Visual Analysis of Eye-Tracking Data." In *Eye Tracking and Visualization: Foundations, Techniques, and Applications*, edited by Michael Burch, Lewis Chuang, Brian Fisher, and Albrecht Schmidt, ETVIS 2015: 3–22. Cham: Springer International Publishing.

Laner, Barbara. 2022. "'Guck mal der Baum' – Zur Verwendung von Wahrnehmungsimperativen mit und ohne mal." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 23: 1–35.

doi Mondada, Lorenza. 2018. "Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality." *Res. Lang. Soc. Interact.* 51: 85–106.

- doi Mondada, Lorenza. 2019. "Contemporary Issues in Conversation Analysis: Embodiment and Materiality, Multimodality and Multisensoriality in Social Interaction." *Journal of Pragmatics* 145: 47–62.
- Morek, Miriam. 2016. "'watt soll ich dazu Sagen' – (Dis)Alignment bei der interaktiven Manifestation epistemischer Asymmetrien." In *Wissen in institutioneller Interaktion*, edited by Alexandra Groß and Inga Harren. Forum Angewandte Linguistik 55, 145–175, Frankfurt am Main: Peter Lang.
- Oben, Bert, Clarissa de Vries, and Geert Brône. This volume. "Mobile eye-tracking and mixed-methods approaches to inter-action analysis".
- doi Park, Innhwa. 2010. "Marking an Impasse: The Use of Anyway as a Sequence-Closing Device." *Journal of Pragmatics* 42(12): 3283–3299.
- doi Pekarek Doehler, Simona, Hilla Polak-Yitzhaki, Xiaoting Li, Ioana Maria Stoenica, Martin Havlík, and Leelo Keevallik. 2021. "Multimodal Assemblies for Prefacing a Dispreferred Response: A Cross-Linguistic Analysis." *Frontiers in Psychology* 12: 1–24.
- Rasmussen, Gitte and Elisabeth Dalby Kristiansen. This volume. "The influence of the specificities of gaze behaviour on emerging and ensuing interaction – A contribution to the discussion of the use of eye-tracking recordings for EMCA analysis."
- doi Rayner, Keith. 2009. "Eye Movements in Reading: Models and Data." *J Eye Mov Res.* 2(5): 1–10.
- doi Robinson, Jeffrey. D. 2020. "One Type of Polar, Information-Seeking Question and Its Stance of Probability: Implications for the Preference for Agreement." *Res. Lang. Soc. Interact.* 53: 425–442.
- Rossano, Federico. 2013. "Gaze in Conversation." In *The Handbook of Conversation Analysis*, ed. by Jack Sidnell and Tanya Stivers, 308–329. Chichester: Wiley-Blackwell.
- doi Sacks, Harvey. 1987. "On the Preferences for Agreement and Contiguity in Sequences in Conversation." In *Talk and Social Organisation*, ed. by Graham Button and John R. E. Lee, 54–69. Clevedon: Multilingual Matters.
- doi Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A Simplest Systematics for the Organisation of Turn-talking in Conversation." *Language* 50 (4): 696–735.
- doi Steensig, Jakob. 2019. "Conversation Analysis and Affiliation and Alignment." In *The Encyclopedia of Applied Linguistics*, ed. by Carol A. Chapelle, 1–6. John Wiley & Sons.
- doi Stivers, Tanya. 2008. "Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation." *Research on Language & Social Interaction* 41 (1): 31–57.
- doi Stivers, Tanya. 2010. "An Overview of the Question-Response System in American English Conversation." *Journal of Pragmatics* 42 (10): 2772–2781.
- doi Stivers, Tanya, and Jeffrey D. Robinson. 2006. "A Preference for Progressivity in Interaction." *Lang. Soc.* 35(03).
- doi Stivers, Tanya, Lorenza Mondada, and Jakob Steensig. 2011. "Knowledge, Morality and Affiliation in Social Interaction." In *The Morality of Knowledge in Conversation*, ed. by Tanya Stivers, Lorenza Mondada, and Jakob Steensig. Cambridge/New York: Cambridge University Press.
- doi Stokoe, Elizabeth, Bogdana Humă, Rein O. Sikveland, and Heidi Kevoe-Feldman. 2020. "When Delayed Responses are Productive: Being Persuaded Following Resistance in Conversation." *Journal of Pragmatics* 155: 70–82.

Stukenbrock, Anja. 2018a. "Blickpraktiken von SprecherInnen und AdressatInnen bei der Lokaldeixis: Mobile Eye Tracking-Analysen zur Herstellung von joint attention." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 19: 132–168.

 Stukenbrock, Anja. 2018b. "Forward-Looking: Where Do We Go with Multimodal Projections?" In *Time in Embodied Interaction. Synchronicity and Sequentiality of Multimodal Resources*, ed. by Arnulf Deppermann and Jürgen Streeck, 31–68. Amsterdam: John Benjamins.

 Stukenbrock, Anja. 2018c. "Mobile Dual Eye-Tracking in Face-to-Face Interaction: The case of deixis and joint attention." In *Eye-Tracking in Interaction: Studies on the Role of Eye Gaze in Dialogue*, ed. by Geert Brône and Bert Oben, 265–300. Amsterdam, Philadelphia: John Benjamins.

 Stukenbrock, Anja, and Anh N. Dao. 2019. "Joint Attention in Passing: What Dual Mobile Eye Tracking Reveals About Gaze in Coordinating Embodied Activities at a Market." In *Embodied Activities in Face-to-face and Mediated Settings: Social Encounters in Time and Space*, ed. by Elisabeth Reber and Cornelia Gerhardt, 177–213. Cham: Springer.

Tobii Pro. 2007. "Eye Tracker Data Quality Report: Accuracy, Precision and Detected Gaze Under Optimal Conditions – Controlled Environment." *Tobii Pro Glasses 2 firmware v1.61* (Methodology version: 1.0), 1–16.

 Vranjes, Jelena, Hanneke Bot, Kurt Feyaerts, and Geert Brône. 2018. "Displaying Reciprocity in an Interpreter-Mediated Dialogue: An Eye-Tracking Study." In *Eye-Tracking in Interaction: Studies on the Role of Eye Gaze in Dialogue*, ed. by Geert Brône and Bert Oben, 301–321. Amsterdam, Philadelphia: John Benjamins.

 Weiß, Clarissa. 2018. "When Gaze-Selected Next Speakers Do Not Take the Turn." *Journal of Pragmatics* 133: 28–44.

Weiß, Clarissa. 2019. "Blickverhalten des nicht-blickselegierten Sprechers während Korrekturen und Elaborierungen." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 20: 1–28.

Weiß, Clarissa, and Peter Auer. 2016. "Das Blickverhalten des Rezipienten bei Sprecherhäsitationen: eine explorative Studie." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 17: 132–167.

 Zima, Elisabeth. 2020. "Gaze and Recipient Feedback in Triadic Storytelling Activities." *Discourse Processes* 57 (9), 725–748.

 Zima, Elisabeth, Clarissa Weiß, and Geert Brône. 2019. "Gaze and Overlap Resolution in Triadic Interactions." *Journal of Pragmatics* 140: 49–69.

Zima, Elisabeth, Peter Auer, and Christoph Rühlemann. This volume. "Why research in gaze in social interaction needs mobile eye tracking."

Pupil size indicates planning effort at turn transitions in natural conversation

Mathias Barthel & Christoph Rühlemann

Leibniz-Institute for the German Language | University of Freiburg

The study investigates the cognitive demands of speech planning in unrestricted, natural conversation. Focusing on question-answer sequences in triadic interactions, we analyse whether answerers, compared to not-answerers, exhibit increased cognitive effort during turn transitions. Using pupil size data from the Freiburg Multimodal Interaction Corpus, we find that answerers indeed show greater pupil dilation than not-answerers, suggesting heightened processing load during speech planning at transition-relevance places. This finding supports the hypothesis that speech planning is a primary contributor to increased cognitive effort during turn transitions, highlighting the value of pupillometry in the study of naturalistic conversation. The findings offer insights into the cognitive dynamics of multiparty social interaction, bridging the gap between controlled experiments and ecologically valid conversational settings.

Keywords: interaction, conversation, triads, turn-taking, turn-transitions, pupillometry, processing load, speech planning, question-answer sequences, corpus analysis

1. Introduction

When talking to each other in everyday interactions, be it chit-chat among a group of friends, price negotiations between business representatives or planning a family holiday trip at the dinner table, conversational partners take turns at talk. The turn-taking system is a fundamental component among the organising principles that participants of a conversation adhere to so as to bring order and continuity into their interaction dynamics (Sacks et al. 1974). This system incorporates a set of rules for turn allocation that interlocutors follow to regularly switch the roles of being the speaker at one time and the listener at another time during the conversation. When a turn by one participant approaches possible completion, a transition to the next turn by another participant becomes rele-

vant (Ford & Thompson 1996, Ford et al. 1996, Selting 2000). Even though turns at talk are mostly rather short, transitions between them are generally quite fast, with the most frequent case being a very short gap between turns of only a few hundred milliseconds (Heldner & Edlund 2010, Levinson & Torreira 2015). In a corpus containing recordings from ten typologically distinct languages, Stivers et al. (2009) found that, while differing in a number of details, this universal tendency for timely aligned turn transitions pertains to all the studied languages, and is thus a very good candidate for a universal characteristic in conversational language use. Stivers et al. furthermore showed that question-answer sequences, which they primarily focused on in their study, serve as a very good proxy for the timing patterns of turn transitions in general, i.e., for turn-timing in other sequences in the corpus.

The turn-taking system and the rapid turn transitions put considerable time pressure on the next speaker who wants to initiate their next turn right at the time when (or very shortly after) the current turn by their interlocutor approaches its point of completion. One fundamental reason for the time pressure lies in the systematics of the rule set of the turn-allocation component. Whenever the next speaker is not selected by the current speaker during the current turn, the participant speaking up next when turn transition is relevant gains the rights to produce the next turn (Sacks et al. 1974). Contrary to earlier accounts, which have emphasized the role of self-selection, more recent research suggests that numerous cases of apparent self-selection are in fact cases of next-speaker selection by gaze (Auer 2021a, 2021b, see also Zima et al. (2019) on the role of gaze in resolving simultaneous starts).¹ Irrespective of the selection mechanism at any given turn transition, being late when intending to start a next turn might lead to missing out on the chance to take the floor, either because the current speaker might continue to take another turn or because another listener might speak up next. Another reason why the next speaker would aim to keep turn transition times short lies in the “universal semiotics of delayed response” (Stivers et al. 2009: 10591). Leaving a markedly long gap before initiating the next turn might be interpreted by conversational partners in numerous ways, for instance as hesitation (e.g. in response to an invitation), reluctance (e.g. in response to a request), or disagreement (e.g. in response to an evaluating statement) (Fox Tree 2002, Henetz 2017, Roberts & Francis 2013, Roberts et al. 2006). To plan a relevant contribution to the ongoing exchange of turns and to prepare it for articulation, time is thus very limited between the moment when planning can reliably begin (i.e., when the message of the current turn is understandable or at least sufficiently predictable) and the

1. The effects of speaker selection are beyond the scope of the present study, but see Rühlemann & Barthel (under revision).

moment when articulation of the next turn is intended to start (Barthel et al. 2016, 2017, Bögels et al. 2015; Levinson & Torreira 2015). Next speakers thus need to focus their planning effort to a rather small time window around the transition space between two turns-at-talk, the so-called “crunch-zone” of conversational turn-taking (Roberts & Levinson 2017). The time pressure of turn planning conceivably increases in multi-party conversations as compared to dialogical situations, as there is more than one potential next speaker who might want to claim the right to produce the next turn by speaking up first when turn transition becomes relevant. Indeed, turn transitions are found to generally be shorter in multi-party conversations than in dialogue, arguably due to the competition between potential next speakers in combination with the “first-starter takes the turn” principle (Holler et al. 2021).

While being a generally well-practiced and mundane task, planning a turn-at-talk in conversation has been shown to take up cognitive resources, e.g. reducing the ability to drive or perform other visuo-motor tasks (Boiteau et al. 2014; Drews et al. 2008; Kubose et al. 2006). Recent studies have explored the use of physiological markers as indicators of cognitive processing load during speech production and utterance planning (Bögels 2020; Bögels & Levinson 2017; Bögels et al. 2015; Rühlemann & Barthel 2024; Papesh & Goldinger 2012, Sauppe 2017, Sevilla et al. 2014). A very promising one of these markers is pupil size. While pupil size is affected by a number of factors, both external to the mind and nervous system of the speaker, like lighting conditions, as well as internal, like drug consumption, pathological states, and emotional arousal (Bradley et al. 2008, Laeng et al. 2012, Mathôt 2018), it has repeatedly been shown to be a reliable indicator of processing load in a number of different cognitive tasks such as arithmetic computation (Hess & Polt 1964, van der Wel & Steenbergen 2018), face recognition (Wu et al. 2012), as well as language processing tasks (Engelhardt et al. 2010, Just & Carpenter 1993, Koch & Janse 2016, Schmidtke 2014, Tromp et al. 2016, Zekveld et al. 2010, see also Kahneman & Beatty 1966, Beatty 1982, Sirois & Brisson 2014).

In the area of speech production research, pupil size changes have first been studied in monological settings in highly restrictive experimental tasks. These studies find speakers’ pupil dilations to increase with language planning tasks that are increasingly difficult, for instance due to non-canonical word order or infrequent semantic role assignment (Sauppe 2017, Sevilla et al. 2014) or due to the production of infrequent words (Papesh & Goldinger 2012). More recent studies also successfully applied pupil size analyses to interactive tasks such as task-oriented dialogue and free conversation. Based on large corpora of naturalistic conversation, Rühlemann and Barthel (2024) find speakers’ pupil size to be correlated to word frequency patterns in turns-at-talk. As the frequency of words used within a turn decreases, speakers’ pupil size increases, and the rate of decline of

word frequencies is predictive of the rate of increase of pupil size, arguably due to the increasing mental effort to retrieve and encode less frequent words (Levelt et al. 1999, Jescheniak & Levelt 1994, Indefrey & Levelt 2004). In an experimental dialogue study, Barthel & Sauppe (2019) used a collaborative list-completion task to measure the effects of speech planning at turn transitions on speakers' pupil size. Using different sentence types, they manipulated at what point in time next speakers were able to (and did) start planning their upcoming turn, as was observable in the timing of participants' gaze movements for speech planning. The authors found that next speakers' pupils started to dilate earlier or later depending on the time at which they started to plan their turn. Moreover, they found the rate of pupil size increase to depend on the difficulty of the planning task, with intensified and prolonged pupil responses in more difficult planning conditions (planning in overlap, including interference with incoming speech) as compared to easier planning conditions (planning in silence, without interfering incoming speech). These results show that pupillometry is a promising candidate for a reliable measure of processing load during interactive language use.

We build on these experimental and naturalistic corpus findings and use changes in pupil size as a proxy for processing load in the vicinity of turn transitions in natural conversational settings. The analysed conversations were recorded as part of the Freiburg Multimodal Interaction Corpus (FreMIC, see Rühlemann & Ptak 2023). We examined eleven conversations with three conversational interactants each. Studying triadic conversations, we can treat a highly unrestricted interactive situation as a natural experiment, since any incidence of interest (a turn-transition with a new speaker taking a next turn) contains measures in a critical condition (of a current listener who becomes the next speaker, planning a turn) as well as in a control condition (of a current listener who stays in the listener role during the next turn). These situations in triadic interaction allow us to estimate the effects of speech planning on mental processing load (operationalised as changes in pupil size) against a baseline condition that features the same input in connection with the absence of speech planning.

The analysed interactional sequences contain question-response pairs in which the current speaker asks an information-seeking question and the next speaker answers that question in the next turn. Question-answer sequences are an ideal site to start testing the effects of speech planning in natural conversation for two reasons. Firstly, questions make a response normatively relevant in the next turn (Stivers & Rossano 2010), making the upcoming sequential move after turn transition strongly projectable (Auer 2021a, 2021b). And secondly, the turn-timing of question-answer sequences has been found to be representative of turn-timing in conversation in general (Stivers et al. 2009). Restricting our analysis to question-answer sequences thus both yields a representative sample of turn transitions contained in the corpus

and at the same time allows us to increase homogeneity within the analysed cases while keeping data processing efforts for this study within manageable bounds.

We analyse the extracted sequences to investigate whether answering question recipients show increased processing load at turn transitions as compared to not-answering question recipients. If so, answerers should show an increase in pupil size relative to not-answerers while they are planning their next turn.

2. Methods

2.1 Data collection

The study draws on data from the Freiburg Multimodal Interaction Corpus (FreMIC; Rühlemann & Ptak 2023). FreMIC is a multimodal corpus of unscripted dyadic and triadic conversational interactions. The interactions are transcribed both orthographically as well as conversation-analytically (e.g. Jefferson 2004) in ELAN (v6.7, The Language Archive 2023). Annotations are built on inter-pausal units (IPUs) and separated whenever a speaker pauses for more than 180 ms. This threshold reflects the threshold for detection of acoustic silences in humans, which lies between 120 ms and 200 ms (Walker & Trimboli 1984) and has been used in a number of previous studies (e.g. Heldner 2011, Heldner & Edlund 2010, Levinson & Torreira 2015, Roberts et al. 2015). The onsets and offsets of the IPUs were determined through inspection of waveforms and spectrograms using Praat (v5.3.56, Boersma & Weenik 2015). In FreMIC, pupil size measurements (and other non-verbal observations) are synchronized with these IPUs. For the present analyses, questions consisting of more than one IPU as well as the pupil size values associated with them were concatenated into larger strings. The concatenation of question IPUs is exemplified in Extract (1).²

- (1) a. 1 ID01.C what's the big city in:
2 (0.519)
3 ID01.C South Carolina=
4 (0.018)
5 ID01.A =(v: laughs)) w(h)e don't have one=
b.³ 1 ID01.C what's the big city in: { (0.519) { South Carolina=
2 (0.018)
3 ID01.A =(v: laughs)) w(h)e don't have one=

2. See Appendix for transcription conventions.

3. Curly brackets indicate IPU boundaries.

In (1a), speaker C's question *what's the big city in:* in line 1 clearly has not yet reached its completion and is therefore not yet transition-ready. The completion is achieved only after the turn-internal pause in line 2 and the addition of the local reference *South Carolina*= in line 3. The question turn as a whole is represented in (1b).

For the recording sessions, participants were seated in an equilateral triangle, facing each other (F-formation; Kendon 1967). Prior to the start of the recording, participants were equipped with Dikablis Glasses 3 eye-tracking glasses by Ergoneers. These eye-tracking glasses, after being calibrated by the experimenter, recorded both the participants' visual field as well as the direction of their gaze (see Figure 1) and their pupil size (in 60 Hz).⁴ Participants' audio was recorded with a scene microphone positioned in the centre between the participants. Participants were given the freedom to converse on any topics they wanted to for approximately 30–45 minutes until the recording was stopped. Written informed consent was obtained from all participants to publish transcripts and images of the recorded interactions. Therefore, ethical approval did not need to be obtained from the ethics committee at Freiburg University.

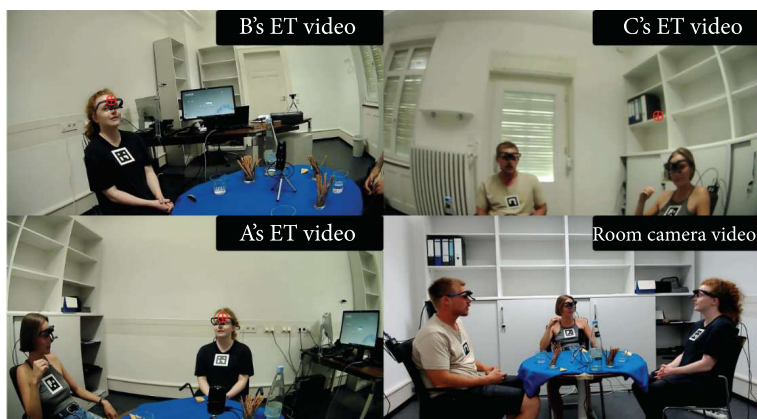


Figure 1. Still taken from a split-screen video of a tradic conversation in the FreMIC corpus. Three tiles show participants' eye-tracking (ET) video (top left, top right, and bottom left), one tile shows the room camera perspective (bottom right).

Red cursors on participants' eye-tracking videos indicate participants' gaze direction

4. Pupillometric data were collected during unrestricted, informal conversation with limited control over lighting conditions, which may lead to noise in the pupil data. Despite this potential noise, the data are robust enough to allow for the identification of differential patterns of pupil size development in listeners of questions, see Section 3 Results.

The eye-tracking and pupil data are integrated with the conversational data and the corresponding transcriptions into a comprehensive corpus structure in which all data are immediately available for inspection and analysis (cf. Rühlemann & Ptak 2023). The data selected from the corpus for the present study are based on the question-answer (QA) sequences occurring in the 11 triadic conversations that have been fully transcribed and annotated to date, covering a total of over 9 hours of conversation.⁵ This set of sequences was developed as part of an ongoing research project⁶ and is restricted to information-seeking questions (instead of, for example, tag questions, ironic questions, rhetorical questions, self-directed questions, or repair-initiating questions) and to the presence of a type-fitting answer given in the next turn (instead of, for example, an unanswered question or a comment on the question).

This original pool of QA sequences consisted of 360 sequences. For the present study, sequences were excluded if the response was exclusively non-verbal or if pupil size measurements were missing for one of the question recipients.⁷ Given our aim to compare answerers vs. not-answerers, sequences were also discarded if both question recipients co-constructed one answer or produced answers in overlap, as in Extracts (2)–(3).⁸

- (2) 1 ID08.A so what's your work
 2 (0.905)
 3 ID08.B we:=
 4 (0.060)
 5 ID08.C =we both work at Fraunhofer

In (2), participant A is asking *so what's your work* while gaze-selecting recipient B (Auer 2021a). That recipient, however, lowers his gaze and only after a long silence in line 2 starts to answer in line 3. Participant C, who is a work colleague of B's and therefore equally eligible to answer, takes over providing the co-constructed answer *we both work at Fraunhofer*.

5. The files include: Fo1, Fo4, Fo7, Fo8, F12, F16, F18, F19, F20, F22, and F23.

6. DFG grant number 497779797; cf. <https://gepris.dfg.de/gepris/projekt/497779797>

7. File F20 was thus excluded, due to a technical failure in pupil size recording in one of the participants.

8. While in the vast majority of cases, questions are addressed to one recipient specifically, sequential environments that license both recipients to produce answers include what are called collective and distributive questions (cf. Auer 2021a). Using question type as an additional factor in the statistical analysis was beyond the scope of the present study, but see Rühlemann (2024) and Rühlemann and Barthel (under revision).

- (3) 1 ID04.A tch but !why:!! eye trackers >like I don't get it< (.) why:
 2 (0.382)
 3 ID04.B [cos it's RELATE- RELATED to your cognition]
 4 ID04.C [cause they wanna see where] }
 5 (0.315) } like your language processing

In (3), participant A is inquiring about the reason they are wearing eyetrackers in the recording in line 1. Since all three participants are wearing the device, the question does not select a particular participant as answerer (Hayashi 2012). Both participants are eligible to answer, which they do with their respective answers in extended overlap in lines 3–4.

The remaining sequences contain an information-seeking question that is responded to with an answer in next position. Information-seeking questions can be considered prototypical questions in that they exhibit the clearest knowledge-asymmetry (Heritage 1984: 250). A question such as *What's the time?* indicates a complete knowledge gap on the part of the questioner, whereas in a polar question such as *Is it seven?*, the knowledge gap is tentatively filled and the questioner merely seeks confirmation or dis-confirmation of the underlying proposition *It is seven*. Consider the following examples of QA-sequences in the collection underlying this study in Extracts (4) and (5):

- (4) 1 ID01.A >like I don't understand< sorry like how old's your mom?
 2 (0.855)
 3 ID01.C eh six:ty::-one=
 (5) 1 ID07.B what does steam punk mean [°exactly°]?
 2 ID07.A [steam punk] is:: a
 3 (.) >a sort of< Art !style but it's also (.)
 4 it has its own scene like a medieval scene,=

In Extract (4), subject A asks a *wh*-question in line 1, asking about the recipient's mother's age. After a longish gap in line 2, subject C gives her mother's age in next position in line 3. In Extract (5), subject B's question about the meaning of "steam punk" in line 1 terminally overlaps with subject A's extended answer in lines 2–4.

A total of 328 QA-sequences were thus selected from 22 individual participants (questioners and question recipients) for this study.

2.2 Data Pre-processing and statistical analysis

For each QA-sequence, the time lines of the talk and pupil size measurements of all participants were synchronised. Given our aim to compare pupil size developments of answerers and not-answerers, pupil size measurements were subsetting to

the two recipients of the question in each sequence. Pupil sizes were averaged over both pupils for each recipient (cf. Barthel & Sauppe 2019). During blinks (where pupil size cannot be measured), pupillometric values were linearly interpolated.

It has been shown that next speakers can start to plan their next turn already before the end of the incoming turn by their interlocutor (Barthel et al. 2016, 2017, Barthel & Levinson 2020; Bögels 2020, Bögels et al. 2015). The reason why next speakers engage in planning in overlap is likely that early planning helps them to meet the time challenge of rapid turn taking, especially in multi-party conversation (Sacks et al. 1974, Holler et al. 2021). In order to be able to compare pupil size changes of answerers with those of not-answerers during the time window in which answerers are most likely to plan their response, pupil measurements were time-locked to the offset of the question and modelled from 600 ms before question offset until answer onset. Thus, sequences in which the answer was initiated in more than 600 ms overlap with the end of the question were removed from the present analysis.

The final analysis was carried out on a total of 291 QA-sequences extracted from 10 triadic conversations with a total of 21 individual question recipients.

A mixed effects regression model was built using the *R*-package *lme4* (Bates et al. 2015) to model participants' pupil sizes. Participant Role (answerer vs. not-answerer) was included as a dummy coded fixed effect, with Answerer as the reference level. Orthogonal polynomial predictors of Time (from 600 ms before question offset until answer onset) were computed using the *poly()* function from the *R*-package *stats* and added as fixed effects to the model, together with their interactions with Participant Role. Random intercepts as well as random slopes by Role were added by ParticipantID. Additionally, random intercepts were added by ConversationID. Statistical significance of single predictors and interactions in the form of *p*-values were obtained with the *R*-package *lmerTest* (Kuznetsova et al. 2017). After checking that the linear Time term (modelling a linear development of pupil size over time) as well as its interaction with Participant Role was significant in a model without a higher-order Time term, a quadratic Time term (modelling a curved (accelerated) trend in pupil size over time) was added to the model. Since the quadratic term as well as its interaction with Role was also significant, a cubic Time term was added to the model. Since neither the cubic term nor its interaction with Role was significant, we will present and analyse the quadratic model in Section 3.

3. Results

Pupil sizes of answerers and not-answerers of questions have been modelled across time from 600 ms before question offset until answer onset in a mixed effects regression model with Participant Role (answerer vs. not-answerer) as a predictor (see *Section 2.2 Data Pre-processing and Statistical Analysis*). Table 1 shows the model output. The linear Time term is significant, indicating that answerers’ pupil size tends to increase during the time window of analysis. The interaction of Role and the linear Time term is found to be significant as well, indicating that across the analysed time span, answerers’ pupil size increases more than not-answerers’ pupil size. Additionally, the model shows the quadratic Time term to be significant as well, indicating that the increase in pupil size in answerers accelerates across the time window of analysis. The interaction of Role and the quadratic Time term is also found to be significant, indicating that the increase in pupil dilation accelerates more in answerers than in not-answerers. Taken together, the model shows that answerers’ pupils dilate more than not-answerers’ pupils in the vicinity of turn transitions in the examined QA-sequences (see Figure 2).

Table 1. Output of mixed effects regression model on pupil sizes in question recipients of selected question-answer sequences. Pupil size data are time locked to question offset. Formula = pupilSize ~ poly(Time, degree = 2) * Role + (1 + Role | ParticipantID) + (1 | ConversationID)

Fixed effects	β	SE	<i>t</i>	<i>p</i>
Intercept	1159.76	135.61	8.552	
Time	4092.38	737.23	5.551	< .001 ***
Time ²	6569.55	716.05	9.175	< .001 ***
Role_notAnswerer	-56.95	22.95	-2.481	.024 *
Time : Role_notAnswerer	-3456.88	1033.19	-3.346	< .001 ***
Time ² : Role_notAnswerer	-5947.62	1011.69	-5.879	< .001 ***

4. Discussion

This study tested whether answerers to a question in multiparty conversations show increased processing load as compared to not-answering recipients of the same question due to their preparation of the response. Using changes in pupil size as an indicator for processing load, we found that answerers indeed show a greater

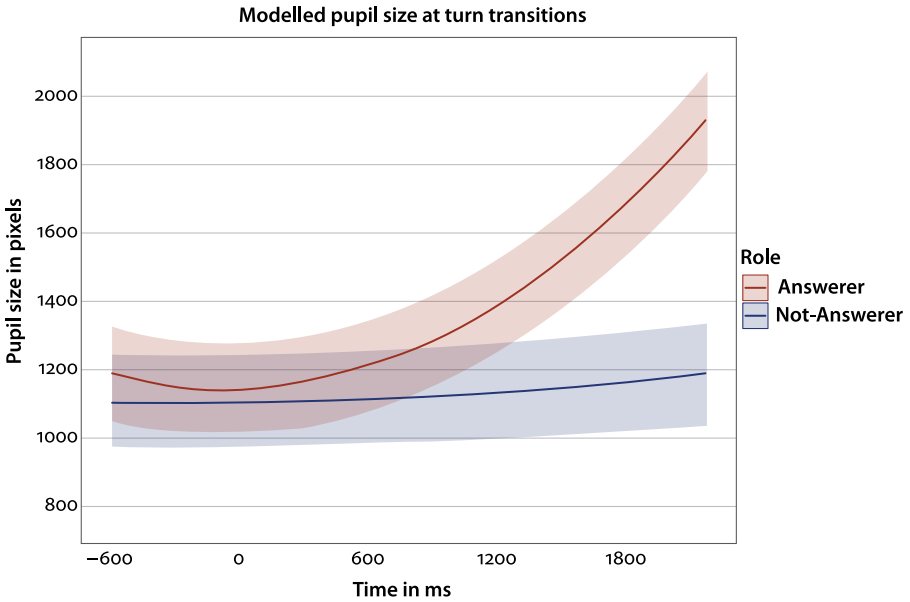


Figure 2. Modelled pupil sizes in question recipients in triadic question-answer sequences time-locked to the offset of the question. Shaded areas indicate one standard error of the mean

increase in pupil size in the vicinity of turn-transitions than not-answerers, likely due to the planning and preparation of a response to the question in answerers as compared to no (or at least reduced) planning in not-answerers, who show no or very little increase in pupil size before the answer to the question. These findings are in line with earlier pupillometric research showing that pupil size of next speakers increases when speech planning begins at turn transitions, an increase that is more intense in situations of higher processing load due to speech planning (Barthel & Sauppe 2019).

In contrast to previous studies using controlled experimental conditions, our investigation delves into real-world conversations characterized by unrestricted speech that interlocutors produce without any reference to an experimental task (see also Oben et al. this volume). Utilizing data extracted from the FreMIC corpus, this study contributes insights into the cognitive demands associated with turn-taking in naturalistic conversational settings, increasing the ecological validity of the available analyses on the processes underlying speech planning in interaction. To the best of our knowledge, pupillometric measurements have not been used for the investigation of processing effort in non-scripted conversational interaction before, even though pupillometry provides a non-invasive and comparatively unrestrictive method of measurement and is thus well suited for naturalistic

interaction research that seeks to increase the ecological validity of the obtained data.⁹ The presented findings show that pupil size can be used in the analysis of unrestricted multiparty conversation as an indicator of processing load at turn transitions.

The turn transition space is a crunch zone of speech processing, with comprehension and planning running in parallel, which can lead to peaks in (verbal) processing load (Roberts & Levinson 2017, Barthel 2020, 2021). While previous investigations into processing load at turn transitions show that processing load in next speakers increases in the vicinity of turn transitions (Barthel & Sauppe 2019, Boiteau et al. 2014), these studies, by design, could not investigate the relative contributions of speech comprehension and speech planning to the observable increase in processing load, as they lack a matched control condition in which speech planning is absent. Since in the present study answerers' and not-answerers' pupil sizes are measured in the same situations and in response to the same questions, this control condition is naturally present and available for comparison. The finding that listeners who do not answer the question in the next turn do not show an increase in processing load that is comparable to recipients who answer the question as next speakers supports the hypothesis that speech planning is the main cause of the local spike in cognitive effort that has been observed in the present study as well as in earlier studies. Arguably, next speakers seem to prioritize speech production processes in their allocation of cognitive resources as the incoming turn is approaching completion (Barthel 2021), possibly due to the time pressure inherent in the conversational turn-taking system (Sacks et al. 1974). This time pressure might conceivably be even more pronounced in multiparty conversations as compared to dyadic interactions, while for question-answer sequences in particular, no difference in turn-transition times has been found between dyadic and triadic conversations (Holler et al. 2021).

A limitation to the present study is that it does not account for who is being selected by the current speaker as the answerer and how that selection might impact speech planning processes in the selected and the not-selected participant. A common practice used by current speakers to select a particular next speaker is gazing (Sacks et al. 1974, Lerner 2003). In roughly a fourth of German *ihr*-questions (2nd person plural questions) analysed in Auer (2021a), the last-gazed-at recipient was not the answerer. Such cases can readily be found in our data as well. Consider Extract (6):

9. The only drawback remaining for now being that participants do have to wear the eye-tracking glasses and need to be calibrated before recording.

- (6) 1 ID07.C isn't the Mundenhof by the Dreisam?
 2 C-gaze -----A---|...|-----B---->
 3 (0.050)
 4 ID07.A it's not [fa:r from the Dreisam]=
 5 ID07.C [fairly *close*]

The participants are talking about a location called “Mundenhof” near Freiburg. Speaker C asks *isn't the Mundenhof by the Dreisam?* in line 1 (the Dreisam being the river in the Freiburg area), a collective question that both recipients (as residents of Freiburg) are eligible to answer. As indicated in line 2, C gazes towards participant A at question onset and then shifts toward participant B before the completion of the question.¹⁰ The answer to the question, however, is produced in line 4 by participant A, who was not gazed at turn-finally. If we accept the notion that gaze is effectively used as a next-speaker selection method in 3 out of 4 cases, as shown by Auer (2021a), we would expect to see increases in pupil size in both the (not gaze-selected but self-selecting) answerer as well as the (gaze-selected) not-answerer, as in these cases the not-answering recipient is likely to also start planning a response. Given that the number of such selection mismatches (current speaker selects one recipient as next-speaker but the other recipient self-selects) is by no means negligible, future investigations of speech planning in conversation should factor in next-speaker selection as an additional key variable.

5. Conclusion

In sum, this study has taken a pathway to examining speech planning that is novel in two ways. Firstly, it has used pupil size measurements collected not in experimental settings but in naturalistic conversation, and secondly, it has examined speech planning not in dyadic but in triadic interaction.

Our findings show that pupil size can be used as a reliable indicator of processing load during speech planning in natural triadic conversation. While pupil size might potentially be influenced by a number of factors at turn transitions, including increased attention, memory retrieval of relevant information to be encoded in the response, processes of response formulation, preparation for articulation, or a combination of any of these, the presented comparison of answerers and not-answerers shows that (aspects of) speech planning are the most likely explanatory factor for the attested increase in processing load at turn transitions.

10. Gaze alternation is typical of collective and distributive questions like the one in Excerpt (6) (Rühlemann 2024).

The findings of this study may have implications not only for our understanding of turn-taking dynamics but also for the broader understanding of how cognitive processes shape and are shaped by complex communicative interactions. Through this exploration, we aspire to bridge the gap between controlled experimental settings and the intricacies of everyday conversation, offering a nuanced understanding of the cognitive demands inherent in (multiparty) interactions, as the presented methodology can serve as a road map for future mixed-methods analyses of unrestricted social interaction.

References

-  Auer, Peter. 2021a. “Gaze Selects the next Speaker in Answers to Questions Pronominally Addressed to More than One Co-Participant.” *Interactional Linguistics* 1 (2): 154–82.
-  Auer, Peter. 2021b. “Turn-Allocation and Gaze: A Multimodal Revision of the ‘Current-Speaker-Selects-next’ Rule of the Turn-Taking System of Conversation Analysis.” *Discourse Studies* 23 (2): 117–40.
- Barthel, Mathias. 2020. “Speech Planning in Dialogue — Psycholinguistic Studies of the Timing of Turn Taking.” PhD Thesis, Nijmegen: Radboud University Nijmegen. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3185609
- Barthel, Mathias. 2021. “Speech Planning Interferes with Language Comprehension: Evidence from Semantic Illusions in Question-Response Sequences.” In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue*, 1–14. Potsdam, Germany.
-  Barthel, Mathias, and Stephen C. Levinson. 2020. “Next Speakers Plan Word Forms in Overlap with the Incoming Turn: Evidence from Gaze-Contingent Switch Task Performance.” *Language, Cognition and Neuroscience* 35 (9): 1183–1202.
-  Barthel, Mathias, Antje S. Meyer, and Stephen C. Levinson. 2017. “Next Speakers Plan Their Turn Early and Speak after Turn-Final ‘Go-Signals.’” *Frontiers in Psychology* 8: 393.
-  Barthel, Mathias, and Sebastian Sauppe. 2019. “Speech Planning at Turn Transitions in Dialog Is Associated with Increased Processing Load.” *Cognitive Science* 43 (7): e12768.
-  Barthel, Mathias, Sebastian Sauppe, Stephen C. Levinson, and Antje S. Meyer. 2016. “The Timing of Utterance Planning in Task-Oriented Dialogue: Evidence from a Novel List-Completion Paradigm.” *Frontiers in Psychology* 7: 1858.
-  Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using Lme4.” *Journal of Statistical Software* 67 (1).
-  Beatty, Jackson. 1982. “Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources.” *Psychological Bulletin* 91 (2): 276–92.
- Boersma, Paul, and David Weenink. 2015. *Praat: Doing Phonetics by Computer [Computer Program]*. Version 5.3.56, <http://www.praat.org/>
-  Bögels, Sara. 2020. “Neural Correlates of Turn-Taking in the Wild: Response Planning Starts Early in Free Interviews.” *Cognition* 203 (October): 104347.

- doi Bögels, Sara, and Stephen C. Levinson. 2017. "The Brain Behind the Response: Insights Into Turn-Taking in Conversation From Neuroimaging." *Research on Language and Social Interaction* 50 (1): 71–89.
- doi Bögels, Sara, Lilla Magyari, and Stephen C. Levinson. 2015. "Neural Signatures of Response Planning Occur Midway through an Incoming Question in Conversation." *Scientific Reports* 5 (12881): 1–11.
- doi Boiteau, Timothy W., Patrick S. Malone, Sara A. Peters, and Amit Almor. 2014. "Interference between Conversation and a Concurrent Visuomotor Task." *Journal of Experimental Psychology: General* 143 (1): 295–311.
- doi Bradley, Margaret M., Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. 2008. "The Pupil as a Measure of Emotional Arousal and Autonomic Activation." *Psychophysiology* 45 (4): 602–7.
- doi Drews, Frank A., Monisha Pasupathi, and David L. Strayer. 2008. "Passenger and Cell Phone Conversations in Simulated Driving." *Journal of Experimental Psychology: Applied* 14 (4): 392–400.
- doi Engelhardt, Paul E., Fernanda Ferreira, and Elena G. Patsenko. 2010. "Pupillometry Reveals Processing Load during Spoken Language Comprehension." *Quarterly Journal of Experimental Psychology* 63 (4): 639–45.
- Ford, Cecilia E., Barbara A. Fox, and Sandra A. Thompson. 1996. "Practices in the Construction of Turns: The 'TCU' Revisited." *Pragmatics* 6 (3): 427–54.
- doi Ford, Cecilia E., and Sandra A. Thompson. 1996. "Interactional Units in Conversation: Syntactic, Intonational, and Pragmatic Recourses for the Management of Turns." In *Interaction and Grammar*, edited by Elinor Ochs, Emanuel A. Schegloff, and Sandra A. Thompson, 134–84. Cambridge: Cambridge University Press.
- doi Fox Tree, Jean E. 2002. "Interpreting Pauses and Ums at Turn Exchanges." *Discourse Processes* 34 (1): 37–55.
- doi Hayashi, Makoto. 2012. "Turn Allocation and Turn Sharing." In *The Handbook of Conversation Analysis*, edited by Jack Sidnell and Tanya Stivers, 1st ed., 167–90. Wiley.
- doi Heldner, Mattias. 2011. "Detection Thresholds for Gaps, Overlaps, and No-Gap-No-Overlaps." *The Journal of the Acoustical Society of America* 130 (1): 508–13.
- doi Heldner, Mattias, and Jens Edlund. 2010. "Pauses, Gaps and Overlaps in Conversations." *Journal of Phonetics* 38 (4): 555–68.
- Henetz, Tania. 2017. "Don't Hesitate! The Length of Inter-Turn Gaps Influences Observers' Interactional Attributions." PhD Thesis, Stanford: Stanford University.
- Heritage, John. 1984. *Garfinkel and Ethnomethodology*. Repr. Cambridge: Polity Press.
- doi Hess, E. H., and J. M. Polt. 1964. "Pupil Size in Relation to Mental Activity during Simple Problem-Solving." *Science* 143 (3611): 1190–92.
- doi Holler, Judith, Phillip M. Alday, Caitlin Decuyper, Mareike Geiger, Kobin H. Kendrick, and Antje S. Meyer. 2021. "Competition Reduces Response Times in Multiparty Conversation." *Frontiers in Psychology* 12: 693124.
- doi Indefrey, Peter, and Willem J. M. Levelt. 2004. "The Spatial and Temporal Signatures of Word Production Components." *Cognition* 92 (1–2): 101–44.
- doi Jefferson, Gail. 2004. "Glossary of Transcript Symbols with an Introduction." In *Pragmatics & Beyond New Series*, edited by Gene H. Lerner, 125:13–31. Amsterdam: John Benjamins Publishing Company.

- doi Jescheniak, Jörg D., and Willem J. M. Levelt. 1994. "Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20 (4): 824–43.
- doi Just, Marcel A., and Patricia A. Carpenter. 1993. "The Intensity Dimension of Thought: Pupillometric Indices of Sentence Processing." *Canadian Journal of Experimental Psychology* 47 (2): 310–39.
- doi Kahneman, Daniel, and Jackson Beatty. 1966. "Pupil Diameter and Load on Memory." *Science* 154 (3756): 1583–85.
- doi Kendon, Adam. 1967. "Some Functions of Gaze-Direction in Social Interaction." *Acta Psychologica* 26 (January): 22–63.
- doi Koch, Xaver, and Esther Janse. 2016. "Speech Rate Effects on the Processing of Conversational Speech across the Adult Life Span." *The Journal of the Acoustical Society of America* 139 (4): 1618–36.
- doi Kubose, Tate T., Kathryn Bock, Gary S. Dell, Susan M. Garnsey, Arthur F. Kramer, and Jeff Mayhugh. 2006. "The Effects of Speech Production and Speech Comprehension on Simulated Driving Performance." *Applied Cognitive Psychology* 20 (1): 43–63.
- doi Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82 (13).
- doi Laeng, Bruno, Sylvain Sirois, and Gustaf Gredebäck. 2012. "Pupillometry: A Window to the Preconscious?" *Perspectives on Psychological Science* 7 (1): 18–27.
- doi Lerner, Gene H. 2003. "Selecting next Speaker: The Context-Sensitive Operation of a Context-Free Organization." *Language in Society* 32 (2): 177–201.
- doi Levelt, Willem J. M., Ardi Roelofs, and Antje S. Meyer. 1999. "A Theory of Lexical Access in Speech Production." *Behavioural and Brain Sciences* 22 (01): 1–75.
- doi Levinson, Stephen C., and Francisco Torreira. 2015. "Timing in Turn-Taking and Its Implications for Processing Models of Language." *Frontiers in Psychology* 6 (731): 10–26.
- doi Mathôt, Sebastiaan. 2018. "Pupillometry: Psychology, Physiology, and Function." *Journal of Cognition* 1 (1): 16.
- doi Papesh, Megan H., and Stephen D. Goldinger. 2012. "Pupil-BLASH-Metry: Cognitive Effort in Speech Planning Reflected by Pupil Dilation." *Attention, Perception, & Psychophysics* 74 (4): 754–65.
- doi Roberts, Felicia, and Alexander L. Francis. 2013. "Identifying a Temporal Threshold of Tolerance for Silent Gaps after Requests." *The Journal of the Acoustical Society of America* 133 (6): EL471–77.
- doi Roberts, Felicia, Alexander L. Francis, and Melanie Morgan. 2006. "The Interaction of Inter-Turn Silence with Prosodic Cues in Listener Perceptions of 'Trouble' in Conversation." *Speech Communication* 48 (9): 1079–93.
- doi Roberts, Seán G., and Stephen C. Levinson. 2017. "Conversation, Cognition and Cultural Evolution: A Model of the Cultural Evolution of Word Order through Pressures Imposed from Turn Taking in Conversation." *Interaction Studies* 18 (3): 402–42.
- doi Roberts, Seán G., Francisco Torreira, and Stephen C. Levinson. 2015. "The Effects of Processing and Sequence Organization on the Timing of Turn Taking: A Corpus Study." *Frontiers in Psychology* 6.

- doi Rühlemann, Christoph. 2024. "Gaze alternation predicts inclusive next-speaker selection: Evidence from eyetracking." *Frontiers in Communication* (Sec. Multimodality of Communication) 9:1396925.
- doi Rühlemann, Christoph, and Mathias Barthel. 2024. "Word frequency and cognitive effort in turns-at-talk: Turn structure affects processing load in natural conversation." *Frontiers in Psychology* (Sec. Psychology of Language) 15.
- Rühlemann, Christoph, and Mathias Barthel. Under revision. "Speech planning depends on next-speaker selection: Evidence from pupillometry in question-answer sequences in naturalistic triadic conversation."
- doi Rühlemann, Christoph, and Alexander Ptak. 2023. "Reaching beneath the Tip of the Iceberg: A Guide to the Freiburg Multimodal Interaction Corpus." *Open Linguistics* 9 (1).
- doi Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. "A Simple Systematics for the Organization of Turn-Taking for Conversation." *Language* 50 (4): 696–735.
- doi Sauppe, Sebastian. 2017. "Symmetrical and Asymmetrical Voice Systems and Processing Load: Pupillometric Evidence from Sentence Production in Tagalog and German." *Language* 93 (2): 288–313.
- doi Schmidtke, Jens. 2014. "Second Language Experience Modulates Word Retrieval Effort in Bilinguals: Evidence from Pupillometry." *Frontiers in Psychology* 5: 137.
- doi Selting, Margret. 2000. "The Construction of Units in Conversational Talk." *Language in Society* 29: 477–517.
- doi Sevilla, Yamila, Mora Maldonado, and Diego E. Shalóm. 2014. "Pupillary Dynamics Reveal Computational Cost in Sentence Planning." *Quarterly Journal of Experimental Psychology* 67 (6): 1041–52.
- doi Sirois, Sylvain, and Julie Brisson. 2014. "Pupillometry." *Wiley Interdisciplinary Reviews: Cognitive Science* 5 (6): 679–92.
- doi Stivers, Tanja, Nick J. Enfield, Penelope Brown, C. Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, et al. 2009. "Universals and Cultural Variation in Turn-Taking in Conversation." *Proceedings of the National Academy of Sciences* 106 (26): 10587–92.
- doi Stivers, Tanya, and Federico Rossano. 2010. "Mobilizing Response." *Research on Language & Social Interaction* 43 (1): 3–31.
- The Language Archive. 2023. *ELAN* (version 6.7). Nijmegen: Max Planck Institute for Psycholinguistics. <http://www.lat-mpi.eu/tools/elan/>
- doi Tromp, Johanne, Peter Hagoort, and Antje S. Meyer. 2016. "Pupillometry Reveals Increased Pupil Size during Indirect Request Comprehension." *Quarterly Journal of Experimental Psychology* 69 (6): 1093–1108.
- doi Walker, M. B., and C. Trimboli. 1984. "The Role of Nonverbal Signals in Co-Ordinating Speaking Turns." *Journal of Language and Social Psychology* 3 (4): 257–72.
- doi Wel, Pauline van der, and Henk van Steenbergen. 2018. "Pupil Dilation as an Index of Effort in Cognitive Control Tasks: A Review." *Psychonomic Bulletin & Review* 25 (6): 2005–15.
- doi Wu, Esther Xiu Wen, Bruno Laeng, and Svein Magnussen. 2012. "Through the Eyes of the Own-Race Bias: Eye-Tracking and Pupillometry during Face Recognition." *Social Neuroscience* 7 (2): 202–16.

-  Zekveld, Adriana A., Sophia E. Kramer, and Joost M. Festen. 2010. "Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility." *Ear and Hearing* 31 (4): 480–90.
-  Zima, Elisabeth, Clarissa Weiß, and Geert Brône. 2019. "Gaze and Overlap Resolution in Triadic Interactions." *Journal of Pragmatics* 140 (January): 49–69.

Appendix

Transcription conventions underlying the Jeffersonian transcription in the Freiburg Multimodal Interaction Corpus (FreMIC).

Sequential:

overlap: [...]

latching: =... or ...=

Temporal:

decelerated speech: <...>

accelerated speech: >...<

minimal pauses: (.)

Phonological:

intensity:

loud voice: caps, e.g., HEY

quiet voice: °...°

emphasis: !syllable!

vowel stretching: colon, e.g., dra:ft

truncation: dash, e.g., spring-

intonation:

full rise: ?

half rise: ˆ

sentence-like drop: .

continued intonation:,

high pitch: ↑...↑

low pitch: ↓...↓

'scale' upward: /.../↑

'scale' downward: /.../↓

voice quality:

tremulous voice: ≈...≈

creaky voice: ʔ...ʔ

smiley voice: £...£

Laughter:

within-word laughter: laughter pulses, e.g., ok(h)ay

freestanding laughter: as event, e.g., ((v: laughs))

Mobile settings

Laughter and gaze among talkers on a walk

Peter Auer & Barbara Laner

University of Freiburg

While the functions of gaze in (other types of) F-formations have been the focus of a considerable amount of research, the gaze patterns found in side-by-side constellations have remained largely unexplored. In this chapter, we look at a particularly frequent, but also highly complex type of side-by-side constellation, i.e. people walking and talking together. Whereas participants in a circular or vis-à-vis F-formation typically and frequently look at each other during verbal interaction, gazing at the co-walker, and even more so mutual gaze between walkers, is the exception rather than the rule. This exceptional characteristic of gazing at a co-participant raises the question of when and why participants diverge from their usual pattern of gazing forwardly while talking. We zoom in on one of the most recurrent patterns in our data, i.e. speaker-gaze at recipient in the context of a candidate laughable. The study is based on 9 dyadic walks through a national park and uses mobile eye-tracking for recording the walkers' verbal and nonverbal practices.

Keywords: laughter in interaction, laughables, speaker-gaze, walking and talking, mobile interactions, side-by-side F-formation, eye-tracking

1. Gaze patterns in side-by-side constellations

When two or more persons are engaged in a focused interaction in Western culture, they regularly display the fact that they are currently part of a “with” (Goffman 1963) by orienting their bodies to each other in such a way that they create an interactional space (De Stefani & Mondada 2014, Stukenbrock & Dao 2019) in which they “sustain a spatial and orientational relationship” (Kendon 1990:209). In Kendon’s words, they are in a F(acing)-formation (Kendon 1972, 1990 etc.) in which the “interactional segments” in front of them overlap. Kendon distinguishes between vis-à-vis (face-to-face), side-by-side, and L-shaped F-formations; further types have been added (see Figure 1).

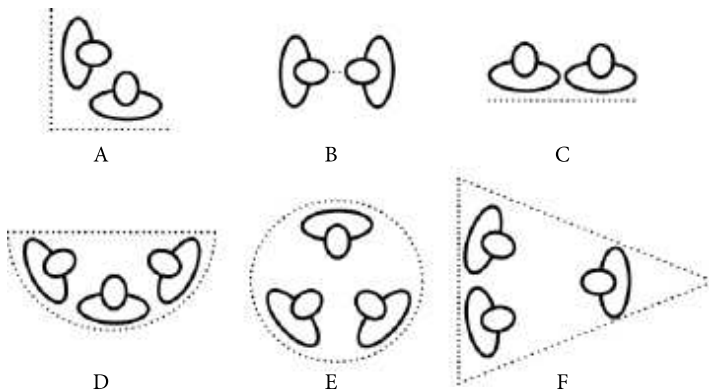


Figure 1. Types of F(acing)-formations: L-shaped, vis-à-vis, side-by-side, semi-circular, circular, triangular (from Tong et al. 2016)

Kendon argues that the participants' overlapping transactional spaces form a so-called o-space; this space is both physically (for gesturing or manipulating) and visually accessible to each participant. For non-participants, this type of arrangement makes it easy to identify the boundaries of the interactional space (cf. Mondada 2009), and they will refrain from interfering with people in an F-formation accordingly. Thus, participants form an interactional "unit of behaviour" (Kendon 1990: 212).

However, already a superficial inspection of the constellations depicted in Figure (1) reveals that there is one type of F-formation to which this description only applies in a restricted fashion at best; this is the side-by-side constellation. It is in fact an open question whether this constellation should be included among the F-formations at all, as the transactional segments of the participants only peripherally overlap in this case. Their bodies are not oriented towards each other, but into the space before them. Unless they turn their heads, the center of their vision (the area of foveal vision, see below) is not directed towards the co-participant. This forward orientation indicates an openness to the social and physical world outside the interactional space which is absent in the other types of F-formations. Note also that side-by-side constellations in Western culture are not as strongly linked to focused interaction as other F-formations are; for instance, strangers may sit or stand close to each other in crowded places (such as subway trains) in a side-by-side constellation without starting a focused interaction, while this may not be possible in a vis-à-vis constellation, proxemics unchanged. Side-by-side constellations also allow more than other F-formations for participants' co-presence in an "open state of talk" (Goffman 1981: 134–5); they seem to lie halfway on the continuum between dispersed seating or standing arrangements without or with very little bodily co-orientation among participants (as

described e.g. by Gardner & Mushin 2015 for Garrwa) and circular or vis-à-vis arrangements which are the least conducive to an “open state of talk”.

Another difference between side-by-side configurations and other F-formations is that the latter require participants to remain more or less stable in one place; moving together in space means that participants have to dissolve the circular, vis-à-vis or L-shaped F-formation and re-arrange themselves (see Ciolek & Kendon 1980 for details). Side-by-side configurations, on the other hand, are perfectly suited for (forward) moving the transition from ‘standing together’ to ‘moving together’ is therefore regularly accompanied by a shift into a side-by-side configuration, and vice versa.¹

Of course, participants in a side-by-side arrangement can turn and look at each other and also establish eye contact. However, they only rarely do. The typical gaze patterns found in a side-by-side dyad and the other dyadic F-formations are therefore very different.

Auer & Zima (2021) investigated the total amount of mutual and unilateral gaze during talk in two stationary dyadic F-formations (vis-à-vis) and in two (walking) side-by-side dyads. In the vis-à-vis constellation, the total time of mutual gaze between the two participants amounted to 55% and 78% of the total speaking time in the two dyads, respectively. When analyzed separately, there was an imbalance between speaker-gaze and recipient-gaze, which supports Kendon’s (1967) finding, later confirmed by Brône et al. (2017), that recipients look more and longer at speakers than vice versa: the total duration of speaker-gaze at recipient was on average 16.8% shorter than recipient-gaze at speaker (see also Stukenbrock & Zima, this volume).

In contrast to these findings on stationary vis-à-vis constellations, mutual gaze in two dyads of walkers occurred in only 9.3% and 11.4% of the speaking time, respectively (i.e., excluding ‘open states of talk’, cf. Auer & Zima 2021: 405). In addition, the quantitative dominance of recipient-gaze over speaker-gaze did not hold here. The recipients looked at the speakers on an average of 19.63% of the total speaking time, which was the same as the average speakers’ gaze duration on their recipients (19.51%).

1. See Broth & Mondada (2013) on walking away; vom Lehn (2013) on moving in a museum from one picture to the next together; Mondada (2014), Stukenbrock & Birkner (2010) and de Stefani & Mondada (2014) on walking and stopping on guided tours. Of course, other constellations are possible as well while walking together (such as face-to-back).

2. Gaze and walking

Although these findings suggest that gaze in (mobile) side-by-side constellations is organized in fundamentally different ways than gaze in (other) F-formations, these gaze patterns have not been analyzed in the interactional literature so far. This is true although the structure of interaction in mobile settings has attracted some attention recently. In particular, research has dealt with the ways in which moving “withs” (walkers, or car occupants) find their way together, navigate in space, and notice objects and events in their changing surroundings (see, among many others, Haddington, Mondada, Nevile (eds) 2014, Goodwin & Goodwin 2012 on car driving; Deppermann 2018 on driving lessons; Stukenbrock & Dao 2019 and De Stefani 2011 on shopping in a (super-) market; Mondada 2017 on guided tours; Laner 2022 and Auer, Laner, Pfeiffer & Botsch 2024 on noticing while walking). In this chapter, we focus on gaze in walking dyads.

Walking together is in itself a complex embodied achievement which requires the meticulous coordination of the walkers’ bodies. They need to monitor and anticipate each other’s locomotive movements and establish a bodily rhythm without which it is impossible to walk in synchrony and maintain a walking group’s “we-ness and “we-body” (Meyer & Wedelstaedt 2017:15, with respect to doing sports together). However, merely monitoring the others’ bodily movements and synchronizing with them is not sufficient; walkers also have to monitor their environment, adjust to its physical structure and other moving agents that might interfere with the walking group’s movements (such as other walkers, cars, animals, etc.; see Ryave & Schenkein 1984, Weilenmann, Normaker & Laurier 2013, Mondada 2014, Steger 2019). All these bodily movements and adjustments require visual attention and cognitive resources (cf. Mayor & Bangerter 2013).

People walk together for various reasons. They may walk for pleasure (as in the case of the forest hikes investigated in this chapter), sometimes not even with a specific destination in mind, or in order to transfer as quickly as possible from one location to another. They may walk with the intention of visiting a specific location or to see things on the way. They may walk over short or long distances. They may know the way (routine walks) or may be in need of finding a new way (which requires conscious way-finding activities). But in all cases, silent walking is the exception and at least some talking is the rule. This talk may be “situated”, linked in some way or other to the current surroundings (for wayfinding or the inspection of objects/events in these surroundings), or it may be “displaced”, dealing with referents not sensually accessible in the situation (Auer 1988). Walking and talking thereby becomes a specific type of “multiactivity” (Mondada 2016), in which the activity of walking often recedes into the background. The verbal exchanges need to be adapted to the coordinated bodily activities which may

limit, disturb or even interrupt the ongoing talk (cf. Relieu 1999). Situated and displaced talk may interfere with each other as well, with situated talk taking priority.

In this chapter, we investigate co-walkers' gaze patterns when the referents they discuss are not visually accessible in their physical surroundings, as we are interested in participants' gaze at each other. Deictic practices require different gaze patterns which include gazing at the referential objects (Laner 2022 & 2025, Stukenbrock 2018, 2020).

When co-walkers neither focus on objects in the surroundings nor direct their gaze at the co-participant, their unmarked gaze behaviour while listening to the other seems to be looking either into the distance or at the ground, slightly lowering their heads (see Figure 2, right walker).



Figure 2. Gaze of two walkers during displaced speech. The view of the right walker is shown on the screenshot from the scene camera on the right, the view of the left walker is shown on the left. In this moment the walker on the left is producing a laughable while looking at the recipient who is walking on the right side

Apart from the fact that such downward-looking may alert the walkers to obstacles on the ground (which is particularly relevant in nature walks), this gaze pattern seems to be particularly suited for displaying to the co-walker that they are fully attentive recipients, not ‘distracted’ by the environment. In contrast, head (and gaze) movements to the side, particularly away from the co-walker, may signal that the looker has perceived something noticeable in the surroundings which might have attracted their attention. This may foreshadow a transition into situated talk (Botsch et al., in this volume).

When looking ahead, co-walkers' fields of vision can be visualized as in Figure 3. The human eye enables us to extend our vision to appr. 100° to the left and right of the visual center line (Zhisheng et al. 2019). However, the extreme 10° of the peripheral field of vision (the “monocular vision area”) is mainly restricted to the perception of movements. The “focus recognition area” (foveal vision) only extends appr. 20° to each side of the center line. Eye movements without head movements occur in an area of maximally 60° to both sides of the center line.

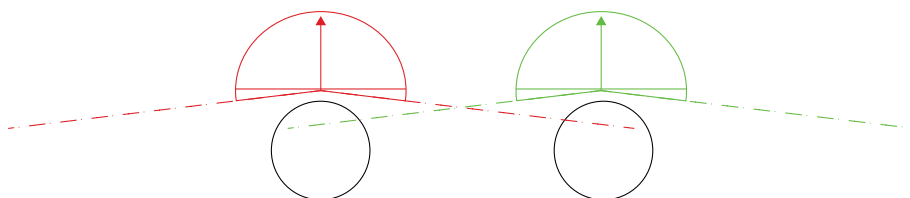


Figure 3. Fields of vision in a side-by-side constellation. The dash-dotted lines represent the limits of the two walkers' peripheral vision, when they are looking straight ahead. The horizontal line parallels the walkers' shoulder lines

This means that under the conditions depicted in Figure 3 (no head movement, walking exactly side-by-side), co-walkers will be able to perceive larger movements of the torso, hand gestures, or even pronounced head movements of the co-walker in their peripheral vision, but they will not be able to perceive details such as co-participants' eye-movements, small head-movements or mimics, including smiling. Thus, co-walkers can determine in which direction the other turns the body and (with restrictions) head, but need to turn towards the co-walkers in order to perceive smaller details. Note that even without vision, spatial hearing allows co-walkers to perceive whether the speaker is oriented to them or forward oriented.

3. Laughter and laughables

If looking at the co-participant is the exception rather than the rule in side-by-side constellations, it is of some interest to investigate those interactional moments in which gaze at a co-participant does occur. An investigation of these moments can tell us more about the functions of gaze and the dynamics of F-formations as it allows us to identify the most important functions of gaze; those that are so central that co-participants employ gaze even under circumstances that require additional bodily investment when compared to (other) F-formations (i.e. head and torso movements), and which potentially distract from other important tasks of vision while walking. An inspection of our data tells us that these functions are different for the recipient and the speaker. For instance, recipients regularly gaze at speakers when the latter's speech is marked by major hesitations. Recipients thereby not only display their continuous attention and co-participation, but also monitor the speaker's gaze in order to identify invitations to help out (Auer & Zima 2022). Here, however, we will focus on speaker-gaze. Again, there is a variety of interactional contexts in which speaker-gaze at recipient occurs in a recurrent fashion. An important and strikingly recurrent one is (joint) laughter.

The approach to laughter taken in conversation analysis, starting with early work by Jefferson (1984), and developed further by Glenn (2003, Glenn & Holt 2013), assumes that laughter is indicative of a (candidate) laughable, even though this laughable may not always be easy for the analyst (and the recipient) to identify. There is general agreement among researchers that laughables are not restricted to 'humorous' verbal or non-verbal happenings or facts (see the discussion in Partington 2006: 13–19). Glenn & Holt (2013: 2), for instance, argue that laughter occurs in two major environments, "celebrations" (showing appreciation and affiliation, including the appreciation of humor) and "trouble" (where laughter is "a resource for aligning, modifying actions, and mitigating meanings"). On the other hand, it is also true that not all forms of humor need to be accompanied (or followed) by laughter; it is enough that the shift into a humorous frame is appreciated, which can also be done, for instance, by a humorous uptake (Hay 2020).

Jefferson (1979) assimilates the structure of joint laughing to that of action sequences by distinguishing between laughter invitations or "offers" (laughing in "first position", i.e. together with the production of the laughable) and "laughter acceptance" (subsequent speakers joining in in "second position") or "rejection". But speakers may also produce a laughable without laughing, in which case recipients' laughter is "volunteered". Although first laughter is very often followed by co-participants' second laughter, this is not always the preferred option of responding. For instance, second participants will not join in the laughter when it was produced in the context of first speaker's trouble talk (see Glenn 1991, Jefferson 1984, Vöge 2010). The preference for laughter in second position therefore only holds when the laughable is part of certain action types.

We follow Jefferson's and Glenn's approach by treating laughter as a conversational activity which indexes a laughable (rather than starting with a semantic/pragmatic approach to the identification of laughables), and concur with their view that first laughter is often, though not always, expected to be taken up by recipient's laughter or some other appreciative next action. However, instead of assimilating laughter to verbal activities and their sequencing, we assume that laughter is a way of keying (framing) the interaction, and follow Ford & Fox (2010: 344) in their observation that "laughable practices are regularly distributed across strips of activity rather than discretely bounded in single units". Laughter is a metapragmatic cue which steers the interpretation of the interaction, by pointing to (indexing) a laughable. It can occur free-standingly, i.e. before or after the (formulation of the) laughable, as a phonetic distortion of the speech signal (which is interspersed with in- and exhalations), or as a particular type of voice quality (so-called smile voice) (Potter & Hepburn 2010). The new frame may start during or after the laughable, but can also occur before it, for instance before the beginning of a humorous story, as an announcement of a particular way to receive

it (cf. Spielmann 1988:206). Just like verbal action sequences, laughter is sequentially organized, but not in a strict one-speaker-at-a-time fashion; it preferably occurs in overlap. Once the frame has shifted from serious into humorous talk, it does not seem to be tied to specific sequential positions.

(Hearable and visible) laughter by the recipient signals to the speaker that their laughable has been identified; but there are alternative, visual cues that may function in the same way. Among them are facial cues such as lip-spreading, which is seen as smiling (Gironzetti et al. 2016). Smiles can be a weaker alternative to laughter, and they can – just like other facial cues such as headshakes or crinkling of eyes – organize the transition into laughter (Glenn 2003:66). They may also occur in environments in which laughter would not be adequate (Hanaka 2010). Speakers who produce a candidate laughable should therefore pay attention to the recipients' face in order to monitor their reception of a candidate laughable. This explains the enhanced need to monitor the recipient visually and is evidenced by Brône's finding (2021) that speakers shift gaze between multiple recipients of a humorous utterance more often and look into the open space less than in a serious key. He also found that in multi-party interaction, recipients look at each other more frequently and for longer periods during humorous turns in order to check whether their reception of the speaker converges with that of the others.

The following Section 4 will briefly introduce our data and explain the eye-tracking part of our study. Sections 5.1 will give a quantitative overview of (joint) laughter and speaker-gaze as well as mutual gaze. We then discuss some extracts in detail in Sections 5.2–5.4.

4. Data and methods

Our analysis is based on a corpus of nine recordings of couples walking through the Black Forest National Park, each with a duration of 80–120 minutes. All participants were L1 German speakers. They wore eye-tracking glasses² which allow for a precise analysis of their gaze behavior while talking and walking through nature. Participants were recruited via various platforms (social media) on which we asked for couples that had not visited the Black Forest National Park before and were ready to participate in an easy hike through the park. Upon their arrival they were instructed to follow a certain trail that we had selected beforehand. The glasses which they wore on their walk are light and unobtrusive and they hardly impede the peripheral vision because the glasses have no rims and because the eyeglass temples (the arms connected to the frame of the eye-tracking glasses on the sides) are thin and positioned higher than in regular glasses (see Figure 2).

2. Tobii Pro Glasses 2: <https://www.tobiiipro.com/de> (last accessed on December 8, 22).

The eye-tracking glasses have two in-built infrared cameras that track the movements of the participants' pupils through the reflections on the cornea (installed underneath the two lenses) and thus capture foveal vision. Additionally, a scene camera that is positioned in the middle of the glasses, records part of the wearer's field of vision. Peripheral vision is not fully captured and restricted to 90 degrees. Very pronounced movements of the eyes to one side therefore risk not being tracked (gazing underneath the temples of the glasses on the sides) since the 'cursor' will not be visible on the scene camera.

We decided against a co-walking, video recording investigator (external camera perspective) who would have been highly obtrusive, transforming a walk among two friends into a triadic interaction with a stranger walking ahead or after the walking couple.³

All participants gave informed written consent to the publication of transcripts and screenshots from the recordings.

For our analysis, the videos of the scene cameras and the eye-tracking data were overlaid. These resulting data were then synchronized for each walking dyad and arranged in a split screen view using Adobe Premiere Pro. We then identified all occurrences of laughter or laughing particles of at least one participant. We found 270 instances of potential laughables; these are the focus of this paper (a detailed quantitative overview is given in Section 5.1).

All sequences were transcribed following GAT-2 conventions (Selting et al. 2009, see Appendix A). In line with Jefferson 1979 (see also Glenn 2003:42ff), laughter was transcribed phonetically. Multimodal transcription follows Mondada (2017, see Appendix B) and Merlino & Mondada (2018). For the transcription of gaze, special conventions were developed (see Appendix of this chapter). Two pentagons above the verbal transcripts illustrate the physical orientation and gaze behavior of the two participants during their walks (and conversation).

5. Laughables and gaze during mobile interaction

5.1 Overview

Most of the examples in our data collection can be subsumed under two types which approximate the two types identified by Jefferson (1979). In type 1, the laughter is produced by a speaker in the same turn in which the (candidate) laugh-

3. See the chapter by Zima et al. in this volume for further methodological discussion and Rasmussen & Kristiansen for a critical discussion on the limitations of using external cameras in mobile settings.

able occurs (see extract a. below). Here, the recipients are in a relatively comfortable position; even if they are not able to identify the laughable, they can laugh along, thereby accepting the key without responding to the semantics of the preceding turn.

In the second case (type 2), the speaker formulates the laughable without providing laughter as a key for its interpretation (see extract b.).⁴ In this case, the recipients need to identify the (candidate) laughable on the verbal level, i.e. they have to recognize an ironic remark, a pun, an allusion, etc., without the speaker providing laughter in order to help with the new keying. We only included cases in this group in which we were able to detect these verbal cues. Not identifying the laughable can be face-threatening for the recipient, who may risk being seen as lacking wit (see Sacks 1974); however, it also puts the speaker in a potentially difficult position. If the candidate laughable remains undetected, the utterance may be misunderstood; for instance, an allusion may be taken as serious, a hyperbole as the truth, a tease as an insult (Drew 1987), etc.

In addition to these two types, we also found cases in which the speaker neither keys his utterance as containing a laughable (by laughter/laughing particles), nor can the verbal utterance be identified in a straightforward way as containing a laughable on semantic grounds (type 3, see extract c.). There are no grounds on which the speaker can be heard as intending to produce a laughable. Still, the recipient retrospectively treats the utterance as if it had contained a laughable by laughing ‘in response’. Obviously, it may be the case that we, as analysts, are lacking resources which are available to the co-participants to identify the laughable. Yet the quantitative gaze patterns described below suggest that the participants converge with our analytical categorization and treat this type as different from type 2.

- a. → 01 Gisa da wirts mir scho WARM wenn ich(h)hier rAU(h)fgeh(h);
 I'm already getting hot walking up here;
 02 Hans hehe[hehe]
 03 Gisa [°h< hehEH]
 b. 05 Jule ich habe gedacht es wäre vIEL KÄlTer;
 I thought it would be a lot colder;
 06 (1.5)
 → 07 Finn but- you can leave your hat on.
 08 Jule °h a:h ha ha (.) he °h

4. Note that speakers' smiles are not sufficient as a cue for the co-walker in this case, as normally the recipients do not look at the speakers and therefore cannot monitor them. Since the scene camera in the recipients' trackers will not capture the facial mimics of the speakers either, we have no data beyond those available to the recipient to analyze whether smiling occurs in type 2.

- c. 01 Anna ich will_n TEE;;
I want some tea;
02 Lars ja. (1.6)
→ 03 Lars TEE und den BACKcamembert;
tea and the fried camber (cheese);
04 Anna hh J(h)A(h) heheh *h

Figure 4 gives an overview of the gaze patterns accompanying these three types.

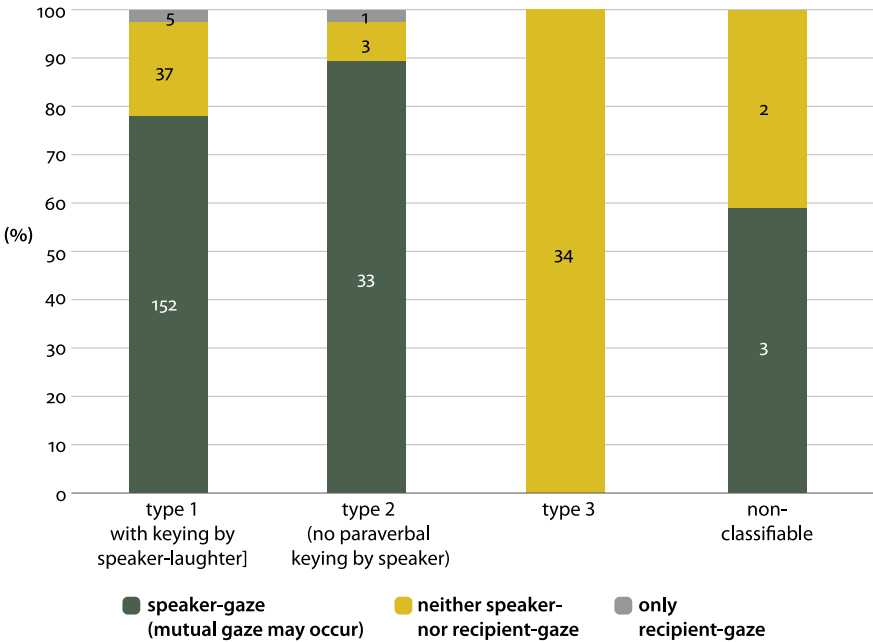


Figure 4. Gaze behavior while producing different types of laughables

As can be seen in Figure 4, type 1 laughables are the most frequent ($n=194$). In the large majority of cases (almost 80%) they are accompanied by a speaker-gaze at the addressee.⁵ 37 laughables are produced without laughter by the speaker within the same turn (type 2). Almost all of these (about 90%) are accompanied by speaker-gaze at the addressee. In contrast, type 3 laughables, of which we counted 34, are never produced with a speaker-gaze. This suggests that the subsequent laughter by the co-participant is not elicited by the speaker and that the first turn is only treated by the recipient, but not the speaker, as containing a laughable. Five cases could not be assigned to one of the three categories, because they were either

5. “Gaze at the addressee” was defined in a restrictive way here and means that the speaker’s gaze reaches the facial region of the addressee.

crossover cases, in which the speaker first produced a type 2 laughable and then reformulated the utterance with in-speech laughter (type 1), or because the keying was done verbally, i.e. the speaker starts the turn saying 'the funny thing is...'

In addition to the cases with or without speaker-gaze at recipient, there were six cases in which the recipient gazes at the speaker but not vice versa. These gazes occur during long pauses/hesitations or when the speaker produces (big) gestures. The gaze seems to be linked to these phenomena in the speaker's turn and is probably not connected to the laughable itself.

A first conclusion is, then, that most laughables are accompanied by speaker-gaze at the addressee. In the majority of instances in which no speaker-gaze was observed ($n=37$), the speakers' bodies were nevertheless oriented towards the recipient, i.e. their heads and gaze moved (slightly) in the co-walkers' direction (28 cases). In the remaining nine cases a humorous key was already in place at the point where the crucial laughable occurred. Hence, the speakers could be relatively sure that the new utterance would also be understood in this key and there was perhaps no need to monitor the recipients' understanding. All in all, there is very strong evidence that the production of laughables is a locus in mobile side-by-side conversations in which speakers turn towards and usually gaze at recipients.

Why is this the case? The quantitative overview suggests two possible interpretations. One is that the speaker needs to check whether the laughable has been detected by monitoring the co-walker's bodily response. In addition to audible laughter, this may also be done by non-audible, but visible cues such as smiling for the monitoring of which the speaker needs to turn to the recipient. This interpretation is supported by the fact that in type 2 laughables (where no audible keying by speaker's laughter occurs and the detection of the laughable by the recipient is even more an issue than in type 1) the percentage of speaker's gaze is slightly higher than in type 1. A second possible interpretation is that gaze at the recipient elicits or invites such a response. This interpretation is supported by the observation that speakers (slightly) turn in the direction of the recipient even when they do not look in their faces; this bodily movement will be perceivable by the recipient in most cases in peripheral vision or via voice direction recognition. In this latter case, the monitoring function is less likely. We return to this question in the final discussion.

Out of the 185 clear instances in which speakers turned their gaze at the recipient, 36 led to the establishment of mutual gaze (with approximately the same percentage in types 1 and 2). This is especially frequent when recipients laugh out loudly, or grin broadly at the speaker (Figure 5). Grinning at the speaker as a response to their potential laughable appears to be the main reason for a recipient-gaze. By turning to and showing their face to the speakers the recipients display their understanding of the laughable by presenting this response to the speaker.

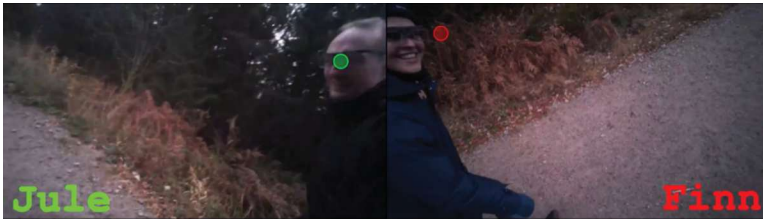


Figure 5. Speaker's (right) and recipient's (left) fields of vision (recipient grins broadly at speaker)

Some further details on the placement of recipients' responses are of interest. In most cases of type 1 laughables, "shared laughter" (Glenn 1991) occurs: either the recipient laughs subsequently to the speaker (most cases) or they both laugh at (roughly) the same time after the turn containing a laughable (19 cases). In contrast, there were no cases of shared laughter following a type 2 laughable, confirming Glenn's observation that shared laughter occurs most commonly when the speaker of the laughable laughs first (2003:101).

While almost all cases in our collection of laughables were followed by recipients' acknowledgment of the laughable by laughing or at least grinning, there are three cases in which the recipients continue with normal phonation (two of them following ironic remarks). This does not necessarily mean that the recipients did not detect the laughable; they may also have declined the offer to laugh along (cf. on such declines, see Jefferson 1979:84 and Drew 1984 in the position after teases).

5.2 Type 1 – speaker's laughter combined with gaze at recipient

In this section, we discuss type 1 laughables in greater detail by presenting some data extracts. In the first example, the two walkers are talking about dinner plans with their friends. They are discussing whether they should invite them to their (own) place or whether they should meet them at a restaurant.

Extract (1) Dinner with friends (VPo708, 00:22:17)

- 01 Luis laden wir die zu UNS ein?
are we going to invite them to our place?
- 02 (0.9)
- 03 Mara ja:==
yes
- 04 Luis =oder gehen wir nach MARburg.
or are we going to Marburg?
- 05 (1.1)

06 Mara ja wir könnt_n ja auch in diesem BRAUhaus zum BURgeressen gehen.
well we could also go to that 'brauhaus' (name of a restaurant) to eat burgers

07 Luis AH stimmt.
oh right

08 (0.4)

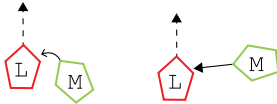


09 Mara {oder?
or

10 Luis wenn der wieder offen hat JA;
if that (place) is open again yes

11 Mara wir könn_n se auch GERne zu Uns einladen;
we can also invite them to our place

12 =da müssen wir am frEItag aber noch PUTzen.}
but we must clean on Friday then



13 Luis {m}



{HM-}

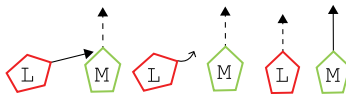


14 {(0.8)}

{(0.8)}



15 also BRAUhaus} {(hh)hö(h)rt
well 'brauhaus' sounds



si[ch] {#1 GU(h)T} {an.} {eh hh]
 REFL good PTCL(verb)
well 'brauhaus' sounds good

16 Mara [eh hh (.)] heh;}



#1

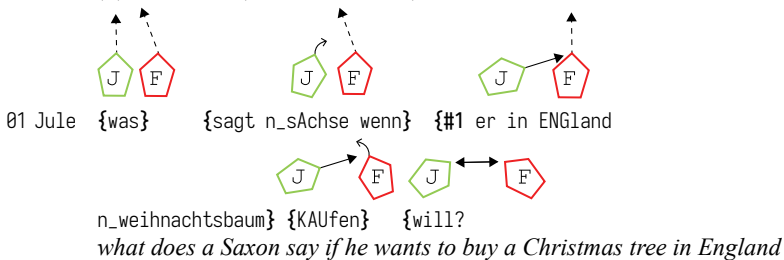
At the beginning of this sequence, Luis asks Mara if they should invite their friends to their place (line 01) or if they should meet them in the city (line 04). Mara answers that they could go to a burger place (line 06), a proposal which is acknowledged by Luis as a good idea he hadn't thought of (cf. the change-of-state token *ah* in line 07), but not accepted wholeheartedly; rather, after a silence, Luis increments the turn constructional unit with an if-clause, which suggests that the place might not be open. This makes Mara come back to Luis' alternative proposal to invite the friends to their place (line 11). However, she puts a condition on her acceptance of this alternative, which arguably also serves as an account for her not having chosen it in the first place: they have to clean (their apartment) beforehand (line 12). During the entire sequence, the two walkers have not looked at each other but ahead (Mara) or slightly down to the ground (Luis). But while Luis acknowledges Mara's remark in line 13, she turns to look at him for a short while. We do not discuss this look in detail here as it does not involve laughter, but it is clear that it triggers a number of inferences (such as: Luis doesn't like cleaning, Mara does not want to do the cleaning alone, if Luis wants to invite the friends to their place he also has to take care of the cleaning, etc.); it even brings a slightly challenging tone into the interaction. After a considerable silence (1.6 seconds, line 14), Luis responds with 'well, BRAUhaus sounds good' (line 15), thereby settling the issue by opting for the 'pub' alternative. He does not comment on Mara's innuendo explicitly, but by inserting several laughing particles within his response, he can still be heard to respond to it. Luis knows that Mara knows that his choice of the pub is not based on its qualities but is due to the fact that he doesn't want to clean their apartment. This also constitutes the laughable.

While producing this response turn, Luis starts to turn to Mara right before his first in-speech laughing particle in line 15; his gaze reaches her face at the end of *gut*. He continues to gaze at her while she laughs in response, starting two syllables after his first laughing particles (line 16). Her laughter responds to Luis' laughter and it starts at a "recognition point" (Jefferson 1979: 82), where she seems to already anticipate the rest of his utterance. There is a short period of simultaneous laughter, which closes the sequence (Spielman 1988: 211).

This extract shows the typical pattern of type 1 laughables in mobile side-by-side configurations in our data: One of the participants produces an utterance with a laughable which includes either laughing particles or full-fledged laughing. More or less exactly at the same time the speaker turns to gaze at the addressee.

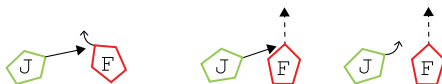
In our second example of type 1 laughables, the laughable is clearly recognizable on the semantic level, i.e. a (funny) conundrum (based on a pun).

Extract (2) Fir tree (VP2728, 26:04)



#1

02 (0.4)}



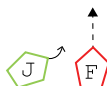
03 Finn {ich hab} % {keine AHnu} {ng;
I have no idea

Finn %grins->

04 (0.1)} %



05 Jule {adDENSchen} {#2 please; [heh-hh]}
 ɛdɛ:nʒn ɸli:s
 (dialectal for:) *a fir tree please*



06 Finn

[OH;] {hh-hh}



#2

The sequence starts in a typical riddle format in line 01: Jule asks a question which cannot be answered. The riddle contains two referential items, the personal reference *ein Sachse* ('a Saxonian'), which in the present context of reported speech ('what does a Saxon say when...') evokes the Upper Saxonian dialect, and the local reference 'in England', which evokes English as a language.⁶ During this question, Jule looks at Finn, as questioners regularly do, sometimes even when walking. More remarkable is that she keeps her gaze on him almost until the end of the following standard response of the addressee that he cannot answer the question. After a short gaze movement away (end of line 03 and line 04), Jule then gazes back at Finn when she resolves the riddle. She does so by producing a pun, which is the laughable in this extract (line 05). The pun is based on the phonetic similarity between English *attention* [ə'tɛnʃən] and the Upper Saxon pronunciation of Std.G. *ein Tännchen* ('a fir tree') [ɛdɛ:nʒn].

Even though the design of the speaker's turn as a riddle clearly makes a laughable expectable, the speaker additionally keys her utterance containing the resolution of the riddle, and hence the laughable, by laughter. In this example, the laughter only starts after the end of the turn, in overlap with the recipient's *oh*-exclamative (line 06), which is his immediate response to the solution of the riddle (Golato 2012). Slightly after Jule's laughter, Finn laughs, too.

There are also examples where the potential laughable is not 'funny' and still, the same type 1 pattern can be observed. In Extract 3, one of the participants makes fun of herself by indirectly hinting at not being in good shape. 'Laughing at self', although not funny, regularly invites the addressee to laugh along (Jefferson 1984, Glenn 2003:101f).

Extract (3) Not being in shape (VP0910, 00:05:40)

01 Gisa {da wirts mir scho WARM} {wenn ich(h) hier}
there it-gets myself already warm when I here

02 Hans hehe[hehe]

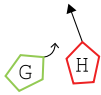
03 Gisa [°h hehEH]}

{#1 rAU(h) fgeh(h)
up walk
I'm already getting hot walking up here}

6. The Upper Saxon dialect is one of the most stereotyped ones in Germany. Among its most stereotypical features are the coronalization of the fricative [ç] and the lenition of stops.



#1



- 04 { (0.6) }
- 05 dann brauch ich NIX mehr *h.
then I won't need anything anymore
- 06 (0.7)
- 07 Hans ansonsten willste_n TEE habn;=
or would you like some tea
- 08 =bissl wie a WÄRMflasch[n oder wie,]
a bit like a hot water bottle or so
- 09 Gisa [OH go]tt nei.
oh god no

In line 01, Gisa remarks that she is already getting hot from walking 'up here'. The semantics of this utterance do not per se constitute a laughable. However, Gisa keys her utterance with several laughing particles within her speech. Shortly before producing the laughing particles, she starts to turn to the recipient. Her gaze reaches Hans' face two syllables before the end of her turn. Hans starts to laugh subsequently (line 02) and Gisa continues to laugh (line 03). As in the last example, recipient's co-laughter only sets in after the speaker's utterance, marked by intermittent laughter particles, is finished; it is clearly responsive. Speaker's gaze moves to the addressee more or less during the onset of her laughter.

In the next example of type 1, laughter is again essential to key the utterance as the story told by Finn is not per se funny or amusing. Rather, it tells about a face-threatening event in which Finn was involved. Finn is telling Jule about his new job for which he can work remotely and that he asked his friends if they had a good internet connection so that he could visit them and work at their place.

Extract (4) Remote Work (VP2728, 00:04:46)

01 Finn: (also) ich hab gEstern in meiner freundesgruppe eben erzählt dass: (--)
 ich den: (-) HOME office (--) äh jOb hab-
 (so) yesterday, I told my (group of) friends tha:t (--) I got tha:t (-) remote
 work job (--) uh job

02 (1.6)

03 u::nd-
 a::nd

04 (1.2)

05 dann n_PAULi gefragt wie gut sein INternet is.
 then I asked Pauli how well his internet works

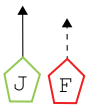
06 also (.) hInsichtlich dass ich ihn mal beSÜchen könnt für n_paar-
 you know concerning that I could visit him for a few

07 (0.3)

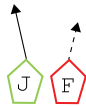
08 für ne WOche oder was auch immer,
 for a week or whatever

09 (0.9)

10 dann hat er gesagt=Ah s_internet is halt ziemlich SCHLECHT;
 then he said=ah the internet is pretty bad



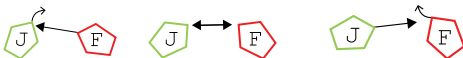
11 { (0.9) }



12 {sin gleich zwEi andere in die BRESche gesch gesprungen;
 two others stepped into the breach right away



13 =und ham_mir} {angeboten dass} {#1 ich (h) bei (h)Ihnen}
 and AUX me offered that I at their-place

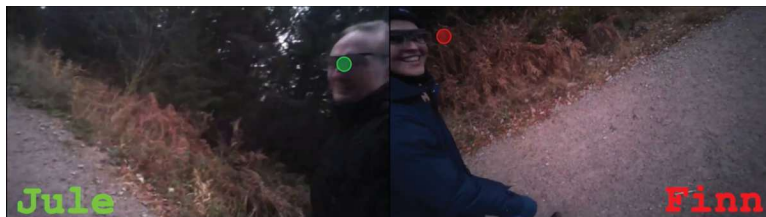


%{arbei} {#2 ten ka(h)} {nn(h), (--)}
 work can
 and offered that I could work at their places

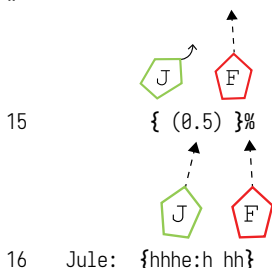
Jule: %grinning-->



#1



#2



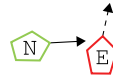
Finn tells Jule that he asked his friend Pauli how well his internet connection works (line 05–09), implying that he might come to his place to work. When his friend stated that his internet isn’t the best (line 10), Finn’s request was indirectly rejected. But two other friends ‘jumped in’ and told Finn that he could work at their place (line 12–13), thereby rescuing Finn’s face. The speaker marks this utterance as constituting a laughable, turning it into an amusing event, towards the end of its production by inserting laughing particles right after having directed his gaze at Jule. This transformation into an amusing story mitigates the face-threat to Finn.

Jule gazes back and grins at him as well; later, laughter follows (line 16). Just as in the last extract, speaker-gaze at addressee in parallel with laughter can both serve to monitor the recipient’s response and to elicit her co-laughter.

Our last example for type 1 laughables shows one of the rare cases, in which the described pattern does not lead to recipient-laughter. As already mentioned, this “decline” of laughter (Jefferson 1979) only occurs in three cases in our data. In Extract 5, the two participants are talking about apartments that Ella’s family owns. Nina, for her part, is searching for a new apartment.

Extract (5) Cabbage Farmers (VP0102, 00:13:21)

- 01 Nina habt ihr SO viele WOHNung[en?]
do you (PL) own that many apartments
- 02 Ella [d]ie von Opa ja,
the one from grandpa yes
- 03 Nina ja gut aber die hat ja HUNdert quadratmeter oder?
yeah well but that one has a hundred square meters right
- 04 (0.8)
- 05 die_s RIEsig;
it's huge
- 06 Ella ä::hm- (0.6) JA.
uhm yes
- 07 *(7.4)
Ella *squats in front of mushrooms and takes pictures-->
- 08 Nina cool. *-->
- 09 Nina vor allem wie das LICHT da noch so [reinfällt;=
especially the way the light pours in here
- 10 Ella [is RIChtig schön.]
it's really beautiful
- 11 Nina =voll SCHÖ]N.
very beautiful
- 12 (1.3) *
- 13 cool.
- 14 (0.6)
- 15 +genau Ähm- (0.9) die von Opa in harthausen sind_s DREI.
exactly uhm (with) the one from grandpa in harthausen three
Ella +starts to walk again
- 16 Nina ach KRASS.
oh sick
- 17 Ella mHM.
- 18 Nina {(na)ihr seid ja Übel die }{#1 rItschen KRAUT} {bau#2ern.
well you(PL) are PTCL awfully the rich cabbage farmers
you (guys) are awfully rich cabbage farmers





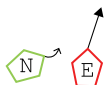
#1



#2

19 he h h[hhh he hhh
heh he °h

] he (.)



20 Ella [(ja) KRAUTbauern vor a }{llem.]]
Yeah cabbage farmers in particular

At the beginning of this extract, the participants are still standing in front of several toadstools of which Ella is taking photos. We observe a shift from displaced speech (line 01–06) about Ella's family and the apartments they own, to situated speech about mushrooms (line 08–13), and back to displaced speech in line 15, when co-participants also resume walking (side-by-side). Ella now explains that her family owns three apartments, which Nina assesses in line 16 as 'oh sick'. After Ella's *mHM*, which can be heard as a confirmation, Nina produces her ironic remark 'well you guys are awfully rich cabbage farmers'. (The remark is ironic as they are no cabbage farmers any longer these days and as cabbage farming cannot be the foundation of economic wealth in modern agriculture.) While producing *ritschen Krautbauern*, she starts to turn to Ella (line 18) and then produces a very long laughing while keeping her gaze at Nina (line 19). Although we see the same gaze pattern as in the previous extracts, the recipient does not laugh (nor does she establish eye contact with Ella); instead, she responds to the semantics of Ella's utterance and comments on it by 'yeah cabbage farmers in particular' (line 20).

The syntax and prosody of the response puts a focus on the expression 'cabbage farmers' and is again ironic: 'in particular' here means 'surely not'. The social category 'cabbage farmer' is thereby rejected. But the meaning of Ella's comment as a whole (i.e., that her family is very rich) is not commented on and not disagreed with. This is in line with the sequence preceding the laughter (lines 16/

17) where Ella had evaluated the information given by Nina that her grandfather owns three appartments as *krass* ('sick'), which Nina had confirmed.

The absence of responsive laughter in this extract may be due to the teasing character of Ella's comment (Drew 1984). Oben & Brône (2022) show that speakers always look at their addressee when delivering a tease, while the "teasing target" looks away. Heritage (1987) observes that teases are often not taken up in a humorous mode but rejected without the recipient showing an orientation to the humorous frame introduced by the teasing party. All these features of teases also apply in the present case.

5.3 Type 2 – no speaker's laughter but gaze at recipient

The second type, in which speakers do not key their utterances as containing a laughable by laughing particles, is much rarer in our data ($n=37$). This strategy is on the one hand potentially more face-threatening for speakers, as the recipients may not be able to understand the keying, and on the other hand requires more work from the recipients since they must recognize the laughable on the verbal level without any additional hints.

In our first type 2 example, the two co-walkers are talking about the temperature.

Extract (6) Joe Cocker song (VP2728, 00:15:09)

01 Finn HUH=es wird WÄRmer.
huh it's getting warmer

02 (0.7)

03 Jule ehe;

04 (3.5)

05 ich habe gedacht es wäre vIe:l KÄlter;
I thought it would be a lot colder

06 { (1.5) }

07 Finn {<1but-> (-)} {+you#1 can leave your HAT}{on.}
 Finn +pointing gesture at Jule-->



#1



08 Jule {#2} (°ha:h)+ {ha ha (.)he °h)}



#2

While the two walkers go up a hill, the sequence begins with Finn's comment that it is getting warmer (line 01). Jule agrees with him (line 03), adding that she thought it would be much colder (line 05). This implies that her clothes are too warm for the occasion. As a comment to that, Finn quotes the famous Joe Cocker's song 'You Can Leave Your Hat On' (line 07). The quote is marked by codeswitching into English. The comment is of course ironic, as Jule doesn't wear a hat. In the movie '9 ½ weeks', the song accompanies a striptease scene. The contrast between the movie scene and the actual scene adds to the humorous effect.

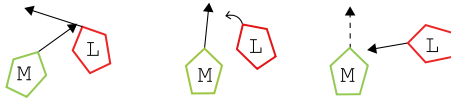
Finn does not include any laughing particles in his turn, but gazes at the recipient already very early during its emergence (line 07). It is this gaze that invites Jule to detect a feature of his talk that needs special attention, and if it is a laughable, to produce a laughing as a response.

Jule indeed starts to laugh right after the turn is finished and gazes back at the speaker (#2); she can therefore see that while she is laughing (line 08), Finn is smiling at her. This confirms her interpretation of Finn's comment as containing a laughable.⁷

In Extract 7, the temperature is again indirectly the topic of the conversation, but here, one of the participants is joking about 'nose warmers', which in his opinion should officially be accepted in the fashion world.

7. Finn's quotation is accompanied by an open hand pointing gesture at Jule. Gestures regularly attract co-participants' gaze in side-by-side constellations. Jule's gaze at Finn can therefore be linked to Finn's gaze just as well as to his gesture.

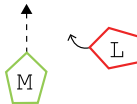
Extract (7) ‘Nose warmer’ (VP0708, 00:00:54)



01 Luis {ich bin ja immer} {noch dafür} {#1 dass es
offiziell so_n Nasenwärmer oder so [gibt;]
I still think nose warmers or something like that should ‘officially’ be available

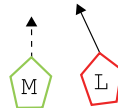
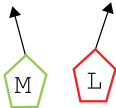


#1

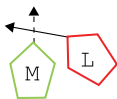
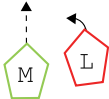


02 Mara

[ehh] {hh}}



03 Luis hhh hhh {so was dann auch KEine (.)} {Mödesünde oder}



{keine} {SCHANde is wenn man_s (.)} {trÄgt;
(so that) it isn't a fashion sin or (that) there is no shame in wearing it

The sequence-initiating comment in line 01 contains as a laughable the absurd idea of wearing a ‘nose-warmer’ and making their use ‘official’ (whatever the speaker may mean by this term). As in the previous extract, the speaker starts to gaze at the recipient early in his turn (line 01) but produces no laughing particles; he holds his gaze on Mara until she starts to laugh in line 02. He can monitor how the laughable is perceived and how the recipient responds to it. At the same time, his gaze elicits a response from Mara. Right after Mara starts to laugh, Luis turns away and after two outbreaks, which can be heard as laughing particles in third position, elaborates on how it should not be a ‘fashion sin’ or ‘disgrace’ to wear it (line 03). During this elaboration, he does not gaze at the recipient anymore, even though the humorous idea of nose-warmers is further

developed. Monitoring is no longer necessary as the recipient has already understood the humorous keying.

Laughables of type 2 seem to fall in the folk category of 'dry humor', which among the walkers in our study is a specialty of Finn. Therefore, we present another example of this speaker. In this sequence, the two participants are talking about popular science books.

Extract (8) Books (VP2728, 01:16:00)

01 Jule ICH hab h° noch aus: dE:r habe noch (.) GANZ ganz_äh (.)
 tolle JUgendbücher zu hause; (0.6)
I still have from: that I have very very great young adult books at home

02 ä:hm (.) allgeMEINwissen;=
uhm general knowledge

03 =alles was [du WISs]en musst,
everything you should know

04 Finn [ah;]

05 ja.
yes

06 (0.5)

07 Jule äh (.) da is halt äh- (1.3) eins chemIE,
uh there is PTCL uh one (in) chemistry

08 (.) EIns is-
one is

09 Finn ah: was ist [WAS? (.)
*oh how and why*⁸

10 Jule [ähm; [::;]
ehm

11 Finn [(.) oder so]was in der ART;
or something like that

12 Jule !JA_NE:!=
well no

13 =es sind halt (.) richtig so dicke Büch[er,=]
they are really thick books like that

14 Finn [o=]
 =[kE;]

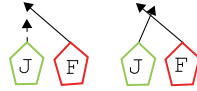
15 Jule =[wo du] dann halt auf (.) ZWEI seiten im grunde
 zuSAMmengefasst hast=
in which you have two pages basically summarizing

16 =über (.) marie cuRIE: oder;
about Marie Curie or

17 Finn AH: ja.
oh yes

18 Jule äh (--) berühmte persÖnlichkeiten;
uh famous personalities

19 (1.5)



20 Finn [<<french pronunciation>{freddi mEr}]{cuRIE> zum beispiel;
for example

21 Jule [äh;;
uh]



22 (0.8)} {(0.3)} {#1 (0.3)}



#1



23 Jule {(h)was} {(h)hab} {ich NOCH(hh),} (0.1)
what have I else
what else do I have



äh: (-) {DICHter,=
eh poets

24 =sind glaube hab ich auch ein BAND;}
are also in the volume I think



25 Finn {ja;}
yes

Jule is telling Finn about the books for young people which she still keeps at home and which contain ‘everything you need to know’ (line 01–02). She starts to list these books, but is interrupted by Finn who asks whether they are like the “Was ist was?” (‘How and why?’) series (line 09). Jule denies and further explains what

8. ‘How and why’ is a well-known book series designed to teach science and history to kids and young teenagers.

kinds of books she has (line 06), that they are rather thick books (line 12–13), and that they often contain summaries of two pages about people like *marie curie* (line 15–16) or other famous personalities (line 18). This ends her turn. After a long pause (line 19), Finn adds *freddi mercurie* as an example for famous personalities. He pronounces the name with a French accent and with a stress on *cuRIE*, thereby establishing a phonetic similarity with *marie cuRIE*. The laughable is the pun achieved by this forced phonetic similarity together with the fact that Freddie Mercury obviously doesn't fit in the list of famous scientists. Finn starts to turn to Jule (line 22) and then gazes at her (#1) while she smiles. Simultaneously with the beginning of Finn's pun, Jule embarks on expanding her turn, starting with a hesitation marker (line 21), but then leaves the turn to Finn.

After Finn's turn containing the laughable (but no laughing particles), Jule continues her utterance; she does not comment on the laughable verbally but continues with describing her books (line 23). However, the beginning of her turn is accompanied by laughing particles, signaling that she has understood the laughable but doesn't want to terminate her talk about the books she cherishes so much.

Considering the gaze behavior of the speaker of the laughable more closely, two aspects stand out: (1) Jule produces her in-speech laughing particles immediately after Finn gazes at her while continuing her talk about her books, and (2) Finn turns away right after the first laughing particle is produced (line 23). Arguably, Finn's gaze at Jule not only monitors her response to his pun, but also invites such a response. Compared to the previously discussed extracts of type 2 laughables, the extract demonstrates that laughter as a response to a laughable can occur simultaneously with the next verbal activity which is not related to this laughable on the content and sequential level.

5.4 Type 3 – no speaker's laughter and no gaze, but recipient laughter

In all the sequences discussed so far, we saw that speaker-gaze accompanying a laughable can work as a monitoring as well as a mobilizing device for a response to the potential laughable. We now turn to the third category (cf. Figure 4, type 3) in which recipients laugh in response to an utterance which cannot be identified as a laughable in a straightforward way (no laughable can be recognized on the semantic level and the turn does not contain any laughter or within-speech laughing particles). The following extract exemplifies this case.

In this example the two participants are talking about what they would like to drink and eat (no gaze transcriptions are given, as the participants simply gaze along the walking trail).

Extract (9) Fried Camembert Cheese (VP0506, 01:16:25)

01 Anna ich will n TEE;;
I want some tea

02 Lars ja. (1.6)
yes

03 TEE und den BACKcamembert;
tea and the fried camembert (cheese)

04 Anna hh J(h)A(h) heheh °h
yes

This sequence starts with Anna's statement that she wants some tea (line 01) (at the near-by chalet, where the two have decided to take a rest). Lars acknowledges her wish in line 02 and, after a pause, expands the list by 'fried camembert' (line 02). In response to this utterance Anna laughs while saying 'yes' (line 04). While the utterance has the potential to be heard as a laughable, given the unusual combination of tea and camembert, it refers back to the beginning of the hike in this case, when camembert cheese had been a serious topic (cf. the definite article). Although the back-referencing and the absence of speaker-gaze suggest that the speaker had no intention of achieving a funny effect, the recipient still treats it that way and laughs. With Anna's laughter the sequence ends.

6. Conclusions

In this chapter, we have investigated speakers' gaze behavior in walking dyads, taking hikes in the forest as an example. We were interested in the 'marked' case of the current speaker turning toward the addressee. One of the contexts in which such marked behavior was found to be very frequent are turns that contain a potential laughable (cf. also Brone et al., in this volume who show that speakers direct their gaze toward recipients more frequently during ironic statements than during non-ironic ones). The study is based on a collection of 270 sequences containing such laughables, defined by the occurrence of laughter in at least one of the co-walkers, usually the speaker. The recipient often joins in or responds by laughing as well. The large majority of these turns was accompanied by speaker-gaze at the addressee (but only occasionally resulted in mutual gaze). Speaker-gaze at addressee occurred very often (in more than 80% of all instances) during turns which were keyed by speaker's laughing as containing a laughable. When the laughable had to be identified on semantic grounds and was not keyed by the speaker's laughter, speaker-gaze occurred in almost all cases (in only 4 cases there was no speaker-gaze towards the recipients face detectable). When speakers did not mark their turn by laughter nor present it as a laughable by its wording, they

did not turn their gaze to the co-walker. In this case the laughable is established by the recipient alone, and there is no evidence that it responds to a verbal or visual cue in the preceding turn.

On the basis of a fine-grained interactional analyses, we discussed two interactional functions of gaze, a regulatory⁹ and a monitoring one. We surmise that these are the two basic functions of gaze in interaction during displaced speech.¹⁰ The regulatory function of gaze, which includes turn-final gaze in order to select next speakers (cf. Auer 2018, 2021), invites or even elicits the co-participant's response to the current speaker's action. The monitoring function serves the speaker to check on how the current turn was received by the addressee. These two functions are of course related and not mutually exclusive: seeing that the speaker monitors possible responses by the addressee is also an invitation for the addressee to display this response to the speaker.

In many cases, both functions combine. Establishing their specific contribution needs a detailed investigation of the relative timing of gaze towards the addressee, speaker's laughter (interspersed, preceding or following the verbal turn components), and recipients' responses, in each case. For instance, in Extract 1, recipient's responding laughter already starts slightly before the speaker gazes at her. This suggests that in this example, gaze primarily serves to monitor the response to the turn containing the laughable. In other cases, particularly in type 2 laughables, in which the speaker does not key the utterance as containing a laughable by laughter, it seems to be of topmost importance for the speaker to monitor the addressee's response in order to understand how the recipient has understood the turn containing the laughable.

There are other interactional contexts in which current speakers gaze at their co-walkers in a side-by-side constellation, such as during the production of certain question types; here, the elicitation function prevails. There are also interactional contexts in which recipients gaze at current speakers, such as during longish hesitations in their speech, or while speakers are performing large gestures; in this case, the monitoring function is in the foreground. A fuller investigation of gazing at the co-participant in a constellation in which it is the exception rather than the rule can give us a better understanding of why we gaze at each other.

9. The regulatory function of gaze includes its role in turn-taking, turn allocation, response mobilization and sequence organization (Rossano 2013: 315).

10. I.e. as long as it is not part of an "environmentally coupled" package (Goodwin 2007) including gestures (for instance accompanying deixis), or other bodily activities. For gaze in the context of pointing and establishing joint attention, see Stukenbrock (2020).

Acknowledgements

Both authors contributed equally to this chapter. This work originated from a joint project with the *Nationalpark Schwarzwald* (special thanks to Kerstin Botsch).

References

- doi Auer, Peter. 1988. "On deixis and displacement." *Folia Linguistica* XXII/3–4, 263–292.
- doi Auer, Peter. 2018. "Gaze, addressee selection and turn-taking in three-party interaction." In *Eye-tracking in Interaction. Studies on the Role of Eye Gaze in Dialogue*, ed. by Geert Brône, and Bert Oben, 197–231. Amsterdam: John Benjamins.
- doi Auer, Peter. 2021. "Gaze selects the next speaker in answers to questions pronominally addressed to more than one co-participant." *Interactional Linguistics* 1 (2): 154–182.
- Auer, Peter, and Elisabeth Zima. 2021. "On word searches, gaze, and co-participation." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 22: 390–425.
- doi Auer, Peter, Barbara Laner, Martin Pfeiffer, and Kerstin Botsch. 2024. "Noticing and assessing nature: A multimodal investigation of the format 'perception imperative + exclamative' based on mobile eye-tracking data." In *New Perspectives in Interactional Linguistic Research*, ed. by Dagmar Barth-Weingarten, and Margret Selting. Amsterdam: John Benjamins, 245–277.
- doi Brône, Geert. 2021. "The multimodal negotiation of irony and humor in interaction. On the role of eye gaze in joint pretense." In *Figurative Language – Intersubjectivity and Usage*, ed. by Augusto Soares da Silva, 109–136. Amsterdam: John Benjamins.
- doi Brône, Geert, Bert Oben, Annelies Jehoul, Jelena Vranjes, and Kurt Feytaerts. 2017. "Eye gaze and viewpoint in multimodal interaction management." *Cognitive Linguistics* 28 (3): 449–483.
- doi Broth, Matthias, and Lorenza Mondada. 2013. "Walking away. The embodied achievement of activity closings in mobile interactions." *Journal of Pragmatics* 47: 41–58.
- doi Broth, Matthias, and Fredrik Lundström. 2013. "A walk on the pier. Establishing relevant places in mobile instructions." In *Interaction and Mobility*, ed. by Pentti Haddington, Lorenza Mondada, and Maurice Nevile, 91–122. Berlin/Boston: De Gruyter.
- doi Ciolek, T. Matthew, and Adam Kendon. 1980. "Environment and the spatial arrangement of conversational encounters." *Sociological Inquiry* 50: 237–271.
- De Stefani, Elwys. 2011. *'Ah petta ecco, io prendo questi che mi piacciono'. Agire come coppia al supermercato*. Roma: Aracne.
- doi De Stefani, Elwys. 2013. "Rearranging (in) space: On mobility and its relevance for the study of face-to-face interaction." In *Space in Language and Linguistics*, ed. by Peter Auer et al., 411–433. Berlin/Boston: De Gruyter.
- De Stefani, Elwys, and Lorenza Mondada. 2014. "Reorganizing mobile formations: When 'guided' participants initiate reorientations in guided tours." *Language and Culture* 17 (2): 157–175.
- doi Deppermann, Arnulf. 2018. "Instruction practices in German driving lessons: differential uses of declaratives and imperative." *IJAL* 28 (2): 265–282.
- doi Drew, Paul. 1987. "Po-faced receipts of teases." *Linguistics* 25: 219–253.

- doi Ford, Cecilia E., and Barbara Fox. 2010. "Multiple practices for constructing laughables." In *Prosody in Interaction*, ed. by Dagmar Barth-Weingarten, Elisabeth Reber, and Margret Selting, 339–368. Amsterdam: John Benjamins.
- doi Gardner, Rod, and Ilana Mushin. 2015. "Expanded transition spaces: the case of Garrwa." *Frontiers in Psychology* 6 (251): 1–14.
- doi Gironzetti, Elia, Lucy Pickering, Meichan Huang, Ying Zhang, Shigehito Menjo, and Salvatore Attardo. 2016. "Smiling synchronicity and gaze patterns in dyadic humorous conversations." *Humor* 29 (2): 301–324.
- Glenn, Philipp. 1991. "Current speaker initiation of two-party shared laughter." *Research on Language and Social Interaction* 25: 139–162.
- Glenn, Philipp. 2003. *Laughter in Interaction*. Cambridge: CUP.
- Glenn, Philipp, and Elisabeth Holt. 2016. "Introduction." In *Studies of Laughter in Interaction* ed. by Philipp Glenn and Elisabeth Holt, 1–22. London: Bloomsbury Publishing.
- Goffman, Erving. 1963. *Behavior in Public Places*. New York: The Free Press.
- Goffman, Erving. 1981. *Forms of Talk*. Pennsylvania: University of Pennsylvania Press.
- doi Goodwin, Marjorie, and Charles Goodwin. 2012. "Car talk: integrating texts, bodies, and changing landscapes." *Semiotica* 191: 257–286.
- doi Haakana, Markku. 2010. "Laughter and smiling: notes on co-occurrences." *Journal of Pragmatics* 42: 1499–1512.
- Haddington, Pentti, Lorenza Mondada, and Maurice Nevile (eds). 2014. *Interaction and Mobility*. Berlin/Boston: De Gruyter.
- doi Hay, Jennifer. 2020. "The pragmatics of humor support." *Humor* 14 (1): 55–82.
- Heritage, John. 1987. Ethnomethodology. In *Social Theory Today*, ed. by Anthony Giddens, and Jonathan H. Turner, 224–272. Cambridge: Polity Press.
- Hepburn, Alexa, and Scott Varney. 2016. "Beyond ((laughter)). Some notes on transcription." In *Studies of Laughter in Interaction*, ed. by Philipp Glenn, and Elizabeth Holt, 25–38. London: Bloomsbury Publishing.
- Jefferson, Gail. 1984. "On the organization of laughter in talk about troubles." In *Structures of Social Action. Studies in Ethnomethodology*, ed. by Max Atkinson, and John Heritage, 346–369. Cambridge: Cambridge University Press.
- doi Kendon, Adam. 1967. "Some functions of gaze direction in social interaction." *Acta Psychologica* 26: 22–63.
- doi Kendon, Adam. 1972. "Some relationships between body motion and speech: An analysis of an example." In *Studies in Dyadic Communication*, ed. by Aron Wolfe Siegman, and Benjamin Pope, 177–210. Headington: Pergamon Press.
- Kendon, Adam. 1990. *Conducting Interaction: Patterns of Behaviour in Focused Encounters*. Cambridge: Cambridge University Press.
- Laner, Barbara. 2022. "'Guck mal der Baum' – Zur Verwendung von Wahrnehmungsimperativen mit und ohne mal." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 23: 1–35.
- doi Laner, Barbara (2025). Mobile stance-taking in nature: An exploration of gaze patterns during assessments of objects in nature. *Frontiers in Psychology*.

- Mausner, Claudia. 2005. "Capturing the hike experience on video: a new methodology for studying human transactions with nature." In *Proceedings of the 2005 Northeastern Recreation Research Symposium*. Gen. Tech. Rep. NE-341, ed. by John G. Peden, and Rudy M. Schuster, 168–177. New York: U.S. Forest Service, Northeastern Research Station.
- Mayor, Eric, and Adrian Bangert. 2013. "Coordinating turning while walking and talking." *Proceedings of the Annual Meeting of the Cognitive Science Society* 35 (35): 2002–2007.
- doi Mayor, Eric, and Adrian Bangert. 2016. "Flexible coordination of stationary and mobile conversations with gaze: Resource allocation among joint activities." *Frontiers in Psychology* 24 (7): 1582.
- McIlvenny, Paul. 2013. "Vélobike formations-in-action: Biking and talking together." *Space and Language* 17 (2): 137–156.
- doi Merlino, Sandra and Lorenza Mondada. 2018. "Crossing the street. How pedestrians interact with cars." *Language & Communication* 65: 131–147.
- doi Meyer, Christian, and Ulrich von Wedelstaedt. 2017. "Intercorporeality, enaction, and interkinesthesia: New perspectives on moving bodies in interaction." In *Moving Bodies in Interaction — Interacting Bodies in Motion: Intercorporeality, Interkinesthesia, and Enaction in Sports*, ed. by Christian Meyer, and Ulrich von Wedelstaedt, 1–23. Berlin: De Gruyter.
- doi Mondada, Lorenza. 2009. "Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space." *Journal of Pragmatics* 41 (10): 1977–1997.
- doi Mondada, Lorenza. 2014. "Bodies in action: multimodal analysis of walking and talking." *Language and Dialogue* 4 (3): 357–403.
- doi Mondada, Lorenza. 2016. "Challenges of multimodality: Language and the body in social interaction." *Journal of Sociolinguistics* 20 (3): 336–366.
- doi Mondada, Lorenza. 2017. "Walking and talking together: Question/answers and mobile participation in guided visits." *Social Science Information* 56 (2): 220–253.
- Oben, Bert, and Geert Brône. 2022. *Gaze in teasing sequences*. [Oral presentation]. University of Freiburg.
- doi Partington, Alan. 2006. *The Linguistics of Laughter. A corpus-Assisted Study of Laughter-Talk*. London: Routledge.
- doi Potter, Jonathan, and Alexa Hepburn. 2010. "Putting aspiration into words: 'laugh particles', managing descriptive trouble and modulating action." *Journal of Pragmatics* 42: 1543–1555.
- doi Relieu, Marc. 1999. "Parler en marchant. Pour une écologie dynamique des échanges de paroles." *Langage et Société* 89: 37–68.
- Rossano, Federico. 2013. "Gaze in conversation." In *The Handbook of Conversation Analysis*, ed. by Jack Sidnell, and Tanya Stivers, 308–329. Oxford: Blackwell.
- Ryave, A. Lincoln, and James N. Schenkein. 1974. "Notes on the art of walking." In *Ethnomethodology*, ed. by Roy Turner, 265–274. München: Penguin.
- Sacks, Harvey. 1974. "An analysis of the course of a joke's telling in conversation." In *Explorations in the Ethnography of Speaking*, ed. by Richard Bauman, and Joel F. Sherzer, 337–353. Cambridge: Cambridge University Press.
- Selting, Margret et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung* 10, 353–402.

-  Steger, Antonia. 2019. "Aneinander-Vorbeigehen — eine Interaktionsanalyse flüchtiger urbaner Begegnungen." *Zeitschrift für Germanistische Linguistik* 47 (2): 313–336.
-  Stukenbrock, Anja. 2018. "Forward-looking. Where do we go with multimodal projections?" In *Time in Embodied Interaction*, ed. by Arnolf Deppermann, and Jürgen Streeck, 31–68. Amsterdam: John Benjamins.
-  Stukenbrock, Anja. 2020. "Deixis, gaze practices, and the interactional achievement of joint attention." *Frontiers in Psychology* 11: 1779.
- Stukenbrock, Anja, and Karin Birkner. 2010. "Multimodale Ressourcen für Stadtführungen." In *Deutschland als fremde Kultur: Vermittlungsverfahren in Touristenführungen*, ed. by Marcella Costa, and Bernd Müller-Jacquier, 214–243. München: Judicium.
-  Stukenbrock, Anja, and Anh Nhi Dao. 2019. "Joint attention in passing." In *Embodied Activities in Face-to-Face and Mediated Settings*, ed. by Elisabeth Reber, and Cornelia Gerhardt, 177–213. London: Palgrave Macmillan.
-  Tong, Lili, Audrey Serna, Simon Pageaud, Sébastien George, and Aurélien Tabard. 2016. "It's not how you stand, it's how you move: F-formations and collaboration dynamics in a mobile learning game." *Mobile HCI '16: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 318–329.
-  Vöge, Monika. 2010. "Local identity processes in business meetings displayed through laughter in complaint sequences." *Journal of Pragmatics* 42: 1556–1576.
-  vom Lehn, Dirk. 2013. "Withdrawing from exhibits: The interactional organization of museum visits." In *Interaction and Mobility*, ed. by Pentti Haddington, Lorenza Mondada, and Maurice Nevile, 65–90. Berlin: De Gruyter.
-  Weilenmann, Sandra, Daniel Normark, and Eric Laurier. 2013. "Managing walking together: The challenge of revolving doors." *Space and Culture* 17 (2): 122–136.
- Weiß, Clarissa, and Peter Auer. 2016. "Das Blickverhalten des Rezipienten bei Sprecherhäsitationen: eine explorative Studie." *Gesprächsforschung — Online-Zeitschrift zur verbalen Interaktion* 17: 132–167.
- Zhisheng, Wang, Yukari Nagai, Dan Zhu, Jiahui Liu, and Nianyu Zou. 2019. "Based on creative thinking to museum lighting design influences to visitors emotional response levels theory research." *IOP Conference Series: Materials Science and Engineering* 573: 1–7.

Conventions for the transcription of gaze

	Pentagons above the verbal transcript iconically represent the two walkers (with their initials) and their bodily orientation
	Arrows indicate the walkers' gaze directions (in cases of momentary tracker malfunctions, no arrows are shown)
	Dashed arrows indicate that the person is gazing downwards
	Curved arrows indicate bodily reorientations (e.g., turning towards the object of reference)
	Various icons represent the objects of reference
<code>1 {verbal trans}{cript}</code>	Curly brackets indicate the scope of the iconically illustrated gaze behavior above the verbal transcript

When the establishment of joint attention becomes problematic

How participants manage divergent and competing foci of attention

Anja Stukenbrock & Angeliki Balantani

University of Heidelberg | University of Lausanne

In the past decades, a substantial amount of research has studied how joint attention is collaboratively accomplished in social interaction. By contrast, divergent and competing foci of attention have remained largely unexplored. Our study investigates how participants establish, or refrain from establishing, joint attention in the face of attentional divergence and competition. When participants summon their co-participants' attention on an object, the preferred response is to reorient and share attention. However, for various reasons, addressees may not always follow the invitation to share attention. One of the instances in which they may not (immediately) respond by reorienting is when they are themselves engrossed in something and prefer not to give it up for the sake of attention sharing.

Using the methodological principles of Conversation Analysis and a corpus of naturally occurring interactions recorded with video cameras and mobile eye tracking glasses, we examine the use of deictics and embodied practices to invite joint attention in open states of talk when the co-participant's attention is diverging. The recordings enable us to zoom in on how gaze (eye tracking data) and embodied orientation (data from external cameras) index and contribute to how sequences of divergent, competing, and joint attention unfold. Preliminary observations suggest, first, that the participants' spatial configuration contributes to how the problem of competing foci of attention is handled, and second, that participants deploy different verbal and embodied practices to pursue joint attention in the face of competing sites of interest. These practices are sensitive to, and reflexively constitute the participants' spatial configuration and range on a continuum from less to more response mobilising.

Keywords: competing foci of attention, joint attention, gaze, deixis, social interaction, mobile eye tracking

1. Introduction

When humans talk and interact in co-presence, they regularly look at each other as well as away (Kendon 1967). Looking away can mean two things: “just” looking away without attending to visible phenomena in the surrounds, or looking specifically at something else. Phenomena that participants look at individually may or may not become the focus of joint attention in the course of the interaction. In the default case, participants neither continuously sustain mutual gaze (Goodwin 1981, Kendon 1967, Streeck 2014) nor do they constantly share visible attention on the same phenomenon. Instead, divergent foci of attention are part and parcel of our everyday experience and deeply grounded in the embodied egocentricity of the lived body (Auer & Stukenbrock 2022). Joint attention is an interactional accomplishment that involves two (or more) participants who mutually coordinate to establish a triadic relationship with an object or event (Clark 1996, Stukenbrock 2015, 2020).

Coordination may not always run smoothly so that some interactional work needs to be done in order for joint attention to be achieved. Our chapter is concerned with interactionally delicate moments in which the establishment of joint attention becomes problematic. It investigates instances where the participants are attending to different phenomena and one of them invites the other to share attention on something he or she finds interesting.

We distinguish between *divergent* and *competing foci of visual attention*. We define *divergent foci of attention* as a situated state or moment in which co-participants, i.e. ratified participants in social interaction, are attending to different phenomena and are therefore potentially not available for attention sharing with others. This regularly occurs when participants are in an “open state of talk” (Goffman 1981: 134) and does not necessarily pose a problem. In contrast, *competing foci of attention* arise when, growing out of a *state of divergent foci of attention*, one of the participants summons the other to share attention on the phenomenon that the summoner finds interesting but that requests from the other, in order to comply with the request, to give up his or her own attentional focus. As Goodwin and Goodwin (2012) noted, addressees who are “engrossed in materials of their own” may display reluctance or even “refuse to co-participate in the way projected by the initiator” (275). The analysis presented here focuses on how participants resolve the problem of competing foci of attention that emerges when, in states of divergent foci of attention, a joint attentional sequence is initiated but treated as dispreferred by the co-participant.

The following extract¹ from a corpus of mobile eye tracking recordings (cf. Section 3 for details) illustrates the phenomenon we are interested in. It documents how competing foci of attention may arise in social interaction, and exemplifies the interactional work done to resolve this problem. Two friends, Petra (P) and Anna (A), are visiting a university museum. They are in an open state of talk and several meters apart. While Anna is reading a text on moulage on an information board (Figure 1, left), Petra's visual attention is focused on a gynaecological model (Figure 1, right).

Extract 1. “model vs. information board” (Uniseum01_03_00:32:26)



#1
 01 P-vb *hh KUMma wie d#1as AUSschat;
 look PTCL what that looks like
 P-gz -model----->
 A-gz >>reading---->
 02 (2.3)
 A-gz -reading---->
 03 A-vb +WArte?+
 wait
 P-gz +-A-----+....>
 A-gz -reading---->



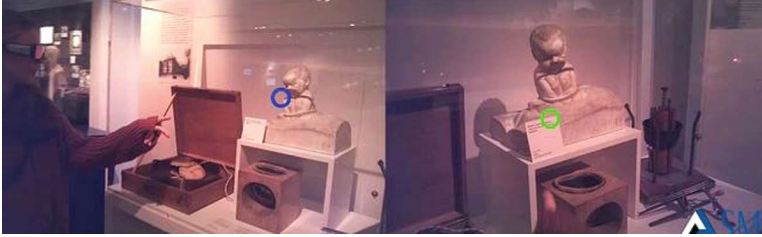
#2

1. In the split-screen figures, Petra's perspective is displayed on the right and Anna's on the left. Talk is transcribed according to GAT₂ (Selting et al. 2011; see Appendix A). The following abbreviations are used: “vb” for the verbal tier, “gz” for gaze, “ge” for gesture, “bd” for bodily behaviour. Multimodal annotations adapt the conventions developed by Mondada (2019; see Appendix B).



#3

- 04 (6.8) \$*#2(1.1)(0.2) *#3(1.2)
 A-gz -reading--*...to P-----*-searching---->
 A-bd \$-turns, walks to P----->
- 05 A-vb W+AS,
 what
 A-gz -searching---->
 P-gz .+-model----->



#4

- 06 (0.8) & +*#4(0.2)
 P-ge &-points----->
 P-gz -model-----+label----->
 A-gz -searching-*model---->
- 07 P-vb gUck mal \$DEN *hier;&
 look at that one here
 P-ge -points-----&
 P-gz -label----->
 A-gz -model-----*-label->
 A-bd -----\$stops----->
- 08 (0.3) +(1.0)
 P-gz -label+-model-->
 A-gz -label----->
- 09 P-vb das moDELL;
 the model
 P-gz -model----->
 A-gz -label----->
- 10 (1.0) *
 A-gz -label*
 P-gz -model-->
- 11 P-vb hhh° °h [hh
- 12 A-vb [ja;
 yes

P-gz -model----->
A-gz -model----->
13 P-vb Irgendwie KRASS;
 somehow weird
P-gz -model----->
A-gz -model----->
14 (2.0)
15 ((laughing together, joint departure))

The participants are in a *state of divergent foci of attention* and not available for mutual engagement when Petra summons Anna to share attention on the gynaecological model (l.01). The summons is formulated with a perceptual directive (KUMma/ ‘look PTCL’) and a demonstrative (das/ ‘that’) both of which request addressee gaze (Goodwin & Goodwin 2012, Laner 2022, Stukenbrock 2020), and a wh-exclamative that treats the object as remarkable (Auer et al. 2024, Pfeiffer 2016). Anna, however, instead of complying and reorienting to Petra, keeps her gaze on the information board and, after a noticeable pause (l.02), requests her friend to wait (l.03). The request displays Anna’s unavailability while simultaneously projecting a sequential resolution to the problem of competing foci of attention. Her embodied orientation displays her current, competing focus of attention to Petra, who, in the course of Anna’s request, shifts gaze to her (l.03) (cf. Stukenbrock 2015, 2020 on addressee monitoring). After Anna has finished reading, she shifts gaze to Petra (Figure 2, left), unsuccessfully follows her gaze direction (Figure 3, left), moves towards her and initiates repair (l.05: WAS,/ ‘what’). Petra incrementally redesigns her initial action in such a way as to help Anna find the object, first, by producing a pointing gesture (l.07, Figure 4, left), and subsequently, by incrementing her turn with a noun phrase (l.09: das moDELL;/ ‘the model’). Sequentially, a response from Anna is expected, who, after another pause (l.10), produces a minimal acknowledgement token (l.12: ja;/ ‘yes’). The sequence comes to close after Petra formulates an assessment (l.13), which, without receiving a verbal response (l.14), is followed by shared laughter and joint departure from the exhibit (l.15).

The extract exemplifies the interactional work done to resolve the problem of competing foci of attention that may emerge when participants, on the move through perceptually rich environments, temporarily separate to explore things on their own. We saw that the summoner’s initial request to share attention was met by a dispreferred response (noticeable pause, request to wait, embodied display of addressee’s own focus of attention). This was followed by an other-initiated repair sequence in which the summoner mobilised additional resources (gesture, spatial deictic, incremented noun phrase) to establish joint attention with the addressee.²

2. Note that it is beyond the scope of this chapter to analyse in detail the reasons for the addressee’s minimal response (see, however, Stukenbrock 2020).

Extract (1) departs from how reference and joint attention, in its default sequential structure, is interactionally achieved. To this we will turn in the next section.

Our chapter is structured as follows: After providing the relevant background on how joint attention is accomplished in social interaction (Section 2), we introduce the data and methodology and explicate the particularities of using mobile eye tracking in naturally occurring interaction (Section 3). The analysis (Section 4) covers different ways of dealing with the problem of competing foci of visual attention. We start with the sequential resolution in which two competing objects are attended to subsequently (Section 4.1). Next, we present sequences in which one object is jointly attended to while the other is abandoned (Section 4.2). This is contrasted with those rare instances in which participants, in a state of competing foci of attention, interactionally display that they give preference to their own focus of attention and refrain from establishing joint attention altogether (Section 4.3). The chapter concludes with a discussion of our findings.

2. Background: On joint attention and how it is accomplished in social interaction

The ability to share attention is crucial for and reflects the evolution of the communicative and socio-cognitive skills that are considered to be uniquely human (Tomasello 2008). In the course of history, human languages have developed a wide range of attention-directing devices, most importantly, demonstratives. Demonstratives have grammaticalised for the specific purpose of establishing joint attention (Diessel 2006, Diessel & Coventry 2020). In face-to-face interaction, demonstratives are often accompanied by pointing gestures (Clark 2003, Eriksson 2009, Fricke 2007, Goodwin 2003, Hindmarsh & Heath 2000, Kendon 2004, Mondada 2012, Streeck 2002, Stukenbrock 2015, 2020). Whether pointing is done verbally or gesturally or in conjunction, the triadic relation it instantiates between two (or more) participants and a third entity could not come about without the participants jointly attending to that entity by directing their gaze to it. Joint attention can thus be understood as two or more participants in a social “encounter” (Goffman 1963) focusing on the same object and being mutually aware of it (Clark 1996, Clark & Marshall 1981, Moore & Dunham 1995).

Conversation analytic work on the multimodality of demonstrative reference has shown that participants draw on a large array of embodied practices (Clark 2003, Eriksson 2009, Kendon 2004, Mondada 2014b, Streeck 2009, Stukenbrock 2014, 2015) as contextually embedded, temporally adapted and recipient-designed attention-directing devices. Furthermore, it has highlighted the critical role of space as an interactional resource (Hausendorf, Mondada & Schmitt 2012, Mon-

dada 2013a), the materiality of particular environments and participants' situated activities in those environments (Goodwin 2003, 2007; Hindmarsh & Heath 2000). Studies on joint attention in mobile settings (Auer et al. 2024, Botsch et al. this volume, De Stefani & Deppermann 2021, Goodwin & Goodwin 2012, Laner 2022, Stukenbrock 2023, Stukenbrock & Dao 2019) and in Virtual Reality (Hindmarsh, Heath & Fraser 2006, Keating 2015, Luff et al. 2003) have revealed the challenges posed by fleeting and hybrid relations between lived bodies and space (Auer & Stukenbrock 2022).

Joint attention enables participants to share objects, actions and events, to coordinate and mutually adapt their activities based on visible, embodied projections of socially intelligible lines of action. Coordinating joint actions and shared attention is less successful when participants have limited or no visual access to the same objects and to one another (Clark & Krych 2004). This is the case in spatial configurations where, for lack of being in one another's immediate presence, participants' "transactional segments" do not overlap (Kendon 1990) and an "eye-to-eye ecological huddle" (Goffman 1963: 95) is not established.

In various activities, engaging in joint attention is strongly invited, but may also become problematic. For instance, shopping together regularly entails establishing joint orientation towards commercial objects (De Stefani 2014, Stukenbrock & Dao 2019). In a similar vein, visiting a museum together calls for sharing attention on the exhibits as a way of enhancing togetherness (Vom Lehn 2013). Yet, these activities also lend themselves to individual exploration (Stukenbrock 2023) and may give rise to divergent and competing foci of attention. Consequently, it takes additional interactional work to co-ordinate, co-orient and co-operate (Hausendorf 2013) in order to achieve reference and joint attention.

Joint attention is an interactional accomplishment that involves two (or more) participants who mutually coordinate to establish a triadic relationship with an object or event (Clark 1996). Studies on the multimodal format of demonstrative reference and joint attention (Stukenbrock 2015, 2018a, 2020) have shown that joint attention is initiated by a summons-answer sequence within the turn (cf. also Goodwin 1981: 169): i.e. joint attention is initiated by a gesturally used demonstrative which requests addressee gaze. In a simplified version, this may be summarised as follows:

1st position summons to share attention (i.e. request for addressee gaze)

1. demonstrative + embodied pointing device
2. addressee monitoring (optional)

2nd position answer

1. embodied re-orientation to speaker
2. visual attention to target and referent

3rd position acknowledgement of a joint focus of attention

1. display of shared perception
2. documentation of understanding

In first position, a summons invites the addressee to share attention on an object. The summons can take the form of noticings (Hindmarsh & Heath 2000, Stukenbrock 2023, Stukenbrock & Dao 2019), perceptual directives such as *look* (Goodwin & Goodwin 2012, Laner 2022) and, most prominently, demonstratives (Diessel 2006, Stukenbrock 2015, 2020). In Extract (1), Petra uses a perceptual directive in combination with an exclamative and a demonstrative to summon her addressee's attention. In the default case, the summons gets responded to by bodily reorientation to the speaker and visual attention to target and referent.³ However, as we have seen, this was not the case in Extract (1). Speakers can often be observed to engage in addressee gaze monitoring in order to check whether joint attention has been achieved. In side-by-side configurations, however, this is less frequent (Stukenbrock 2020). In Extract (1), addressee monitoring coincided with the addressee's request to wait (l.03). By shifting gaze to the addressee, the speaker gained visual evidence for the reason of the addressee's current unavailability. The joint attentional sequence comes to a close when the addressee displays successful reference in the subsequent turn. At this point, speaker and addressee both know that they are sharing attention on the phenomenon in question. Individual perception is transformed into mutually known perception, i.e., joint attention (Clark 1996). However, addressees may also display trouble in second position, initiate repair, or not comply with the request to share attention. This was the case in Extract (1) where compliance with the request was delayed and joint attention only established after repair resolution (on the difference between target and referent repair cf. Stukenbrock 2015).

The generic sequential ordering summarised above is based on the assumption that requests to share attention (i.e. addressee gaze summoning) receive a complying response. Although this is the preferred response, for various reasons, addressees may not follow the invitation to jointly attend to the object that the speaker wants to share. One of the instances in which addressees do not (immediately) respond by reorienting is when they are themselves attending to something else and prefer not to give it up for the sake of attention sharing. These are the cases that our study is concerned with. It investigates instances where participants are attending to different phenomena and one of them invites the other to share attention on something they find interesting. In particular, our analysis focuses on

3. On the distinction between target and referent (cf. Clark, Schreuder & Buttrick 1983, Quine 1960, Stukenbrock 2015, 2020).

how participants handle the problem of competing sites of interest that emerges when, in states of divergent foci of attention, a joint attentional sequence is invited but currently treated as dispreferred by the co-participant.

3. Data and methodology

The study has been conducted using the methodological principles of Conversation Analysis (Sacks 1992, Schegloff 2007) and Interactional Linguistics (Couper-Kuhlen & Selting 2018), which in terms of multimodality investigate how embodied resources such as gesture, gaze, body movements and the physical surroundings are used concurrently with talk in the performance of social action (Deppermann 2013, Deppermann & Streeck 2018, Goodwin 2017, Mondada 2014a, Streeck, Goodwin & LeBaron 2011). Data recording was conducted with the use of mobile eye tracking glasses (Tobii Pro Glasses 2) worn by the participants in naturally occurring interaction. Additionally, a third camera was used to document embodied conduct not visible in the eye tracking data. The recordings were synchronised into one split-screen video using Adobe Premiere Pro and imported into ELAN (Wittenburg et al. 2006) for transcription and multimodal annotation. Talk was transcribed according to GAT2 (Selting et al. 2009, Couper-Kuhlen & Barth-Weingarten 2011 see appendix A of this volume); the annotation of embodied conduct represents a simplified version of Mondada (2019; see appendix B of this volume). All participants gave written informed consent to publish transcripts and images from the recordings. Ethical review and approval were not required for this study in accordance with the local legislation and institutional requirements.

Mobile eye tracking glasses nowadays allow researchers to take this hitherto bulky experimental technology out of the lab and “into the wild” (Stukenbrock 2018a: 268), where it allows for unobtrusive, *in-situ* recordings of mundane social interaction that meet the conversation analytic criteria for data collection (Mondada 2013b). Yet, to date, most studies still rely exclusively on video recordings, which do not allow to zoom in on the details of gaze (cf. Zima, Auer & Rühlemann this volume, for a critical assessment). Consequently, conversation analytic studies on reference, joint attention and gaze based on robust, reliable eye gaze data are scarce (see, however, Balantani 2021, Balantani & Lázaro 2021, Stukenbrock 2018a, b, 2020, 2023, Stukenbrock & Dao 2019, as well as Auer & Laner this volume, and Botsch et al. in this volume). Our study fulfils the desideratum to complement video data by high-precision eye gaze data, and contributes to existing research on reference and joint attention by examining participant’ visual coordination in the face of divergent and competing foci of attention.

The data are part of a larger corpus⁴ of naturally occurring interactions in German and Swiss German collected in different settings. For the purpose of the current study, a collection of 39 instances has been assembled where participants are not attending to the same object and one of them invites the other to share attention. The data for this chapter come from a sub-corpus of recordings undertaken at the Swiss museum of games. The museum contains a selection of old and new games accompanied by descriptions. Participants walked freely around the museum in dyads. They passed through halls that exhibit different games and ended up in a room full of games that visitors can pick from and play.

4. How participants manage divergent and competing foci of attention

In what follows, we will present a collection of cases where participants are initially not attending to the same object and one of them invites the other to share attention on a phenomenon. We will show that there are different ways in which the problem of divergent and competing foci of attention arises and is treated in the unfolding interaction. In the first Section (4.1), we will focus on instances where the problem of competing foci of attention is resolved sequentially. Initially, participants have their visual attention on different phenomena. Both participants first share attention on the summoner's phenomenon and subsequently on that of the summoned participant. Another way the sequence may unfold is that, while both participants project and invite a different focus of attention, one of them prevails and the other gets abandoned (4.2). Finally, we present those rare instances where participants refrain from establishing a joint focus of attention altogether (4.3).

4.1 Sequential resolution: Both objects attended to

Extract (2) exemplifies how the participants interactionally resolve the problem of competing foci of attention. Two friends, Anna and Mike, are visiting the Swiss museum of games and are currently in front of a wall with a puzzle game. It consists of a big world map and magnetic cards picturing games of the world. Players have to match the games with their respective countries of origin by placing the cards on the correct location on the world map. In the beginning of this extract, the participants are positioned side-by-side and attend to different games. Mike

4. We thank Stefanie Lázaro, Letizia Manco and Sonja Salerno for data collection, transcription and help with data management. We gratefully acknowledge funding of the project "Deixis and Joint Attention: Vision in Interaction" (DEJA-VI; grant number: 10001F_179108) by the Swiss National Science Foundation.

(M) notices that the game scrabble has been matched to the wrong country by previous visitors, while Anna (A) picks up a new card. The participants first focus their attention on Mike's and then on Anna's game. In the figures, Mike's perspective is displayed on the left, Anna's on the right.

Extract 2. "scrabble vs. bilboquet" (SMo2_Rundgang_09:37-09:59)

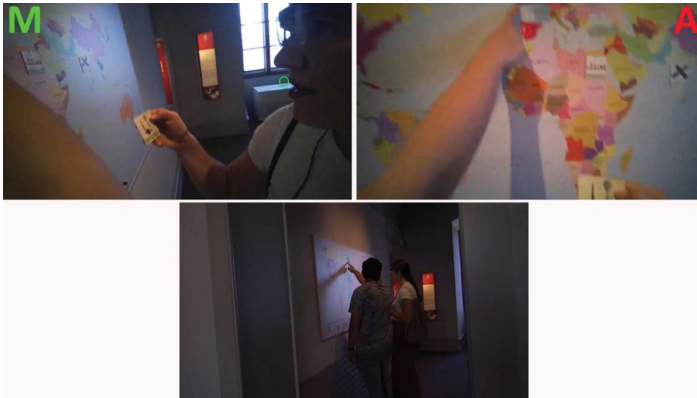
01 A-vb <<p> was isch denn +DAS ei[gentlich,>.]
what is that actually

02 M-vb [finds LUSTig]& dass sie-
I find it funny that they

a-gz >>bilboquet card----->
a-bd >>picks up, holds bilboquet card----->
m-gz +-scrabble card----->
m-ge &-points to scrabble
card--->

03 (0.3) \$ (1.0)

a-gz -bilboquet card----->
a-bd -holds card---\$-lowers hand->
m-gz -scrabble card----->
m-ge -freezes PG ----->



#5

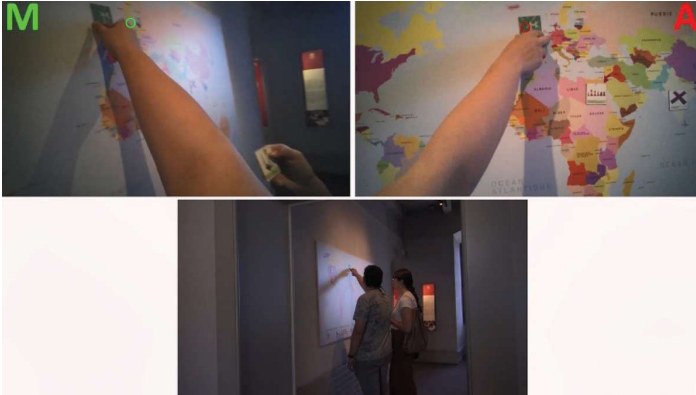
04 M-vb SCRABble vo dä <<:-)> usa+* uf> #5&
(put) scrabble from USA

a-gz -bilboquet card-----*.....
m-gz -scrabble card-----+.to A->
m-ge -----&

05 &(0.1)*+(0.3)+

a-gz*-scrabble card->
m-gz -----+.....+
m-bd &-manipulates card-->

5. "Bilboquet" is a game that consists of a spindle made of wood connected by a cord to a ball. The purpose of the game is to catch the ball on the tip of the spindle.



#6

- 06 A-vb +#6((lacht))
 ((laughs))
 a-gz -scrabble card---->
 m-gz +-scrabble card---->
 07 M-vb ÄNGlan*d <<:-> ta hä[nd];>
 on England
 08 A-vb [
 oh yes
 a-gz -----*
 m-gz -scrabble card-----+
 m-bd -manipulates scrabble card-----&, , ,----->
 09 A-vb *h h°& +EI+ ei ei;
 oh oh oh
 m-bd -----&
 m-gz +-bilboquet card+
 10 *(0.4)
 a-gz *-bilboquet card---->
 11 A-vb KENNSCH du dAs,
 do you know this
 a-gz -bilboquet card---->



#7

- 12 +#7(0.2)\$ (0.3)*
 a-bd \$OM---->
 a-gz -----*
 m-gz +bilboquet card---->
- 13 A-vb was Isch DAS?
 what is that
 a-bd -OM----->
 m-gz -bilboquet card---->
- 14 (0.4)\$ (1.6)
 m-gz -bilboquet card---->
 a-bd -----\$
- 15 M-vb <<flüsternd> BILboquet;>
 ((whispering)) bilboquet
 m-gz -bilboquet card---->
- 16 (0.1)
 m-gz -bilboquet card---->
- 17 M-vb *h a:: da+s isch das-
 oh this is the
 m-gz -----+
- 18 M-vb wo d muesch ä:-
 where you have to eh
- 19 (0.2)
- 20 M-vb d CHUGle ufe,
 the ball up
- 21 (0.6)
- 22 A-vb [oKE: ,]
- 23 M-vb [<<all> aso du HÄSCH so->]
 so you have sort of
- 24 A-vb [joo ich WEISS,]
 yes I know
- 25 M-vb [d CHUGle het so äs] 10ch und [nochär muesch s irgendwie Uffoo;]
 the ball has sort of a hole and then you have to somehow catch it
- 26 A-vb [aber guet ich has woorschinli] in
 ere NEU:artige version gmacht=-
 but okay I have probably done it in a novel version
- 27 A-vb =e bitzli AI<<dim>facher als da;>
 a bit easier than this
- 28 M-vb hehe

The participants are in a side-by-side configuration with their attention directed to different sites of interest. When Anna, in the course of picking up a card with an unfamiliar game, asks her co-participant what it represents (l.01: was isch denn DAS eigentlich,/ ‘what is that actually’) and thereby requests his visual attention, which is on another card, competing foci of attention emerge. Anna’s interrogative does not receive a response from Mike who has already initiated a pointing

gesture on the card that he is looking at. In overlap with Anna's turn, he begins a multi-unit turn (1.02: finds LUSTig dass sie-/ 'I find it funny that they') that delivers an assessment of the card's position on the map and invites Anna to shift gaze to it. Thus, two overlapping invitations to share attention compete. Note that Mike's pointing gesture is kept on the scrabble card until the end of his turn. However, his gesture is not attended to by Anna who keeps gazing at her own card. With his pointing gesture held on the card until Anna orients to it, Mike shifts gaze to Anna and monitors her visual attention (Figure 5, left). At this point, Anna shifts gaze from her own card to the one Mike is pointing at (Figure 5, right). Mike can thus see that she is now sharing attention on the erroneously placed card that he pointed out to her (Figure 5, bottom) (on gaze practices and joint attention cf. Stukenbrock 2015, 2018a, b, 2020). He then shifts gaze back to the erroneous card (Figure 6, left) that Anna is now looking at as well (Figure 6, right), picks it up and places it on the USA. Anna utters an agreement token followed by laughter (1.08) and a humorous response cry mocking trouble (1.09: *Ei ei ei*), thus displaying affiliation with Mike's stance.

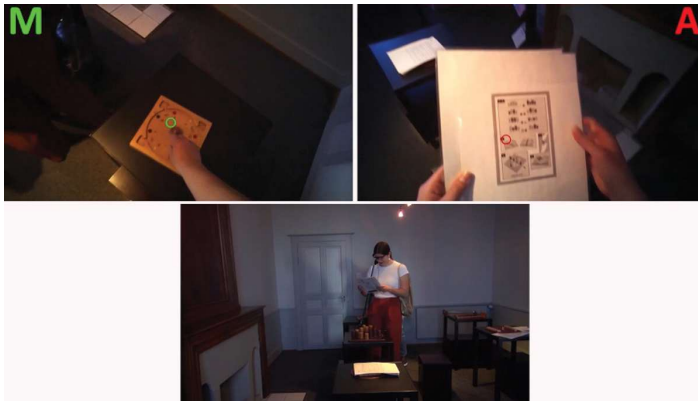
In the course of Anna's response cry, Mike now orients his gaze for 0.3s on Anna's card (1.09) before shifting it to the other cards. Anna takes this as an opportunity space and makes a second attempt to summon Mike's attention to her own card by formulating another interrogative (1.11: *KENNSCH du dAs*,/ 'do you know this'). She succeeds in summoning his gaze and joint attention on her card is successfully established (fig 7, left and right). Anna hands the card over to Mike while partially repeating her initial enquiry from 1.01 (1.13: *was Isch DAS?*/ 'what is that'). After 2.0s of inspecting the card (1.14), Mike provides an extended response to her question (1.15–25).

In sum, the participants resolve the problem of competing foci of visual attention by successively orienting to both objects. While they initiate different trajectories at the beginning of the extract, each orienting to a different card and simultaneously requesting the other's attention, Mike prioritises his own over that of his co-participant. He summons his co-participant's attention on a card on the map with an assessment and a pointing gesture that is held until Anna perceives it. Anna temporarily suspends her own trajectory in favour of Mike's and resumes it at a later stage. She affiliates (1.08) with Mike's amusement (1.02–07) that the card is on the wrong country, thus closing the sequence before summoning his attention on her own card.

Extract (3) offers another instance where the problem of competing foci of visual attention is resolved sequentially. However, in contrast to the previous extract, only one participant invites the other's visual attention. Nonetheless, they then orient to both objects one after the other. We join our two friends from the previous extract, Mike and Anna, as they are moving into an exhibition room full

of old games, looking around at different games and occasionally commenting on them. At the beginning of the extract, Anna focuses her attention on a game called “Quarto”. At that moment, he is in a face-to-back orientation with Anna (Figure 8, bottom) who then slightly turns round to establish an L-formation with him (Figure 9, bottom). In the figures, Mike’s perspective is displayed on the left, Anna’s on the right.

Extract 3. “Quarto vs. spinning top game” (SM02_Spielen_01:12–01:34)



#8

01 A-vb m mol zersch DÜreluege was mr alles [hÄ#8n;]
PTCL first we look through what we have

02 M-vb [o:h] DAS
gseht lustig *us;* oh that

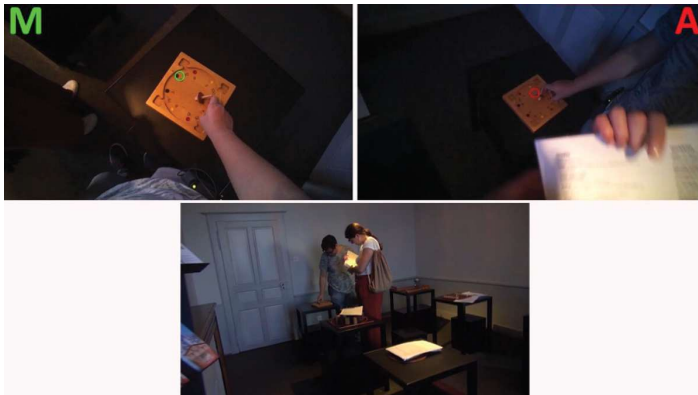
looks funny

a-gz -instructions-----*Quarto*

a-bd >>holds instructions up-->

m-gz -spinning top game-->

m-bd -playing with sp top game-->



#9

03 \$(1.0)*(1.1)#9(0.8)\$ (1.3)
a-gz *-spinning top game----->
a-bd \$-turns-----\$puts instr on chest->
m-gz -spinning top game----->
m-bd -playing with sp top game----->

04 A-vb h?:: ähm-
a-gz -spinning top game---->
a-bd -holds instr on chest->
m-gz -spinning top game---->
m-bd -playing with game---->

05 A-vb und was muess ma do <<lachtend> MAche>,
 and what do we have to do there
a-gz -spinning top game----->
a-bd -holds instr on chest----->
m-gz -spinning top game----->
m-bd -playing with game----->

06 A-vb ((lacht)) [((lacht))]
 ((laughter)) ((laughter))

07 M-vb [kei AHnig_wahrscheinlich-]
 no clue probably
a-gz -spinning top game----->
a-bd -holds instr on chest----->
m-gz -spinning top game----->
m-bd -playing with game----->

08 (1.4)&(3.0)\$ (0.2)*(0.3)
a-gz -sp top game-----*instructions----->
a-bd -----\$
m-gz -spinning top game----->
m-bd -----&

09 M-vb ((lacht))+((lacht))
 ((laughter))
m-gz -----+instructions---->

10 (1.3)+ (0.4)+ (0.8) +
m-gz -----+Quarto+instructions+



11 M-vb +&ah das isch D#10A+S;&+
 oh this is that
 m-gz +-Quarto-----+instructions+
 m-ge &....-points to Quarto&

Anna's utterance at line 01 is still part of the previous sequence; it closes down an exchange with the researcher. Simultaneously, Anna is already looking at the instructions of the game "Quarto" (Figure 8, right). In overlap with Anna's turn (l.01), Mike makes a first attempt to summon Anna's attention on the spinning top game he is gazing at (Figure 8, left, Figure 9, left). He formulates an assessment (l.02: h o::h DAS gseht lustig us;/ 'oh that looks funny') prefaced by the response cry "oh" (Goffman 1978; cf. Anna and Pfeiffer 2021 on children's use of oh-prefaced exclamatives). Anna then suspends her line of action. She puts the information card on her chest, thus indexing that she holds it back to read later, and shifts gaze to Mike's game (Figure 9, right). She formulates an interrogative (l.05: und was muess ma do <lachend>> MACHE,/'and what do we have to do there') that displays to Mike that her attention is on his object of interest.

Mike's response displays lack of knowledge (l.07: kei AHnig/ 'no clue'); concurrently, he plays around with the game trying to figure out how it works. Anna withdraws her attention and reorients to her own game (l.08). Mike then also shifts gaze to Anna's game (Figure 10, left) and delivers a noticing prefaced by a change-of-state token (Heritage 1984) (l.11: ah das isch DAS/ 'oh this is that'). The demonstrative DAS ('that') and the concurrent pointing gesture refer to Anna's game and thus signal that Mike is now sharing attention with her (Figure 9).

Thus, once again, the problem of competing foci of visual attention is resolved sequentially. While Mike invites attention sharing, Anna, by contrast, refrains from summoning the co-participant's attention. Nevertheless, they jointly attend to each other's respective games. By first sharing attention on Mike's game and then on Anna's, they delicately orient to and manage competing foci of attention, thereby displaying and enacting togetherness.

In this section, we have examined instances of competing foci of visual attention that are dealt with sequentially. Participants achieve a joint focus of attention on one phenomenon, temporarily suspending attention on the competing phenomenon, only to jointly return to it subsequently. However, sequential management of two different foci of attention may not always be the case. As we shall see in the next section, participants may also abandon one attentional focus and interactional trajectory in favour of the other.

4.2 Sequential resolution: Sharing attention on one object and abandoning the other

This section is concerned with cases in which participants summon their co-participant's visual attention on a phenomenon when the co-participant has established a divergent focus of attention that they may either also want to share or to sustain for themselves. In order to solve the problem, participants choose to share attention on only one phenomenon while abandoning the other. Extract (4) exemplifies how a participant, upon being summoned to share attention with the other, gives up her own attentional focus to co-orient with the summoner. In the extract, another pair of friends, Jasmin (J) and Melania (M), are at the Swiss museum of games. Positioned side-by-side, they have been reading the description of the game "Jassen"⁶ when Jasmin moves away from the description and orients to the exhibits while Melania remains in front of the description. Hence, the participants move from a side-by-side to a back-to-back configuration. Melania summons her friend's attention back to the description on two occasions. On the first, Jasmin shifts gaze to the description and then back to another exhibit. This is not perceived by Melania, who has her back turned on her interlocutor. On the second occasion, Jasmin not only shifts gaze to the description but also withdraws from the exhibit altogether to walk over and position herself next to Melania, thus establishing joint attention on the description. In the figures, Jasmin's perspective is displayed on the left, Melania's on the right.

Extract 4. "Jassen decks vs. X" (SMo3_Rundgang 06:12–06:31)

01 M-vb HÄ-
 m-gz >>info board---->
 j-gz >>exhibit case-->

02 (0.5)

03 M-vb sind DAS &hier die: die DEutschen⁷;=
 are these here the German
 m-ge &.....-points to info board->
 m-gz -info board----->
 j-gz -exhibit case----->

04 M-vb =und DAS hier* die französischen;=
 and these here the French.
 m-ge -freezes PG----->
 j-gz -----*...-shifts gaze.....>

6. The game "Jass" or "Jassen" is a card game that is considered as the national card game of Switzerland. It consists of a deck of 36 cards and can be played with Swiss-French or Swiss-German cards.

- 05 M-vb =o*der wie?
or *what*
j-gz ..*-gaze to M---->
m-ge -frozen PG----->
- 06 (0.2)\$(0.1)
j-gz -M----->
j-bd \$steps towards M->
m-ge -frozen PG----->
- 07 J-vb JA,
yes
m-ge -frozen PG->
j-gz -M----->
j-bd -steps towards M---->
- 08 J-vb es SCH(t)EINT so;
it seems like that
m-ge -frozen PG----->
j-gz -M----->
j-bd -steps towards M->
- 09 (0.1)&
m-ge -PG--&
j-gz -M----->
j-bd -steps towards M--->
- 10 J-vb nach den FARBe\$n;=
according to the colours
j-bd -steps towards M-\$-stops->
j-gz -M----->
- 11 J-vb =*JA,
yes
j-gz *-shift to exhibit case...
- 12 (0.4)*(1.7)
j-gz*exhibit case->
- 13 J-vb m?
- 14 (0.2)
- 15 M-vb °h aber es GIBT doch auch öh::m-
but there are also
j-gz -exhibit case----->
- 16 (0.4)*(0.7)\$(0.4)*(0.3)
j-gz -----*.....*-to M----->
j-bd \$-steps towards M->

7. Melania's question refers to the German variant of the card game "Jass" (see footnote 6).



#11

- 17 M-vb (jA&ssen: de\$cks in) DEN fa#11&rben hier.
(Jassen decks in) these colours here
m-ge &....-PG to info board---&
j-bd ----->\$
j-gz -to M----->>
- 18 M-vb =oder?
right
j-gz -to M---->
- 19 (0.8)
- 20 J-vb üf::
- 21 (0.2)
- 22 J-vb fragst du MICH.*
you are asking me
j-gz -to M-----*
- 23 (0.8)
- 24 J-vb ich kenn das alles äh [Überhaupt nicht-]
I do not know all this at all
- 25 M-vb [ich frag daNACH mal,]
I will ask about it later

While Jasmin has moved away to the exhibits, Melania is still standing in front of the description they have been reading. The participants thus have no visual access to each other and their attention is on different entities. Prefaced by a response cry (Goffman 1978) indexing a problem of understanding (l.01: HÄ-), Melania produces a yes/no interrogative (Raymond 2003). It consists of two TCUs formulated with the demonstrative DAS ('that') (l.03–05: sind DAS hier die: die DEutschen; und DAS hier die französischen;=oder wie?/ 'are these here the German and here the French or what').⁸ The demonstratives are used gesturally

8. The demonstratives refer to two different decks of Jass cards, the Swiss German and the French variant, see footnote 6.

(Fillmore 1997[1971]): i.e. they are accompanied by a pointing gesture and request addressee gaze (Stukenbrock 2018a, 2020).

The pointing gesture is held until Jasmin responds by producing an epistemically modalized affirmation (1.07–08). Note that in order to respond, Jasmin has to turn round and visually identify what her friend is referring to. At the second demonstrative (1.04), Jasmin shifts gaze from the exhibit to Melania and, in the course of her response, takes a few steps towards her before turning away to another exhibit case. Melania, who has kept her back on Jasmin, can only infer from verbal (the response) and auditory cues (Jasmin's voice sounding closer to her), that her pointing gesture has been perceived and joint attention been established.

A few seconds later, Melania re-initiates talk on the description and produces a broken-off reformulation of the problem (1.15: *‘h aber es GIBT doch auch öh::m-/ ‘but there are also’*). Since Jasmin has turned her gaze to the exhibit case next to her (1.12), the participants' visual attention is on divergent sites again.

0.4s after Melania's unfinished turn (1.15), Jasmin shifts gaze to Melania (1.16) and walks towards her. At this point, Melania continues the broken-off utterance and delivers the projected direct object (1.17: (*jAssen: decks in*) DEN *farben hier./ ‘(Jassen decks in) these colours here’*). Her turn is designed in such a way – gesturally used demonstrative (DEN/ ‘these’) with concurrent pointing gesture – as to request the co-participant's gaze. Jasmin, who is by now standing next to Melania (Figure 11, bottom), has re-oriented her gaze to the description (fig. 11, left). Jasmin's subsequent response, while documenting that joint attention has been achieved, also displays lack of knowledge about the game (1.20–24) and thereby closes the sequence. Melania aligns and proposes to inquire later (1.25).

To sum up, while the two participants have their visual attention on different sites of interest, when Melania summons her co-participant's attention, Jasmin abandons her objects of interest, shifts gaze to her co-participant, moves towards her and responds. She thus orients to her co-participant's means of mobilising a response and requesting visual attention for it.

In Extract (4), the summoned participant abandoned her focus of attention in order to comply with her friend's request to share attention on a different object as part of the conditional relevance established by questions on that very object. The next example exemplifies another instance where participants, instead of co-orienting to both sites of interest successively, only share attention on one object. In contrast to the previous extract, however, the summoned participant, instead of abandoning her focus of attention altogether, suspends it to share attention with the co-participant and then returns to her own object of attention. The participants, Carola and Torsten, have finished visiting the exhibition at the Swiss museum of games and are now in the game room full of board games to choose from in order to play together. In the figures, Carola's perspective is displayed on the left, Torsten's on the right.

Extract 5. “Fishing game vs. Helvetiq” (SMo1_Spielen_02:22–02:34)

01	C-vb	<<cresc> oh gott ist das alles komplizIE:RT,> oh god this is all complicated
	c-gz	>>helvetiq----->
	t-gz	>>fishing game----->
02	T-vb	<<dim> CArola ich hab das Ultima*tive Spiel gefunden;> Carola I have found the ultimate game
	c-gz	-helvetiq-----*fishing game ----->
	t-gz	-fishing game----->
03		(0.1)



```

#12
04 C-vb #12ja +was IST +es?
    yes what is it
    c-gz -fishing game----->
    t-gz -----+shelf--+fishing game->

05 (1.2)*
    c-gz -fishing game*

06 C-vb ((lacht)) [ ((lacht)) ] ((lacht))*((lacht))
    ((laugter))
    c-gz *-fishing game->

07 T-vb [ein Angelspiel;]
    a fishing game

08 (0.3)

09 C-vb a+HA;*
    c-gz -fishing game*
    t-gz --scanning shelves-->

10 (0.3)*(0.1)
    c-gz *-helvetiq----->

11 C-vb die sehen alle so komplIZIERT aus,
    they all look so complicated

12 (0.2)

13 C-vb find_s du NICHT?
    don't you think

```

The participants are in a back-to-back configuration several steps apart orienting to different games they have picked up. They are in an open state of talk when Carola formulates an assessment about the games being too complicated (1.01), which does not receive a second assessment. Instead, Torsten summons Carola's attention to the game he is holding and looking at (1.02: *Carola ich hab das Ultimative Spiel gefunden*/ 'Carola I've found the ultimate game'). Carola responds with a body torque (Figure 12, bottom; cf. Schegloff 1998) and shifts gaze to the game Torsten is holding (Figure 12, left). While keeping her lower body oriented to her own game and holding on to the instructions in her hand, she asks for more information (1.04: *ja was IST es?*/ 'yes what is it'). However, before Torsten responds (1.07), she begins to laugh and thus displays that she does not treat Torsten's game as a legitimate candidate for joint play but as a joking response to her assessment (1.01) that the games seem too difficult. Subsequently, she reorients to her own game and reformulates her assessment of the games (1.11).

This extract exemplifies that participants do not necessarily abandon their focus of attention in favour of that of their interlocutor altogether but may temporarily suspend their line of action only to resume it subsequent to the attention sharing sequence. Note, first, that the attention sharing sequence in this extract is much shorter than the one in Extract (4), and, second, that Carola's body torque in response to her co-participant's summons projects the return to her own focus of attention.

In this section, we have investigated instances in which the sequential resolution of competing foci of attention does not involve participants mutually orienting to each other's objects subsequently as in Section 4.1. Instead, one participant either abandons her own focus of attention (Extract (4)), or temporarily suspends it in favour of attention sharing, only to return to it, however, without reciprocal co-orientation from her co-participant (Extract (5)). The last section demonstrates that participants with divergent foci of attention may orient to the emergence of competing sites of interest by disattending to the invitation of attention sharing in favour of their own focus of attention.

4.3 Lack of attention sharing

The last two excerpts illustrate the rare instances in which the problem of competing foci of attention does not get resolved interactionally in favour of attention sharing. Instead, the participants display to each other moment-by-moment that they are and continue to be engrossed in something else.

Extract (6) comes from the same recording as Extract (4). The participants, Jasmin and Melania, are in a back-to-back configuration (Figure 13, bottom) that emerged as Jasmin turned away from the showcase in the centre that they were

looking at before, and walked to two showcases on the wall to the right (Figure 14). The participants are in a state of divergent foci of attention when Jasmin invites her co-participant's attention by formulating a noticing. Melania, however, first delivers a noticing herself before responding to Jasmin. Hence, both participants treat their own phenomenon as more interesting and give it priority over the invitation to share attention with the co-participant.

Extract 6. "old cards vs. bride" (SM03_Rundgang_o6:47-07:01)



- #13
 01 J-vb OH;#13
 j-gz >>cards left exhibit case on wall->
 m-gz >>exhibit case room centre----->
 02 J-vb voll schöne alte KARTen hier,
 very nice old cards here
 j-gz -cards left exhibit case on wall->
 m-gz -exhibit case room centre----->
 03 (1.9) *
 j-gz -cards left exhibit case on wall*
 m-gz -exhibit case room centre----->
 04 M-vb *<<laughingly> BRAUT.>
 bride
 j-gz *cards right exhibit case on wall->
 m-gz -exhibit case room centre----->
 05 (0.5)
 j-gz -cards right exhibit case on wall->
 m-gz -exhibit case room centre----->
 06 M-vb °hh
 j-gz -cards right exhibit case on wall->
 m-gz -exhibit case room centre----->
 07 (3.5)+(0.4)
 m-gz -----+.....
 j-gz -cards right exhibit case on wall-->



#14

- 08 M-vb +W#140?
 where
 m-gz +-exhibit cases on the wall----->
 j-gz -cards right exhibit case on wall---->
- 09 (0.4)
 j-gz -cards right exhibit case on wall---->
- 10 M-vb &AH &ja?.
 oh yes
 j-gz -cards right exhibit case on wall---->
 m-bd &turns&steps towards cards----->
- 11 (0.4*\$(0.1)\$
 j-gz -----*
 j-bd \$turns\$
- 12 J-vb \$m?
 j-bd \$-moves away---->>

In this example, the participants do not establish a joint focus of attention. Instead of mutually synchronising their movement, bodily orientation, and attention to promote joint attention and a shared understanding, they remain oriented to their own objects of interest. They are in a back-to-back configuration (Figure 13, bottom) with Melania still focused on the game in a showcase (Figure 13, right) they have been looking at together when Jasmin, who turned away towards a showcase on the wall (Figure 13, left), delivers an *oh*-prefaced (l.01) noticing (cf. Anna & Pfeiffer 2021) that assesses exhibits in a new showcase (l.02: *voll schöne alte KARTen hier, / 'very nice old cards here'*). However, her noticing does not receive a response. Instead, Melania's visual and bodily orientation is kept on the card game in the previous showcase (Figure 13, right). After a pause of 1.9s, Melania, in turn, verbally marks the relevance of her own site of interest by naming one of the cards (l.04: <<laughingly> *BRAUT.*> / 'bride'). While thus displaying her unavailability for attention sharing, the accompanying laughter may also be heard as a (counter-)invitation to share attention and affiliation. This, however, is not

acknowledged by her co-participant who neither responds nor shifts gaze back to Melania (Figure 13–14, left).

Hence, the competing foci of attention that emerge from summoning the other in a state of divergent foci of attention is not being addressed by either participant. Only after 4.4s does Melania acknowledge Jasmin's noticing (l.o1) by initiating repair (l.o8: WO?/ 'where'), turning her upper body (Figure 14, bottom) and shifting her gaze to the exhibits behind her (Figure 14, right). However, when Melania turns her attention to the exhibit case on the wall, Jasmin is about to disengage from it and is moving away.

To conclude, since Melania's response to Jasmin's noticing comes late and in a sequentially non-adjacent position, thus violating the preference for progressivity and contiguity in favour of her own trajectory of action, and since Jasmin neither co-oriens with Melania nor waits until Melania is ready to co-orient with her, the participants do not establish joint attention.

The last example offers another instance of competing foci of attention that emerge from a state of divergent attention and do not get resolved. We join the participants, Carola (C) and Torsten (T), on their way through the Swiss museum of games. Like the participants in the previous extract, they have to manage withdrawal from and dissolution of joint attention on the previous exhibit and transition to the next. In contrast to Extract (6) where joint departure is not achieved and one participant remains oriented to the previous exhibit, the participants in Extract (7) both move on.

Extract 7. "playing area vs. exhibit on the wall" (SMo1_Rundgang_o8:54–09:09)

01 T-vb °hhh*
c-gz >>exhibit case*
t-gz >>exhibit case->
02 (0.8)*+(1.0)
c-gz*-playing area----->
t-gz -----+exhibit on wall-->



- 03 C-vb ah und &HIER#15& kann man dann SPIEL*en;
 ah and here you can then play
 c-gz -playing area-----*
 c-ge &points-&
 t-gz -exhibit on wall----->
- 04 (0.2)
- 05 C-vb <<p>oKEE;>
 okay
- 06 (7.8)
- 07 C-vb was war denn !DEIN!+ lieblingsspiel Torsten,
 what was your favourite game Torsten
 t-gz -exhibit on wall---+

The participants have been sharing attention to objects displayed on an exhibit case in the middle of the room. Subsequently, they withdraw from it with Carola taking the lead and Torsten following her in a face-to-back orientation (Figure 15, bottom). While Carola orients to a small playing area in front of her (Figure 15, right), Torsten, in contrast, turns to the left to look at an exhibit case on the wall (Figure 15, left). The open state of talk ends as Carola launches an *ah*-prefaced noticing (1.03: *ah und HIER kann man dann spielen*; / ‘ah and here you can then play’). The deictic (1.03 *HIER* / ‘here’) refers to the playing area; it is accompanied by a pointing gesture and invites joint attention. Torsten, however, does not respond. Instead of shifting gaze to identify the referent, he keeps looking at exhibits on the wall. He neither acknowledges Carola’s noticing, nor does he abandon his focus of attention in favour of the space made relevant by Carola and the category-bound activity implied by it. After a short pause (1.04) and a sequence-closing *okay* (1.05) (cf. Mondada & Sorjonen 2021), Carola moves into the next room.

This section has demonstrated that in rare instances, participants do not respond to the summons of sharing attention by abandoning or suspending their own focus of attention in favour of the summoner’s. Instead, they may prefer to keep their visual attention on their own site of interest. Instead of sequential resolutions that imply either both participants attending to competing sites of interest subsequently or sharing attention on one site only while disregarding the other, competing foci of attention may be sustained and joint attention prevented by delay (Extract (6)) or lack of uptake altogether (Extract (7)).

5. Discussion

The chapter has investigated how participants manage the establishment of joint attention in the face of competing foci of attention. Specifically, it has examined instances where the participants are not attending to the same object and invite the co-participant to share attention on a phenomenon that they find interesting. We have shown that there are various ways in which competing foci of attention sequentially unfold from states of divergent attention. Participants may resolve the conflict by cooperating and coordinating their lines of action in order to share attention on one phenomenon and then on the other. This was the case in Extracts (2) and (3). Significantly, in these instances, the embodied configuration of the participants – side-by-side in Extract (2), L-formation (Kendon 1990) in Extract (3) – afforded minimal gaze shifts between the two competing foci of attention and was therefore favourable to a sequential resolution where joint attention was accorded to both sites of interest successively. The bodily configuration also enabled addressee gaze monitoring (Extract (2)) as an interactional resource participants may draw on to confirm that shared perception is emerging.

Another way in which the sequence can unfold is by sharing attention on one phenomenon and abandoning the other. In Extract (4), the addressee's attention was summoned by her co-participant on two occasions. In contrast to Extracts (2) and (3), the participants were several metres apart in a back-to-back-configuration with no visual access to each other. On the first occasion, the addressee shifted gaze to the summoner's object and delivered an epistemically modalized response; only on the second occasion did she give up her own line of action altogether to move towards the summoner and look in more detail at the object in question. In the course of the entire sequence, the summoner remained firmly oriented to her site of interest without bodily turning to her co-participant. Addressee (gaze) monitoring did not occur. The same holds for Extract (5). At the outset, the participants were in a back-to-back configuration several meters apart. While the summoning participant upheld his attentional focus without turning to and monitoring the addressee, the latter complied with the request to share attention with a body torque, gaze shift, and a verbal response. In contrast to Extract (4) where the second summons occasioned that the addressee abandon her own line of action altogether and move over to the speaker, the addressee in Extract (5) returned to her own object of interest afterwards. In these instances, participants jointly attend to only one site of interest, either by abandoning their own line of action as in Extracts (1) and (4), or by suspending it as in Extract (5).

Finally, in rare instances, there may be no resolution to the competing foci of attention. Each participant may treat their own site as more interesting or pressing; participants thus refrain from establishing joint attention on either of the

relevant targets altogether. This was exemplified in Extracts (6) and (7). Like in Extracts (4) and (5), the participants were in spatial configurations that bodily reflected and established individual involvement: back-to-back in Extract (6), and face-to-back in Extract (7). In contrast to the previous extracts, the participants did not only negotiate competing sites of interest. Moreover, the problem of competing sites of interest is intricately connected to co-operatively managing withdrawal from exhibits (Vom Lehn 2013), transitions and trajectories from mobile to stationary phases. While the addressee in Extract (7) did not respond at all, thereby treating his co-participant's noticing on the gaming-area as interactionally inconsequential, the addressee in Extract (6) responded with significant delay, a feature that was also observed in Extract (1). However, in contrast to Extract (1) where joint attention finally did occur, the participants in Extract (6) failed to coordinate embodied actions, visual attention as well as affect and stance. In spite of the resemblances – both Extract (1) and Extract (6) exhibit temporal delay and subsequent repair –, they differ with respect to the participants' co-ordination, co-orientation, and collaboration. In Extract (1), the summoning participant waits until the addressee, who accounts for the delay and projects its end, joins her, the summoner in Extract (6) withdraws precisely when the addressee finally turns to her. Their trajectories cross, but do not meet, and neither do attention, affect, and cognition (Kita 2003).

With respect to turn design, the instances we have presented are formulated with strong response mobilising components, such as gesturally used deictics (Extracts (1), (2), (3), (4), (6), (7)) that request addressee gaze (Stukenbrock 2020), perceptual directives (Extract (1)), response cries (Extracts (3), (4), (6), (7)), noticings (Extracts (1), (2), (3), (4), (6), (7)), assessments (Extracts (2), (3), (5), (6)), and questions (Extracts (2), (4)) – combined with pointing gestures (Extracts (1), (2), (3), (4), (7)) or object manipulations (Extracts (2), (3), (5)). There are also instances in our collection with less response mobilising features, such as naming an object (not presented in this chapter). While these mark an attentional focus that is also bodily displayed, they are less likely to receive a response.

Based on our empirical observations on the interactional work required to establish joint attention in the face of competing sites of interest, we propose that there might be a continuum of more to less response mobilising practices. While gesturally used demonstratives no doubt request visual attention (i.e., gaze allocation) in order for addressees to identify the referent, understand the speaker's action, and deliver an adequate response (Stukenbrock 2015, 2020), the gaze mobilising quality of deictics is further enhanced by way of turn design and social action format such as, e.g. assessments and questions, which make a type-related response (Schegloff 2007) conditionally relevant. By contrast, namings as well as noticings that lack additional gaze/response mobilising features, could also be

treated as self-talk or interactionally not pertinent. However, more research is required to confirm this proposal.

Instances of divergent and competing foci of attention constitute conspicuous sites for the investigation of practices that participants consider suitable to solicit and re-engage the co-participant (Stukenbrock 2023). Applying mobile eye tracking technology to the analysis of these moments has revealed details of the participants' gaze behaviour as constitutive of divergent, competing, and joint attentional sequences. Participants' verbal and embodied practices exhibit their orientation to attention sharing as a central ability, social tool, and foundational building block of the cooperative infrastructure of human communication. To complement the picture, the situated choice of less attention-mobilising practices can be viewed as embodying deference to the co-participant's involvement with their own objects of interest.

References

- Anna, Marina, and Martin Pfeiffer. 2021. "Die frühkindliche Verwendung des Exklamativformats oh+X: Noticings und Problemmitteilungen." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 22: 1–35.
-  Auer, Peter, Barbara Laner, Martin Pfeiffer, and Kerstin Botsch. 2024. "Noticing and Assessing Nature: A Multimodal Investigation of the Format "Perception Imperative + Exclamative" based on Mobile Eye-Tracking Data." In *New Perspectives in Interactional Linguistic Research*, ed. by Margret Selting, and Dagmar Barth-Weingarten, 245–275. Amsterdam, Philadelphia: Benjamins.
-  Auer, Peter, and Anja Stukenbrock. 2022. "Deictic Reference in Space." In *Pragmatics of Space*, ed. by Andreas H. Jucker, and Heiko Hausendorf, 23–61. Berlin, Boston: Mouton De Gruyter.
-  Balantani, Angeliki. 2021. "Reference Construction in Interaction: The Case of Type-indicative "so"." *Journal of Pragmatics* 181: 241–258.
-  Balantani, Angeliki, and Stefanie Lázaro. 2021. "Joint Attention and Reference Construction: The Role of Pointing and "so"." *Language and Communication* 79: 33–52.
- Botsch, Kerstin, Peter Auer, Barbara Laner and Martin Pfeiffer. This volume. "Joint attention without language? On intersubjectivity and the joint experience of nature."
-  Clark, Herbert H. 1996. *Using Language*. Cambridge, MA: Cambridge University Press.
- Clark, Herbert H. 2003. "Pointing and Placing." In *Pointing. Where Language, Culture, and Cognition Meet*, ed. by Sotaro Kita, 243–268. Hillsdale NJ: Erlbaum.
-  Clark, Herbert H., and Meredyth A. Krych. 2004. "Speaking while Monitoring Addressees for Understanding." *Journal of Memory and Language* 50 (1): 62–81.
- Clark, Herbert H., and Catherine R. Marshall. 1981. "Definite Reference and Mutual Knowledge." In *Elements of Discourse Understanding*, ed. by Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag B, 10–63. Cambridge: Cambridge University Press.

- doi Clark, Herbert H., Robert Schreuder, and Samuel Buttrick. 1983. "Common Ground and the Understanding of Demonstrative Reference." *Journal of Verbal Learning and Verbal Behaviour* 22: 245–258.
- Couper-Kuhlen, Elizabeth, and Dagmar Barth-Weingarten. 2011. "A System for Transcribing Talk-in-Interaction, Translated and Adapted for English by Elizabeth Couper-Kuhlen and Dagmar Barth-Weingarten." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 12: 1–51.
- doi Couper-Kuhlen, Elizabeth, and Margret Selting. 2018. *Interactional Linguistics. Studying Language in Social Interaction*. Cambridge: Cambridge University Press.
- doi Deppermann, Arnulf. 2013. "Multimodal Interaction from a Conversation Analytic Perspective." *Journal of Pragmatics* 46 (1): 1–7.
- doi Deppermann, Arnulf, and Jürgen Streeck (eds). 2018. *Time in Embodied Interaction: Synchronicity and Sequentiality of Multimodal Resources*. Amsterdam: Benjamins.
- doi De Stefani, E. 2014. "Establishing Joint Orientation towards Commercial Objects in a Self-service Store: How Practices of Categorisation Matter." In *Interacting with Objects*, ed. by Maurice Nevile, Pentti Haddington, Trine Heinemann, and Mirka Rauniomaa, 271–294. Amsterdam, Philadelphia: Benjamins.
- doi De Stefani, Elwys, and Arnulf Deppermann. 2021. "Les gestes de pointage dans un environnement changeant et éphémère: les leçons de conduit." *Langage et Société* 2: 141–166.
- doi Diessel, Holger. 2006. "Demonstratives, Joint Attention, and the Emergence of Grammar." *Cognitive Linguistics* 17: 463–489.
- doi Diessel, Holger, and Kenny R. Coventry. 2020. "Demonstratives in Spatial Language and Social Interaction: An Interdisciplinary Review." *Frontier in Psychology* 11.
- doi Eriksson, Mats. 2009. "Referring as Interaction: On the Interplay between Linguistic and Bodily Practices." *Journal of Pragmatics* 41 (2): 240–262.
- Fillmore, Charles. 1997[1971]. *Lectures on Deixis*. Stanford, CA: University of Chicago Press.
- doi Fricke, Ellen. 2007. *Origo, Geste und Raum*. Berlin, New York: De Gruyter.
- Goffman, Erving. 1963. *Behaviour in Public Places. Notes on the Social Organization of Gathering*. New York: Free Press.
- doi Goffman, Erving. 1978. "Response Cries." *Language* 54 (4): 787–815.
- Goffman, Erving. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Goodwin, Charles. 1981. *Conversational Organization: Interaction between Speakers and Hearers*. New York: Academic Press.
- doi Goodwin, Charles. 2003. "Pointing as Situated Practice." In *Pointing. Where Language, Culture, and Cognition Meet*, ed. by Sotaro Kita, 217–241. Mahwah, NJ: Erlbaum.
- doi Goodwin, Charles. 2007. "Environmentally Coupled Gestures." In *Gesture and the Dynamic Dimension of Language. Essays in Honor of David McNeill*, ed. by Susan D. Duncan, Justine Cassell, and Elena T. Levy, 195–212. Amsterdam, Philadelphia: Benjamins.
- doi Goodwin, Charles. 2017. *Co-Operative Action*. Cambridge: Cambridge University Press.
- doi Goodwin, Marjorie Harness, and Charles Goodwin. 2012. "Car Talk: Integrating Texts, Bodies, and Changing Landscapes." *Semiotica* 191 (1): 257–286.

-  Hausendorf, Heiko. 2013. "On the Interactive Achievement of Space — and its Possible Meanings." In *Space in Language and Linguistics. Geographical, Interactional, and Cognitive Perspectives*, ed. by Peter Auer, Martin Hilpert, Anja Stukenbrock, and Benedikt Szmrecsanyi, 276–303. Berlin, Boston: De Gruyter.
- Hausendorf, Heiko, Lorenza Mondada, and Reinhold Schmitt (eds). 2012. *Raum als interaktive Ressource*. Tübingen: Narr Francke Attempto.
- Heritage, John. 1984. "A Change of State Token and Aspects of Its Sequential Placement." In *Structures of Social Action*, ed. by J. Maxwell Atkinson, and John Heritage, 299–345. Cambridge: Cambridge University Press.
-  Hindmarsh, Jon, and Christian Heath. 2000. "Embodied Reference: a Study of Deixis in Workplace Interaction." *Journal of Pragmatics* 32 (12): 1855–1878.
-  Hindmarsh, Jon, Christian Heath, and Mike Fraser. 2006. "(Im)Materiality, Virtual Reality and Interaction: Grounding the 'Virtual' in Studies of Technology in Action." *The Sociological Review* 54 (4): 795–817.
- Keating, Elizabeth. 2015. "The Role of the Body and Space in Digital Multimodality." In *The Routledge Handbook of Language and Digital Communication*, ed. by Alexandra Georgakopoulou, and Tereza Spilioti, 259–272. London: Routledge.
-  Kendon, Adam. 1967. "Some Functions of Gaze-direction in Social Interaction." *Acta Psychologica* 26: 22–63.
- Kendon, Adam. 1990. *Conducting Interaction. Patterns of Behaviour in Focused Encounters*. Cambridge: Cambridge University Press.
-  Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press, Cambridge.
-  Kita, Sotaro (ed.). 2003. *Pointing: Where Language, Culture and Cognition Meet*. Erlbaum, Mahwah, NJ.
- Laner, Barbara. 2022. "Guck mal der Baum" — Zur Verwendung von Wahrnehmungsimperativen mit und ohne 'mal'. *Gesprächsforschung — Online-Zeitschrift zur verbalen Interaktion* 23: 1–35.
-  Luff, Paul, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Oyama Shinya. 2003. "Fractured Ecologies: Creating Environments for Collaboration." *Human-Computer Interaction* 18 (1–2): 51–85.
-  Mondada, Lorenza. 2012. "Deixis: An Integrated Interactional Multimodal Analysis." In *Prosody and Embodiment in Interactional Grammar*, ed. by Pia Bergmann, Jana Brenning, Martin Pfeiffer, and Elisabeth Reber, 173–206. Berlin, Boston: De Gruyter.
-  Mondada, Lorenza. 2013a. "Interactional Space and the Study of Embodied Talk-in-Interaction." In *Space in Language and Linguistics. Geographical, Interactional, and Cognitive Perspectives*, ed. by Peter Auer, Martin Hilpert, Anja Stukenbrock, and Benedikt Szmrecsanyi, 247–275. Berlin, Boston: De Gruyter.
-  Mondada, Lorenza. 2013b. "The Conversation Analytic Approach to Data Collection." In *The Handbook of Conversation Analysis*, ed. by Jack Sidnell, and Tanya Stivers, 32–56. Wiley-Blackwell, Chichester.
-  Mondada, Lorenza. 2014a. "The Local Constitution of Multimodal Resources for Social Interaction." *Journal of Pragmatics* 65: 137–156.
-  Mondada, Lorenza. 2014b. "Instructions in the Operating Room: How the Surgeon Directs their Assistant's Hands." *Discourse Studies* 16: 131–161.

- Mondada, Lorenza. 2019. "Conventions for Transcribing Multimodality." Accessed 13 March 2023. Available at: <https://www.lorenzamondada.net/multimodal-transcription>
-  Mondada, Lorenza, and Marja-Leena Sorjonen. 2021. "OKAY in Closings and Transitions." In *OKAY across Languages: Toward a Comparative Approach to its Use in Talk-in-Interaction*, ed. by Emma Betz, Arnulf Deppermann, Lorenza Mondada, and Marja-Leena Sorjonen, 94–127. Amsterdam, John Benjamins.
- Moore, Chris, and Philip. J. Dunham (eds). 1995. *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum Associates, Inc.
-  Pfeiffer, Martin. 2016. "The Deictic Dimension of Exclamations: On the Use of wh-exclamatives in German Face-to-Face Interaction." *Revue de Sémantique et Pragmatique* 40: 35–57.
- Quine, Willard van Orman. 1960. *Word and Object*. Cambridge, MA: Massachusetts Institute of Technology Press.
-  Raymond, Geoffrey. 2003. "Grammar and Social Organization: Yes/No Interrogatives and the Structure of Responding." *American Sociological Review* 68 (6): 939–967.
- Sacks, Harvey. 1992. *Lectures on Conversation, Vol. I*, ed. by Gail Jefferson. Oxford: Basil Blackwell.
- Schegloff, Emanuel A. 1998. "Body Torque." *Social Research* 65, 535–596.
-  Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg R. Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock, and Susanne Uhmman. 2009. "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)" *Gesprächsforschung* 10: 353–402. Available at: <http://www.gespraechsforschung-ozs.de/heft2009/heft2009.html>
-  Streeck, Jürgen. 2002. "Grammars, Words, and Embodied Meanings: on the Uses and Evolution of so and like." *Journal of Communication* 52: 581–596.
-  Streeck, Jürgen. 2009. *Gesturecraft: The manu-facture of meaning*. Amsterdam / Philadelphia, John Benjamins Publishing.
-  Streeck, Jürgen. 2014. "Mutual Gaze and Recognition: Revisiting Kendon's 'Gaze Direction in Two-person Conversation.'" *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*, ed. by Mandana Seyfeddinipur, and Marianne Gullberg, 35–56. Amsterdam: John Benjamins.
- Streeck, Jürgen, Charles Goodwin, and Curtis LeBaron (eds). 2011. *Embodied Interaction: Language and Body in the Material World*. Cambridge: Cambridge University Press.
-  Stukenbrock, Anja. 2014. "Take the Words out of my Mouth: Verbal Instructions as Embodied Practices." *Journal of Pragmatics* 65: 80–102.
-  Stukenbrock, Anja. 2015. *Deixis in der face-to-face-Interaktion*. Berlin, Boston: De Gruyter.
- Stukenbrock, Anja. 2018a. "Mobile Dual Eye-Tracking in Face-to-Face Interaction: The Case of Deixis and Joint Attention." In *Advances in Interaction Studies*, ed. by Geert Brône, and Bert Oben, 265–300. Amsterdam, Philadelphia: Benjamins.

-  Stukenbrock, Anja. 2018b. "Forward-looking: Where Do We Go with Multimodal Projections?." In *Time in Embodied Interaction. Synchronicity and Sequentiality of Multimodal Resources*, ed. by Arnulf Deppermann, and Jürgen Streeck, 31–68. Amsterdam, Philadelphia: Benjamins.
-  Stukenbrock, Anja. 2020. "Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention." *Frontiers in Psychology* 11:1779.
-  Stukenbrock, Anja. 2023. "Temporality and the Cooperative Infrastructure of Human Communication: Noticings to Delay and to Accelerate Onward Movement in Mobile Interaction." *Language and Communication* 92: 33–54.
-  Stukenbrock, Anja, and Anh Nhi Dao. 2019. "Joint Attention in Passing. What Dual Mobile Eye-Tracking Reveals about Gaze in Coordinating Embodied Activities at the Market." In *Embodied Activities in Face-to-Face and Mediated Settings*, ed. by Cornelia Gerhardt, and Elisabeth Reber, 177–213. Basingstoke: Palgrave Macmillan.
-  Tomasello, Michael. 2008. *Origins of Human Communication*. Cambridge, MA: MIT Press.
-  Vom Lehn, Dirk. 2013. "Withdrawing from Exhibits: The Interactional Organisation of Museum Visits." In *Interaction and Mobility: Language and the Body in Motion*, ed. by Pentti Haddington, Lorenza Mondada, and Maurice Nevile, 65–90. Berlin, Boston: De Gruyter.
- Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Kalssmann, and Han Sloetjes. 2006. "ELAN: a Professional Framework for Multimodality Research." In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, 1556–1559.

Joint attention without language?

On intersubjectivity and the joint experience of nature

Kerstin Botsch,¹ Peter Auer,² Barbara Laner²
& Martin Pfeiffer³

¹ Black Forest National Park | ² University of Freiburg |

³ University of Potsdam

We discuss the question of whether the intersubjective experience of nature requires language or whether it can be achieved by bodily means alone. Based on mobile eye-tracking data and audio recordings from walks in nature, we focus on noticings. We claim that two levels of intersubjectivity are involved in noticings. On the first level, co-participants can establish joint attention without language, by observing each other's bodily behaviour, such as gaze, body movements and bodily orientation. Following Schütz's concepts of common sense thinking and typification, we argue that in such cases walkers rely on shared knowledge, for instance based on previous experiences. On the second level, we show that language is necessary to take co-participants from joint attention to joint experience.

Keywords: joint attention, noticings, intersubjectivity, walking and talking, experience of nature

1. Introduction

It is generally assumed that joint attention, one of the most basic features of human interaction and human sociability (cf. Tomasello 2005, 2008, H. Clark 1996, E. Clark 2015: 332), is established in a complex interplay of linguistic resources, such as demonstratives and perception imperatives, and bodily means, such as gaze, pointing and bodily orientation. This interplay is assumed to be crucial for directing the recipient's visual attention to an object in the surroundings through noticings, showings, demonstrations, and other actions. In our chapter, we discuss the question of whether joint attention is dependent on language or whether it can be achieved by bodily means alone.

Although joint attention without language has been described in pre-verbal child development (e.g. Scaife & Bruner 1975, Flom et al. 2007), we are not aware of any empirical study that systematically considers this possibility in spontaneous interaction among adults. In this chapter, we focus on noticings which appear to solely rely on the visual monitoring of co-participants, among walkers on a nature walk. Walker A, for example, may slow down, stop, turn the head, and gaze in one direction for a longer time. This bodily behaviour is observable and interpretable by walker B as indicating that A has discovered, is observing, finds interesting, and so on, a feature of the natural surroundings. In turn, walker B can, for instance, follow walker A's gaze, look in the same direction, and – again only using bodily resources – indicate to A that they are looking at the same feature, i.e. that joint attention has been achieved. The following verbal utterance provides evidence that joint attention has been established by presupposing it.

Establishing joint attention without language raises a number of questions, compared to noticings that also rely on verbal means.

- a. While verbal noticings can be assumed to be intentionally meaningful, the bodily attention one person gives to an aspect of the surroundings (for instance by prolonged gaze) is not *eo ipso* addressed to the co-walker. The question then is: how is 'seeing something' transformed from a subjective into an intersubjective event? Which forms of bodily behaviour are used by the noticing participant to make the co-walker understand that their noticing of something is an invitation addressed to them to focus on this feature of the surroundings as well? Can the bodily behaviour of the noticing participant perhaps even be seen as a 'first activity' making the search for the feature a projectable 'second' activity, and how is this projection achieved and different from the subjective act of seeing something?
- b. How can the second participant know what the first participant has noticed when they depend on an inspection of the first participant's bodily behaviour only? While inspecting this behaviour can lead to the establishment of a shared "domain of scrutiny" (Stukenbrock 2020:5, Goodwin 2003) with the co-participant, this domain of scrutiny is not the same as joint attention as demonstrated by Stukenbrock's work (2015, 2018, 2020) on the use of deixis and pointing gestures.

Stukenbrock also distinguishes between the "domain of pointing" projected by the first participant and the "domain of scrutiny" that is established when the second participant orients to the projected domain of pointing (Stukenbrock 2015:56–72). In contrast to Stukenbrock's studies, however, we will investigate joint attention in noticings that do not involve deictic expressions or pointing gestures. In this case, the co-participant cannot be sure that

the noticer has the communicative intention of ‘showing something’. In other words, there is no equivalent to Stukenbrock’s domain of pointing in our examples and the noticing of the first person is ambiguous between an individual and a communicative act.

- c. How does the first participant know that the second participant focusses on the same object as they do, i.e. that joint attention has been established?

In our discussion of these questions, we will rely on recordings made with eye-tracking glasses which give us access to the fine-grained details of participant’s visual behaviour. The trackers record individual vision and allow to reconstruct subjective attention on features of the surroundings. But in addition, they also allow us to check (within certain methodological limitations) whether both walkers indeed look at the same object and whether their displays of joint attention (if any) are justified. If it is the case (as in most examples), it needs to be asked how participants can be successful in establishing joint attention without disposing of the information we, the analysts, have on the basis of eye-tracking.

In order to answer the three questions above, we will resort to Schütz’ notion of common-sense thinking as part of what he calls the natural attitude (*natürliche Einstellung*) towards the lifeworld. This natural attitude makes two “idealizations”, which Schütz subsumes under the “general thesis of reciprocal perspectives” (Schütz 1953: 8). The “idealization of interchangeability of positions” (ibid.: 13) refers to the assumption that if I were at my counterpart’s place, I would experience things from the same perspective, and perceive the same typical aspects as they do. The “idealization of congruence of relevance systems” (ibid.) means that I can assume – and thereby assume that my counterpart also assumes – that differences regarding our biographical backgrounds are not relevant for the present practical purposes, but that we act and agree on the premise that the objects we are encountering have an identical significance for both of us. These assumptions hold as long as no contradictions arise.

In order to answer the central question of how the walkers can know, without using language, that they are focusing on the same object, we will refer to what Schütz (1953: 11) calls “typifications”. Typifications provide the common ground that is needed to make sense of social situations and to cope with new experiences. They include knowledge about typical courses of action or social motives for action. This means that even when the subjectively intended meaning of somebody’s action cannot be fully understood on the basis of its behavioural features, it can be grasped on the basis of typifications.

In the following, we discuss previous work on noticings as the background against which we introduce our phenomenon (Section 2), before turning to a description of our data and methodological approach (Section 3). We will then

describe practices for achieving intersubjectivity without language as well as the sequential patterns they are embedded in (Section 4). Finally, we discuss our findings and draw a conclusion (Section 5).

2. Joint attention and the experience of nature

Joint attention can be understood as a triadic relationship between (at least) two persons and an object (see Clark 1996, Stukenbrock & Balantani, this volume). In order to achieve joint attention, the co-participants must focus on the same object and be aware that this is the case (Clark & Marshall 1981). Previous research shows that there are two major groups of resources that can be used to establish joint attention: verbal and bodily resources.

One practice of establishing joint attention are (verbal) “noticings” (cf. Sacks 1992, 1995, Schegloff 1988, 2007), which make an object in the surroundings relevant and direct the recipient’s focus of attention to it in response. The response relevance established by noticings has been shown to be less rigid than that of first pair parts in an adjacency pair (Goodwin & Goodwin 2012, Stukenbrock 2020), and to be dependent on the local context (Keisanen 2012). For doing noticings, co-participants have been shown to employ a range of multimodal resources, including perception imperatives (e.g. German *guck/schau* ‘look’, Laner 2022, Auer et al. 2024; Finnish *kato* ‘look’, Siitonen et al. 2021), response cries (e.g. *oh*, Anna & Pfeiffer 2021, Pfeiffer & Anna 2021), pointing and deictics (e.g. *hier* ‘here’, *da* ‘there’, Stukenbrock 2015), categorizations and descriptions (e.g. referential nominal phrases such as *a fire*, Goodwin & Goodwin 2012), certain syntactic structures (e.g. polar interrogatives, Laanesoo & Keevallik 2017), as well as gaze shifts, body shifts, facial expressions, and head movements (Kääntä 2014).

The existing work on noticings shows that their design is sensitive to the local context in which they occur. Several studies have used data from mobile interaction. Mobile settings provide fertile grounds for the study of noticing sequences, since “vehicular units” (Goffman 2010: 6) are exposed to a changing visual environment which permanently allows for the discovery of new noticeables deemed worthy of being shared interactively. For instance, Goodwin & Goodwin (2012) and Keisanen (2012) investigate the use of noticings while traveling by car. Both studies point to the tension between the ever-changing surroundings and the fact that noticings involve an indexical relationship to the referential object: summoning another passenger to focus on an object external to the car requires the speaker to produce the noticing as early as possible after the object’s appearance in the environment. As a consequence, noticings often interrupt other passengers’ talk as well as the ongoing sequence.

Compared to verbal interaction during traveling by car, talking while walking together involves other challenges for the co-participants, namely the need to establish and maintain a high degree of mutual orientation and coordination (De Stefani 2010, Mondada 2017, Auer & Laner, this volume). In their study of noticings embedded in the activity of shopping at a farmer's market, Stukenbrock & Dao (2019) show how verbal referential expressions, pointing, and gaze practices are employed in order to introduce and establish a joint focus of attention on a buyable, while passing a market stall. Within this activity, noticings lead to a local negotiation of whether the co-participants should stop for a closer inspection of the identified object or continue walking towards the next stall. Kesselheim et al. (2021) investigate how visitors of a science centre make "joint discoveries". They raise the question of how noticings as a general mechanism are adapted to the local context they are tied to. In their data, the central tasks consist in mutually agreeing on the spectacular character of the discovery and contextualising it as a scientific phenomenon.

The latter two studies demonstrate that noticings and the ensuing state of joint attention are no end in themselves.¹ This also holds for walks in nature which usually are pleasure walks. The walkers do not engage in walking because they want to transfer from one location to the next as fast as possible, but rather because they want to enjoy nature and – on joint walks – because they want to make sure that their experience of nature resonates with that of the co-walking companion. It is for this reason that walkers often display their positive stance towards the noticeable once joint attention has been established. The noticing just provides the grounds for and secures the topic of an assessment or evaluation of the noticeable.

Stukenbrock (2020: 20) distinguishes between two types of inferences participants draw when following another person's gaze, namely "*what* he or she is looking at, and *why*" (italicised in the original). In line with this distinction, we claim that two levels of intersubjectivity are involved in noticings. These inferences take the co-participants from joint attention to joint experience. The first level of intersubjectivity refers to the establishment of joint attention, which involves a *what*-inference and builds on walker B inspecting walker A's bodily behaviour and vice versa. At this level, the walkers establish a "joint attentional frame", that is, focus on an object that they "know [is] part of the attentional focus of both of them" (Tomasello 2005: 22). The second level of intersubjectivity, which involves a *why*-inference, is the process of making sense out of the observed feature. A crucial task for walkers consists in ascribing meaning to the joint perception, that is, in figuring out why each of them is looking at the respective object, and what they find remarkable about it. For experiencing nature together, both levels are essential.

1. See also Stukenbrock (2023) who demonstrates that noticings in museums can be used to delay or accelerate onward movement.

In previous work, we have analyzed these two dimensions of intersubjectivity with reference to those cases in which already the first level involves language (in addition to bodily resources). For instance, in Auer et al. (2024) we investigate verbal practices on both levels: the format of a perception imperative ('look') is used in order to establish joint attention and an exclamative ('how nice!') in order to assess the object. Laner (2022) provides evidence that the second dimension of intersubjectivity can remain implicit. She points out that perception imperatives, sometimes accompanied by deictics and/or pointing (e.g. *guck mal HIER* 'look here'), can be employed without a subsequent explicit account of the noteworthiness of the discovery: Walker A notices something remarkable, for instance a rare plant, and wants to share this discovery with walker B. After having shown it to walker B he assumes that walker B shares the same stance towards this object, once she has perceived it as well.

In this study, we turn to the third case in which the *what*-inference is dealt with without verbal means, while the *why*-inference is dealt with verbally. We therefore investigate three-part sequences in which only the third step is verbal:

- First part: Walker A shows a bodily orientation to a feature of the scene (usually an object in the surroundings), minimally by gazing at it for a prolonged time. Other features such as turning to the object and stopping are additional, even stronger displays of the walker's individual perception of something noteworthy. This first activity is functionally similar to verbal noticing and can be called retrospectively oriented (establishing a "retrosequence" in the sense of Schegloff 2007), as the participant's attention to the object implies that it has caught their attention. However, compared to verbal noticing and the use of deictics, which imply communicative intention (cf. Stukenbrock 2015), bodily orientations to objects in the environment of this kind are ambiguous between a 'private' and a 'social' interpretation. They can be seen as an invitation to walker B to share walker A's noticing, but it is also possible for the co-walker to ignore them and assume that A's gaze was not intended to get the co-walker involved in the first walker's perception.
- Second part: Co-walker B turns toward A and re-orientates their body, head, or (minimally) gaze in the same direction as A, establishing a domain of scrutiny and trying to identify the object of A's attention (Stukenbrock 2015: Chapters 4.6 and 4.7). Walker B uses A's body as a "semiotically structured physical space of expression in order to find an entity in another semiotic space" (Stukenbrock 2015: 60, our translation). In turn, B's re-orientation provides the basis for A to understand that B is trying to identify the object in A's visual attention, i.e. as an attempt to establish joint attention.

- Third part: a verbal sequel such as a verbal description or assessment of the object, or a comment on it, follows or is elicited. This sequel can be produced by either A or B and explicitly deals with the second dimension, that of the ‘why’ of the noticing. At the same time, it provides interactional evidence that joint attention has been established successfully. Participants presuppose that the topic of this verbal third element is the object made available by bodily resources in the first and second part, even though it has not been identified by linguistic means. Via the sequel, the two walkers establish a joint experience.

In our data, we found only three examples in which no verbal sequel followed as a third part. The reason why sequences without a verbal third part are rare seems to be that the *why*-inference is important for the joint experience of nature. The establishment of joint attention provides the basis for socially meaningful interaction, but is not socially relevant in itself.

3. Data and methods

Our analysis is based on a corpus of 12 recordings of couples walking on a loop trail through the Black Forest National Park, each with a duration of 80 to 120 minutes. They had never been to the Black Forest National Park. All participants were L1 speakers of German.

All individuals involved in the study provided written informed consent regarding the collection, the use, and storage of their data. Approval by an ethics committee at the University of Freiburg or the Black Forest National Park was not required.

Each walker wore eye-tracking glasses (Tobii 2). These glasses are light and unobtrusive and do not have large side brims that could restrict the wearers’ peripheral vision. The walkers were asked to follow a certain route which led them through less wild as well as natural/wild areas.

The glasses include a scene camera on the bridge, a microphone and two trackers per eye which record pupil movements. On the basis of the trackers, the algorithm calculates the centre of vision (later visualised by a marker/cursor in the recording). For analysis, this marker is overlaid on the picture of the scene camera. It shows the walkers’ gaze (more exactly, the centre of foveal vision) on the images of the scene camera. The two walkers’ eye-tracking recordings were synchronised and arranged on a split screen using Adobe Premiere Pro CC. The split-screen video and the audio file were then imported into the video analysis

software ELAN (Wittenburg et al. 2006), where speech was transcribed and durational measurements were done.

Transcripts follow GAT 2 conventions (Selting et al. 2009), which were supplemented by transcriptions of multimodal behaviour in separate lines underneath the verbal transcript (following Mondada 2017; see also Merlino & Mondada 2019), as well as an iconic representation of the walkers' gazes and their bodily position as seen from a bird's eye perspective. In this representation, dashed arrows indicate downward-looking, straight arrows stand for looking ahead. Curved arrows indicate that the participants re-orient their gaze and body in the direction of the arrow. The extension of the iconically depicted bodily and gaze constellation is indicated by curled brackets within the verbal transcript (a more detailed description of the symbols used in the transcripts is attached to this chapter, cf. also Laner 2022). At positions marked by "#n", screenshots with the same number show the view from the two participants' perspectives as captured by their scene cameras at this point in time (the cursor-like circles indicate the area of foveal vision).

Eye-tracking allows us to record the walkers' gaze behaviour and to reconstruct their vision in a way that would not be possible using video-recordings from an observer's perspective (see also Zima et al., this volume); the latter are almost impossible during longer walks, since co-walking video-recording investigators are highly obtrusive (particularly, as video-recording investigators would need to walk ahead of the recorded participants in a small distance in order to capture their gaze/head movements). Non-verbal noticings crucially concern the transformation of an individual's subjective experience into shared experience. In order to reconstruct this transformation, we profit from the fact that eye-tracking documents the two walkers' individual vision from which we can reconstruct the orderly ways in which co-participants organise this transformation. Although the participants themselves cannot retrace the other's foci of attention with the same precision as the analysts, the eye-tracking data provide us with an external tool for reconstructing the transformation from individual to joint perception. For instance, it is only by virtue of eye-tracking that we as analysts can gain certainty about whether both participants are focusing on the same object or not, and when exactly this is the case. This allows us to analytically relate individual perception to social displays of perception and to reconstruct the process of how joint perceptual experience is achieved. Hence, eye-tracking makes a thorough investigation of gaze during walking and talking possible. It thereby opens up new avenues for research.

In the study of intersubjectivity, a crucial methodological postulate is to adopt an emic perspective, that is, putting oneself (as a researcher) in the perspective of the co-participants. Recordings of eye-tracking glasses bring us closer to this goal of reconstructing the co-participants' perspectives (cf. Zima et al., this volume,

Rassmussen & Kristiansen, this volume). In contrast to traditional video recordings with an external camera, recordings of mobile eye-trackers enable the researcher to take an analytical position that comes close to representing the participant's "origo" (Bühler 1934). That is, researchers can observe a section of reality that individuals have produced by being in the world and engaging with it, as well as with each other. Eye-tracking data, together with recordings of the scene camera of other co-occurring bodily and verbal behaviour, provide the basis for reconstructing joint experience.

However, even though eye-tracking can be ascribed more ecological validity than classical video recording, the data it produces must not be equated with what the participants see from their perspectives. It can be argued that eye-tracking data provide both less and more information than recordings with an external video camera. On the one hand, eye-tracking glasses provide researchers with less information than is available to the participants of the interaction event. For technical reasons, the trackers cannot fully capture what the human eye is able to recognise (as the angle of the scene camera is limited to 90°). This concerns particularly perception in the peripheral, outermost part of the field of vision (at 94°-108° per eye, cf. Zhisheng et al. 2019), which is specialised for moving objects. It also needs to be kept in mind that the scene camera cannot emulate the complex perceptual process of the human eye. For instance, what and how we see depends, among other factors, on whether we are moving or not. Interestingly, recent studies show that our peripheral vision is improved when we are walking (Cao & Händel 2019). Thus, parts of what participants can perceive through the peripheral vision constitutes a 'blind spot' for eye-tracking and, therefore, for the analysis of social interaction. For instance, a walker may peripherally see the co-walker slowing down, stopping or turning sideways without turning the head, and this may not be captured by the scene camera.

On the other hand, dual eye-trackers provide researchers with more information than is available to the co-participants. Analysts can track the participants' gaze directions at any given moment in time. A participant, in contrast, can only know where the other participant is looking by applying "meta-perceptive gaze practices" (Stukenbrock 2020), that is, by gazing at the other's eyes in order to reconstruct the gaze focus, by following the other's gaze. This has to be taken into account when analysing eye-tracking data from a conversation analytic perspective.

In addition, we will argue that multimodal analysis benefits from drawing on the phenomenological sociology of Alfred Schütz, which originally inspired ethnomethodological conversation analysis. This approach's focus on intersubjective processes in social interaction can provide fruitful impulses for the analysis of sense-making in nature.

4. Joint attention without language

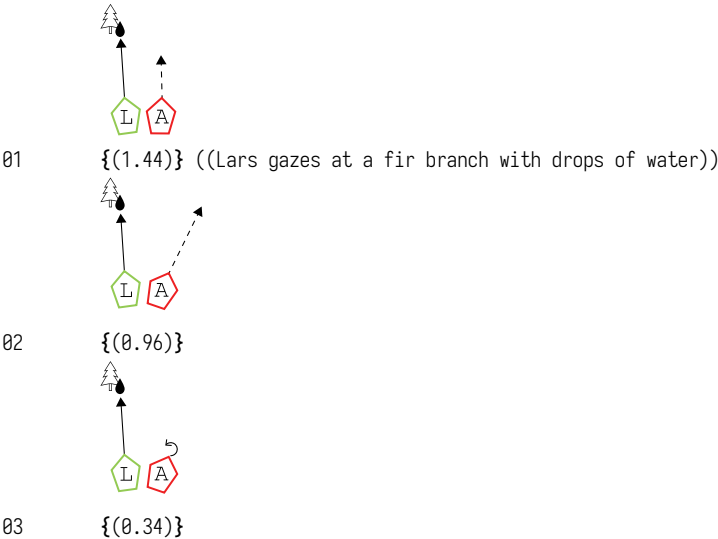
In the following, we will show how the bodily conduct of one participant can become socially relevant for the interaction by eventually leading to joint attention and how sequel actions lead to the shared experience of nature.

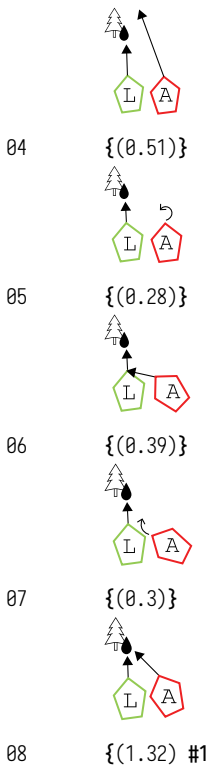
4.1 Walker B bodily co-oriens with walker A and produces a verbal uptake

We first discuss sequences in which participant A bodily orients himself at an object, minimally by gazing at it; subsequently participant B also gazes at the object, before she verbalises either what they are seeing (together), or which stance she takes towards the object they are both inspecting. Through this, she confirms the noteworthiness of the object to which joint attention has been established.

At the beginning of our first extract (line 01), Lars (walking on the left side) is gazing at a fir branch with drops of water, while Anna (walking on the right) is looking down at the path (line 1) and to the sides of the path (lines 2–4). This changes in line 5 when Anna turns to look at Lars's face (line 6) and then follows his gaze (line 7) to the fir branch. Note that they saw another fir tree before on their walk, and they talked about how pretty the drops of water on the branches look.

Extract 1. Fir branch (VP0506, #Zweig_1–00:33:19)





#1

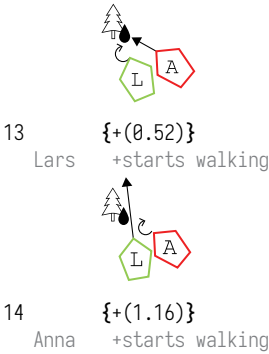
09 Anna HM_hm_hm_hm;

10 $+(0.47)\}$
Lars +stops

11 Anna $\{+v011 SCHÖN.$
very beautiful

Anna +stops

12 $(2.43)\}$



The eye-tracking camera shows us that in the beginning of the extract, Lars’s gaze is fixated at a fir branch, and it is likely that he is looking at this branch. His visual attention to the fir branch ahead of him remains on an individual level, however, for some time (line 01–06); only then, Anna follows his gaze in line 06, i.e. she appears to have perceived his perception of something and looks at the same object from line 08 onward.² We cannot know exactly (and neither does Lars) why Anna started to look towards the left side; her gaze may simply wander around, turning sometimes to the right side (as in line 02) and sometimes to the left. But it is likely that at some point in the 850 ms period of lines 03 and 04, she peripherally perceived Lars’s prolonged gazing ahead. However, it is only by turning her upper body towards him, and her gaze to his face (line 05–06) and then from his face directly to the fir branch (line 07–08) that she orients to Lars’s individual act of seeing as a possible act of showing. Our tracking glasses prove that both walkers then gaze at the fir branch during 1.32 seconds of silence, following Anna’s subsequent vocalization (‘*HM_hm_hm_hm*’, in line 09). This vocalization seems to function as a recognition marker, since the co-walkers have stopped and inspected water drops on a tree on their walk before. Another 0.47 seconds of silent joint inspection of the branch follow, during which Lars stops walking (line 10), and so does Anna (line 11) in a phase of inspection (Mortensen and Wagner 2019). Then Anna assesses the presupposed object of their joint attention as *very beautiful* (‘*vOll SCHÖN*’, line 11). This assessment does not make the assessable explicit and thus presupposes that joint attention has already been established without verbal means before. The fact that Anna doesn’t use any deictics while uttering her assessment (cf. line 11 *very beautiful*) and that Lars does not initiate repair to clarify which object Anna is assessing provides evidence that joint attention on the fir branch is presupposed by the participants. They keep

2. Previous work on gaze following and perceived perception (Hausendorf 2003, Stukenbrock 2020) is based on Luhmann’s (1972:54) concept of “mutual perception” (“wechselseitige Wahrnehmung”).

gazing at the fir branch for another 2.43 seconds, before Lars dissolves the joint inspection of the fir twig by turning away and starting to walk again (line 13), to which Anna bodily aligns by also turning towards the path and starting to walk (line 14).

In this extract, joint attention can be assumed to have been established before one of the co-walkers talks about the object. Lars's prolonged gazing at the object is the first step in the establishment of joint attention, and Anna's gaze-following and gazing at the branch the second. Anna's assessment in line 11, the third step, reflects her assumption that she and Lars have been looking at the same object. The fact that no verbal response by Lars follows (such as a second assessment) suggests that it is heard by Lars as being in a responsive position, i.e. in reaction to Lars' prolonged gazing at the fir branch. Anna formulates an account on behalf of both of them why they gazed at the twig and stopped their walk to inspect it.

The first activity by Lars (his prolonged gaze at the object) hovers between individuality and sociability, as it is typical for nonverbal noticings. Lars not only does not use linguistic resources (such as perception imperatives) to draw Anna's attention to the branch; his bodily behaviour does not display such an intention either. He does not point at the object with his hand, nor can his gaze be understood as an instance of gaze-pointing (cf. Stukenbrock 2015:177–192; Wilkins 2003). He just gazes at the branch, without securing that Anna can perceive this gaze (by looking at her); Anna only sees Lars's gaze at the twig when her gaze happens to turn left by chance.

Two questions come to mind. The first is: how can the co-participants (who, other than the analyst watching the tracking, cannot know what exactly the co-walker is looking at) come – quite 'correctly', in this case – to the conclusion that they are looking at the same object and hence have established joint attention? The second is: How can the second walker give an assessment which expresses not only her own account for why the object was noteworthy, but also one that is shared by the first walker, although this first walker has not produced any verbal utterance which might betray his stance toward it?

Regarding the first question, we might remember Schütz (1953:12): "For each partner the other's body, his gestures, his gait, and facial expressions are immediately observable, not merely as things or events of the outer world but in their physiognomical significance, that is as symptoms of the other's thoughts." The co-participant's bodily conduct is more than behaviour, it is a window for the others into the observed person's mind, the basis of inferences on cognitive processes and states. Anna follows Lars' prolonged gaze and on the basis of this behavioural observation *infers* what he is looking at (cf. Stukenbrock 2020:20). She cannot do this by following his gaze alone; in her position, she will not be able to calculate the angle of his vision with precision, and even if she were able to do this calcula-

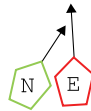
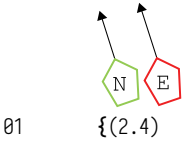
tion precisely, she would only be able to determine the direction of his gaze, not the object of his attention (which could be any feature of the scene located in this direction).

To identify this object of attention, she additionally needs to know which object might be *significant* for him. This touches the second question. If Anna wants to know what Lars might find noteworthy in nature, she can rely on Schütz' reciprocity of perspectives, and particularly his "idealization of congruence of relevance systems", assuming that what has significance for her will also be significant to him. There is cultural knowledge shared by the two which makes a branch with sparkling water drops in autumn potentially look 'picturesque'. On the other hand, they can also build on specific biographical background, i.e. the fact that they had talked about how beautiful they find the water drops on another fir branch discovered earlier on their walk.

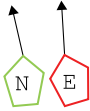
Anna's assessment of the fir branch as *very beautiful* (line 11) shows that she assumes that they are both looking at the fir branch *because* it is beautiful, and not at any other aspect of the scenery, which is not contradicted by Lars. This emphasises that the first level of intersubjectivity (i.e., shared attention) can be reached without verbal means, building on the assumption that there are shared reasons *why* an object is looked at (second level of intersubjectivity). In retrospect, then, the assessment re-invokes an evaluation which presupposes that the object of joint attention in the present situation is the (unexpressed) argument over which this evaluation is predicated. Hence, it displays participants' understanding that they have attended to the same object.

In our second example, we find the same three-part structure, but the bodily orientation to the joint object of attention is stronger, and the sequel in the third position is not an assessment; rather, the co-walkers verbalise what it is that they are looking at.

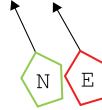
Extract 2. Blueberries (VP0102 #Freudenstadt_Heidelbeeren 00:15:42)



02 Ella SCHWARZwald isch} halt einfach {schön.}
Black Forest is PCTL simply beautiful



03 $\{(3.2)\}$



04 aber hier in der ecke war ich auch Echt noch {[NIE.]}
 but I haven't been to this corner before

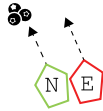
05 Nina [ne:..]
no



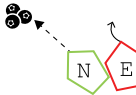
06 {also JA.}
well yeah



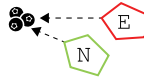
07 $\{(2.8)\}$



```
08      {in FREUdenstadt war ich mal=}
      I've once been to Freudenstadt
```



```
09      {=aber halt IN der stAdt.}
      but PCTL in the city
```

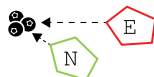


10 Ella +ja {freudenstadt is auch} {&rIchtig [HÄSSlich.]
yeah *Freudenstadt is PCTL really ugly*

Nina +walks towards the blueberry bushes-->

Elia & stops

11 Nina [ähm-]
ehm



12 NÖ: fand ich #1 ei[gentlich+} {%nIch; *}
no I didn't really think so

Nina %stops *walks closer to bush-->

13 Ella [das sind HEI%#2]delbeeren.*
these are blueberries
 Nina %points at blueberry bushes-->

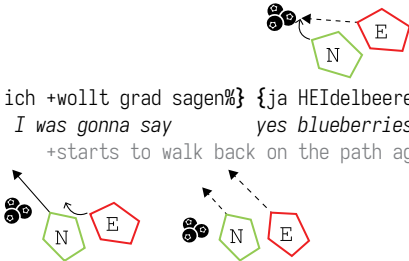


#1



#2

14 Nina ich +wollt grad sagen%} {ja HEIdelbeeren.}
I was gonna say yes blueberries
 Nina +starts to walk back on the path again



15 Ella {aber ohne} {%HEIdelbeeren;
but without blueberries
 Ella %starts to walk again

16 Nina ((clears her throat)) ja;}
yes

In the beginning of this extract (lines 1–12), Nina (walking on the left side) and Ella (walking on the right side) are still engaged in a different topic. They assess the Black Forest in general (line 03) and talk about the fact that they have not been ‘in this corner’ before (line 04). While Nina begins to talk about a close-by town (lines 8 and 9), which she has visited before, she gazes to the left at a bush of blueberries and (in line 10) starts to walk towards it, slightly off the hiking trail. Ella responds to Nina with a negative assessment of this town (line 10). During this verbal assessment, Ella also turns to the bush and gazes at it before stopping to walk altogether.

Nina contradicts Ella's assessment of the town with a blunt rejection³ ('no I didn't really think so', line 12), while both walkers' gazes continue to be directed on the blueberry bush. At the same time, Nina stops, walks closer to the bushes, and stops again right in front of the bushes (lines 11–12). Interrupting Nina's disagreement, Ella says in overlap with Nina's talk that *these are blueberries* (line 13). Almost at the same time, Nina points at the bushes and subsequently adds that she *was gonna say, yes, blueberries* (line 14). She starts to walk back to the path again, and Ella follows her while saying *but without blueberries* (line 15), hinting at the fact that the bushes do not carry any blueberries at the moment.

In this extract, joint attention is again established (and, as the trackers prove, the two walkers factually look at the same object), before the co-walkers talk about the object (and one of them points at it), presupposing joint attention; the delay of the verbal response to the perceived object in this case is due to the verbal exchange on a different topic which continues during the process of bodily co-orientation. Nina's gazing at the bushes and walking towards them provides the basis for the process of establishing joint attention. Ella assumes from Nina's bodily behaviour that she has noticed something of interest. According to Schütz' "idealization of interchangeability of positions" (1953:15), Ella can assume, based on observing Nina's actions (which are understood as meaningful), that she will be able to make the same observation if she takes the same spatial position and perspective. She reacts accordingly and follows by bodily co-orientation, walking to the object, and by gaze. Surely, this co-orientation is based on the assumption that (a) walkers first of all walk, i.e. interrupting the walk (slowing down or stopping) is a marked activity which invites an inference — such as the inference that something noteworthy has been discovered, and (b) walkers on a nature walk want to experience nature together, inviting the inference that A wants to show B something of potential interest to both of them. But these rather strong cues of bodily reorientation notwithstanding, there rests some ambiguity: it is possible that Nina's behaviour is not intended as a first activity to which Ella is invited to deliver a second.

Ella's 'what is the other looking at-inference' converges with the evidence the eye-tracking provides (focus on the bushes); as in the first example, this 'what'-inference is not only based on observing Nina's behaviour — particularly following her gaze —, but also builds on a congruence of relevancies which is culturally shared: (edible) berries in nature are noteworthy, and Nina can be assumed to have been attracted by them (perhaps because she wants to taste

3. Note that the hesitation marker in line 11 starts in overlap with Ella's evaluative term and therefore is not a preface to Nina's disagreement but rather a preface to the activity of talking about the blueberries.

them). But as no berries can be seen, the ‘why-inference’ is not so unproblematic. In this sequence, no common stance towards the object is expressed as a sequel to the mutual gaze at the object in question. Rather, Nina and Ella both engage during this sequel in subsuming the object in question under a verbal category, i.e. in naming it. In fact, categorization seems to have been the very reason for which Nina inspected the object (‘why are we looking at it-inference’).

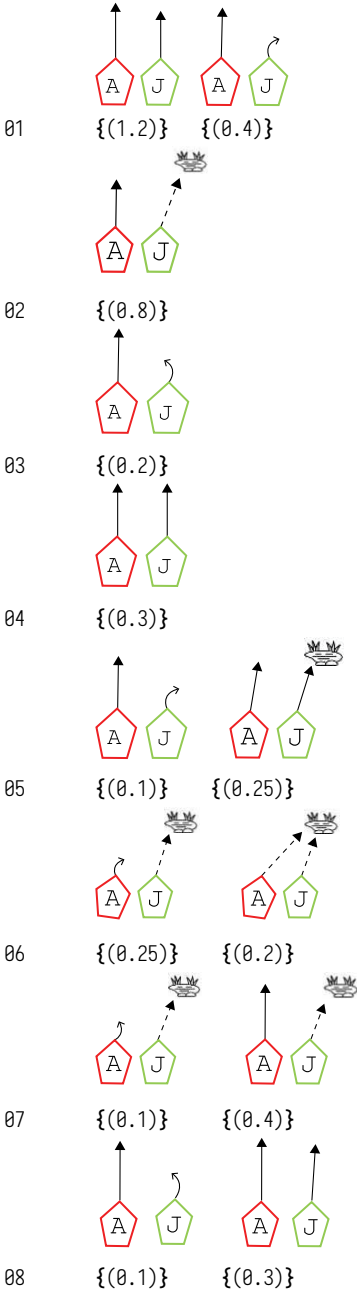
Ella named the bushes, before Nina could do so herself. With the meta-communicative phrasing she uses to respond to Ella’s naming (‘I was gonna say yes blueberries’), Nina expresses that an earlier sequential position would have been adequate for her own naming, claiming that she intended to name the blueberries before her co-walker (Küttner & Raymond 2022 on the use of *I was gonna say* in English). When the naming is successful and the two walkers agree on ‘blueberries’ as the correct description, the sequel accounts for the noticing and in retrospect displays that both participants have been looking at the same object.

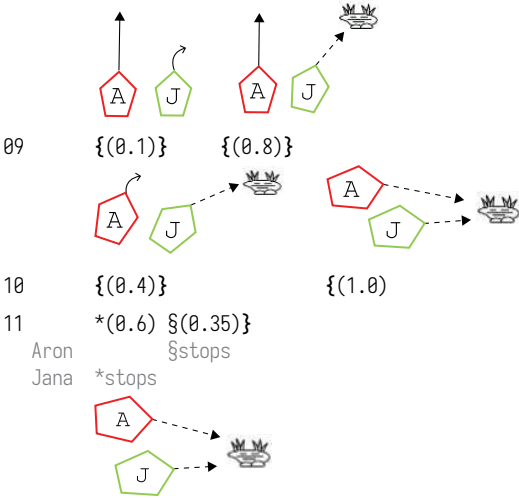
4.2 Walker B bodily co-oriens with walker A, walker A produces a verbal account

In the examples discussed in the last section it was walker B who co-oriented with walker A’s bodily orientation in such a way that joint attention was achieved, and it was also walker B who confirmed the noteworthiness of the object of attention. We now turn to a slightly different variant of this pattern. Again, joint attention is established without verbal means, but in the sequel, it is walker A who accounts for it. This may seem like a small difference. Yet it points to a different way in which the ‘why-inference’ is dealt with and joint experience is established. In the extracts discussed in the previous section, walker B who followed walker A’s bodily orientation and gaze to the object of interest also expressed an understanding of this object’s significance for both participants; the identification of the object was enough to understand the reason for which walker A looked at it. In the example discussed in this section, walker A – the ‘first noticer’ that walker B co-oriens with – gives a verbal account of why s/he did so. The object as such, so walker A seems to assume, is not necessarily sufficient to make walker B understand why it was worth looking at.

In the following extract, the walk takes place in winter; the walkers have talked before about the scenery being snow-covered, which they had not expected. What becomes the object of joint attention in the following is a small waterhole which was probably formed by melted snow. Note that the object is not perceptually well defined, although the lack of snow in this area is a possible visual anchor.

Extract 3. Biotope (VP 1920, #Biotop, 01:07:55)

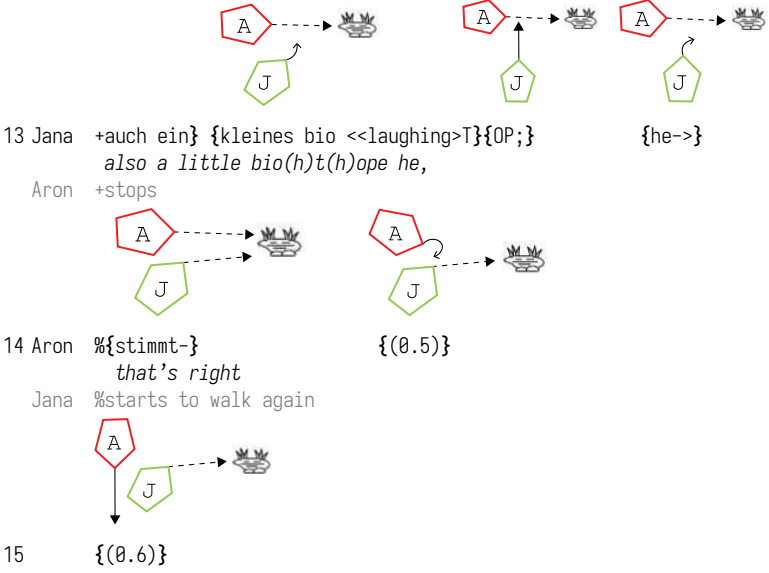


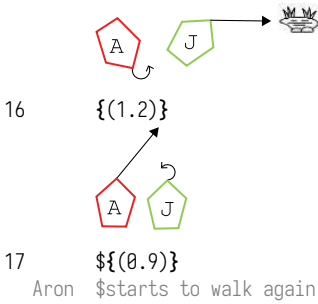


12 $\&\{(0.7)\}\#1$ $\&$
Aron $\&$ moves one step further toward the object $\&$



#1





As we can see from the eye-tracking data, the person walking on the right side (Jana) looks at this spot several times already from the distance (lines 01–09), before her co-walker’s gaze is also drawn to it in line 10. There is even one point during this period in which Jana already seems to notice the waterhole (line 06) while Aron’s gaze also turns in its direction, and both participants’ gazes seem to meet there for a very short period of some 0.2 seconds. However, there is no evidence that what is shown by the eye-trackers also becomes part of the interaction. Aron’s gaze hardly rests on the waterhole, and it is unlikely that Jana is able to see his gaze direction, let alone understand it as a sign of Aron’s focusing on the same object. It is also unclear whether Aron’s gaze is shortly fixated on the spot because he has perceived Jana’s bodily orientation to the right side, or whether it is part of an independent scanning of the scenery. None of them shows to the other that he has perceived the other’s perception of something noteworthy in the surroundings.

This changes in line 10. Already at the end of line 09, Jana not only starts to direct her gaze at the emerging focus of their joint attention, but also turns her body to it. This movement, which Aron is presumably able to observe in the periphery of his vision, leads to him following her gaze. When Jana additionally stops walking, it is clear that he can see her ‘looking at’ something. He also stops and looks in the same direction. To do so, he has to turn slightly against the direction of walking (line 12). As he now is standing a little behind Jana, he is able to follow her gaze quite well. In order to be able to inspect the object even better, he moves one step forward in her direction. The trackers show that both of them are looking at the waterhole now, but during the entire period of their inspection, their gaze wanders around in a small space on the ground which is filled with water. There seems to be an object of joint attention, but compared to Extracts (1) and (2), it seems more difficult to understand why Jana stopped to inspect it, and there is no cultural or contextual cue that can be recognised visually. With the “general thesis of reciprocal perspectives” (Schütz 1953: 8), Jana can be assumed to put herself into Aron’s perspective: If she were in his place, she seems to assume, it would be difficult for her to draw the ‘why are we looking at it’-inference. In this situation, Jana offers an account, by calling the small waterhole they are inspect-

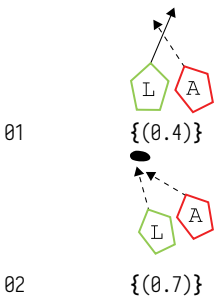
ing a ‘little biotope’, implying that it might be the home of animals and plants even though they cannot be seen. Calling the waterhole a ‘biotope’ establishes its relevance and introduces a potentially interesting additional perspective which is not visible by merely looking at it. While she produces her account, Jana already starts to move away from the object of joint attention, but as Aron seems to have no intention to follow her but rather continues inspecting the waterhole, she also directs her gaze at it again. At the same time, Aron agrees with her description of it as a ‘little biotope’ (*stimmt*, ‘that’s right’). The description of the observable establishes its noteworthiness for both participants and acknowledges Jana’s account as valid. Indirectly, the sequel also confirms that joint attention has successfully been established by non-verbal means.

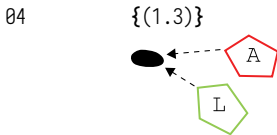
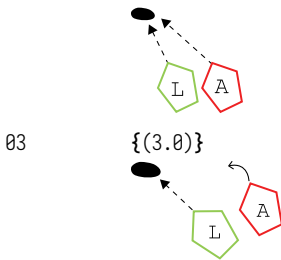
4.3 Walker B bodily co-oriens with walker A and initiates repair

In the example presented in the last section, the walker who gazes at the object first seems to assume that her co-walker may not be able to draw the ‘why are we looking at it’-inference without language. She provides a verbal account, making explicit why the object they are looking at has caught her attention, which leads to the successful establishment of the second level of intersubjectivity and closes the sequence. In this section, we turn to cases in which the noticer does not provide a verbal account. In these examples, too, joint attention (first level of intersubjectivity, ‘what are we looking at’-inference) is established successfully. However, in contrast to the unproblematic achievement of intersubjectivity presented in Section 4.1, the second walker has trouble making sense out of the observed object (second level of intersubjectivity, ‘why are we looking at it’-inference). To resolve the problem, she initiates repair.

In the following example, Lars and Anna are in an “open state of talk” (Goffman 1981: 134–5), walking side-by-side on a trail and letting their gaze wander across the scenery, before a grey stone partly covered by moss becomes the object of joint attention.

Extract 4. Moss (VP0506, Moos_1 – 1:09:22)





05 Anna +{(1.2) hm? (0.3)}

Lars +stops

Lars +stops

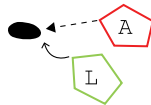
06 Lars das sieht toll %aus mit dem(.) #1 MOOS- %
that looks great with the moss.

Lars %palm-down open hand gesture%



#1

07 (2.9)

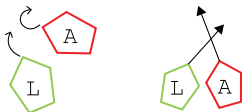


08 *(0.3)+ +wie_s} {so- (0.1)}

how it like

Lars *starts to walk again

Anna +starts to walk again



09 {so den} {STEIN umschlingt.}
like engulfs the stone

10 Anna ((laughs))

11 (0.9)

12 ((produces a 'sucking sound'))

While they are walking, Lars can be seen to gaze at a large grey stone on the left side of the trail which becomes visible already from a certain distance. It stands out as a perceptually salient object from the ground mainly covered by small plants (line 02). His gaze remains focused on the stone, leading him to turn his head to the left as they are approaching it. This may have been perceived by Anna in her peripheral field of vision, who is walking on Lars's right side and slightly behind him while gazing to the left (line 02). Shortly afterwards, she follows Lars in directing her gaze at the stone and changes her body orientation to the left as they reach it (line 04). Both of them stop at the same time, looking down at the stone (line 05). Since Anna is standing somewhat behind Lars, she is likely able to see where Lars is looking, who is standing closer to the stone, and Lars can perceive in his peripheral vision that Anna has turned around and stopped; we can assume that both walkers know that they are sharing a joint focus of attention, i.e. the what-inference was successful.

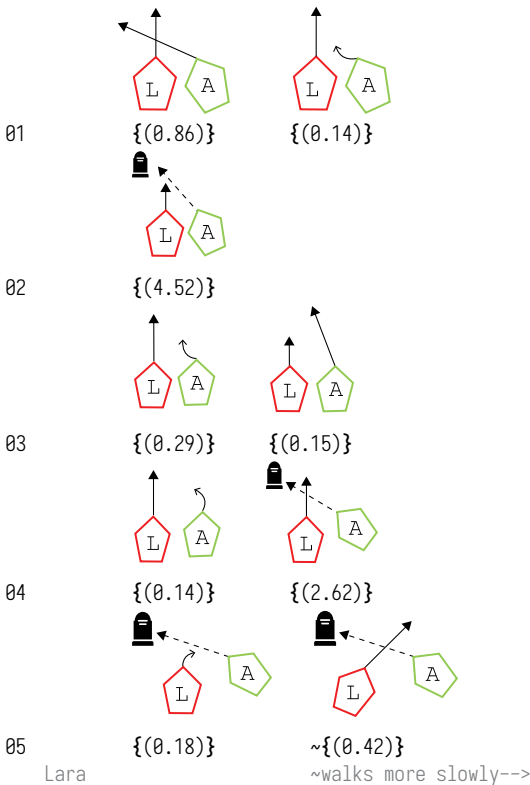
The following sequential trajectory shows that, at this point, the why-inference cannot yet be drawn by Anna. In line 05, she initiates repair using the “‘open’ class” (Drew 1997) repair initiator *hm?* with rising intonation, indicating that a problem has arisen without specifying the kind of trouble encountered. While up to this moment Lars's behaviour has been ambiguous regarding its status as either an individual noticing or a noticing aiming at intersubjectivity and inviting the co-walker to respond, Anna's repair initiation resolves this ambiguity by treating Lars's looking and stopping as a social event relevant for interaction. Given that repair initiations are usually placed in vicinity of the repairable, and since the walkers have not been engaged in talking, the repair initiator can be interpreted as referring to the stopping associated with extended looking at an object, adjacent to which it is positioned. The fact that it is Anna who initiates repair shows that she holds Lars accountable for stopping and gazing at the stone. Lars's ensuing evaluation (*that looks great with the moss*, line 06) provides evidence that he has understood Anna's repair initiation as referring to his stopping as a result of having noticed something noteworthy. He offers an assessment of it which explains its noteworthiness (for him) and at the same time a description of what attracted his attention. This utterance is combined with a palm-down open hand gesture reaching out towards the part of the stone covered with moss (see figure #1). (The gesture does more than pointing at the stone; the spread-out fingers and the revolving motion iconically represent the ‘engulfing’ grip of the moss on the stone, which will be the object of a verbal description in lines 08/09.) This turn design shows that he treats his stopping and looking at the stone as in need of explanation, orienting to the second level on which intersubjectivity has not yet been established.

A long pause of almost three seconds (line 07) ensues during which both co-walkers continue to look at the stone. Lars then expands his account (*how it*

like like engulfs the stone, lines 08/09), turning the gesture in line 06 into a verbal description. He uses a metaphorical description to specify what exactly he finds remarkable, namely the moss in relation to the stone. It is the way the moss is twining around the stone that “looks great”. Lars then turns around and continues walking; Anna immediately joins him. In response to Lars’s metaphorical description, she laughs and produces a sound which portrays the act of ‘engulfing’, a sound reminiscent of a ‘slimy’ animal devouring something, for instance an octopus sucking in his prey (line 10). She thereby agrees with Lars’s metaphor of the moss engulfing the stone (cf. Auer & Laner, this volume). Hence, Lars’s specification of the account was successful in establishing intersubjectivity.

In the next example, too, the walkers are successful in achieving joint attention without language, that is, they both seem to know that they are gazing at the same object based on monitoring each other’s bodily behaviour. But again, the ‘why are we looking at it’-inference is more difficult for the second walker to draw. Lara and Alex are approaching an elongated upright stone that becomes visible, jutting up from the ground. Just like in Extract 4, the two walkers are not engaged in talking in the beginning of the sequence.

Extract 5. Tombstone (VP0304 #Grabstein 00:46:56)





Already from quite a far distance, Alex, who has been scanning the scenery on his left-hand side, directs his gaze to the upright stone on the left side of the path (line 01). His gaze remains fixated on the stone for a long period of time (until line 09), only interrupted by a short glance to the path further ahead (lines 03/04). While Alex is looking at the stone, Lara is gazing to the path in front of her (lines 01–04). It is unlikely that she has perceived where Alex is looking, and the eye-tracking data does not provide evidence to support such a claim. Up to this point, his perception of the stone, as suggested by his prolonged gaze at it, is merely on the subjective level.

This changes in line 05, when Lara directs her gaze to the right side of the trail. This change in gaze direction seems to have provided her with new perceptual information from her peripheral field of vision, from which she seemingly has noticed his prolonged gaze to the left side of the path. Immediately afterwards, she turns to the right and gazes at Alex' face (line 06), then turns back to the left to follow his line of regard, and looks at the stone (line 07). Both walkers can be assumed to mutually know that they are sharing a joint visual focus on the stone. Lara has perceived where Alex is looking, and the perceptual prominence of the stone which stands out against the background provides an additional cue that this is the object that has caught Alex's attention. Furthermore, we can safely assume that Alex has perceived that Lara has seen where he is looking, since Lara is standing between him and the stone with her face positioned right next to the centre of his foveal vision when she turns around to look at him, and then at the stone. In other words, joint attention has been achieved as a result of "perceived perception" (Stukenbrock 2018).

However, mutually knowing that a joint visual focus has been established on a certain object only means that the first level of intersubjectivity has been achieved, but it does not seem to be immediately clear why the object is being attended to. After both walkers have been looking at the stone for a period of 0.81 seconds (during which Lara stops while Alex continues to walk, line 07), it becomes evident that the second level of intersubjectivity has not yet been reached. Just like Anna in Extract 4, Lara initiates repair using the token *hm?* (line 08), turning Alex's prolonged looking at the stone into interactionally relevant behaviour in need of explanation. Her repair initiation indicates that she is not yet able to understand what meaning he ascribed to the joint perceived object. The difference between the perspectives of ego and alter, which is of fundamental importance here, can be related to Schütz' concept of motives for action. If Lara wants to understand the meaning Alex' actions have for himself, and she is

4. The cursor is not visible in this short segment because Lara gazes too far to the side to Alex (see #1).

unable to do so based on what is meaningful to her (reciprocity of perspectives), she needs to resort to repair initiation as an attempt to achieve an intersubjectively shared understanding of their experience in nature.

Immediately after Lara's repair initiation, Alex turns his gaze away from the stone, turns to the left (line 09), and looks back onto the path (where other walkers are approaching) while continuing to move in the same direction (i.e. walking backwards). During this time, he responds to the repair initiation with 'nothing' (0.18) *I thought this is a tombstone* (lines 10/11). His response shows that Alex has understood the repair initiation as targeting the significance of the stone, i.e. the question of why it is worth looking at it for so long. He does not explicitly name the object or point to it, implying that he presupposes a joint focus of attention. With his response, Alex negates the significance of the object of joint attention (*nothing*). Since the stone – contrary to his original assumption – turned out not to be a tombstone, he retrospectively acknowledges that it is not noteworthy. The account also shows that he treats the problem as having been on his, not on Lara's side. Lara dissolves her gaze focus right after Alex has produced *nothing* and turns back to the path to continue to walk. Intersubjectivity has been reached in the sense that both walkers agree that the object of joint attention is not remarkable and, thus, discarded as a candidate for joint experience.

5. Discussion and conclusion

According to Schütz and Luckmann (1979: 63), people have a special interest in the sector of the lifeworld that immediately surrounds them in time and space. An activity in which the direct surrounding plays a particularly prominent role is the joint experience of nature, for instance while walking through the forest. As an important part of this activity, walkers are exposed to a multitude of sensory impressions and are constantly noticing, individually or together, objects in or aspects of their immediate environment. In our study, we used audiovisual data from pairs of walkers in the forest recorded with mobile eye-trackers in order to investigate social practices for constructing intersubjectivity while experiencing nature. These practices crucially involve monitoring each other's bodily behaviour, including gaze.

We have argued that joint attention can be achieved without language. Our claim goes beyond the assumption that the walkers are successful in establishing a shared domain of scrutiny, which would amount to mutually knowing, for instance, that they are looking in the same direction, or are scanning the same area of the woods. Instead, we argue that the walkers are able to achieve joint attention on a specific object based on monitoring each other's bodily displays

and by relying on Schütz' common-sense constructs. The crucial empirical evidence we have provided for this argument are the verbal sequels in third position. These sequels presuppose that reference to the respective object has already been established by both participants. Thus, they retrospectively prove successful establishment of joint attention. As the participants do not have the knowledge we have via the eye-trackers, the convergence of gaze provides us only with additional external evidence for successful joint object identification besides the verbal sequel.

The central question we addressed is how the establishment of joint attention is possible without language. Our explanation comprises two components: first, the deployment and mutual monitoring of certain bodily practices and, second, Schütz' notion of common-sense as the basis for drawing inferences about where the other is looking, and why.

In our data, the co-walkers mobilise bodily resources to a much greater extent than for verbal noticings. For instance, gaze-following, which does not occur very often in noticings that rely on language (Auer et al., 2024), is used in most of our examples. Gaze-following ensures that not only does the second walker know that the first walker is looking at a certain object, but also that the first walker knows that the second walker knows that he is looking at it. This process transforms subjective seeing into social seeing. Gaze-following is often preceded by the first walker's prolonged gazing at the object which is, on the subjective level, the result of a sustained focus of visual attention. Prolonged gazing regularly co-occurs with reducing walking speed and stopping. On the interactional level, the co-walker regularly treats prolonged gazing as an invitation to follow the first walker's gaze, and stopping as an invitation to stop for joint inspection. The second walker orients to these bodily actions by the first walker as indications that something noteworthy has been discovered. Our data also show that stopping is commonly used as a practice for displaying that the object in focus has meaning for the walker. Extract 5 is an interesting exception in this respect. Here, the first walker had been constantly gazing at a stone for a longer period of time, which led the co-walker to follow his gaze. The co-walker even stopped as a result of gaze-following, which only happens rarely in our data. However, the first walker, rather than stopping for a closer inspection of the object of joint attention, dissolves his gaze and continues to walk. In this case, the reason for not stopping seems to be that he negates the significance of the stone, retrospectively treating it as not worth attending to. In the few cases in which gaze-following does not occur, approaching an object together seems to be important. At least in certain cases, jointly moving towards a 'target' for closer inspection seems to help the second walker identify the object of interest.

In addition, we claim that establishing joint attention without language relies on common-sense constructs, described by Schütz' "general thesis of reciprocal perspectives" and his concept of "typifications" as part of the walkers' "natural attitude" towards the lifeworld. Gaze direction, gestures (for instance depicting objects), and other bodily movements (such as changed walking speed, stopping, turning around) are immediately observable by co-walkers and are understood as "symptoms of the other's thoughts" (Schütz 1953: 12), i.e. as meaningful within the context of nature. The second walker inspects the first walker's bodily behaviour and uses assumptions about typical social motives and typical courses of action to make sense of it. This, we argue, provides the background for 'what'- and 'why'-inferences.

Consider again Extract 1 ('fir branch') as an example that demonstrates how Schütz' concepts, in particular the idealization of congruence of relevance systems, can explain how the establishment of joint attention is possible without language. In this extract, intersubjectivity regarding the object of joint attention ('what') as well as its significance ('why') is achieved smoothly. Although the 'what'- and the 'why'-inference can be separated analytically, this example shows that they are often intertwined in situ. Trying to figure out why the co-walker directs his gaze in a certain direction for an unusually long period of time may result in searching the domain of scrutiny for possible noticeables in order to find the relevant object and establish joint attention on it ('what'-inference). In the context of nature, certain objects are typically more likely to be considered noteworthy than other objects. Co-walkers can draw on the idealization of the congruence of relevance systems and on these typifications in order to identify the object of interest, i.e. to move from a domain of scrutiny to an object of joint attention which can be ascribed meaning. This is what happens in Extract 1. The successful and unproblematic establishment of joint attention is due to, on the one hand, the typification of the object (sparkling water drops on a branch) which can be considered a typical instance of what can count as 'beautiful' and 'worth seeing' during a pleasure walk. On the other hand, since the walkers have already been talking about the drops on a branch they had noticed earlier and how pretty they are, this object can easily be identified on the basis of the idealization of congruence of relevance systems, assuming that what the second walker considers to be significant must also be considered significant by the first walker.

Extract 3 ('biotope') showed how a walker who notices an object first presupposes joint attention ('what'), but provides an account for looking at the object, preempting co-walker's potential problems of ascribing meaning to it ('why') based on visual information only. Indeed, there are cases in which joint attention has been successfully established without verbal means, but in which the second walker has trouble making sense out of the observed object (Extracts 4 'moss' and

5 ‘tombstone’). On the basis of the observable actions of the first walker, his ‘in order to’-motives remain unclear. Thus, she initiates repair in order to elicit an account that supports her ‘why’-inference, which is successful in both cases and leads to the establishment of intersubjectivity on the second level.

In all the examples shown, the verbal sequel in third position proves that joint attention has already been established without using language. However, language seems to be required to move from the first level (‘what’-inference, establishment of joint attention on an object) to the second level of intersubjectivity (‘why’-inference, making sense out of the joint perception). According to Schütz, the meaning of shared experiences must be socially negotiated before it can be ‘stored’ on a stock of previous experiences. That language is crucial in this process can be seen from the fact that there are only three examples of bodily co-orientation in which no language is used at all. The sequential trajectories regularly show that a verbal uptake (in third position) is necessary for subjective meaning ascribed to an object to become a shared meaning. For this reason, although the establishment of joint attention as the first step towards intersubjectivity can be based on bodily resources alone, language is necessary to take the walkers from joint attention to the joint experience of nature.

Acknowledgements

This work was carried out as part of a research project: “Looking, noticing, and talking: How walkers experience the Black Forest National Park” (Auer/Botsch), partly funded by the Black Forest National Park.

References

- Anna, Marina, and Martin Pfeiffer. 2021. “Die frühkindliche Verwendung des Exklamativformats ,oh+X’: Noticings und Problemmitteilungen [Young children’s use of the exclamative format ‘oh+X’: Noticings and reports of trouble].” *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 22: 1–35.
-  Auer, Peter, Barbara Laner, Martin Pfeiffer, and Kerstin Botsch. 2024. “Noticing and assessing nature: A multimodal investigation of the format “perception imperative + exclamative” based on mobile eye-tracking data.” In *New Perspectives in Interactional Linguistic Research*, ed. by Margret Selting, and Dagmar Barth-Weingarten, 245–275. Amsterdam: John Benjamins.
-  Auer, Peter and Barbara Laner, this volume. “Laughter and gaze among talkers on a walk” in *Mobile Eye Tracking: New avenues for the study of gaze in social interaction*. ed. by E. Zima and A. Stukenbrock (Amsterdam: John Benjamins).

- Bühler, Karl. 1934. *Sprachtheorie: die Darstellungsfunktion der Sprache* [Language theory: the representational function of language]. Jena: Gustav Fischer.
-  Cao, Liyu, and Barbara Händel. 2019. "Walking enhances peripheral visual processing in humans." *PLoS Biology* 17(10): e3000511.
-  Clark, Eve V. 2015. "Common ground." In *The Handbook of Language Emergence*, ed. by Brian MacWhinney, and William O'Grady, 328–353. John Wiley & Sons.
-  Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Clark, Herbert H., and Catherine R. Marshall. 1981. "Definite reference and mutual knowledge." In *Elements of Discourse Understanding*, ed. by A.K. Joshi, B. Webber, and I. Sag, 10–63. Cambridge: Cambridge University Press.
- De Stefani, Elwys. 2010. "Reference as an interactively and multimodally accomplished practice. Organizing spatial reorientation in guided tours." In *Spoken Communication*, ed. by Massimo Pettorino, Antonella Giannini, Isabella Chiara, and Francesca Dovetto, 137–170. Newcastle: Cambridge Scholars Publishing.
-  Drew, Paul. 1997. "'Open' class repair initiators in response to sequential sources of trouble in conversation." *Journal of Pragmatics* 28: 69–101.
- Flom, Ross, Kang Lee, and Darwin Muir (eds). 2007. *Gaze-Following: Its Development and Significance*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Goffman, Erving. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Goffman, Erving. 2010. *Relations in Public. Microstudies of the Public Order*. New Brunswick/London: Transaction Publishers.
- Goodwin, Charles. 2003. "Pointing as situated practice." In *Spoken Communication*, ed. by Kita Sotaro, 217–241. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Goodwin, Marjorie, and Charles Goodwin. 2012. "Car talk: integrating texts, bodies, and changing landscapes." *Semiotica* 191: 257–286.
-  Hausendorf, Heiko. 2003. "Deixis and speech situation revisited: the mechanism of perceived perception." In *Deictic Conceptualisation of Space, Time and Person*, ed. by F. Lenz, 249–269. Benjamins.
-  Kääntä, Leila. 2014. "From noticing to initiating correction: students' epistemic displays in instructional interaction." *Journal of Pragmatics*. 66: 86–105.
- Kesselheim, Wolfgang, Christina Brandenberger and Christoph Hottiger. 2021. "How to notice a tsunami in a water tank: joint discoveries in a science center." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 22: 87–113.
-  Küttner, Uwe-A., and Chase Wesley Raymond. 2022. "I was gonna say...: on the doubly reflexive character of a meta-communicative practice." In *Sprachreflexive Praktiken: Empirische Perspektiven auf Metakommunikation*, ed. By Florian Busch, Pepe Droste, and Elisa Wessels, 51–73. Berlin: J.B. Metzler/Springer.
-  Keisanen, Tina. 2012. "'Uh-oh, we were going there': environmentally occasioned noticings of trouble in in-car interaction." *Semiotica*, 191, 197–222.
-  Laanesoo, K., and Keevallik, L. 2017. "Noticing breaches with non-polar interrogatives: Estonian Kes 'who' ascribing responsibility for problematic conduct." *Research on Language and Social Interaction*. 50, 286–306.
- Laner, Barbara. 2022. "'Guck mal der Baum' – Zur Verwendung von Wahrnehmungsimperativen mit und ohne ‚mal' ['Look PTCL (at) that tree' – On the use of perception imperatives with and without 'mal']". *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 23: 1–35.

- doi Luhmann, Niklas. 1972. "Einfache Sozialsysteme [Simple social systems]." *Zeitschrift für Soziologie* 1 (1): 51–65.
- doi Merlino, Sara, and Lorenza Mondada. 2019. "Crossing the street: how pedestrians interact with cars." *Language & Communication* 65: 131–147.
- doi Mondada, Lorenza. 2017. "Walking and talking together: questions/answers and mobile participation in guided visits." *Social Science Information* 56 (2): 220–253.
- Mortensen, Kristian, and Johannes Wagner. 2019. "Inspection sequences — Multisensorial inspections of unfamiliar objects." *Gesprächsforschung — Online-Zeitschrift zur verbalen Interaktion* 20: 399–443.
- doi Pfeiffer, Martin, and Marina Anna. 2021. "Recruiting assistance in early childhood: longitudinal changes in the use of 'oh+X' as a way of reporting trouble in German." *Research on Language and Social Interaction* 54 (2): 142–162.
- doi Rasmussen, Gitte and Elisabeth Dalby Kristiansen, this volume. "The influence of the specificities of gaze behaviour on emerging and ensuing interaction — A contribution to the discussion of the use of eye-tracking recordings for EMCA analysis" in *Mobile Eye Tracking: New avenues for the study of gaze in social interaction*. ed. by E. Zima and A. Stukenbrock (Amsterdam: John Benjamins).
- Sacks, Harvey. 1992. *Lectures on Conversation* Oxford: Blackwell.
- doi Sacks, Harvey. 1995. *Lectures on Conversation*. Oxford: Basil Blackwell.
- doi Scaife, M., and J.S. Bruner. 1975. "The capacity for joint visual attention in the infant." *Nature* 253 (5489): 265–266.
- Selting, Margret et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 10: 353–402.
- doi Schegloff, Emanuel A. 2007. "Retro-sequences." In *Sequence Organization in Interaction. A Primer in Conversation Analysis*, 217–219. Cambridge: Cambridge University Press.
- Schegloff, Emanuel A. 1988. "Presequences and indirection: applying speech act theory to ordinary conversation", *Journal of Pragmatics* 12 (1): 55–62.
- doi Schütz, Alfred. 1953. "Common-sense and scientific interpretation of human action." *Philosophy and Phenomenological Research* 14 (1): 1–38.
- Schütz, Alfred, and Thomas Luckmann. 1979. *Strukturen der Lebenswelt [Structures of the Life-World]*. Frankfurt/Main: Suhrkamp.
- Siitonen, Pauliina, Mirka Rauniomaa, and Tiina Keisanen. "2021. Language and the moving body: directive actions with the Finnish Kato 'look' in nature-related activities. *Frontiers in Psychology* 12: 66178.
- doi Stukenbrock, Anja. 2015. *Deixis in der face-to-face-Interaktion [Deixis in Face-to-face Interaction]*. Berlin: de Gruyter.
- Stukenbrock, Anja. 2018. "Blickpraktiken von SprecherInnen und AdressatInnen bei der Lokaldeixis: Mobile Eye Tracking-Analysen zur Herstellung von joint attention [Gaze Practices of Speakers and Addressees in Local Deixis: Mobile Eye Tracking Analyses on the Establishment of Joint Attention]" *Gesprächsforschung — Online-Zeitschrift zur verbalen Interaktion* 19: 132–168.
- doi Stukenbrock, Anja. 2020. "Deixis, meta-perceptive gaze practices, and the interactional achievement of joint attention." *Frontiers in Psychology* 11 (1779): 1–23.
- doi Stukenbrock, Anja. 2023. "Temporality and the cooperative infrastructure of human communication: noticings to delay and to accelerate onward movement in mobile interaction." *Language & Communication* 92: 33–54.

 Stukenbrock, Anja and Anh Nhi Dao. 2019. "Joint Attention in passing: what dual mobile eye tracking reveals about gaze in coordinating embodied activities at a market." In *Embodied Activities in Face-to-Face and Mediated Settings. Social Encounters in Time and Space*, ed. by Elisabeth Reber, and Cornelia Gerhardt, 177–213. Cham: Palgrave Macmillan.

 Tomasello, Michael. 2005. *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge: Harvard University Press.

 Tomasello, Michael. 2008. *Origins of Human Communication*. MIT Press.

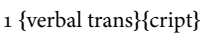
Wilkins, David. 2003. "Why pointing with the index finger is not a universal (in sociocultural and semiotic Terms)." In *Pointing: Where Language, Culture, and Cognition Meet*, ed. by Sotaro Kita, 171–215. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wittenburg, Peter, Brugman, Hennie, Russel, Albert, Klassmann, Alex and Sloetjes, Han (2006). "ELAN: a professional framework for multimodality research". In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy: European Language Resources Association (ELRA).

Zhisheng, W., Y. Nagai, D. Zhu, J. Liu, and N. Zou. 2019. "Based on creative thinking to museum lighting design influences to visitors emotional response levels theory Research." *IOP Conference Series: Materials Science and Engineering* 573: 1–7.

 Zima, Elisabeth, Peter Auer and Christoph Rühlemann, this volume. "Why research in gaze in social interaction needs mobile eye tracking" in *Mobile Eye Tracking: New avenues for the study of gaze in social interaction*. ed. by E. Zima and A. Stukenbrock (Amsterdam: John Benjamins).

Description of the iconic transcripts

	Pentagons above the verbal transcript iconically represent the two walkers (with their initials) and their bodily orientation
	Arrows indicate the walkers' gaze directions
	Dashed arrows indicate that the person is gazing downwards
	Curved arrows indicate that the participants reorient their gaze and body in the direction of the arrow (e.g., turning towards the object of reference)
	Various icons represent the objects of reference. They are only represented in the transcript if at least one of the participants focuses on the object.
	Curly brackets indicate the scope of the iconically illustrated gaze behaviour above the verbal transcript

Appendix A

Transcription convention of Gat-2 (Gesprächsanalytisches Transkriptionssystem 2)

Sequential structure

[] overlap and simultaneous talk

[]

= fast, immediate continuation with a new turn or segment
(latching)

Other segmental conventions

: lengthening, by about 0.2–0.5 sec.

:: lengthening, by about 0.5–0.8 sec.

::: lengthening, by about 0.8–1.0 sec.

? cut-off by glottal closure

In- and outbreaths

°h / h° in- / outbreaths of appr. 0.2–0.5 sec. duration

°hh / hh° in- / outbreaths of appr. 0.5–0.8 sec. duration

°hhh / hhh° in- / outbreaths of appr. 0.8–1.0 sec. duration

Pauses

(.) micro pause, estimated, up to 0.2 sec. duration appr.

(-) short estimated pause of appr. 0.2–0.5 sec. duration

(--) intermediary estimated pause of appr. 0.5–0.8 sec. duration

(---) longer estimated pause of appr. 0.8–1.0 sec. duration

(0.5)/(2.0) measured pause of appr. 0.5 / 2.0 sec. duration

(to tenth of a second)

Other segmental conventions

and_uh cliticizations within units

uh, uhm, etc. hesitation markers, so-called “filled pauses”

Laughter and crying

haha

hehe

hihi

syllabic laughter

((laughs))

((cries)) description of laughter and crying

<<laughing> > laughter particles accompanying speech with

indication of scope

<<:-)> so> smile voice

Continuers

hm, yes, no, yeah monosyllabic tokens

hm_hm, ye_es, bi-syllabic tokens

no_o

?hm?hm with glottal closure, often negating

Other conventions

((coughs)) non-verbal vocal actions and events

<<coughing> > ...with indication of scope

() unintelligible passage

(xxx), (xxx xxx) one or two unintelligible syllables

(may i) assumed wording

(may i say/let us say) possible alternatives

((unintelligible, unintelligible passage with indication of
appr. 3 sec)) duration

((...)) omission in transcript

→ refers to a line of transcript relevant in the argument

Accentuation

SYLlable focus accent

!SYL!lable extra strong accent

Final pitch movements of intonation phrases

? rising to high

, rising to mid

– level

; falling to mid

. falling to low

Other conventions

<<surprised> > interpretive comment with indication of scope

Pitch jumps

↑ smaller pitch upstep

↓ smaller pitch downstep

↑↑ larger pitch upstep

↓↓ larger pitch downstep

Changes in pitch register

<<l> > lower pitch register

<<h> > higher pitch register

Intraline notation of accent pitch movements

`SO falling

'SO rising

~SO level

^SO rising-falling

~SO falling-rising

↑` small pitch upstep to the peak of the accented syllable

↓' small pitch downstep to the valley of the accented syllable

↑~SO bzw. ↓~SO pitch jumps to higher or lower level accented syllables

↑↑`SO bzw. ↓↓'SO larger pitch upsteps or downsteps to the peak or valley of the accented syllable

Loudness und tempo changes, with scope

<<f> > forte, loud

<<ff> > fortissimo, very loud

<<p> > piano, soft

<<pp> > pianissimo, very soft

<<all> > allegro, fast

<<len> > lento, slow

<<cresc> > crescendo, increasingly louder

<<dim> > diminuendo, increasingly softer

<<acc> > accelerando, increasingly faster

<<rall> > rallentando, increasingly slower

Changes in voice quality and articulation, with scope

<<creaky> > glottalized

<<whispery> > change in voice quality as stated

Appendix B

Conventions for multimodal transcriptions (Mondada 2017)

- * * Gestures and descriptions of embodied actions are delimited between
- + + two identical symbols (one symbol per participant)
- △ △ and are synchronized with corresponding stretches of talk.
- *---> The action described continues across subsequent lines
- >* until the same symbol is reached.
- >> The action described begins before the excerpt's beginning.
- >> The action described continues after the excerpt's end.
- Action's preparation.
- Action's apex is reached and maintained.
- ,,,, Action's retraction.
- ric Participant doing the embodied action is identified when (s)he is not the speaker.
- fig The exact moment at which a screen shot has been taken
- # is indicated with a specific symbol showing its position within the turn at talk.

Index

A

action 67–69, 72, 74–75, 77–78,
92, 94–95, 99, 165–167, 171, 173,
178, 181–184, 249, 270, 279, 303,
314
 joint 3, 102, 122, 166, 249
addressee monitoring 162, 247,
249–250, 272
alignment 15, 18, 165–166, 175,
178, 181–184, 125, 186
AOI *See* area of interest
attention 3–4, 11, 13, 65–66, 70,
72–73, 134, 211–212, 243–245,
247–253, 255–257, 259–260,
262–263, 265–272, 280–282,
290, 300

B

backchannel relevance space
135, 163
bodily orientation 242, 267, 277,
282, 290, 293, 294, 297, 307, 310

C

camera perspective 25, 39, 93,
103
continuer 134, 136, 165
co-walkers 212–213, 236–237,
285, 288–289, 293–294, 298,
300, 305–306

D

deixis 20, 98, 187, 252, 275
disalignment 11–12, 18, 88,
165–169, 171, 173, 175, 177–179,
181–185, 187
domain of scrutiny 278, 282,
304, 306
dyadic interactions 3, 136, 199

E

embodied
 action 67, 70, 75, 92, 94,
271, 314

 conduct 69–70, 72, 77, 86,
94, 251
 interaction 10, 14, 16, 19, 93,
187, 241, 273, 275–276
EMCA 24–26, 28–29, 32, 62,
67–68, 93
ethnomethodology 96, 184, 202,
239–240
experience of nature 14,
280–283, 286 304, 307
external video recordings 69,
72–73, 75, 77, 80, 83, 85–86,
90–91, 94
eye contact 10, 109, 111–112,
114–118, 120–123, 147, 155, 159,
210, 229
 breaking of 10–11, 109–111,
115–118, 120–123
eye movements 6, 31, 35–37, 54,
56, 70–72, 109, 168–170, 174

F

face-to-face interaction 1, 20, 98,
100–101, 108–109, 122, 163, 238,
248
feedback 11, 132–133, 135,
138–139, 142, 144, 149, 151,
160–162
Feedback Relevance Space
 (FRS) 11, 132, 134–135, 139,
147–149, 151, 155, 157, 159–163
field of vision 7, 34, 62, 169, 212,
216, 285, 303
F-formations 9, 25, 32–33, 42,
193, 208–209, 211, 213
fixations 6, 68–72, 74, 77, 81,
86–87, 91, 94, 111–112, 114–115,
139, 170
focus of attention 103, 247–248,
252, 263, 265, 269
 competing 13, 20, 66,
243–245, 247, 252–52, 255,
259, 265, 269–270
 divergent 13, 244, 247, 251,
265–266, 268

foveal vision 7–8, 31, 178, 182,
209, 212, 216, 283–284, 303
Freiburg Multimodal Interaction
Corpus (FreMIC) 19, 188,
191–192, 204–205

G

gaze aversion 107–109, 115,
117–120, 140, 143, 157, 165,
167–169, 171, 173, 175, 177,
179–181, 183, 185, 187
gaze constellations 140, 284
gaze contact 9, 43, 48, 51–52, 54,
58–59
gaze cueing 121
gaze direction 18–19, 30–32,
62–63, 121, 124, 239, 297, 303,
306
gaze-following 17, 305, 308
gaze practices 4–5, 9, 11–12, 166,
173, 184, 241, 256, 281
gaze shifts 3–5, 10–12, 102–108,
122, 138, 140, 142, 145, 147–148,
151, 160–162
gaze synchronisation 106, 109,
114–117, 121–123
gaze targets 9, 24, 30, 103, 105,
140, 166, 168, 171
gaze transcription 24–25, 39, 59,
235
gaze window 11, 133, 137, 153
general thesis of reciprocal
 perspectives 14, 279, 297, 306

H

head movements 30–31, 34–35,
37, 56, 62, 102, 106–107, 117, 122,
212–213, 280
humor 173, 214–215, 219,
230–233, 256

I

idealization 279, 290, 293, 306
Insight Interaction Corpus 103,
110

interactional impasse 12, 166,
173, 177, 182–183

intersubjectivity 2, 5, 14, 279,
282–284, 290, 298, 300–307

intonation phrase 11, 134–135,
138, 142, 144–145, 149, 153, 155,
157, 159–160, 312

IP boundary 135, 142, 144, 160

irony 10, 100–102, 104, 106–107,
124–127, 238

J

joint activities 166, 240

joint attention 4–5, 13–14, 20,
30–31, 33, 66, 109, 126, 187,
243–245, 247–253, 255–257, 259,
261, 263, 265, 267–273, 275–283,
285–291, 293–295, 297–299, 301,
303–307, 309

joint perception 281, 284, 307

L

laughable 12, 14, 208, 212,
214–215, 217, 219–224, 230–237

laughter 13, 106–107, 205, 208,
213–216, 218, 222, 224–225, 227,
229, 235–237, 239–241, 256, 258,
264, 267, 311–312

M

mixed-methods approaches 10,
15, 18, 100–101, 103, 105, 107,
109, 111, 113, 115, 117, 119,
121–123, 125, 127, 186

mobile interaction 11–12, 20,
28–29, 68, 95, 208, 216, 238,
276, 280, 309

multimodal interaction 30, 68,
78, 93, 95, 273

multiparty conversation 197,
199, 202

mutual gaze 4, 11–12, 24, 36–37,
39–40, 42–44, 46, 49, 54,
56–59, 61–63, 104, 106–109,
132–134, 136–151, 153–155, 157,
160–162, 182–184, 210

mutual orientation 307

N

natural attitude 27, 279, 306

next-speaker selection 189, 200,
204

noticings 278, 282, 305

P

participation 3, 15, 17, 28–29,
124, 167, 174–175, 182, 185

participation framework 15, 165,
167

perception 8, 27, 30–31, 69, 250,
270, 277, 281–282, 284, 288,
297, 303

peripheral vision 7, 37, 213,
215–216, 219, 283, 285, 300, 303

perspective

bystander's 24–25, 32, 39,
56, 62–63, 168

observer's 5–6, 8–9, 24–25,
39, 48, 57–59, 63, 139, 284

external camera 103, 110,
216

pointing gestures 13, 230,

247–248, 256, 263, 269, 271, 278

practice

embodied 14, 20, 243, 248,
272, 275

response mobilising 13, 271

pupillometry 188, 191, 198

pupil size 12, 16, 188–193,

195–201, 203, 205

processing load 188, 190–192,
197–200

progressivity 166–170, 178,
182–184, 268

Q

quantitative analysis 10, 101, 108,
116, 118, 122, 159

question-answer sequences 4,
12, 136, 188–191, 197–199

R

recipient feedback 21, 66, 127,
132, 137, 148, 164, 187

recipient token 134–136,
138–140, 142–151, 153, 155, 157,
159–162

S

scene camera 7, 31, 169, 212,
216–217, 283–285

search activities 67, 80, 82–83,
86, 88, 92

self-involvement 165, 175,
177–180, 182–183

side-by-side constellation 33,
35–37, 208–209, 211, 213, 231,
237

speech planning 188, 191,
198–200

stance-taking 101, 104, 107, 132,
165, 256, 271, 281–282, 286, 294

synchronisation 101, 108, 110,
113–117, 121–123

T

Transition Relevance Place
(TRP) 107–108, 135, 155, 161,
188

Turn Construction Unit (TCU)
70, 99, 133–135, 142, 202, 262

turn transitions 12, 108–109, 120,
188–192, 197–200

V

video camera, external 69, 71,
78, 85–87, 92–93, 139, 285

visual attention 244–245,
248–250, 252, 255–256,
259–260, 263, 269, 271, 277,
288, 305

visual orientation 14, 179, 182

W

walking and talking 208, 211,
284

Situated within the flourishing domain of pragmatics, this volume explores the crucial role of gaze in human interaction, with a particular focus on the potential of mobile eye tracking to advance our methodology and understanding of multimodal communication. Readers will find a comprehensive, balanced exploration of the benefits and challenges associated with taking eye tracking out of the lab to record authentic interaction in real-life settings. By integrating insights from pragmatics, the contributions highlight the function of gaze as a resource for coordination, cooperation and joint sense-making in human interaction. The chapters are written by leading scholars in the field as well as younger researchers. They offer in-depth methodological discussions alongside detailed case studies from static and mobile interaction settings. The book makes a strong case for the use of mobile eye tracking in addition to video cameras. It provides researchers with a solid and state-of-the-art foundation on which to make informed choices about recording technologies for their own work. The volume is a must-read for scholars in multimodal conversation analysis, interactional linguistics, as well as cognitive linguists, linguistic anthropologists, and psychologists with a strong interest in new ways of studying gaze in social interaction.

ISBN 978 90 272 1993 0



John Benjamins Publishing Company