Tampere University

SAKARI ILOMÄKI

Distant but Present

Rebuilding intersubjectivity in video-mediated interaction

Tampere University Dissertations 699

Tampere University Dissertations 699

SAKARI ILOMÄKI

Distant but Present

Rebuilding intersubjectivity in video-mediated interaction

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Social Sciences
of Tampere University,
for public discussion in the Väinö Linna auditorium
of the Linna building, Kalevantie 5, Tampere,
on 3 December 2022, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Social Sciences
Finland

Responsible supervisor and Custos	Professor Johanna Ruusuvuori Tampere University Finland	
Pre-examiners	Associate professor Wyke Stommel Radboud University Netherlands	Associate professor Brian Due University of Copenhagen Denmark
Opponent	Professor Christian Licoppe Institut Polytechnique de Paris France	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Copyright ©2022 author

Cover design: Roihu Inc.

ISBN 978-952-03-2630-2 (print)
ISBN 978-952-03-2631-9 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-2631-9

Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2022

ACKNOWLEDGEMENTS

This thesis would not have come to be without data to work with and time to dig into it. Those who participated in the data collection provided priceless input for the thesis. Thank you for opening the doors of your apartments and offices and thus making this study possible. I thank the following funders for their support: The Juho Vainio Foundation, the Strategic Research Council at the Academy of Finland (Grant Numbers 303430 and 336277), the Finnish Work Environment Fund (grant number 117151) and the Faculty of Social Sciences of Tampere University.

I am delighted that Professor Christian Licoppe accepted the invitation to act as my opponent. Professor Licoppe's work and thinking on technology-mediated interaction have influenced my work in many ways, providing detailed examples on how to study the relationship of human action and the possibilities that technology affords. It has always been a pleasure to talk with him about what technological mediation means for human conduct. Thank you for devoting your time and for your interest in my work.

I thank my pre-examiners Wyke Stommel and Brian Due for providing thorough, insightful and constructive feedback on the thesis. They helped me clarify the formulation of the research problem, elaborate my findings and connect my work better to existing research and professional practice. They pushed me to epistemological considerations which led me to the borders of my current understanding, where I had to admit to myself that I do not know. I think that this is what a review process at its best can offer. Their input improved the thesis significantly; any remaining mistakes and shortcomings are my own.

My supervisor, Professor Johanna Ruusuvuori, has undeniably played an enormous part in my work and this thesis. She has spent countless hours reading and commenting on my drafts, bringing the core of the findings into daylight, providing viewpoints that made it possible to communicate my ideas to outside world and noting examples of where I could improve my analysis. On many occasions she has offered much needed emotional and practical support. Johanna has been my

supervisor since I was working on my bachelor's thesis, guiding me when needed and letting me seek my own path when possible. After all these years under her thoughtful supervision, I feel confident to carry on more independently.

In addition to my supervisor, I had the pleasure of working with other authors when writing my articles. Jaana Laitinen provided the data from group health counselling and offered valuable reflections on the relevance of the findings for professional practice. Collaborating with Melisa Stevanovic on the topic of problem solving in hybrid interaction was a delight. Melisa's critical, precise and theoretically superb comments and suggestions helped shape the article into what it came to be. Thank you both!

I conducted my research as part of two research projects. I thank the Promo@Work – Evidence-Based Health Promotion at Work consortium for an interesting multidisciplinary experience: Jaana Laitinen, Kirsi Heikkilä-Tammi, Matti Muhos, Harri Oinas-Kukkonen, Salla Ahola, Heli Hannonen, Janne Jaako, Markku Kekkonen, Eveliina Korkiakangas, Aija Logren, Sirpa Lusa, Jaana-Piia Mäkiniemi, Satu Mänttäri, Juha Oksa, Tuula Oksanen, Anne Punakallio, Jouko Remes, Jani Ruotsalainen, Johanna Ruusuvuori, Anna-Mari Simunaniemi, Sanni Tiitinen, Piiastiina Tikka and Jos Verbeek. I also thank the research team of Healthcare Workers in the Eye of the Digital Turbulence: New Forms of Cooperation and Customer Orientation: Päivi Husman, Johanna Ruusuvuori, Onni Aalto, Tiina Koivisto, Inka Koskela, Elisa Kurtti, Virve Peteri and Elina Weiste. They provided a valuable learning experience for me on project work and the opportunity to develop as an independent researcher. I especially want to thank Tiina, who in many ways put in the effort to get me access to the field to collect data and set an example on how to manage projects efficiently. It was always a pleasure to work with you.

Throughout the years, I have received endless support from the social psychology PhD seminar at Tampere University. Thank you to Teija Ahopelto, Samuli Järvinen, Heidi Kajander, Aku Kallio (I became interested in conversation analysis in the first place because of Aku's talent in floorball, so blame it on him if you don't like the thesis), Julia Katila, Katri Keskinen, Tiina Koivisto, Elisa Kurtti (who also produced the line drawings for Article 3, so extra thanks for those), Aija Logren (who shared an office with me during the final stages of the writing process, offering a companion to discuss ideas and problems I encountered), Eerik Mantere, Miia Marttinen, Juhana Mustakallio, Paula Paajanen, Federica Previtali, Hanna-Leena Ristimäki and Reetta

Riikonen. I especially want to thank Hanna-Leena, with whom I have had the pleasure of working since my master's seminar, for managing the demanding dual role of being both a peer early-career researcher and a role model on how to conduct high-quality analysis. I also thank the people who shared their insights into my data during data sessions at Tampere University and elsewhere. Providing support to one another in this way reveals the true beauty of research as communal practice.

During this process, I have had the pleasure to get to know many inspiring and kind people from the international conversation analyst community. I have shared many fun and thought-provoking data sessions with Jessica Hansen and Simon Bierring Lange. We were all PhD researchers covering the same kinds of topics, and having this small and supportive group of youngsters was very important for me. Jessica and Simon finished their degrees before I did, and I urge you to read their work as well. Thank you both for the companionship!

I would like to thank the community of ethnomethodology and conversation analysis researchers who gathered to discuss video-mediated interaction in early 2019, a discussion that later evolved into the theme issue in which my first article was published: Alexandra Gross, Ann Merrit Rikke Nielsen, Brian Due, Christian Heath, Christian Licoppe, Jeanette Landgrebe, Jessica Hansen, Johannes Wagner, Karin Birkner, Mie Femø Nielsen, Paul Luff, Simon Bierring Lange and Wyke Stommel. Thanks also to Tuire Oittinen, Heidi Jauni, Teppo Jakonen and Lucas Seuren. Inspiring talks and the sense of community that all of you have provided have been essential for my work!

I thank Niklas Norén for coming up with the idea that the question about the use of multiple data sources on video-mediated interaction could be compared to using ethnographic data to support the analysis. My interpretation of this idea, with all its benefits and weaknesses, is my responsibility.

I met Sanni Tiitinen in the very early stages of my career as a conversation analyst, when Sanni commented my bachelor's thesis. While the text was undoubtedly crude, Sanni's comments were always constructive. This combination of sharp-sightedness and supportiveness continued during the PhD process whenever I had the chance to get Sanni's feedback. I later had the privilege of collaborating with Sanni in the Promo@Work research project. Seeing that it is possible to stay humane while producing high-quality research was a truly important lesson in my early career,

especially as life in academia can sometimes be harsh. Sanni is a true role model for me.

Tuuli Turja was a great support throughout the process, asking how it was going when I needed to hear that question and not asking it when I did not want to answer. Anna Kork's warm words in spring 2022 were one of the most humane experiences I have had in academia. I am delighted that I will continue working with Tuuli and Anna in the future. Sanna Raudaskoski has always been an inspirational ethnomethodologist and teacher, warm supporter and enthusiastic tea drinker. Without Sanna's introductory course to ethnomethodology, which showed me that research can and should be a fun endeavour, I probably would not have pursued a career as a researcher in the first place.

My childhood family provided the substrate from which I have built myself. I thank my parents Emma and Eero and my siblings Risto, Immi, Paavo and Timo. Our endless and sometimes heated discussions on society sparked the interest in societal issues in me, which I have taken into my own direction. Thank you also for all the amazing people you have brought into my life by building your own families. A special shoutout goes to my godchild Toivo, whose ability to call into question every members' category shows how the everyday is the most fascinating thing.

A big thank you to all my friends everywhere.

Thanks to the weekly role-playing game group – Teemu, Antti, Satu and Mase – for providing me a weekly getaway from this world, the possibility for deep reflection and a constant supply of under-the-belt mayhem. While we have not yet got to make any sanity rolls, you have definitely been key to keeping me sane during this process.

Thank you, Maiju, for sharing everyday life and celebration with me. You not only offer intellectual debate but also provide me a safe space where I can learn to be me. I hope you have seen me in my darkest days, and I hope to show you our brightest days.

ABSTRACT

During the last few decades, video mediation has become ubiquitous in our social lives. This is due to accessible equipment and internet connections and societal changes that support the adoption of technologically mediated interaction. Research and theorising on video mediation have concentrated on measuring task performance and explaining it with the features of video mediation, often described as *cold* or *poor* compared to face-to-face interaction. Less is known about how interaction unfolds in video-mediated settings and how, despite the different possibilities for action compared to face-to-face settings, people are able to maintain and rebuild shared understandings in video mediation. In this doctoral dissertation, which is comprised of four empirical articles and an integrative chapter, I examine how the processes of intersubjectivity – that is, forming, maintaining and repairing a shared understanding of the ongoing action – are managed in interaction and how video mediation becomes consequential in these processes. I use ethnomethodological conversation analysis to examine how interactants recognise technology-generated ruptures of intersubjectivity, how they repair these ruptures, and how technical mediation becomes consequential for these practices in video-mediated interaction.

I use data from three different settings: 1) video-mediated health counselling groups simultaneously video-recorded from both the group and the leader perspectives, 2) video-mediated tele-homecare visits between single home-dwelling older adults and professionals recorded in either the home environment or the professional's office, and 3) hybrid tele-consultations recorded in a general practitioner's office. The varied nature of the data in terms of institutional context, number of participants, technological settings and perspectives recorded enable analysing two recurring phenomena that have been recognised in ethnomethodological research on video-mediated interaction: a) how transmission delays produce non-mutual interactional realities where the timing of actions differ from different perspectives of action, and b) how limited video frames produce fractured ecologies which hinder the co-ordination of body movements and the use of artefacts.

Based on the analyses of the data, I argue that as in face-to-face interactions, participants recognise the ruptures of intersubjectivity against the sequential relevance of actions. Whether or not the interactants recognise something as potential trouble for intersubjectivity is contingent on the distant participant's ability to participate in an appropriate way in a given situation; that is, to produce sequentially relevant next actions. When resolving the ruptures of intersubjectivity, interactants need to make their perspectives available to others and have those others' perspectives available to themselves. The interactants achieve this by two intertwined practices: *verbal explications* and *physical demonstrations*. When employing these practices, individuals orient themselves to the technological mediation as relevant by fitting their conduct to the media available to the other participants. Video mediation shapes the conditions of both evaluating sequential relevance and making perspectives salient, as it distorts both the timing of turns and the space in which bodily interactions are produced and received. Thus, when repairing intersubjectivity, interactants fit their repair practices to the affordances of the technological medium and the nature of a given misunderstanding.

The study contributes to conversation analytic research on video-mediated interaction by examining the relationship of verbal explications and physical demonstrations with the more general topic of repairing intersubjectivity in video-mediated interaction. Based on these analyses, I suggest that the broader field of computer-mediated communication would benefit from the action-centred and context-sensitive mode of analysis offered by ethnomethodological conversation analysis, as this would highlight creative and diverse ways of using communication technologies and offer a more robust theoretical understanding of the relationship between human conduct and communication media.

TIIVISTELMÄ

Viimeisten vuosikymmenten aikana videovälitteisestä vuorovaikutuksesta ja tullut Skypen ja Teams-kokousten myötä erottamaton osa jälkiteollisen maailman arkea. Yleisyydestään huolimatta videovälitteinen vuorovaikutus koetaan usein vaillinaisena, ja sosiaalisten vihjeiden, kuten eleiden ja äänensävyn, merkitystä painottavat lähestymistavat korostavat, kuinka videovälitteinen vuorovaikutus on kasvokkaiseen verrattuna *kylmää* tai *epäluonnollista*. Vähemmän kuitenkin tiedetään videovälitteisen toiminnan vuorovaikutusdynamiikasta ja siitä, miten ilmeisistä rajoitteistaan huolimatta, tämä teknologinen alusta tarjoaa mahdollisuuksia jaetun ymmärryksen ylläpitoon ja korjaamiseen. Tässä neljästä tutkimusartikkelista ja yhteenvetoluvusta koostuvassa väitöskirjatutkimuksessa tarkastelen, miten tätä jaettua ymmärrystä meneillään olevasta toiminnasta, intersubjektiivisuutta, rakennetaan ja korjataan videovälitteisessä vuorovaikutuksessa ja miten kyseinen viestintäteknologia tulee toiminnan kannalta olennaiseksi näissä prosesseissa. Hyödynnän työssäni etnometodologista keskustelunanalyysia selvittääkseni, miten teknologian aiheuttamia yhteisymmärryksen ongelmia tunnistetaan ja korjataan sekä miten videovälitteisyys muokkaa näitä prosesseja.

Käytän väitöskirjani tutkimusaineistona videotaltiointeja kolmenlaisista vuorovaikutustilanteista sosiaali- ja terveysalan todellisissa asiakastapaamisissa: 1) videovälitteisistä elintapaohjausryhmistä, jotka taltioitiin sekä ryhmän että ohjaajan näkökulmasta, 2) kotihoidon etäkäynneiltä, jotka taltioitiin joko hoitajien tai kotihoidon palveluja käyttävien ikäihmisten näkökulmasta sekä 3) etäkonsultaatioilta, joissa potilas ja yleislääkäri ovat yhteydessä erikoislääkäriin videoyhteyden kautta. Tämä monipuolinen aineisto mahdollistaa kahden aiemmassa tutkimuksessa tunnistetun videovälitteisen vuorovaikutuksen perusilmiön tutkimisen. Ensinnäkin, teknologiavälitteisyyteen liittyvä viive tuottaa *jakamattomia vuorovaikutustodellisuuksia*, joissa toimintojen ajoitukset suhteessa toisiinsa vääristyvät. Toiseksi, videovälitteisissä kohtaamisissa kehollisia toimintoja tuotetaan ja tulkitaan toisistaan irrallisissa *pirstoutuneissa toimintaympäristöissä*, mikä vaikuttaa näiden toimintojen tulkitsemiseen ja ymmärtämiseen.

Analyysini osoittavat, että kuten kasvokkaisessa myös videovälitteisessä vuorovaikutuksessa yhteisymmärryksen ongelmia tunnistetaan edellisten

vuorovaikutustekojen ja toimintojen luomia toiminnallisia odotuksia vasten: mikäli toinen osapuoli jättää tekemättä odotuksenmukaisen teon tai tekee jotain, mikä ei näyttäydy odotuksenmukaisena, voi toinen kohdella tätä yhteisymmärryksen hajoamisen merkkinä. Ratkoessaan näitä ongelmia, osallistujien pitää tuoda oma näkökulmansa toiselle tiettäväksi ja vastaavasti saada toisen näkökulma omaan tietoonsa. Tämän näkökulmien vastaavuuden aikaansaamiseksi osallistujat hyödyntävät kahdentyyppisiä yhteisymmärrystä korjaavia toimintoja, *sanallisia eksplikointeja* ja *fyysisiä demonstraatioita*. Koska videovälitteisyys irrottaa toiminnan tuottamisen ja tulkinnan ajallisesti ja tilallisesti toisistaan, sekä toimintojen odotuksenmukaisuuden tulkitsemisen että näkökulmien jakamisen edellytykset muuttuvat. Niinpä osallistujat sovittavat vuorovaikutustaan niihin toiminnallisiin mahdollisuuksiin, joita videoneuvotteluteknologia tarjoaa, saadakseen intersubjektiivisuuden palautettua.

Väitöskirjatutkimukseni täydentää videovälitteistä vuorovaikutusta käsittelevää keskustelunanalyyttista tutkimuskirjallisuutta käsittelemällä sanallisten ekpslikointien ja fyysisten demosntraatioiden roolia intersubjektiivisuuden korjaamisen keinoina kolmessa erilaisessa palvelukontekstissa. Analyysini ja yleisen etnometdologisen teoretisoinnin pohjalta ehdotan, että laajempi teknologiavälitteisen vuorovaikutuksen tutkimuskenttä hyötyisi näiden lähestymistapojen tarjoamasta toimintojen kontekstisidonnaista merkitystä korostavasta lähestymistavasta kahdella tapaa: ensinnäkin, tutkimuksen huomio siirtyisi teknologioiden rajoittuneisuuden tarkastelusta ihmisten moninaisten ja luovien toimintatapojen tarkasteluun teknologiavälitteisissä ympäristöissä, ja toiseksi, tämä siirtymä tarjoaisi mahdollisuuden kestävämpien teoretisointien muotoilulle, ja siten ihmisen ja teknologian välisen suhteen tarkemmalle ymmärtämiselle.

CONTENTS

1 INTRODUCTION 1

1.1 Development of video mediation 3

1.2 The social cues approach to video-mediated interaction 6

1.3 The affordance approach to technologically mediated interaction 10

1.4 The ethnomethodological perspective on intersubjectivity as a local achievement 12

2 CONVERSATION ANALYSIS AS THE ANALYTIC PERSPECTIVE 16

2.1 Conversation analysis as a description of members' methods 16

2.2 Key analytic terms in the dissertation 17

2.2.1 Turn-taking and overlap resolution 17

2.2.2 Sequence organisation 19

2.2.3 Repair organisation 20

3 THE CONVERSATION ANALYTIC PERSPECTIVE ON VIDEO-MEDIATED INTERACTION 22

3.1 Openings and closings of video-mediated interactions 24

3.2 Technological disruptions 26

3.3 Visual connection and fractured ecologies 28

4 THE DATA, ANALYTICAL PROCESS AND RESEARCH ETHICS 32

4.1 Data 32

4.2 Analytical process and research questions 36

4.3 Research ethics 39

5 RESULTS 42

5.1 Recognising problems with intersubjectivity 46

5.2 Resolving problems with intersubjectivity 48

6 DISCUSSION 53

6.1 Video mediation creates multiple contexts of action which complicate context-sensitive sense-making 53

6.2 Intersubjectivity in video-mediated settings: From "why that now?" to "when is *now* to whom?" and "what *that* is?" 56

6.3 Understanding social cues as contextual configurations in video-mediated interactions 58

6.4 Practical implications for professional video-mediated interaction 61

6.5 Methodological remarks and directions for future research 63

REFERENCES 70

List of Tables

Table 1. Overview of the data used

Table 2. Initiation and execution of perspective display.

TRANSCRIPTION SYMBOLS

[word]	Brackets: onset and offset of overlapping talk
=	Equals sign: contiguous utterances, second is latched immediately onto the first
(0.2)	Timed interval within or between utterances, measured in seconds and tenths of seconds
(.)	Interval of less than 0.2 s
wo:rd	Colon: extension of the sound or syllable
.	Full stop: falling intonation
,	Comma: continuing intonation
?	Question mark: rising intonation1
↑	Upward arrow: Rising pitch
↓	Downward arrow: Falling pitch
wo-	Dash: abrupt cut-off
WORD	Capital letters: louder volume
>word<	Faster-paced talk than surrounding talk
<word>	Slower-paced talk than the surrounding talk
°word°	Degree signs: quieter volume
hh	Audible aspiration.
hh	Audible inhalation
w(h)ord	Laughter
(----)	Lines in paranthesis: Unclear and unidentifiable talk
((word))	Text in parentheses: transcriber's comments
→	Arrow: Feature of interest

%	Percentile sign: Clients' bodily actions
+	Plus sign: Professionals' bodily actions
…	Line of stops: Preparation of a bodily action
„	Line of commas: Retraction of bodily action
♦	Diamond sign: Unspecified features on a screen or audio system
#	Hash sign: The timing /location of the images

ORIGINAL PUBLICATIONS

Publication I Ilomäki, S., Ruusuvuori, J. & Laitinen, J. (2021). Effects of Transmission Delay on Client Participation in Video-Mediated Group Health Counseling. *Qualitative Health Research, 31*(12), 2328–2339. https://doi.org/10.1177/10497323211010726

Publication II Ilomäki, S. & Ruusuvuori, J. (2020). From appearings to disengagements: Openings and closings in video-mediated telehomecare encounters. *Social Interaction. Video-Based Studies of Human Sociality, 3*(3). https://doi.org/10.7146/si.v3i3.122711

Publication III Ilomäki, S. & Ruusuvuori, J. (2022). Preserving client autonomy when guiding medicine taking in telehomecare: A conversation analytic case study. *Nursing Ethics, 29*(3), 719-732. https://doi.org/10.1177/09697330211051004

Publication IV Ilomäki, S. & Stevanovic, M. (submitted) Distributed cognition in fractured ecologies: Collaborative problem solving in video-mediated interaction.

1 INTRODUCTION

The shared understanding of what is going on – that is, intersubjectivity – may become complicated in video-mediated interaction. Imagine a mundane instance from a video-mediated tele-meeting involving a typical work team. The meeting starts, and after some audio checks and activation of cameras and microphones, the team is ready to proceed with the business at hand. After a brief update on everybody's situation, the team leader starts to talk about the team's goals and working methods for the next quarter before asking the participants' opinions: "What do you think would be the best of these options: A, B or C?" But they have no idea which of the three options the leader is talking about, let alone what aspects relate to which option. After a moment of awkwardness, the team leader recognises that she hasn't turned on screen sharing, apologises and activates screen sharing to reveal a PowerPoint presentation with the information needed. In the blink of an eye, all participants are quite literally on the same page and can see what the leader was talking about by "these options".

This hypothetical but likely familiar scenario provides an example of how a shared understanding of what is going on – that is, intersubjectivity (see section 1.4) – can become complicated in video-mediated interaction. By video mediation, I refer to synchronous interactions in which participants have visual access to one another via web cameras; it is sometimes called *video telephony* or *videoconferencing* and uses software like MS Teams and Zoom. The scenario also shows how, despite being omnipresent, the problems of intersubjectivity do not characterise the whole interaction but become relevant only at certain points. Furthermore, the interactants need to work towards regaining intersubjectivity in the flow of interaction. This raises questions about where in video-mediated interaction problems of intersubjectivity arise, how interactants recognise and resolve them, and how technology shapes these processes. These are the kinds of questions I examine in this doctoral dissertation.

To gain clarity into precisely what video mediation is and how it is relevant for interactants, I adopt an ethnomethodological approach to mediated interaction. This approach considers technological mediation to be

> the way the particular organization and unfolding of activities in definite material settings might constrain or enable or even "afford" … the production of particular forms of accountable responses and shape the criteria to assess their relevance … to be found in the way the production of particular sequences may be accomplished, enabled, constrained, or inhibited. (Arminen et al. 2016, p. 293)

Accordingly, I study the ways in which video mediation technology brings possibilities of and hindrances to maintaining intersubjectivity not by determining actions but affording them (section 1.3). In order to understand how mediation shapes the processes of producing and repairing intersubjectivity, it is vital to analyse how participants in concrete situations orient themselves to mediation in their own perceivable actions. In this dissertation, I use video-recorded data from naturally occurring interactional encounters in three kinds of health and social contexts – group health counselling, tele-homecare and tele-consultations – to study the relationship between communication technology and human conduct; a detailed description of the data appears in chapter 4. Based on this approach, my research questions are as follows: 1) How do interactants display recognition of technology-generated ruptures of intersubjectivity in video-mediated interaction? 2) How do interactants repair these ruptures in intersubjectivity? 3) How does technical mediation become consequential for these practices as interaction unfolds?

My research is based on a conversation analytic examination of different video-mediated encounters. Chapter 2 describes the methodological approach, chapter 3 reviews existing conversation analysis research on the topic, and chapter 4 details the methods and data. Drawing from ethnomethodological theorising to interpret these analyses (see section 1.4), I argue first that as with face-to-face interaction, people display recognition of the ruptures of intersubjectivity against the normative expectations in interaction; that is, if another interactant does not produce an expected turn, this may be treated as signalling a rupture. Second, restoring intersubjectivity is achieved by making one's own perspective available to others (and vice versa) through verbal explications and physical demonstrations. Third, as part of the interactional ecology, video mediation can detach the production and reception of actions from each other, and thus the basis for both interpreting actions as expected or not and the grounds for making one's perspective salient may become obscured. Thus, when rebuilding intersubjectivity, the interactants fit their interactional practices to the technological affordances of the medium. By

comparing my findings to those from earlier research into computer-mediated communication using social cues approaches, namely social presence theory, media richness theory and media naturalness theory (section 1.2.), I go on to suggest that the broader field of computer-mediated communication research would benefit from the action-centred and context-sensitive aspects of ethnomethodological conversation analysis by shifting the analytical focus from the communication media's limited ability to transmit social cues to diverse modes of use and by gaining a more robust theoretical understanding of the relationship between human conduct and communication media.

The structure of the thesis is as follows: after briefly discussing the history of video mediation, which I argue has affected how we study and theorise video-mediated interaction, I devote the rest of chapter 1 to introducing the social cues approach, the prevalent approach to technologically mediated interaction, and discuss some of its limitations before contrasting it with the concept of technological affordances and ethnomethodological theory, both of which I draw on in my empirical analyses. In chapter 2, I describe my research method, multimodal conversation analysis (Mondada, 2019; Sidnell & Stivers, 2012; ten Have, 2007), and its central concepts, followed by a literature review on conversation analytic research on video-mediated interaction in chapter 3. I continue by describing my data and research process in chapter 4. The main body of the thesis is chapter 5, in which I present a synthesis of the results of the published articles to exemplify my main arguments about rebuilding intersubjectivity. Finally, in chapter 6 I discuss my research contribution to the fields of ethnomethodological conversation analysis and the broader research field of computer-mediated communication, the limitations of the work presented and potential avenues of future research.

1.1 Development of video mediation

While video-mediated interaction is almost a century old, growth in the use of video mediation technology was slow until the last decade. The earliest visions of tele-videotelephony arose not long after the advent of the telephone. For example, Mark Twain's science fiction story "From the 'London Times' of 1904" describes the telectroscope, a device which would make "the daily doings of the globe … visible to everybody, and audibly discussable too" (1898, p. 101). The first real-world example of

video mediation occurred in 1927, when a connection between the Secretary of Commerce (and future president) Herbert Hoover in Washington, DC, and a Bell Labs representative in New York was established (Schnaars & Wymbs, 2004). The first commercial video-calling service operated between 1936 and 1940 in Germany, when the first closed-circuit systems for video telephony through video calling booths was introduced (Rintel, in press). After World War II, attempts to introduce video mediation into the corporate world took place in many industrialised countries. For example, in the United States, AT&T began efforts to introduce videoconferencing in the 1960s, first through video calling booths, later by desktop videophones and finally by videophone meeting services (Schnaars & Wymbs, 2004). However, challenges such as low picture quality, incompatibility with standard transmission cables, high prices and a lack of demand contributed to the slow uptake of video conferencing systems (Schnaars & Wymbs, 2004), and video mediation remained largely a curiosity until the first decade of the 21st century (Mlynář et al., 2018; Rintel, in press). However, in geographically remote areas, where long travel was not realistic, videoconferencing remained a feasible option, although usage rates were low (Scollon, 2004).

The development of cheaper technology which was more compatible with existing devices and slow but steady changes in the economic structure during the 20th century gave rise to video mediation. Skype, the first videotelephony application based on Voice over Internet Protocol (VOIP), was introduced in 2003. Compared to earlier videoconferencing technologies, which demanded a separate videophone device or even a dedicated room, videoconferencing now became possible with existing personal computers and off-the-shelf web-cameras, eliminating many of the barriers to adoption. Simultaneously, the prices of the equipment decreased (Statista, 2022a, 2022b). This produced a self-feeding spiral, with lower prices contributing to more sales which further lowered prices and led to more uptake of video mediation. For example, in Finland, the national context of this PhD thesis, only 5% of internet users had made videocalls and 10% other internet-based calls in 2005; the proportion of those making any type of internet-based calls reached 67% in 2019 (Statistics Finland, 2005, 2021). By then, the potential customer base for professional video-conferencing services had also increased. As the end of the 20th century approached, the economic structure evolved, with people moving from farms and factories to offices, creating more potential users and uses for video mediation in professional contexts. It was no longer a few managers who would be the target clientele for video mediation, but ever-expanding cadres of

white-collar workers. Most recently, the COVID-19 pandemic has normalised video-mediated interaction, making it virtually ubiquitous in industrialised societies.

As was the case in business and management, the uptake of video-mediated services in health and social care, the institutional context of this study, has been slow (Hyppönen et al., 2015; Kyytsönen et al., 2021; Reponen et al., 2021). However, the forced adoption of tele-services during the COVID-19 pandemic has accelerated that process. In addition to expected usefulness, ease of use, social influences and the price of video mediation technology (Connolly et al., 2020; Schmitz et al., 2022; Venkatesh et al., 2012) and the more general reasons described above (Schnaars & Wymbs, 2004) problems with interaction can be considered one potential explanation for its slow uptake in health and social care. While video-mediated services can be an efficient way of delivering care in many settings (see, e.g., Laitinen et al., 2010, for video-mediated health counselling, Greene et al., 2010; Latikka et al., 2021; Marquis et al., 2015, for care of older adults, and Aidemark, 2022; Ignatowicz et al., 2019 for video consultations), both professionals and clients have reported negative interactional phenomena in video-mediated services, such as difficulty in non-verbal communication and an overall unnatural feel (Connolly et al., 2020; Dalley et al., 2021; Ignatowicz et al., 2019; Seuren & Shaw, 2022).

Thus, to understand why video-mediated services may be perceived as insufficient, we must turn our focus to interaction. A prominent line of reasoning in the study of computer-mediated-communication is that when interaction moves from face-to-face encounters to technological platforms, important social cues are lost in transmission, causing problems in understanding and building relationships (Walter, 2011; Whittaker, 2003). However, as I discuss below, this medium-centric approach may lead to problems like oversimplified analyses of human action, overlooking possibilities for human choice and an inability to make rigorous and durable theoretical claims about the relationship between technology and human action. This may ultimately hinder our understanding of the central intersubjective processes in video-mediated interaction.

1.2 The social cues approach to video-mediated interaction

The slow adoption of video mediation initially took place in business and management. This shaped academic interest in the topic, and research on video-mediated interaction as part of computer-mediated communication first evolved in the field of organisation studies. Unsurprisingly, that early research largely focused on producing knowledge about issues relevant to that area, such as acceptance of new technologies and their effects on task performance. From the viewpoint of technology acceptance, it has been shown how behavioural intentions and thus actual use of video mediation are formed by people's expectations about its usefulness and ease of use, social influences like norms concerning such use and facilitating conditions such as support, motivation and value for money (Connolly et al., 2020; Schmitz et al., 2022; Venkatesh et al., 2012). Another prominent line of research has concentrated on explaining how technical media's limited ability to transmit social cues affects task performance and relationship building (Walter, 2011; Whittaker, 2003). Within this line of research, there is a persistent notion that an impersonal (Short et al., 1976), cold (Rosetti & Surynt, 1985) poor (Daft et al., 1987; Daft & Lengel, 1983, 1986) or unnatural (Kock, 2004) communication medium like video mediation is able to transmit fewer communication cues than a warm, rich or natural medium. This inability makes video mediation less satisfactory than face-to-face interaction and can lead to decreased task performance. It is possible that the connotations of these terms, especially their appearance in dichotomous pairs (personal–impersonal, warm–cold, rich–poor, natural–unnatural) that ultimately reflect the good–bad dichotomy (Lévi-Strauss, 1980), may hinder our understanding of the technology as not only a limiter but also an enabler of actions.

The first iteration of this social cues approach appears in the foundational book *The Social Psychology of Telecommunication* (Short et al., 1976), in which the authors propose *social presence theory*. Drawing on the idea that visual cues are more suitable for transmitting interpersonal material, whereas aural cues transmit less personal interparty material, social presence theory suggests that changes in a given medium's ability to transmit these different cues alter the perceived social presence of the other and therefore the ways in which subjects interact (Short et al., 1976). This further shapes the effectiveness of interaction, which relies on a match between the medium's level of social presence and the task's requirements of interpersonal involvement. This idea of the medium's ability to transmit social cues is sometimes referred to as the varying *warmth* of the

communication medium, with warmer media like face-to-face contact conveying more social cues than colder ones like video mediation and text messaging (Rosetti & Surynt, 1985). As results studying the immersiveness of the technological medium, social presence and task outcomes have been mixed, it was later suggested that, in addition to a medium's ability to transmit social cues, contextual differences such as task type and individual psychological traits moderate the social presence experienced and task outcomes (Oh et al., 2018).

The idea of reduced cues later evolved into *media richness theory* (Daft et al., 1987; Daft & Lengel, 1983, 1986), according to which the information richness of a medium is conceptualised as its ability to transfer information within a given time frame. The transfer of information can be further divided into the number of social cues and the ability to give feedback (Daft et al., 1987; Daft & Lengel, 1983, 1986). Media richness theory thus added an important element to theorising about video-mediated interaction: the latency, delay or lag time between producing and receiving a message. For example, production and reception of turns happen instantly in face-to-face interaction[1], while by email it can vary from seconds to minutes, depending on the connection, and can extend far longer if the recipient is away from keyboard.

Early research on media richness assumed that the key social cues inherent in face-to-face interaction were missing in computer-mediated communication, whereas more recent dynamic approaches argue that participants can compensate for the lack of social cues in mediated environments (Andel et al., 2020; Roos et al., 2020). The results on the connection of media richness and task performance are mixed (Klitmøller & Lauring, 2013; Suh, 1999), and it has been suggested that different intermediate factors, such as group cohesion and the interaction style of the group, can mediate these effects (Montes López, 1992; Yoo & Alavi, 2001).

From the viewpoint of shared understanding, media richness theory reveals two sources of misunderstanding, or *information contingencies*, as they are conceptualised and called within that framework: uncertainty, which refers to the absence of information, and equivocality, which refers to ambiguity between different interpretations about information that is present (Daft & Lengel, 1986). The theory also suggests a

[1] Of course, there is a slight delay even in face-to-face settings, as it takes about 0.003 seconds for audio waves to travel one metre in air. However, humans have clearly adapted to cope with this.

relationship between the contingency of a given task and the communication medium chosen: the more complex – that is, uncertain or equivocal – an issue is, the richer a medium should be to successfully carry out the information exchange. Thus, while social presence theory suggests that more personal media, including video mediation, are more suitable for personal communication, media richness theory suggests that it is task complexity that should guide the selection of a medium.

Portraying itself as a substitute for social presence and media richness theories, the *psychobiological model* (Kock, 2004), sometimes referred to as *media naturalness theory* (the label I will be using here; see Karl et al., 2022), suggests that when the naturalness of a communication medium decreases, the cognitive effort demanded of interactants increases, which may lead to negative changes in communication outcomes, such as lowered satisfaction and efficacy or choosing some other communication media entirely. Thus, compared to earlier theories, media naturalness theory emphasises the match between the medium and the "biological communication apparatus" (Kock, 2004, p. 329); that is, the human body and the organs taking part in the interaction process.

The naturalness of a medium can be traced to two dimensions and their respective characteristics. The space–time dimension comprises the colocation and synchronicity of the medium, while the expressive–perceptual dimension comprises the ability to produce and observe facial expressions, body language and speech, in short social cues (Kock, 2004). Furthermore, these expressive characteristics are regarded as hierarchical since talk is conceived of as significantly more important than the other elements of the expressive–perceptual dimension.

An interesting addition to earlier theorising is that because matching the communication medium, the biological communication apparatus and the task is central, media naturalness theory suggests that a medium can be *too* rich (Kock, 2004). While media richness theory claims that more is always merrier, media naturalness theory proposes that for certain tasks, media with less expressive power and ability to transmit social cues can actually work better. Accordingly, simple tasks such as checking if one has booked a meeting room can be achieved with media low in both the space–time and expressive–perceptual dimensions, such as email, while more expressive forms could complicate the task.

Despite some differences in detail and methodological emphases, such as media naturalness theory taking the individual as the unit of analysis where social presence and media richness theories usually take groups or organisations, all these approaches follow the basic principle that video mediation is able to transmit fewer communication cues than face-to-face interaction, which perforce makes it less effective than face-to-face interaction. They have built upon one another by expanding the original idea according to which the ability to transmit social cues increases the perceived presence: social presence theory began by systematising the idea that it is the communication medium's (in)ability to transmit social cues that determines the experience of social presence and task outcomes; media richness theory added the idea of synchronicity as an explanatory factor beyond social cues; and media naturalness theory positioned social cues in the expressive–perceptual dimension and synchronicity in the time–space dimension of interaction. Interestingly, while the pioneering work by Short et al. (1976) recognised that if made "aware of the reduced-cue situation, the actor will modify his behavior" (p. 64), researchers in both media richness theory (Andel et al., 2020; Roos et al., 2020) and media naturalness theory (Kock, 2001, 2004) have brought up the idea of fitting one's conduct to the available media as a new and important insight. This fitting of one's actions to the technological medium appears to be something that is repeatedly invented, forgotten and re-invented.

While these approaches provide valuable insights into the kind of phenomena that predict the uptake of video mediation, what happens to performance in video-mediated settings and to some extent why this happens, they say little about *how* it happens. My focus in this dissertation is on this very question: how do people modify their conduct to overcome the technological medium's inability to transfer social cues or the other hindrances that mediation causes in the processes of feedback and colocation to rebuild a shared understanding?

On a more theoretical level, the social cues approach takes on an object- or technologically centred view of mediated interaction. That is, it begins the process of analysis from the technology and its features and explains changes in human behaviour as caused by that technology in a nearly direct causal manner. This approach has several shortcomings, such as the risk of reductionist and oversimplified analyses of human action (de la Cruz Paragas & Lin, 2016), overlooking the possibility of choosing between different lines of action (Wyatt, 2007) and, ultimately, the inability to make robust

theoretical claims about the broader connections between technology and human action due to an emphasis on technological features that are subject to constant change (Flanagin, 2020). In the end, this kind of object-centred approach, which focuses on the new and unique features of technological artefacts, risks overlooking the mechanisms through which technological mediation shapes human interaction, not by forcing and determining social actions but by making some actions more available than others (Hutchby, 2001b). Therefore, it has been suggested (Carr, 2020; Flanagin, 2020; Housley, 2021; Walther, 2009) that research on technologically mediated interaction should focus on the underlying communicative phenomena which span media and the relationship of these phenomena and technological mediation. One way to conceptualise the relationship between communicative phenomena and communication technology is to examine how technologies both limit and enable certain actions through what can be called technological affordances.

1.3 The affordance approach to technologically mediated interaction

In this dissertation, I adopt an affordance approach to the relationship between technology and social action to foreground the ways in which technological artefacts like video mediation equipment can enable and constrain human actions through affordances instead of simply determining them (Hutchby, 2001a, 2001b). The idea of affordances originated in Gibson's (1979) work on the perception of animals in their natural environment (see also Baggs, 2021). Gibson claimed that instead of mere colours, textures and surfaces, animals perceive objects, including the broader surroundings and scenery, based on the possibilities for action they provide or *afford*. For example, a dragonfly might perceive a river through the possibilities it provides for catching prey or reproducing; that is, the affordances it possesses. Furthermore, while affordances are connected to the physical features of objects, they can vary depending on the species that perceives the object. For example, the same river might afford a means of travel to a fish but be an insuperable obstacle for a human. This connectedness to physical features of objects makes affordances non-arbitrary. For example, while a river can offer various different affordances ranging from a place to live to a substance to be consumed, it does not have the affordance of being used as building material (unless it is frozen). Thus, Gibson's approach to affordances emphasises the relational and action-focused nature of the concept: while affordances of material parts of the

environment exist regardless of animals, these affordances actualise in relation to certain species and their actions. For example, "the surface's affordance of walk-on-ability becomes manifest when the animal walks on it" (Gibson, 1982, p. 409).

These principles were later applied to the relationship between human action and technological artefacts by Hutchby (2001a, 2001b). Like objects in the natural environment, technological artefacts also carry functional possibilities and limitations – that is, affordances. Consider showing a physical photograph to another person: in a face-to-face-situation, one can pass the picture to another person, who responsively takes it, leans towards the picture to see it more clearly, points out some detail in the picture and so forth (Aaltonen et al., 2014). Compare this to video-mediated showing: while some of the aforementioned actions are not possible in the video-mediated setting, certain new options emerge, such as taking a screen capture of the picture to inspect it more closely or zooming in to highlight specific parts of the image (Arminen et al., 2016; Rosenbaun & Licoppe, 2019). Thus, as with the affordances of natural objects, technological affordances should not be considered solely as limiting social actions; they are also enablers, depending on which actions individuals produce in relation to them.

Approaching video-mediated interaction from the point of view of the affordances that video mediation technology possesses overcomes the limitations of techno-centric and potentially techno-deterministic models of the relationship between technology and human conduct. While technological affordances limit the scope of some potential actions in video-mediated interaction, they do not dictate what will happen in that context. For example, the affordances of video mediation technology do not determine how a physical examination is conducted in mediated settings; rather, the interactants themselves choose lines of action within the possibilities that the technology affords (Due & Lange, 2020; Pappas & Seale, 2010; Seuren et al., 2020; Stommel, Licoppe & Stommel, 2020, Stommel, van Goor & Stommel, 2020). Furthermore, while affordances are properties of technological artefacts in the sense that they exist regardless of whether somebody acts upon those artefacts, technological affordances only actualise in relation to a human actor or actors[2]. Like different animals, different humans have different goals for their actions, perhaps based on their profession, which shape how

[2] Or, to use an example from a modern automated dairy farm, a cow.

technological affordances support or hinder their central work tasks (Eason & Waterson, 2014). Therefore, technology cannot determine how humans act; rather, people work to fit their actions to the affordances at hand to reach certain (inter)actional goals.

In this dissertation, I use the notion of affordances to study the mundane production of shared understanding – that is, intersubjectivity – and how it is repaired through making one's own perspective accessible to others through verbal explications and physical demonstrations in video-mediated settings. Approaching video mediation through technological affordances and the ways in which interactants fit their practices to those affordances to reproduce shared understandings enables the analysis of the underlying communicative phenomena (Carr, 2020; Flanagin, 2020; Walther, 2009); namely, resolution of overlapping talk, building recognisable lines of action and repairing problems with understanding what single actions are produced to mean, which I discuss in more detail in chapter 2. Before advancing to the central interactional phenomena and the conversation analytic perspective on them, I briefly discuss how ethnomethodology, which serves as the theoretical grounds for conversation analysis and thus this thesis, conceptualises shared understanding as intersubjectivity.

1.4 The ethnomethodological perspective on intersubjectivity as a local achievement

In this dissertation, I adopt the ethnomethodological perspective to study the production, maintenance and repair of intersubjectivity in video-mediated interaction (Garfinkel, 1967; Heritage, 1984; see also Merleau-Ponty, 2013; Schuetz, 1953 for phenomenological accounts, Gillespie & Cornish, 2009; Mead, 1934 for symbolic interactionist views, Iacoboni, 2008 for neuroscientific and Tomasello, 2008, for primatological accounts). Intersubjectivity, a term coined by Husserl, refers to the interchange of thoughts and perceptions between two subjects (Cooper-White, 2014). From the ethnomethodological perspective, intersubjectivity emerges as a side product of people's interactions. When interacting, people draw on practical knowledge to produce actions that are recognisable as having a shared meaning and correspondingly recognise actions of others based on that same contextual practical knowledge (Linell, 2017). Thus, intersubjectivity is a practical achievement of people interacting together:

as "inter-action" rather than "inter-thinking" (Lindström et al., 2021). While some elements of building intersubjectivity, such as language, pre-exist the situations in which they are employed, they still need to be managed locally by interactants (e.g., Simona et al., 2018).

The position that even the pre-existing building blocks of intersubjectivity need to be managed situationally is most clearly exemplified by the concept of the *indexicality* of actions (Garfinkel, 1967). Indexicality refers to the fact that words and verbal expressions possess many potential meanings. Thus, people who are interacting deduce what an expression means in a specific situation based on the direction and phenomonen in which it indexes by drawing upon the contextual knowledge *in situ* (Heritage, 1984). For example, the Finnish word *kuusi* refers to both the number "six" and a "spruce tree"; which of these meanings is correct can be determined only by examining to what it is indexed in a certain context of action (Suoninen, 2001, p. 375).[3] Furthermore, indexicality is a feature of not only words and spoken language but is also invested in any social action. For example, the simple act of pointing can carry many different meanings regarding what is being indicated and, perhaps more importantly, the actions to which pointing contributes in a specific context (Mondada, 2014). This highlights how the ethnomethodological perspective considers language and other forms of human conduct to be contextually inference-rich actions.

The idea of indexicality of actions differentiates ethnomethodological examination of a shared understanding from those of the social cues approach, especially regarding the conception of communication in which these different approaches engage. Compared to the media richness approach to information contingencies, which conceptualises ruptures of intersubjectivity as resulting either from uncertainty caused by missing information or equivocality caused by ambiguous information (Daft & Lengel, 1986), indexicality emphasises that all information is ambiguous and thus requires contextual sense-making and interpretation. Meanwhile, in media naturalness theory it is the misalignment of pre-existing individual schemas, skills and knowledge that are regarded as causing misunderstandings in technology-mediated communication (Kock, 2004), leaving the interaction process unexamined. These differences in analytical focus, that

[3] The idea of indexicality is also at the heart of the following joke: A spouse tells their partner, who happens to be a software engineer, what to get from a convenience store: "Bring two cartons of milk. And if they have eggs, bring *kuusi*". What items and how many does the partner bring?

is, ethnomethodology's action centredness and social cues approaches medium and individual centredness, highlight variations in the conceptions of communication: ethnomethodology views communication as a process of meaning-making, while the social cues approach takes communication as a channel to transmit pre-existing meanings. This difference further emphasises the importance of studying intersubjectivity as local and practical sense-making.

A central feature of this sense-making is that accountability is invested in every action (Garfinkel, 1967, pp. 135–178; Heritage, 1984, pp. 135–178). That is, every action is laden with an implicit intelligibility or explainability that interactants in each situation can reach by examining that action against the context in which it occurred. To reach this intelligibility, interactants ask (usually implicitly and subconsciously) *why that now*? What does that action do in this interactional context? While this accountability becomes apparent at times when an interactant cannot deduce the intelligibility procedurally or simply refuses to do so, as was the case in Garfinkel's (1967) breaching experiments, this methodological work occurs all the time. In interaction, we rely on others to do this kind of interpretative work and are correspondingly expected to do so with regard to others' actions to build a shared understanding of what is going on.

In ethnomethodological conversation analysis, the method I use in this dissertation (see chapters 2 and 4 for more details), the core machinery of intersubjectivity is the social order, which interactants build action by action (Schegloff, 1992). That is, while interactants may have differing perspectives on the world and actions, intersubjectivity is studied by examining how interactants bring forward their understandings in their own actions and how others build their actions on those previous actions and the understandings invested in them (Depperman, 2015; Raymond, 2019). This is referred to as the sequential organisation of actions (Schegloff, 2007): with each subsequent action, interactants display their understanding of the previous action. The interactant who produced this previous action then has, as the third action, an opportunity to confirm or revise the understanding displayed by the producer of the second action. That is, in the second turn interactant B expresses what he or she considers the first-turn action by A to be, and A in the third turn has an opportunity to confirm or repair B's interpretation. This sequential organisation, which is publicly yet often implicitly available to interactants (and analysts) through the multimodal conduct of interactants, serves as a method to display and repair shared understanding of the ongoing action,

moment by moment and action by action (Heritage, 1984, pp. 254–260). Thus, what single actions mean for interactants is a result of these actions' role in larger sequences of actions. This sequential organisation of actions is at the core of ethnomethodological conversation analysis.

These actions linking to other actions to form broader sequences of actions and sequential meaning-making bring forth what Heritage (1984, p. 242) calls the double contextuality of interaction. This means that every action both shapes and is sensitive to its context. Actions are context-sensitive in that when producing actions, interactants take into account what has previously happened – often in the immediately preceding action – and design their actions to fit with those prior actions. At the same time, actions are context renewing since each action modifies the context of future actions. Therefore, building and maintaining intersubjectivity is not only necessary for understanding what others are doing but also for being able to produce relevant actions oneself.

These basic features of building intersubjectivity through interaction can differ between both ends in video-mediated interaction. This is because, as I explain in greater detail in chapter 3 and demonstrate through the results presented in chapter 5, technological mediation detaches the production and reception of actions from each other temporally and spatially. This detachment can hinder the context-sensitive sense-making that is needed to understand how single actions contribute to broader sequences of actions. Before explaining this phenomenon and other relevant findings of conversation analytic research on video-mediated interaction, I briefly discuss conversation analysis and the concepts that are the focus of my analysis.

2 CONVERSATION ANALYSIS AS THE ANALYTIC PERSPECTIVE

2.1 Conversation analysis as a description of members' methods

In this dissertation, I use multimodal conversation analysis as the research method (Heritage, 1984, pp. 232–292; Mondada, 2019; Sidnell & Stivers, 2012; ten Have, 2007). Conversation analysis is an inductive method to study the recurrent patterns of social interaction from interactants' perspective. Drawing from naturally occurring data (as opposed to, say, laboratory experiments or survey questionnaires), conversation analysis aims to describe interactants' perspectives on an ongoing action, how they produce actions that are intelligible to others, and how they make sense of others' actions.

Three central and interrelated analytic principles characterise conversation analysis and differentiate it from other approaches to social interaction. First, it concentrates on interaction as local and collaborative action, analysing single turns and actions in relation to the broader sequences of actions to which they contribute (see section 2.2.2 for further elaboration of sequential analysis), instead of analysing interactants' characteristics or isolated turns of talk and their inherent meanings. Second, conversation analysis focuses on using the emic viewpoint to analyse the ways in which the interactants make action understandable to one another in their publicly accessible actions, compared to various coding schemes (e.g., Brauner et al., 2018, in which the researcher categorises actions from the etic perspective. The interactants' perspectives are obtained through what is called the "next-turn proof procedure" (Edwards, 2004), which means that like the interactants themselves, the analyst can deduce what a previous action was doing by examining how it is treated as the interaction unfolds. Third, conversation analysis ultimately aims at unravelling how this interaction builds social order rather than being its product. By concentrating on locally produced structures of action, analysing that action from the ways in which interactants make action intelligible to one another in their publicly accessible actions, and examining how this interaction builds social order (instead of being its product), the connection between conversation analysis and ethnomethodology becomes explicit.

The term *multimodality* refers to the plurality of ways in which interactants in a given situation employ various ways, means or modes to accomplish certain actions (Mondada, 2018, 2019). This means that when producing recognisable actions, interactants draw on words, prosody, non-lexical vocalisations, gaze direction and body movement and the use of various artefacts and the physical or virtual space in which the interaction takes place not as separate and separatable "languages" but as a united *multimodal gestalts*. Likewise, recipients make sense of the meaning of these actions by drawing upon these multimodal gestalts, not on single modalities. While the importance of multiple modalities was recognised in the foundational work in conversation analysis by Sacks et al. (1974; see also Kendon, 1967), on the use of gaze in interaction)[4], and interest in the body's role in interactions grew in the 1980s with affordable camera technologies (Goodwin, 1981; Goodwin & Goodwin, 1986; Heath, 1986), the embodied turn in ethnomethodological and conversation analytic research in the late 2000s has made it basically impossible to overlook multimodality (Nevile, 2015).

2.2 Key analytic terms in the dissertation

While I use conversation analysis to conduct this research, it should be noted that conversation analysis does not offer a research method in the traditional sense of the word (Heath, 2004; ten Have, 2007). That is, it does not offer a clearly defined series of actions which lead to a scientific outcome. Instead, it offers conceptual and often practical tools for conceptualising and analysing social interaction. For this doctoral study, the concepts most important in my analysis are turn-taking, overlap resolution, sequential organisation and repair, each of which I introduce next.

2.2.1 Turn-taking and overlap resolution

Perhaps the most fundamental structures of interaction relate to managing questions of who should talk at a given moment, when the speaker should change, and how to deal with potential overlapping actions. Interactants usually organise turn-taking so that only

[4] Partially because of that importance, Sacks et al. (1974) initially focused on telephone conversations where the interactants themselves lacked access to bodily features of interaction. Thus, conversation analysis has in a sense always been the study of technologically mediated interaction.

one speaker talks at a time, thus minimising both silence and overlap between turns (Sacks et al., 1974). When interactants reach a transition-relevant place – a point at which a recognisable turn construction unit has come to its end – they draw on a set of methodological rules to determine whether the current speaker selects the next speaker, whether any non-talking participant can self-select, or whether the current speaker could (but does not have to) continue (Sacks et al., 1974).

Overlaps and their management are an integral part of the turn-taking system (Sacks et al., 1974, pp. 706–708). When participants face overlaps, they use an overlap resolution device (Jefferson, 2004a; Schegloff, 2000). The device comprises of resources (hitches and perturbations) that participants employ in specific sets of places in relation to the overlap. Participants solve the overlap beat by beat; that is, syllable by syllable. In doing so, they produce a recognisable "competitive sequential topography" (Schegloff, 2000, p. 11).

While both turn-taking and overlap resolution are multimodal accomplishments (e.g., Mondada & Oloff, 2011; Oloff, 2013), I concentrate in this study on overlapping talk from the perspective of how the perceived location of the overlap in relation to the ongoing stretch of talk shapes overlap resolution (see especially Article 1). Overlaps have different implications for the ongoing action and for overlap resolution, based on their location in relation to the ongoing turn (Drew, 2009; Jefferson, 1984, 1986). They are of one of four different types based on these locations: *transition space onset, last item onset, post-transition onset* and *interjacent overlaps* (Drew, 2009). The most relevant for the present study are the interjacent overlaps, which are turns produced between transition-relevant places far from both the beginning and the end of a turn's construction unit. An interjacent overlap signals that the current speaker ought to give the floor to the producer of the overlapping turn, whether that interruption is hostile interrupting or co-operative (Drew, 2009, pp. 88–91). As discussed in chapters 3 and 5, transmission delay can lead interactants to perceive this location differently, which shapes the building of a shared understanding about who should take the turn that follows the overlap.

2.2.2 Sequence organisation

Sequence organisation refers to how lines of action are put forward through single actions and how these actions form coherent and recognisable sequences of action (Schegloff, 2007). The simplest examples of sequence organisation are *adjacency pairs*, which are sequences comprised of two actions that are situated adjacently, produced by different interactants and organised as the first and second pair parts so that a specific first pair part makes a specific second pair part (or one from a group of potential second pair parts) relevant (Schegloff, 2007, pp. 7–12). For example directing someone to do something and agreeing or disagreeing with that directive comprise an adjacency pair structure (Antaki & Kent, 2012). All actions have some kind of sequential implications for future actions; in adjacency pairs, this sequential implication is especially powerful.

The basic adjacency pair structure can expand in many ways (Schegloff, 2007; Stivers, 2012, pp. 193–200). For the present analysis, the most important type is the extended courses of action. In this kind of expansion, the basic sequence, such as a question–answer adjacency pair, is followed by another, such as providing and receiving advice, which achieves the next phase in a broader course of action (Heritage & Sefi, 1992; Schegloff, 2007). For example, the openings in video-mediated tele-homecare encounters can be viewed as comprising four adjacency pairs that follow one another (see chapter 5 and Article 2). In addition, institutional encounters often have an overall structure with distinct phases (Drew & Heritage, 1992, pp. 43–45), and interactants manage the transitions between those phases multimodally, as through minimal verbal contributions suggesting a sequence closure (Schegloff, 2007) accompanied by moving artefacts relevant to the activity (Nielsen, 2013; Robinson & Stivers, 2001) or through body movement (Broth & Mondada, 2013).

The multimodal approach has specified the basic principles of sequentiality. The core observations are that actions can be carried out through both talk and other means as multimodal gestalts and that talk cannot be assumed to be the most important modality (Goodwin, 2007; Olsher, 2004b, 2004a; Stevanovic & Monzoni, 2016). Another important finding is that different modalities can unfold at different speeds as an interaction progresses; as actions are produced multimodally, these different paces form *multiple temporalities* in interaction (Mondada, 2018). Interactants make sense of the ongoing interaction and adjust their own actions based on these multiple temporalities

to contribute in meaningful ways to the ongoing sequence of action. For example, requesting and offering can be conveyed through both talk and bodily actions (such as looking and pointing at the requested object and handing the object), and these different modalities flow in different temporalities, and thus, the basic ideas of the *firstness* and *secondness* of actions and *simultaneity* and *nextness* need to be considered in new ways (Mondada, 2018). The relationship between sequentiality and multimodality is thus a complex and open question. In video-mediated interactions, this relationship can become even further complicated because, as chapters 3 and 5 demonstrate, video mediation can detach the production and reception of actions from interactants both temporally and spatially.

2.2.3 Repair organisation

The term *repair organisation* refers to methods that interactants use to recognise potential troubles with speaking, hearing and understanding and to repair these troubles (Benjamin & Mazeland, 2012; Fox et al., 2012; Hayashi et al., 2013; Kitzinger, 2012; Schegloff, 1992). I concentrate on repairing troubles with understanding, as they are the focus of this dissertation.

Repair organisation can be divided into two phases: repair initiation and repair proper (Fox et al., 2012). Repair initiation includes both locating the trouble source and recognising the trouble type, such as hearing or understanding. For the present study, the mechanism of recognising troubles of understanding is especially important (Hayashi et al., 2013). Repair proper can be achieved via a number of recurrent repair operations such as repeats, confirmations or corrective replacements of candidate understandings or through further specification of the trouble source. The last of these operations is of particular interest in this dissertation because a common means of repairing understanding in my data is making one's perspective accessible to the other interactant(s) in one way or another.

Both repair initiation and repair proper can be done by the one producing the trouble source (*self*) or any recipient (*other*), thus resulting in four different kinds of initiation–repair relationships (Benjamin & Mazeland, 2012; Schegloff et al., 1977). In face-to-face interaction, there is a preference for self-initiated corrections; even when others initiate

repair, they often refrain from producing the repair proper (Schegloff et al., 1977). However, this dynamic may change in technology-mediated settings, as I detail in the next chapter, where I describe the relevant conversation analytic research on video-mediated interaction.

3 THE CONVERSATION ANALYTIC PERSPECTIVE ON VIDEO-MEDIATED INTERACTION

From the viewpoint of conversation analysis, technological mediation is studied by analysing how it becomes sequentially consequential as an interaction unfolds (Arminen et al., 2016) rather than considering technology as a set of features that directly determine interaction patterns and outcomes. Therefore, to study video mediation, one should examine how the specific material setting, including local ecologies of action that are made available to the participants through video mediation technology, affords the production of meaningful actions and how those actions are interpreted – in short, one should analyse the sequential unfolding of social action in the specific mediated setting (Arminen et al., 2016; Mlynář et al., 2018). Thus, it is crucial to describe how the interactants themselves treat technological mediation as relevant for the production and interpretation of actions: the key finding is not *that* video-mediated and face-to-face interactions differ but *how* – that is, through which processes – they differ[5]. From this viewpoint, two central theoretical concepts have emerged to depict the fundamental ways in which video mediation becomes consequential for interaction: *non-mutual interactional* realities (Ruhleder & Jordan, 2001) and *fractured ecologies* (Heath & Luff, 1992; Luff et al., 2003, 2016).

The notion of non-mutual realities was proposed by Ruhleder and Jordan (2001) to describe the phenomenon in which the timing of the production of an action differs from when it is received in the interaction due to a transmission delay produced at different stages of the information transmission process. Because of this difference, participants interact in non-mutual realities. The authors also point out that it is the relationship between the technological feature of delay and human conduct, especially turn-taking, sequence organisation and repair, that these non-mutual realities are constructed. That is, non-mutual realities are a feature of human–technology interaction, not of technology *per se*. Furthermore, this difference, while being omnipresent in interaction under transmission delay, may remain unnoticed by the participants. Thus, interactants not only interact in non-mutual realities but may also be

[5] This is comparable to the study of institutional interaction (for an overview, see Drew & Heritage, 1992, and Arminen, 2005). While the results from mundane conversations and institutional interactions can be compared to identify differences in these two general settings, the study of institutional interaction still needs to articulate how interactants talk that institution into being.

unaware of this non-mutualness and thus do not necessarily observably work to secure shared understanding.

Interacting in non-mutual realities can lead to both ruptures in the flow of interaction, such as turn-taking, and changes in the meanings of single turns and actions. From the viewpoint of interactional flow, non-mutual realities lead to rephrasing earlier speech due to unintended silences, which anticipates a dispreferred response (for example, declining an invitation), mistimed responses and unintended overlapping talk due to the perceived lack of a response (Ruhleder & Jordan, 2001). Furthermore, changes in the timing of co-participants' turns can lead not only to interactional dysfluencies but also to different interpretations of the meanings of the turns when participants experience the temporality of the interaction differently (Ruhleder & Jordan, 2001). For example, when one interactant receives another's responsive laughter at a different point than when it was produced, what is being laughed at and the meaning of that laughter could differ between interactants (Ruhleder & Jordan, 2001, pp. 131–132)

The second important theoretical discussion on video-mediated interaction revolves around the concept of fractured ecologies (Luff et al., 2003, 2016), which refers to how in video-mediated interaction, bodily conduct is produced in different physical surroundings than those in which the distant participant receives it. In this kind of situation, the finite camera angle limits the possibilities of making sense of this conduct in relation to the action's material surroundings. For example, when someone shifts his or her gaze away from the screen or camera in video-mediated interaction, the distant participant cannot follow that gaze to deduce what is being looked at and what action the gaze shift produces (see Seuren et al., 2020, pp. 3–5, for an example using remote blood oxygen level measuring). Furthermore, participants may be unable to design their bodily conduct to make it understandable to others. Therefore, bodily conduct is fractured both from "the environment in which it is produced and from the environment in which is received" (Luff et al., 2003, p. 55).

This inability to produce bodily actions sensitive to the mediated setting is not a deterministic force; rather, interactants can fit their actions to the technological medium in various ways. A case in point appears in a study on Norwegian sign language users (Hjulstad, 2016), which shows how a signing teacher combines a name sign and pointing to the area on his computer screen where he sees the relevant student to build a *referential map* of the students to allocate turns. Furthermore, since students in a given session remain in the same places on the screen, this pointing always refers to the same student,

and students can recognise the referential map and receive the pointing as meaningful even without the name sign. This establishment of the referential map allows participants to allocate turns and manage participation in a fractured ecology. Once again, we can see how technological mediation, in this case through fractured ecologies, does not determine human conduct but shapes it by both limiting and affording certain kinds of actions.

Next, I present the existing conversation analytic literature on video-mediated interaction that is most relevant to this doctoral research. I use the categorisation employed by Mlynář et al. (2018) in their review of ethnomethodological and conversation analytic research on video-mediated interaction: *openings and closings of video-mediated interactions*, *maintaining a working connection despite omnirelevant potential for technical disruptions*, and *visual contact and fractured ecologies*. While these phenomena have been studied in relation to mobile video-mediated interactions, such as smartphone-based videocalls (e.g., Gan, 2021; Gan et al., 2020) and tele-presence robots (e.g. Due, 2021; Jakonen & Jauni, 2021, 2022; Nielsen, 2020), I largely concentrate on more traditional computer-based video-mediated interaction and video conferencing in meeting rooms, as my data come from those kinds of technological settings.

3.1 Openings and closings of video-mediated interactions

Openings in video-mediated interaction have been shown to develop progressively from multiple appearings through different modalities. For example, in videoconferences, co-located interactants start to orient themselves to the upcoming video-mediated interaction even in the pre-opening phase by adjusting their positions for the cameras and initiating their connections, before the connection opens and participants greet one another (Mondada, 2015; Pappas & Seale, 2009). The central findings regarding openings in video-mediated settings are that they are not as structured as openings on the telephone and that both callers and call-takers can initiate verbal greetings. These findings gave rise to a closer examination of openings as multimodal phenomena. Later, Licoppe (2017b) showed how openings of mundane video-mediated interactions consist of multiple adjacency pairs of appearing and noticing, first by becoming present in the video calling software, then appearing visually when turning on the camera and finally producing a verbal greeting, thus appearing aurally. These multiple appearings also invite multiple noticings that are achieved through both talk and bodily conduct like smiling.

During the opening phase, visibility is treated as important. Participants can refrain from advancing to introduce the topic of the call before the "talking head configuration", in which the head and perhaps some of the upper torso of each interactant are made visible, is established (Licoppe & Morel, 2012). Furthermore, when producing a joint interactional frame in these openings, interactants can recognise and notice "whatever may count as an appearance and thus [make] it interactionally relevant", for example by greeting an interactant who has earlier been visible but occupied with some other activity (Licoppe, 2017b, p. 382). After opening connections, interactants might still orient themselves to the technological mediation, for example by talking about visibility and ensuring that the relevant people are all visible before progressing to the central business at hand (Hansen, 2020; Nielsen, 2020; Pappas & Seale, 2009; Sävenstedt et al., 2005; see also Due, 2021, on visuality and mobility in conversational openings with a tele-presence robot).

In institutional contexts, these basic features of openings can be manipulated to achieve institutional tasks (see Arminen, 2005; Drew & Heritage, 1992) on institutional interaction in general). These might include organising a group of medical professionals so that those highest in the hierarchy appear in the centre of the screen (Mondada, 2015) or, in the context of video-mediated courtroom hearings, establishing who is present and where they appear on the screens, their institutional roles and the institutional and judge-led turn-taking system and the institutional frame of interpretation (Licoppe & Dumoulin, 2010). In medical consultations following a surgery, professionals can use the *how are you?* question to progress from the opening in a way that is both patient-attentive and medically relevant (Stommel et al., 2019).

While openings in video-mediated settings have attracted a reasonable amount of scholarly interest, and there is extensive research on closings of telephone conversations (e.g. Kevoe-Feldman, 2015; Raymond & Zimmerman, 2016; Schegloff & Sacks, 1973), to the best of my knowledge, there is only one published conversation analytic study on closings of video-mediated interactions (see Oittinen, 2022, on closings of audio conferences with document sharing)[6]. That study, which examines the openings and closings of doctors' tele-presence visits in a residential rehabilitation centre, shows how the closings included careful co-ordination of farewells, managing visuality (especially creating visual contact between the tele-presence robot and the individual producing the reciprocal farewells) and mobility, all of which are practices that mirror those in the

[6] Since Due's (2021) study was published after the review by Mlynář et al. (2018), those authors did not include the topic of closings in their categorisation.

openings (Due, 2021). While these results provide important insights for my analysis, the technological affordances of the studied settings differ from my data, especially regarding mobility.

The notions of multiple appearings, the importance of visuality and the potential institutional adaptations of these basic features of conversational openings serve as the starting point for my analysis of the openings in video-mediated homecare. Closings in general have attracted less attention than openings in video-mediated settings, thus making them a relevant object of study in this dissertation.

3.2 Technological disruptions

In video-mediated settings, a central source of repairable troubles are different technological distortions, which can range from the obvious, such as truncated turns or a complete loss of connection (Cipolletta et al., 2018; Rintel, 2015), to freezing images that are initially interpreted as showings (Licoppe, 2017a) to omnipresent non-mutual realities produced by the delay (Hansen & Svennevig, 2021; Ruhleder & Jordan, 2001; Seuren et al., 2021). An interesting feature of repairs in video-mediated interaction are the ways in which technological affordances shape the dynamics of other- and self-initiated repair. While in face-to-face interactions, self-initiated repairs are favoured (Schegloff et al., 1977), in video-mediated setting the one who produces a distorted turn does not have access to how others receive that turn, which shapes the process of initiating repair: either the recipient needs to initiate a repair, or the producer has to deduce that a turn has been distorted based on a missing or inappropriate next turn by a recipient (Rintel, 2013b, 2015). Despite this shift in repair dynamics, recipients may still refrain from initiating repair and prefer that the producer of the trouble source notices it, even when this might lead to repair taking place far from the trouble source (Oittinen, 2020).

One interesting source for disruptions is transmission delay. Despite the omnirelevance of the delay, interactants do not explicitly treat delay as relevant all the time. Whether participants topicalise delay appears to depend on their sequential location (Olbertz-Siitonen, 2015) and their sequential implications (Rintel, 2015; see Oittinen, 2018 on the same phenomenon in audio conferencing augmented with document sharing). In video-mediated interaction, lag time is topicalised in three specific sequential locations: 1) after prolonged multi-word overlaps that lead to silence, 2) in situations where the

participants' turns appear in obscured sequential locations or appear to be absent (e.g., missing answers after a question), and 3) when participants fail to successfully organise turn-taking (Olbertz-Siitonen, 2015; see Rintel, 2015, on turning a technology-generated delay in responsive turns into a joke).

In addition to these specific locations, participants explicate lag and other distortions in situations where it is impossible for them to produce a relevant next turn (Rintel, 2015), as when somebody is asked a question that is insufficiently heard, making it impossible to produce an adequate answer. Furthermore, findings from augmented audio-conferencing show how interactants may prioritise the progressivity of interaction before initiating repair concerning a distant participant's turn: they do not initiate repair if they can produce the sequentially relevant next turn, even though some parts of the previous turn might have gone unheard or been obscured (Oittinen, 2018).

In addition to technological distortions, ruptures of interactional and understanding space can arise from interactions between human conduct and technology, as when audio or screen sharing is turned off (Oittinen, 2020). These ruptures are recognised through adverse or disjointed activities, such as overlapping talk due to a lack of audio connection or references to materials that are not visible to others. These ruptures lead to embodied noticings such as gaze shifts from screens to co-present participants, followed by remedial work; that is, turning on the audio or activating screen sharing.

A special case of technological rupture involves the ways in which transmission delays shape the flow of interaction (Ruhleder & Jordan, 2001). As the overlaps are resolved beat by beat, their resolution is tied to the timing of the turns. As noted above, video mediation can alter the timing of turns and obscure the conditions for turn-taking and overlap resolution (Hansen & Svennevig, 2021; Ruhleder & Jordan, 2001; Seuren et al., 2021). This leads to at least two kinds of problems for turn-taking and overlap resolution: first, interactants perceive silence when talk should occur; second, when facing overlapping talk, interactants have different perceptions of the overlap onset and thus the implications of the overlap for its resolution (Seuren et al., 2021).

When facing overlaps in video-mediated settings, different practices have been recognised. Directing one's gaze towards the screen where the other participant is visible has been recognised a common practice in which interactants engage when facing overlaps in a number of video-mediated institutional contexts (Hansen & Svennevig, 2021; Rusk & Pörn, 2019; Seuren et al., 2021; see Sävendsted et al., 2005, on directing the gaze to nursing home residents when facing lowered participation). In addition,

turns that invite others to contribute, as by asking questions of another participant (Kozar, 2016) or by explicitly allocating a turn to another party (Hansen & Svennevig, 2021; Rusk & Pörn, 2019), have been recognised as ways of dealing with overlaps in video-mediated interactions.

These studies have revealed when distortions are topicalised (Oittinen, 2018; Olbertz-Siitonen, 2015; Rintel, 2015) and how they are treated (Kozar, 2016; Rusk & Pörn, 2019; Seuren et al., 2021). However, many of these studies (e.g. Oittinen, 2020; Rintel, 2015) examine reasonably distinguishable technical distortions, such as freezing images, and the treatment of those distortions rather than the potential troubles they pose for intersubjectivity. In the results section of this integrative chapter, I concentrate on more discrete ruptures of intersubjectivity.

3.3 Visual connection and fractured ecologies

A central feature of video-mediated interaction is the visual connection, which is both limited in comparison to co-located interaction and abundant compared to telephone conversations. Conversation analytic research on visual connections has focused on the role of visibility as a means of attaining and maintaining participation (Hansen, 2020; Licoppe & Morel, 2012; Pappas et al., 2019; Sävenstedt et al., 2005; Seuren et al., 2020; Shaw et al., 2020; Stommel & Stommel, 2021) and making parts of the interactants' local ecology visible to others through showings (Due & Lange, 2020; Licoppe, 2017a; Licoppe & Morel, 2012; Licoppe & Tuncer, 2019; Pappas & Seale, 2010; Seuren et al., 2020; Stommel, Licoppe & Stommel, 2020; Stommel, van Goor & Stommel, 2020). In general, what is shown on the screen in video-mediated interactions is treated as relevant; by the same token, anything that is relevant should be made visible on the screen. This relates to relevant participants in interaction (Hansen, 2020; Licoppe & Morel, 2012; Pappas & Seale, 2010; Sävenstedt et al., 2005; Stommel & Stommel, 2021) and to physical objects that range from body parts (Due & Lange, 2020; Seuren et al., 2020; Stommel, Licoppe & Stommel, 2020; Stommel, van Goor & Stommel, 2020) to artefacts and physical surroundings more generally (Licoppe, 2017a; Licoppe & Tuncer, 2019). These studies highlight how despite the lack of physical co-presence, interactants can use the affordances of video mediation technology to build contextually intelligible actions in video-mediated settings.

As noted above, participants are expected to make relevant things visible on the screen and to treat what is visible as relevant. Participants orient themselves to the maxim of putting the (whole) face of the current speaker (or listener) on screen; that is, they produce the talking head configuration throughout video-mediated interactions, not just in the openings discussed above (Licoppe & Morel, 2012). Participants also orient themselves to the video image as showing something gazeworthy (Licoppe, 2017a; Licoppe & Tuncer, 2019; Stommel & Stommel, 2021), which becomes apparent, for example, when objects that accidentally become visible are treated as important for the ongoing interaction (Licoppe, 2017a, pp. 65–67).

While it is reasonably straightforward to achieve the talking heads configuration in dyadic interactions, this can become more complicated in hybrid multiparty interactions, with some interactants co-present and others involved via video mediation. This is due to the limited and asymmetric visual access that interactants have to each other. This shapes the management of participation. For example, wider and narrower video shots can be produced to create different participation frameworks (Goffman, 1981; Goodwin & Goodwin, 2004; Hutchby, 2014) and to make certain participatory roles salient in multiparty video conferencing (Licoppe, 2015; Licoppe & Veyrier, 2017). Studies on video-mediated interpreting in multiparty hybrid interactions have shown how limited visual access can impede making sense of talk and bodily actions that are used to both offer and take turns, which can further ambiguate the kind of participation that is expected from different participants and when these contributions should be made (Hansen, 2020). Interestingly for the focus of the present study, cases of explicating what kind of participation is expected from the distant participant have been recognised as a way to deal with this ambiguity (Hansen, 2020, pp. 14–16; Hansen & Svennevig, 2021, pp. 154–155). The same type of problems with defining what kind of participation is relevant from whom has also been recognised in hybrid tele-consultations (Pappas & Seale, 2009; Stommel & Stommel, 2021).

One part of managing visibility in video-mediated interaction is making artefacts visible to other interactants by showing, either by moving the object into the camera frame (Due & Lange, 2020; Licoppe, 2017a; Licoppe & Tuncer, 2019; Seuren, 2020) or by moving the camera to show the object (Licoppe, 2015; Pappas & Seale, 2010; Stommel, van Goor & Stommel, 2020). In general, showings can be differentiated into gestural showings, which are recognisably produced as a contribution to some ongoing or projected stretch of talk, and showing sequences, where the showing itself becomes the main action or "topic" of the sequence (Licoppe, 2017a). For the present study, showing

gestures are the more important mode of showing, as they are used in my data for physical demonstrations to make one's perspective salient to others when facing ruptures of intersubjectivity. Gestural showings can further be used as co-expressive gestures, which emerge together with accompanying verbal conduct, and as communicative moves, in which the showing is a standalone action produced in silence (Licoppe, 2017a, pp. 67–71).

Showings may include interactional work in the form of both getting the recipient to properly see the showable object by projecting or prefacing it (Licoppe, 2017a; Licoppe & Tuncer, 2019) or by highlighting its relevant aspects (Due & Lange, 2020; Licoppe, 2017a; Pappas & Seale, 2010) and getting the shower to produce an appropriate showing (Seuren et al., 2020; Stommel, Licoppe & Stommel 2020; Stommel, van Goor & Stommel 2020). Co-expressive showings are projected in the talk they accompany, while standalone communicative showings can be projected as parts of adjacency pair structures, making them understandable (Licoppe, 2017a; see Licoppe & Tuncer, 2019, on prefacing showing sequences). In addition to making the showable visually available, showings may include highlighting (Goodwin, 1994) the relevant parts of the showable object, for example by talking about or pointing at some of its details (Due & Lange, 2020; Licoppe, 2017a) and by providing relevant information about the shown object, such as describing the softness of body parts shown in physical examinations (Pappas & Seale, 2010).

The *activity* of showing – and not just receiving that showing – is also interactionally achieved. In the case of physical examinations in a medical context, adequate showings can be difficult to achieve even when a medical professional directs the showing (Stommel, Licoppe & Stommel, 2020), and the interactants can replace the showing with a talk-based physical assessment of the patient's body if a successful showing is still not achieved (Stommel, van Goor & Stommel, 2020). Furthermore, when interactants produce showings by moving the camera, they may no longer be able to monitor the "vanity screen" (where their own video image is visible to themselves), as often occurs when mobile devices are used as the communication technology (Seuren et al., 2020). In these situations, interactants need to coordinate the showing by providing feedback on what the recipient sees moment by moment. This further highlights how showing objects is an interactional achievement in video-mediated contexts.

Fractured ecologies are a special case of the limited visually shared space that video mediation creates. Fractured ecologies in particular shape the possibilities of interacting with activity-relevant artefacts. As noted above, fractured ecologies detach a bodily action from "the environment in which it is produced and from the environment in which it is received" (Luff et al., 2003, p. 55). Thus, when physical access to activity-relevant artefacts is limited, this shapes the sequential organisation and sense-making around these objects. This has been studied especially in relation to locating and using activity-relevant artefacts. One recognised practice is to decompose the instruction into manageable steps accompanied by mimicable body actions, such as showing someone else which direction to look by turning one's own head (Due et al., 2019). Mimicking and describing items can also be used to help interactants find relevant objects in the distant ecology (Seuren et al., 2020).

These studies show that like face-to-face interactions, participants in video-mediated encounters use visibility to both produce and interpret recognisable forms of participation (Goodwin & Goodwin, 2004). However, limited visibility creates a specific mediated material setting of action into which interactants must fit their actions. Maintaining relevant participation frames may become especially important in multiparty activities around physical objects (Nevile et al., 2014). In this dissertation, this object-oriented interaction and video mediation's effect on it are most prominent during technological problem-solving in hybrid tele-consultations (Article 4) and in distance medicine-taking, with limited visual access to the medicines, as is a common experience in tele-homecare (Article 3).

4 THE DATA, ANALYTICAL PROCESS AND RESEARCH ETHICS

4.1 Data

In this dissertation, I use data from three datasets that are summarised in Table 1. All data are in Finnish and were collected in Finland.

Table 1. Overview of the data used

Type of encounter	Video-mediated group counselling	Video-mediated tele-homecare	Video-mediated tele-consultations
Year of data collection	2006	2018–2019	2019
Amount of data	Three sessions, ~268 minutes of data	14 tele-visits, ~107 minutes of data	5 tele-consultations, ~255 mins of data.
Number of participants	Three groups with a total of 19 participants, along with two nutritionists	Four home-dwelling older adults and three nurses	Five patients, one general practitioner and one specialist
Technological Setting	Hybrid: all group members were co-located and connected to the nutritionist via video link	Traditional: encounters between individual clients and nurses, with no co-presence between participants	Hybrid: patient and general practitioner co-present and connected to the specialist via video link
Analysed interactional phenomena	Overlap resolution (Article 1)	Management of openings and closings (Article 2), managing medicine-taking (Article 3)	Technical problem-solving in a prolonged opening sequence (Article 4)

The first dataset consists of video recordings of video-mediated group counselling sessions for people at high risk for type 2 diabetes. The data were originally collected by the Finnish Institute of Occupational Health (FIOH) for an independent project (Laitinen et al., 2010; Nevanperä et al., 2015), and I gained the opportunity to access them while working on the Promo@Work – Evidence-Based Health Promotion at Work research project. The participating nutritionists were part of the team planning and executing the intervention, and the group members were recruited as part of their regular appointments with nurses. All participants gave informed consent before data collection. The hospital district of Helsinki and Uusimaa approved the collection and use of the data (document number 50/E0/2007).

While the entire corpus contained video recordings of group health counselling in both face-to-face and video-mediated settings, I only used data from video-mediated counselling. Furthermore, from the larger corpus of video-mediated group meetings, I came to work with three encounters that were recorded from both perspectives of the action: the group and the nutritionist. The availability of these two separate but simultaneous perspectives created a valuable opportunity to explore how transmission delay can shape interaction.

I collected the second and third datasets during the Healthcare Workers in the Eye of the Digital Turbulence: New Forms of Cooperation and Customer Orientation research project, a collaboration between Tampere University and FIOH that was funded by The Finnish Work Environment Fund (see Koivisto et al., 2020, for a project summary in Finnish). The Ethics Committee of the Tampere Region provided a positive statement for collecting and using both these datasets for the purposes of research on video-mediated interaction (document number 49/2017).

The second dataset consists of video-mediated tele-homecare encounters between home-dwelling older adults and their nurses. These visits were organised as part of a service pilot in which one daily homecare visit was replaced by a videocall. The clients were given a tablet computer with a simple video calling program that allowed them to answer calls from their nurse (Figure 1). Before data collection began, the project obtained a research permit from the participating municipality. I then visited the nurses' weekly team meeting to recruit individual nurses. The municipality first selected potential clients for the service pilot, and a research assistant and I visited those people to provide information about the project and to interview potential participants about their expectations about the service. As all the potential participating clients had some mild memory deficits, we paid special attention to recruiting them, as is explained in greater detail below. Both the nurses and the clients gave informed consent before data collection began.

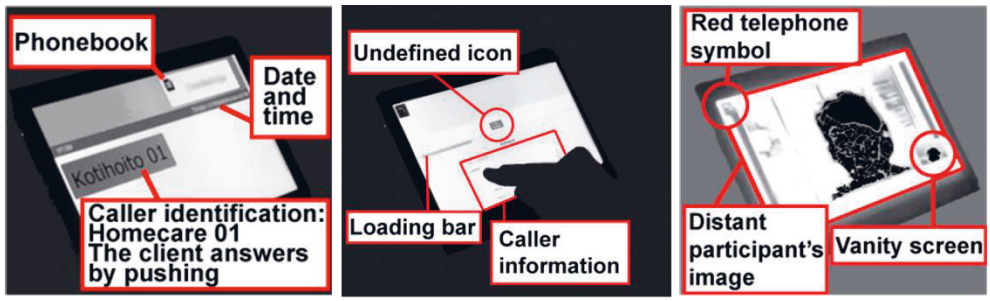


Figure 1. The client user interface in tele-homecare, shown when the nurse is calling.

In total, I gathered 14 tele-homecare encounters. Initially the plan was to collect all the data with two simultaneous perspectives in the action, was the case with the as counselling data. However, this proved impossible to coordinate with the nurses' work schedules. Thus, I recorded two encounters from two clients' homes (one recording in each case) and the other 12 from the office from which the nurses placed the calls (Figure 2).

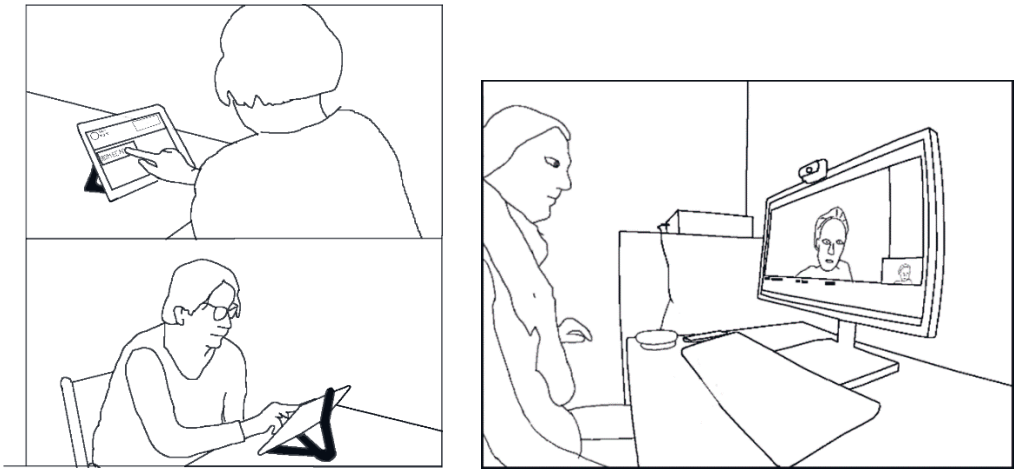


Figure 2. Two perspectives on homecare: a client answering a call at home (left), and a nurse initiating a call from the nurses' office (right).

The third dataset consists of five video-mediated tele-consultations where patients and a general practitioner gathered in the general practitioner's office to connect with a specialist via video link. The details of these encounters were collected as part of a private clinic pilot that tested potential forms of video-mediated services. The participating professionals were recruited during the planning of the Healthcare

Workers in the Eye of the Digital Turbulence project. The general practitioner selected potential patients for the tele-consultation based on their need for consultation from the specialist's field of expertise, told them about the research project and gave them the information leaflet when scheduling a tele-consultation for the patient. I met the patients before their scheduled tele-consultations to see whether they had any questions and were willing to participate. All participants gave informed consent before the data collection.

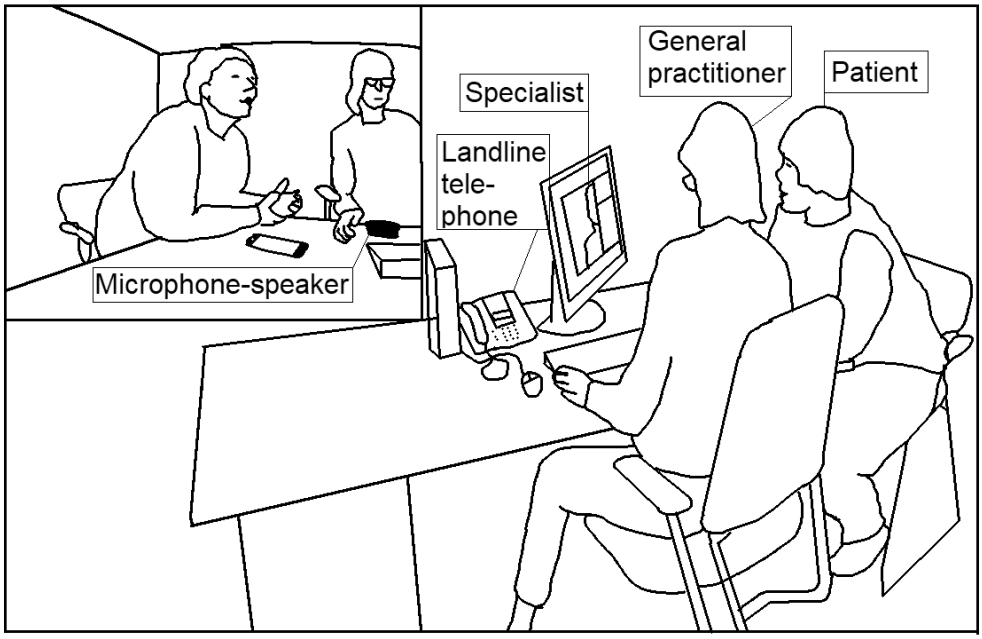


Figure 3. The setting in the tele-consultation data, with two camera angles.

All the data are from health and social care settings; thus, interactants take into account the institutional relevancies that these services bring with them. From a conversation analytic perspective on institutional interaction, the important question is how these institutions are talked into being (Heritage, 1984). Through their lexical choices, turn-design features, orientation to institutionally relevant sequences of actions, such as the overall structure of the specific institutional encounter, and asymmetries regarding knowledge and power, interactants produce the interaction as institutional (Drew & Heritage, 1992). That is, the institutionality of interaction is conceptualised through the ways in which the interactants take into account the institution in their conduct and correspondingly produce that institution as a social fact in that conduct, instead of considering institutionality as an irresistible force that determines their conduct.

Thus, when producing their actions, the interactants in my data take these institutional relevancies into account in one way or another. These relevancies are discussed in the individual articles. In this integrative chapter I have, however, concentrate less on the specific institutional relevancies and aim to describe more general phenomena that span different data. The plurality of the institutional contexts in my data and the potential applicability of my results for institutional practices are discussed in chapter 6.

4.2 Analytical process and research questions

In principle, the analytical processes for all dataset resembled one another. I started by transcribing the data in accordance with conventions familiar from Jefferson (2004c) and Mondada (2001). During the transcription process, I began noticing phenomena that could prove interesting for the analysis. In the counselling data, this was resolution of overlapping talk, which later evolved into the topic of Article 1. With the homecare material, I first became interested in the overall structure of these encounters and then in how the transitions from one phase to another were managed. This resulted in the analysis of openings and closings (Article 2) and medicine-taking (Article 3). In the case of the tele-consultations, the problematic and prolonged opening of the very first consultation caught my attention even at the research site, so I selected it as the topic of Article 4. After defining the main topics for the articles, I analysed them by applying the basic principles provided by ten Have (2007):

1) Select a sequence.

2) Characterise the actions in the sequence.

3) Consider how the speakers' packaging of actions provides for certain understandings of the actions performed and the matters talked about.

4) Consider how the timing and taking of turns provide for certain understandings of the actions and the matters talked about.

5) Consider how the ways the actions were accomplished implicate certain identities, roles and/or relationships for the interactants.

Finally, my co-authors and I connected these analyses to the broader theoretical concepts that best fit the analyses in relation to their target journals: overlapping talk

with client participation (Article 1), organisation of medicine-taking with client autonomy (Article 3) and practices of problem solving with distributed cognition (Article 4). For Article 2, we concentrated on the sequential organisation of openings and closings as standalone phenomena, without this kind of broader theoretical connection outside the conversation analysis community.

In order to form the topic and the research questions for this integrative chapter of my PhD thesis, I approached the individual research articles as if I were conducting a literature review. I read them through and collected the main findings of each article to identify common themes and topics. While the articles covered a wide range of topics (the relationship between turn-taking and client participation, sequential organisation of openings and closings of video-mediated tele-homecare encounters, client autonomy when directing medicine taking in tele-homecare and problem solving in a hybrid-video-mediated setting), they all dealt with a shared understanding about how to proceed with interaction. Thus, I selected intersubjectivity as the overarching theoretical concept under which I could arrange my findings. It is noteworthy that while the topic of intersubjectivity is central in ethnomethodology and conversation analysis, it is only explicitly discussed in Article 2 and only emerged as the topic of this PhD study after the individual articles had been written. I then reread my notes about the articles to specify the aspects of intersubjectivity on which these articles touch. Through this process, I formulated the following research questions for this PhD thesis: 1) How do interactants display recognition of technology-generated ruptures of intersubjectivity in video-mediated interaction? 2) How do interactants repair these ruptures in intersubjectivity? 3) How does technical mediation become consequential for these practices as interaction unfolds?

The data I used were diverse in two ways. First, they varied in relation to technological settings, which were comprised of both "traditional" dyadic videoconferencing and hybrid encounters in which some participants were co-present and others used video mediation technology. Second, the data varied in relation to perspectives on the action; as the analyst, I had varied levels of access to different interactants' perspectives: I could see both perspectives simultaneously in the counselling data, while for the homecare data, some encounters were accessible from the client perspective and others from the nurse perspective. Finally, the tele-consultations were recorded from only one perspective. This diversity of data enabled identifying subtleties in the difference between perspectives, the importance of analysing each perspective in its own right and acknowledging the local ecology of action by comparing similar actions in both

mediated and non-mediated settings, as opposed to explaining observations in the mediated setting merely by contrasting them with earlier research findings. I should note that I started using these three datasets in the order I introduced them above, and it is in retrospect that it became apparent how much their diversity shaped the way I conducted the analyses.

Concurrently working with the counselling data and its two perspectives on action and learning more about Ruhleder and Jordan's (2001) research led me to realise how changes in perspective can be extremely subtle but still be deeply significant for the ongoing action. For example, technological ruptures do not need to be prominent in order to shape the possibilities for upcoming actions and situational meaning-making processes: that is, not only frozen images and fuzzy audio but also less obvious phenomena such as transmission delays need to be taken into account. Furthermore, working with these data showed how the analyst cannot simply assume that data collected from one perspective can cover all the relevant aspects that appear for interactants in the other perspective. This was most salient when I analysed the meanings of different overlaps: while the overlap may appear to one party as "interrupting", it cannot be claimed that the other party is deliberately interrupting – or even can recognise that they have interrupted – if there are no data from the perspective in which the action is produced. Thus, I needed to limit my analytical interpretations of the distant participant's perspective and concentrate instead on how the participants interpreted one another's conduct from their own perspectives. Furthermore, working with these data enabled me to analyse the significant impacts that small changes in overlap timing can have on overlap resolution, thus providing the topic for Article 1.

While the original plan to gather homecare data from two perspectives on the action did not materialise, having worked with the counselling data provided me with a cautionary approach to the analysis described above. This had two important implications for the results. First, both the client and nurse perspectives deserved to be analysed in their own right. Observations made based on one perspective should not be used as a benchmark for the analyses of the other. Instead, the potential for non-mutual realities needed to be assumed at all times, and the different perspectives on interactions needed to be analysed independently. While I did not have access to both perspectives of action simultaneously in an individual encounter, having access to different perspectives in different encounters and recognising the importance of the non-mutuality of perspectives afforded, for example, the analysis of the dual functions of actions in the opening sequences of tele-homecare (Article 2). Second, having these

kinds of data pushed me to search for phenomena and activities on which I would have material from both perspectives, resulting in the analysis of medicine-taking (Article 3). This offered me a clear opportunity to analyse how fractured ecologies, another key concept of video-mediated interaction, shaped the production of intersubjectivity.

Working with the somewhat limited – in the sense that only one perspective on the action was recorded – tele-consultation data benefitted from the analytical insights I had gained while working with the other, more complex data. Thus, while the data themselves came only from one perspective, I could account for notions of both non-mutual realities and fractured ecologies during the analysis. Furthermore, when the data were from a hybrid setting, with one participant distant and the others co-present, the analysis of similar practices was enabled, especially directing others' attention and actions, both when respondents were co-present and when they were distant (Article 4). This provided access to the interactional processes through which the participants made technical mediation interactionally consequential moment by moment.

4.3 Research ethics

During the course of my PhD studies, I have complied with the general research ethical principles of the Finnish Advisory Board on Research Integrity (2012) to protect the privacy, integrity and autonomy of research participants. I protected participant privacy by storing and using the data on password-protected hard drives and employing anonymised data extracts in public presentations (conference presentations, teaching and journal articles); for conference presentations and teaching, I changed the pitch of the sound, removed names from the soundtrack and used video-filters to produce line drawings of the participants, while in the research articles I used anonymised transcripts and line drawings. In data sessions and other occasions where non-anonymised data were used among co-researchers, I collected a pledge of silence from my colleagues. Taking part in the research caused no direct harm to the participants. I told all those who were recruited that their decision to participate or not participate would not affect their status as an employee or the services they received and that the research project was unattached to the service it studied. The autonomy of the research participants was respected: all potential participants were provided with information about the research and their rights as participants before making the decision to take part, including the

possibility of withdrawing from the study at any point. Regarding the data I collected, this information was given both orally and in writing.

While collecting and using data from counselling and tele-consultation contexts did not raise any major ethical dilemmas, the homecare data did raise questions about participant autonomy. Since all the home-dwelling adults had some level of memory deficit, it could be argued that they should not have been recruited for the study in the first place. Recruiting people whose cognitive capabilities deviate from what we consider "normal" raises ethical considerations, especially as to informed consent (Dewing, 2008; McDonald & Kidney, 2012). However, it has also been pointed out that people with memory deficits should have the right to make decisions about participating in research and be regarded as incapable of making these decisions only if there are reasonable doubts about their capability (Mäki-Petäjä-Leinonen, 2006; Wilkinson, 2001).

Thus, in order to ensure that the home-dwelling older adults had the autonomy to participate on one hand and did not have to participate in something they did not understand or want to be involved in on the other, three kinds of measures were implemented. First, the municipality excluded older adults with severe memory deficits from the service pilot, so only people without memory deficit or with mild memory deficit were even eligible to be recruited for the study. Second, when recruiting the participants, I provided information on the research both orally and in writing and set aside a reasonable amount of time for potential participants to ask questions about the study. While drafting the information leaflet and consent form, I ensured that they would be understandable and used plain language. I also explicitly asked the participants whether they understood the content and what was entailed by participation. As some of those who were recruited declined to participate in the entire study and others agreed only to part of it (for example, being interviewed but not video-recorded), I interpreted this as a sign that this was an understandable procedure for the target group in general. Third, the research assistant and I interviewed each client before the video data collection began and excluded people with obvious problems with understanding or remembering the research information (e.g., those who forgot who we were during the interview), despite their initial willingness to participate. Thus, the two more general means of protecting people's autonomy were accompanied by this more individualised measure.

For me, this dilemma was a prime example of how ethical research is far more than a set of guidelines and principles but an often-complex process with a series of practical actions and choices. Furthermore, it showed how not only researchers but also other actors, such as the municipality that conducted the initial screening of participant suitability, and the nurses, who occasionally reminded clients about the data collection, are part of configuring ethically sound research.

I now move to the core part of this integrative chapter: my findings.

5 RESULTS

Based on the analyses of the data, I argue that like face-to-face interaction, people recognise the ruptures of intersubjectivity against the sequential relevance of actions: if interactant A does not produce a sequentially relevant next action or produces a turn that from interactant B's perspective is not relevant, interactant B may treat this as signalling a rupture of intersubjectivity. When resolving ruptures of intersubjectivity, individuals need to make their perspective on the ongoing action available to others and have others' perspectives available to them. The interactants achieve this by two intertwined practices: verbal explication and physical demonstration. By making their perspectives available to one another through these practices, interactants work around the subtle temporal and spatial changes that non-mutual realities and fractured ecologies produce. However, as non-mutual interactional realities and fractured ecologies detach the production and reception of actions from one another, video mediation shapes the conditions of both interpreting the sequential relevance of actions and employing the aforementioned practices of explication and demonstration. This relates to the temporal and spatial organisation of interaction. On one hand, the lag time changes the timing of actions relative to one another from different mediated perspectives; on the other, fractured ecologies detach the production and reception of bodily actions from one another. Thus, when employing these practices, people orient themselves to the technological mediation as relevant by fitting their conduct to the media available to other participants. However, because of the non-mutual and fractured nature of technology mediated interaction, this orientation might not be observable to the other party, at least in the same way as it is for the producer. I now briefly summarise the results of the individual articles.

In Article 1, my colleagues and I analyse how transmission delays can shape overlap resolution, the possibilities of client participation and related positive social processes in video-mediated group counselling. We concentrate on overlaps in action sequences that started with the nutritionists' questions, with a special focus on the nurses' overlap resolution practices. The delay changes the timing of the overlapping turns and of the silences that the interactants hold when facing overlaps at each end of the mediated

counselling[7]; that is, from the group and nutritionist perspectives. These changes produce non-mutual interactional realities (Ruhleder & Jordan, 2001) where the interactional implications of the locations of overlaps and different overlap resolution practices, especially implicitly offering the turn to another by pausing, differ in two local ecologies of action.

In the article, we show how this can lead to three kinds of ruptures in the flow of interaction and a shared understanding of who should talk. First, due to the delay, the interactants have different orientations to the progression of interaction; namely, whether a group member- or nutritionist-initiated action should follow, which could lead to competition over the turn. Second, as the delay alters the timing of the nutritionist's turn-offering silences, which is a common way to implicitly offer the turn to a group member, the group members often interpret implicit turn offerings as within-turn pauses. Combined with the changes in the onset of the overlapping talk and the onset's implications for turn-taking, this makes it difficult to interpret who should take the turn after the overlap; thus, speaker change does not occur. This may pose obstacles to client participation and consequently to positive social processes like providing peer support. Furthermore, due to their non-mutual realities, the interactants remain unaware of the technological root of the misunderstanding. Third, when the nutritionists recognise that a group member has not taken the turn after the turn is implicitly offered, they may use explicit turn offerings – either minimal offerings such as "yeah" with a rising intonation or more elaborate ones such as "go ahead" – to secure the speaker change and client participation. By doing so, they verbally explicate who should talk after the overlap, thus making their perspective salient to the others.

In Article 2, we analyse openings and closings of video-mediated tele-homecare encounters. The openings are organised around a structure consisting of four adjacency pairs: summons–answer, appearing–noticing, greeting–greeting and *how are you?* question–answer. These turns can serve dual functions in interaction, as the first greeting by the nurse who initiates the call operates both as the greeting (the first pair part of the greeting–greeting adjacency pair) and the verbal noticing (the second pair

[7] In the language of conversation analysis, silences within a turn construction unit are called *pauses* and those between turn construction units *gaps*. However, I prefer the term *silence* since the interactants in the data do not share the same understanding of that silence: to one it is a pause within a turn and to others potentially a gap between turns.

part in the appearing–noticing adjacency pair). This dual function is not, however, present from the client's perspective. This shows how the sequential implications of turns differ slightly from these two perspectives.

In the opening sequence, the participants treat the visibility of one another as a relevant part of the appearing–noticing adjacency pair. This is most clear when one (or both) of the participants is unable to appear properly during the opening sequence. In these kinds of situations, participants postpone moving on to asking *how are you?* until a proper visual appearing is achieved. Furthermore, as there can be multiple appearings (aural and visual, the caller's and the call-takers'), and the interactants do not have access to the distant participant's viewpoint, they may have different understandings of when both interactants have adequately appeared. This kind of situation may involve talk about what and when each interactant hears and sees. Through this talk, the interactants verbally explicate their readiness to proceed with the opening.

The closings are often primed with the nurses' positive evaluations, the participants talking about the future arrangement of the care, preclosing tokens and the client expressing gratitude or appreciation for the service. Through these priming practices, the participants progress to the verbal terminal exchange accompanied by mutual visual disengagement; that is, the nurse closing down the connection and the client withdrawing from the screen. Like openings, closings are achieved multimodally. This is most apparent in the ways in which the participants coordinate the terminal exchange and disengage together. While it would be possible for the nurse to end the connection immediately after the terminal exchange, the nurses routinely postpone disconnecting until the client has begun to disengage. Thus, the client's demonstration of readiness to close the connection by physically disengaging is central to the shared understanding that occurs in the closings. Interestingly, the nurses do not necessarily treat the clients' inability to produce some of these pre-closing practices as accountable if they otherwise appear to be aligning with closing and do not produce any misaligning actions.

The topic of Article 3 is the interactional management of medicine-taking in video-mediated homecare, where the nurse and the client have asymmetric visual and physical access to the client's home and thus the medicine. Managing cooperation with physical artefacts in this kind of fractured setting (Luff et al., 2003, 2016), the participants make adjustments in relation to both interactional practices and the arrangement of the

physical objects in the home space. The client needs to simultaneously engage in the interaction with the nurse and to locate the medicine in her physical surroundings. If the video mediation equipment and the medicine are situated far from each other in the home, completing these two tasks simultaneously becomes complicated. This is especially so when the nurse uses straightforward directing into medicine-taking, which can lead to ruptures in the progression of interaction when the client is forced to ask for help in locating the medicine.

Solving this interactional dilemma demands changes in both physical surroundings and interactional practices. One part of the solution is to move the video mediation equipment closer to the medicine, thus enabling the client to simultaneously engage in interaction with the nurse and take the medicine. The second part is that the nurse uses a stepwise progression to medicine-taking, first topicalising the medicine, then waiting for the client's bodily orientation and verbal display of position with regard to the medicine; only after establishing this partially shared orientation does the nurse direct the client to take the medicine. Through this practice, the nurse is able to invite the client to verbally explicate whether they see the medicine and can thus align with an upcoming directive to take the medicine. This stepwise entry into medicine-taking enables participants to build a shared understanding of the medicine in a situation in which they have asymmetric visual access to the medicine. These different arrangements of the home space, technology, other activity-relevant artefacts and interactional practices lay different grounds for client autonomy to emerge: in the case where the medicine and video mediation equipment were far from each other and medicine-taking was introduced promptly forced the client to ask for help, while situating the technology and medicine closer to each other and using a stepwise approach to medicine-taking afforded the client more independence of action.

Article 4 is an examination of problem solving-activity, where the participants in a tele-consultation – a co-present patient and general practitioner and a distant specialist – find themselves in a situation where the specialist cannot hear the others' voices. As the problem-solving progresses, the interactants need to coordinate one another's attention and actions, knowledge of the interactants and technological artefacts (the video mediation equipment and a landline telephone) and their affordances to form a distributed cognitive system (Hollan et al., 2000). Video mediation, by producing fractured ecologies, limits physical access to distant participants' physical ecology, which

complicates this coordination. This is overcome by fitting interactional practices into the communicative medium available at a given moment, for example by producing polar questions that can be answered by gestures (such as a thumbs-up signal or nodding). This enables the participants to invite others to physically demonstrate the relevant parts of their perspective. By gaining access to others' perspectives and providing access to their own perspectives through these demonstrations, the interactants manage to both initiate testing sequences of different options and negotiate whether certain testing approaches would be relevant. By producing interactional practices fitted to the available media, the interactants manage to coordinate human actions and technological affordances to form a distributed cognitive system in which the solution – using a landline telephone to transmit the audio and the video mediation equipment to transmit the image – finally emerges. Fractured ecologies and non-mutual access to both ecologies hinder this coordination. Thus, throughout the problem-solving process, the participants fit their conduct to the media which are available to different participants, taking into consideration one another's epistemic statuses and stances.

While not the explicit topic of any of these articles, problems of intersubjectivity were central to all of them. Based on these analyses, I now discuss the central issue investigated in this dissertation: how participants recognise and resolve ruptures of intersubjectivity. These concrete challenges raise the more theoretical question of how technical mediation becomes interactionally consequential in this process.

5.1 Recognising problems with intersubjectivity

The first main result of my thesis is that people display recognition of ruptures of intersubjectivity against the normative expectations of the sequential relevance of actions: if interactant A does not produce a sequentially relevant next action or produces a turn that from interactant B's perspective is not relevant, interactant B may treat this as indicating a rupture of intersubjectivity. This is the case when not taking a turn that is implicitly being offered (Article 1, Extract 3), appearing inadequately in the opening phase of a tele-homecare consultation (Article 2, Extract 2) and not appearing to take the medicine when directed to do so (Article 3, Extract 1). We excluded the recognition

of rupture from the analysis in Article 4, but the same dynamics – in this case, the inability to produce reciprocal greetings in the opening – are central to that case as well.

Whether or not something can be considered sequentially relevant is interpreted against the broader lines of action to which those missing or irrelevant actions contribute. While interpreting actions as sequentially relevant is closely tied to the basic sequential organisation, for example Initiation–Response–Feedback sequence (Article 1) and different adjacency pair structures such as appearing–noticing (Article 2) and directive–compliance/resistance (Article 3), this interpreting is carried out situationally in relation to broader lines of action and the interactional projects they form. This is exemplified when the tele-homecare nurse continues progressing towards closing despite the lack of reciprocal farewell, instead of treating the client's missing response to her farewell as an inability to produce a sequentially relevant next action and thus a sign of the rupture of intersubjectivity (see Article 2, Extract 5). In the case mentioned, the nurse could treat the missing turn as not accountable because the interactants have already established that no unmentioned mentionables will rise and because the client appears to bodily align with progressing towards the closing. Thus, when the interactants assess this interpreting as sequentially relevant, they can orient to progressivity as more relevant than sequential accountability: as long as the interaction can progress, potential missing turns are tolerated.

This orientation to progressivity can lead interactants to rely on a "non-extraordinary formulation of the event" (Jefferson, 2004b, p. 136): as long as actions can be interpreted as possibly normal and sequentially relevant, the interactants seem to do so and not treat them as potential signs of technology-related ruptures of intersubjectivity (Sacks, 1984). Thus, the technical root cause of a rupture of shared understanding may remain unidentified. Thus, while the analyst with a "god's eye perspective" can observe the emergence of non-mutual realities, interactants themselves do not perceive these ruptures as caused by the technology but as ingrained in the basic dynamics of social interaction; see the Discussion section for methodological remarks on using data from two perspectives. This is most clearly exemplified by delay-caused overlaps and different understandings about overlap resolution practices (Article 1). As any type of overlap onset can be interpreted as interactionally relevant, it is hard for interactants to recognise that overlap onsets might differ from one another, and thus the technological cause for the misunderstanding regarding overlap resolution remains unnoticed. While situations where the one implicitly offered the turn does not take the turn are in some cases

recognised as signs of a rupture of shared understanding about turn-taking, the issue of technology, specifically the delay, was never mentioned. Thus, these ruptures are treated as caused by social rather than technical distortions.

5.2 Resolving problems with intersubjectivity

My second main finding is that when interactants recognise that a shared understanding has broken down, they make their own perspectives available to others and, correspondingly, others' perspectives salient to themselves. This can be achieved by explicating their own perspective through talk (Articles 1, 2, 3) or physical demonstration (Articles 2, 4). To be precise, the interactants do not need to recognise the technological nature of the rupture to employ these practices but can resolve ruptures that they perceive by treating them more as social ruptures of understanding rather than technological ruptures of understanding. Furthermore, as in the case of waiting for the client's physical demonstration of readiness to proceed to closing through disengagement from the screen, these practices could also be used to prevent ruptures in advance.

These practices can be initiated by the one making his or her perspective accessible to the other, or that individual can be invited by another. In my data, self-initiated perspective displays happen when explicitly offering the turn to the other, thus explicating who the interactant considers to be the relevant next speaker (Article 1, Extract 4), when verbalising that the nurse notices the client appearing in the tele-homecare openings (Article 2, Extract 2), and when the homecare client shows that not only the audio but also the video connection work from her perspective (Article 2, Extract 2). Other-initiated perspective displays take place when the nurse invites the client to express readiness to take the medicine (Article 3, Extract 2) and when the specialist invites the general practitioner and the patient to reveal relevant information about the technological artefacts in their physical ecology and their audio connection (Article 4). Table 2 summarises the self- and other-initiated explications and demonstrations in my data.

Table 2. Initiation and execution of perspective display

	Self-initiated	**Other-initiated**
Verbal explication	Offering the turn after overlap (Article 1) Verbalising noticing after a visual appearing (Article 2) Saying that the video connection works after being asked about the audio connection (Article 2)	Explicating the situation with the medicine (Article 3)
Physical demonstration	Getting up to disengage from the encounter (Article 2)	Answering polar-questions by showing the microphone-speaker, producing a thumbs-up gesture or nodding (Article 4)

My data indicate that self-initiated perspective displays are oriented to things that have happened in the past: either to overlapping talk and its resolution (Article 1), to the ambiguity of a multimodal appearing that has just happened during the opening phase of the homecare encounter (Article 2, Extracts 1, 2), or to the nurse's audio-check question that does not address relevant parts of these appearings from the client's perspective (Article 2, Extract 2). This basic dynamic is also in place in the closings of homecare encounters, where the client's physical disengagement demonstrates an understanding of the earlier pre-closing activities as forecasting ending the connection (Article 2, Extracts 4, 5). By undertaking self-initiated perspective displays, these interactants work towards ensuring that others build their future actions to contribute to a shared understanding.

Other-initiated perspective displays, on the other hand, seem to be oriented to things that will happen in the near future. In my data, this means laying the ground for directing the client to take the medicine (Article 3, Extract 2) or finding out whether one should suggest a testing sequence in a distant physical ecology (Article 4). This relates to the doubly contextual nature of actions: in order to be able to form their next action, interactants need to know whether that action would be relevant from the perspective of other interactant(s). Thus, this other perspective needs to be grasped in some way, and if fractured ecologies limit the possibilities of this, others need to be invited to make their perspectives salient.

When employing these practices, people orient themselves to technological mediation as relevant by fitting their conduct to the media available to the other participants. When producing physical demonstrations, the interactants display concrete, physical and showable parts of their perspective. They thus rely on the visual medium of the technology and its associated affordance: the ability to show things. Correspondingly, when producing explications, the interactants need to manage issues that are more abstract and less showable (*the next turn ought to belong to you*, *I have noticed that you have appeared*, *I did not see you a moment ago, but this has changed, and I now see you*). These things are difficult if not impossible to show but can be verbalised with a few words (or in the case of verbal noticing, only one word). Thus, aural medium and talk are employed to displaying these parts of the interactants' perspectives. While employed for different issues and through different media, both explications and demonstrations are used for the same overall purpose: to display one's perspective to others.

This brings us to the third main result of this dissertation: as video mediation detaches the production and reception of actions from each other, the grounds for both interpreting actions as sequentially relevant and for making one's perspective available change. This happens through non-mutual interactional realities and fractured ecologies, which detach the production and reception of actions from each other and make interpreting sequential relevance a difficult task.

Non-mutual realities detach the production of spoken action from their reception, especially as to timing. The most salient example of this appears in Article 1, where the transmission delay is shown to change the locations of overlapping talk and turn-offering silences, leading the interactants to resolve overlapping talk in non-mutual interactional realities (Article 1). Furthermore, single actions can carry different meanings for different participants in video-mediated interactions. This is apparent in the openings of the tele-homecare encounters, where the nurses' verbal greeting carries dual functions, as both the verbal second part in the appearing–noticing adjacency pair and the first part in a reciprocal greeting. These dual functions are, however, not present for the distant participant (Article 2, Extracts 1, 2). Borrowing Mondada's (2019) notion of multiple temporalities, we see that not only can responsive action be produced during initiating actions but also that the same action can simultaneously serve as the responding action (noticing) and the initiating action (greeting). Interpreting these actions demands context-sensitive sense-making and in video-mediated encounters,

actions can appear differently in different contexts of action in relation to their timing in an ongoing stretch of talk (Article 1) and in their sequential context in adjacency pair structures (Article 2). Thus, the sequential contexts of action and the grounds for context-sensitive sense-making can differ with an interactant's perspective.

Fractured ecologies also detach the production and reception of physical actions from each other. Article 3 shows how this becomes prominent when limited access to the medicine makes it impossible for the nurse to make sense of the client not taking the medicine as being caused by not having it at hand: the nurse lacks the visual access to the client's physical ecology and thus cannot take the missing medicine into account when interpreting the sequential relevance of the client's turn. Similarly, Article 4 shows how limited physical access to the local ecology of the general practitioner and patient during problem-solving makes it harder for the specialist to assess the relevance of his own actions, especially whether or not suggesting some solution to the problem would be relevant. Thus, assessing the relevance of these actions becomes difficult, as the one initiating them does not have access to the other interactants' physical reality.

Non-mutual realities and fractured ecologies also shape the repair processes of ruptures in intersubjectivity. As stated above, verbal explications are used with more abstract ruptures, which often relate to non-mutual realities. Ruptures caused by fractured ecologies, on the other hand, can be demonstrated physically and/or explicated verbally. Thus, when repairing intersubjectivity, interactants appear to choose the repair practices that fit both the level of abstraction of the rupture and the medium that affords making one's perspective salient.

In sum, it can be asked how technical mediation becomes interactionally consequential in the process of recognising and repairing ruptures of intersubjectivity. Based on my analyses, I suggest that this happens at three levels. First, technological mediation afforded certain actions for maintaining shared understanding and did not afford others, which in turn was central in the emergence of the ruptures of intersubjectivity. If there were no delay between turns, both interactants would perceive overlaps and their resolution practices similarly, if not identically; likewise, if an interaction took place in the same physical location, interactants could search for activity-relevant objects with their gazes and see where others are looking. Thus, these ruptures of intersubjectivity

would not occur, at least in the same form, if the interactions I studied were not technologically mediated.

Second, technological mediation shapes the processes of recognising these problems. This is due to non-mutual realities and fractured ecologies, which detach the production and reception of both verbal and bodily actions from each other. Interpretation of actions demands context-sensitive sense-making, and since actions in video mediation are produced and interpreted in different contexts, both in relation to time and space, the recognition of these problems and their technological causes can become obscured.

Third, technological mediation also shapes the processes of technological repair. The interactants need to repair intersubjectivity in ways that are accessible to the other. Thus, they fit their actions to the technological media or channel that is both accessible to the other and suitable for the problem. This is achieved especially by verbalising abstract issues of intersubjectivity and showing the concrete ones.

6 DISCUSSION

In this final chapter, I compare my findings with existing conversation analytic research on video-mediated interaction (section 6.1), discuss the relationship between these findings as part of the broader ethnomethodological approach (section 6.2) and computer-mediated communication more generally (Section 6.3), reflect on the limitations of my analysis and outline possible avenues for future research (Section 6.4). My general contribution to conversation analytic research on video-mediated interaction is twofold. On one hand, I have described practices that complement those of interacting in the non-mutual realities and fractured ecologies of video-mediated settings; namely, verbal explications and physical demonstrations. On the other, I have discussed these practices explicitly in relation to the topic of repairing intersubjectivity by showing how these kinds of verbal explications and physical demonstrations can be used to make one's perspective available to others and to gain access to their perspectives. Based on these analyses, I suggest that the broader field of computer-mediated communication would benefit from the action-centred and context-sensitive mode of analysis offered by ethnomethodological conversation analysis, which would replace the limitedness of media with the diversity of modes of use and would offer a more robust theoretical understanding of the relationship between human conduct and communication media.

6.1 Video mediation creates multiple contexts of action which complicate context-sensitive sense-making

As argued in the articles and chapter 5, the interactants recognise the ruptures of intersubjectivity against the sequential relevance of actions: if interactant A does not produce a sequentially relevant next action or produces an action that from interactant B's perspective is not relevant, B may treat this as signalling a rupture of intersubjectivity. When resolving such ruptures, people make their perspectives available to others and others' perspectives available to themselves through verbal explications, visual demonstrations or a combination of the two. When employing these practices, people orient themselves to technological mediation as relevant by fitting their conduct to the media available to other participants. This is most salient with gestures in Article 4; through showing, the interactant not only makes his or her perspective

available to the others but also demonstrates an understanding of technological mediation as causing the rupture of intersubjectivity. The non-mutual interactional realities and fractured ecologies shape the conditions of both the contextual interpreting of actions as sequentially relevant – or not – and the possibilities of making one's perspective salient to others. Interpreting whether or not an action is sequentially relevant can become complicated when the lag time changes the timing of turns in relation to one another and when fractured ecologies detach the production and reception of bodily actions from each other for different mediated perspectives. That is, actions are interpreted in different sequential and physical contexts than where they are produced, which can hinder context-sensitive sense-making. Thus, when resolving these ruptures, interactants fit their perspective-displaying practices to the most suitable technological medium, taking into account the nature and concreteness of the misunderstanding. My analyses indicate that it is not only the fact that the other does not produce a relevant next turn that signals a rupture but also the inability of that missing or irrelevant turn to provide means for building one's own next action.

As the analysis shows, when somebody does not produce a sequentially relevant next action or produces something that is not relevant in a given context, other participants may treat that as signalling a technology-generated rupture of intersubjectivity worthy of repair. This result aligns with existing research on video-mediated interaction, which has emphasised the sequential location of distortions (Olbertz-Siitonen, 2015) and their sequential implications, especially whether distortions make it impossible to produce a relevant next turn (Oittinen, 2018; Rintel, 2015) or interactants produce adverse or disjointed activities (Oittinen, 2020), as important for interactants to determine whether technological distortions are explicitly discussed. The interactants in my data use others' irrelevant and missing turns as signs of ruptures of intersubjectivity, which resonates with earlier findings on how the individual who produces a distorted turn does not have access to how others receive that turn and thus has to deduce the need to initiate a repair from a missing or inappropriate next turn by a recipient (Rintel, 2013a, 2015). Furthermore, as some missing turns can be tolerated if their absence does not create obstacles to the progress of interaction, the interactants appear to orient themselves to progressivity over sequential accountability, as has previously been noted (Oittinen, 2018). Of course, this has to do to with the type of action in which these missing turns ought to participate; thus, missing turns may be less unequivocal in other sequence types than those I have studied.

To solve these problems, the interactants in my data worked to repair intersubjectivity by making their perspective salient to the others, and vice versa, by making others' perspectives salient to themselves. This was achieved by producing and inviting verbal explications and physical demonstrations. These results align in many ways with existing conversation analytic research on video-mediated interaction. Earlier studies have demonstrated how explications can be used to allocate turns after overlaps (Hansen & Svennevig, 2021; Rusk & Pörn, 2019) and projecting one's future actions (Due, 2021) and described ways in which interactants make their instructions understandable to others, as by directing distant participants' bodily actions and attention through mimicable demonstrations (Due & Lange, 2020), naming and mimicking activity-relevant artefacts (Seuren et al., 2020) and making relevant parts of one's local ecology visually available to others by showing (Licoppe, 2017a; Licoppe & Tuncer, 2019; Stommel, Licoppe & Stommel, 2020, Stommel, van Goor & Stommel, 2020). I complement these studies by demonstrating how others can also be invited to reveal their perspectives on the action to help rebuild intersubjectivity. This is achieved through asking questions about the distant ecology of action which can be answered by both explications (Article 3) and showing gestures accompanied with highlighting (Article 4). Furthermore, by gaining access to others' perspectives, interactants are able to create next actions that are intelligible and relevant to their recipients. While explications have been touched upon in previous conversation analytic research (Due, 2021; Hansen, 2020; Rusk & Pörn, 2019), they have not been discussed in relation to the broader idea of repairing ruptures of intersubjectivity. Thus, my analyses offer detailed new findings about both the interactional repair practices of video-mediated interaction and the broader discussion of the relationship between these practices and the more general phenomenon of rebuilding intersubjectivity.

On a more theoretical level, my findings align with the notions of non-mutual interactional realities and fractured ecologies. The ways in which these two features of video-mediated interaction detach the production and reception of actions from each other were relevant in my data both in relation to interpreting actions as sequentially relevant to recognise ruptures of intersubjectivity and in the ways in which the interactants used verbal explications and physical demonstrations to rebuild intersubjectivity. These concepts offer explanatory power for the empirical findings and enable connecting my findings from specific sequential and institutional contexts to the broader field of interactional phenomena.

Based on the analysis, a nuance can be suggested to refine the concept of non-mutual realities. This relates to how the multimodal nature of actions can further complicate the sequential relationship between actions: interpreting what single actions do in video-mediated interaction is shaped not only by their close temporal location in relation to the ongoing action, as originally pointed out by Ruhleder and Jordan (2001), but also in the way in which these actions contribute to adjacency pairs like appearing–noticing or reciprocal greetings and broader lines of action, such as the openings of video-mediated interactions. As discussed above in relation to openings, verbal greetings can have a dual function when produced both as noticing another interactant's appearing and as a greeting, although that may not appear in the same way to the distant participant. Thus, while these turns appear in the same temporal order for both participants, the actions they achieve can differ slightly, depending on perspective.

This has to do with the multiple temporalities of multimodal actions in video-mediated interaction. As actions are put forward multimodally, there are multiple temporalities of talk and other multimodal conduct in interaction (Mondada, 2018), which are shaped separately by video mediation. This can complicate the relationship between the ideas of the *firstness* and *secondness* of actions even more than in face-to-face interactions, as originally suggested by Mondada (2018). The interpretative work that interactants engage in when making sense of these multiple temporalities demonstrates Ruhleder and Jordan's (2001) notion of how non-mutual realities are not just a deterministic technology feature but are actualised in the human–technology relationship.

6.2 Intersubjectivity in video-mediated settings: From "why that now?" to "when is *now* to whom?" and "what *that* is?"

How then, do these findings relate to the general ethnomethodological view on intersubjectivity? In general, my findings align with the central notions of that perspective. Ethnomethodology claims that intersubjectivity emerges as a side product of people interacting together, as they draw from contextual knowledge to produce and interpret indexical actions as intelligible (Depperman, 2015; Garfinkel, 1967; Heritage, 1984; Lindström et al., 2021; Linell, 2017; Simona et al., 2018). In my data, the interactants also draw on contextual knowledge to interpret actions as sequentially relevant. This is achieved by implicitly asking *why that now?* Why does another interactant

produce a specific action at a specific time, and how does that relate to the larger sequence of actions that is being jointly built. While video mediation detaches the production and reception of actions from each other, the interactants still recognise the ruptures of intersubjectivity by interpreting whether these actions are sequentially relevant as parts of the broader sequences they contribute to, as is the case in any other interactional situation.

That said, in addition to showing that basic practices of intersubjectivity appear to hold in video-mediated settings, my results, as part of a growing body of research on video-mediated interaction, offer an important clarification to the ethnomethodological understanding of intersubjectivity. The empirical analysis I have presented offers certain specifications regarding the central question of intersubjectivity: *why that now?* In short, I suggest that by studying intersubjectivity in video-mediated settings, we start to see how neither *now* nor *that* can be taken for granted but need to be appreciated as members' achievements: they may, at least as concepts or members' categories, pre-exist situations where they are employed but they still need to be managed locally by those members.

The study of video-mediated interaction juxtaposes the temporality and spatiality of intersubjectivity with the basic processes of intersubjectivity; that is, accountability for and indexicality of actions. Temporality appears to closely connect to the *now* in the question *why that now?* Since actions may be received differently than they are produced in relation to the temporally unfolding interaction, the basis for answering that question may not be the same as the one to which the producer oriented it. This is precisely the case with overlapping talk appearing differently to different participants (Article 1). Thus, the question is not only *why that now* but also *when is now to whom*.

Accountability and indexicality are also shaped by the new spatiality of video mediation. Fractured ecologies limit access to other participants' perspectives, so that, for example, assessing where one is looking, the action intended to be conveyed through that looking and the kind of context that produces for future actions becomes ambiguous. That is, the grounds for assessing what *that* is can differ among participants. Furthermore, what *that* is can differ between both ends in relation to talk when the first parts of pairs are perceived as second parts due to the time lag. Therefore, what *that* is becomes harder to

determine as interactants lack a shared access to one another's immediate ecologies of action and physical surroundings.

Thus, understanding what both *that* and *now* mean demands contextual sense-making; they are members' achievements. While the basic practices of intersubjectivity – interpreting whether actions are sequentially relevant to recognise ruptures of intersubjectivity and sharing other interactants' perspectives to repair those ruptures – are similar in video-mediated and face-to-face interaction, studying video mediation can offer new insights into the analysis of the key question of intersubjectivity; Carr (2020) and Flanagin (2020) discuss how studying interactional processes in technologically mediated settings can offer new insights into the analysis of co-located interactions.

6.3 Understanding social cues as contextual configurations in video-mediated interactions

Based on the ethnomethodological analysis of the empirical data, I suggest that the current understanding of video-mediated interaction and computer-mediated communication more broadly, which is founded on the social cues approach, would benefit from foregrounding interactional practices. The ethnomethodological and conversation analytic perspective contributes to the understanding of computer-mediated communication in two ways: first, by bringing context to the centre of the analysis and, second, by providing methodological tools that are action-centred rather than object-centred. There are two kinds of benefits to be gained here: on one hand, foregrounding interactional practices can help create more robust, enduring and nuanced theorisation about computer-mediated communication. On the other, interaction-centred analyses can highlight the diversity of the various ways in which people interact through and with communication technologies instead of concentrating on the assumed limitedness of certain media.

By foregrounding the context, we can start to analyse meaningful social actions instead of contextless social cues. Compared to the social cues approach, the ethnomethodological perspective examines social cues as contextual phenomena, as interactional practices that are used to achieve certain actions within certain context. The ethnomethodological approach asks how these cues are parts of meaningful actions

which themselves are parts of larger structures of action. For example, in the medicine-taking case (Article 3), the central object of analytic observation is gaze shift as part of producing a recognisable action; that is, aligning with the suggested medicine-taking, not the isolated gaze shift or more general possibility of mutual gaze *per se*. This stands in stark contrast to the social cues approach, which, in my reading of the key literature, conceptualises the use of social cues without paying attention to their interactional context and functions. That is, they consider different communicative media in relation to how well they can convey gaze, body movements, gestures, prosody et cetera but do not consider to the same extent the purposes for those cues are being used. In line with the ethnomethodological perspective, I have sought to describe how certain actions are achieved in specific sequential and material contexts of action.

This brings us to the second way that ethnomethodological and conversation analytic research can contribute to the broader theorising of computer-mediated-communication: providing methodological tools that are action- rather than object-based. When we consider the key theoretical concepts of conversation analytic research on video-mediated interaction – non-mutual realities and fractured ecologies – we see how these concepts foreground the relationship between human conduct and technology instead of concentrating on the technological features of a given medium. More specifically, they both dissect the relationship between the production and reception of contextually meaningful actions: fractured ecologies by examining the ways in which the production of and perceiving bodily actions take place in different physical surroundings, and non-mutual realities by asking how actions are produced and perceived differently in relation to the temporally unfolding action. They highlight how changes in the spatial and temporal organisation of interaction become meaningful for people as interaction progresses action by action. Here, we can see how these concepts are not so much about the communication devices *per se* but more about how these devices are relevant to the ongoing action. While the central notions of the social cues approach, co-location and synchronicity, first seem to connect nicely to the spatial and temporal changes that video mediation brings to interaction, we can see how both actually foreground the technological medium's abilities, either to allow "individuals engaged in a communication interaction to share the same context" or to "quickly exchange communicative stimuli" (Kock, 2004, p. 333). Foregrounding interactional practices instead of technological artefacts can offer grounds for more detailed, robust

and enduring analyses of computer-mediated communication (Carr, 2020; Flanagin, 2020).

From the more empirical perspective, focusing on interaction highlights on one hand how people can achieve their interactional tasks despite the limitations of a technological medium and, on the other, how they are able to find new ways of interacting that the medium affords. This shifts the analytical focus from the limits of the medium to creative ways of using technology to achieve certain goals. A case in point could be combining the landline telephone's and video mediation equipment's affordances to converse via video even when the audio connection is not functioning: while this technological setting can be considered highly limited (even more so than a working video connection), the interactants are able to scaffold the appropriate affordances to carry out relevant actions (Article 4). As the focus moves from communication technology's assumed limitedness to the interactional practices that technologies afford, the simple and straightforward labels of personal–impersonal, rich–poor and natural–unnatural appear to lack any real explanatory power. Instead, the relationship between human action and technology, in all its fuzziness, becomes apparent and offers more empirically grounded and less deterministic analyses.

In addition to improved empirical analyses of technology-mediated interaction, this shift to action-centred analyses of using technology can offer insights relevant for developing new communication technologies. For example, during the technological problem-solving discussed above (Article 4), directing another person's attention and actions in relation to activity-relevant artefacts, such as a landline telephone and its features, is a relevant interactional task (Article 4). How then could this kind of action of directing another person's attention be implemented in, say, mobile video-mediated interactions or social virtual reality, and could the design of these technological tools somehow support that task? Had the focus of the analysis been on contextless social cues – that is, talk, prosody, pointing and gestures – instead of the action of directing *per se*, this kind of application might not have arisen.

While there are significant methodological and theoretical discrepancies between the ethnomethodological and social cues approaches, there are also linkages between them. The most prevalent is that both are interested in the spatio-temporal context of action. However, adopting action-centred concepts of fractured ecologies and non-mutual

realities would offer computer-mediated communication more precise conceptual tools to analyse the spatio-temporal aspects of interaction. Consider the medicine-taking case in Article 2. Instead of merely comparing rich face-to-face communication to poor video-mediated communication, a whole range of new questions to be considered opens up: is it possible for the participants to simultaneously engage in interaction and manipulate activity-relevant artefacts? Can participants direct one another's attention and actions towards those artefacts? If so, how do they achieve this despite the lack of physical co-presence? In short, actions in context and contexts as actions become available for analysis.

While the importance of adjusting one's actions to the technological medium and its affordances was already recognised in the early social cues literature (Short et al., 1976, p. 64), the field has evolved to become medium-centric, and calls have been made for more action-centred approaches (Carr, 2020; Flanagin, 2020). A recent study on media richness examines the ways in which ambiguity, social regulation, shared understanding and experience of solidarity interact (Roos et al, 2020). However, adopting an action-centred view might require abandoning the conception of communication as information transfer and replacing it with understanding communication as contextual sense-making and replacing the idea of contextless social cues with actions as multimodal gestalts designed for definite material settings.

6.4 Practical implications for professional video-mediated interaction

While this integrative chapter concentrates on general phenomena regarding intersubjectivity and I do not discuss institutional relevancies in relation to intersubjectivity in this integrative chapter, the relationship between video mediation and institutional practice are discussed in the individual articles. Thus, my findings carry practical implications. These implications can be divided into three themes: importance of overall reflection and improvement of interaction skills in video-mediated (care) interactions, specific practices to build intersubjectivity in video-mediated interactions and planning video-mediated services so that the assemblage of interactional practices, interaction spaces (both physical and virtual), communication technology and other artefacts support the central tasks of care.

First, I hope that these results could spark reflection among health care professionals about the effects that video mediation has on interaction. As I have showed in the data, professionals have skills to resolve challenges in the interaction in their professional stocks of knowledge (Vehviläinen & Peräkylä, 2003): the solutions to the difficulties that video mediation (and other communication technologies) bring are nothing new to professionals, but they may need to be highlighted and pinpointed. Thus, I invite practitioners and those who develop health and social care organisations to engage in research and development that draw on encounters with real clients and make the good practices salient within the working community.

The second theme are the specific interactional practices that were used in my data to build and repair intersubjectivity, namely explications and demonstrations. At a general level, these two practices are transferrable to different institutional contexts, not only health counselling, home care and medical consultations. When implementing the practices of explications and demonstrations, institutional and situational specificities need, however, to be taken into account. Consider for example the practice of showing as a way to demonstrate ones perspective: showing a tooth in a tele-dentist consultation might differ from showing a bottle of medicine and thus a dentist inviting a patient to produce this demonstration should take these differences into account. While explications might feel unnatural and, in some situations, lower the experience of social solidarity (see e.g. Roos et al., 2022), explications might in some situations be the only way to rebuild shared understanding.

The third practical implication of the findings is to design video-mediated services so that the assemblage of interactional practices, interaction spaces (both physical and virtual), communication technology and other care-relevant artefacts support the central tasks of care. In order to produce high quality video-mediated services, the central tasks of a given service should be recognised, the ways in which the video mediation technology affords the actions of these task should be analysed and relevant changes be made to the service. As more services are, at least partially, visioned to be provided directly to clients' homes, the affordances of home spaces need to be taken into consideration – and vice versa, the actions that the homespace should afford may need to be taken into consideration when designing and renovating living spaces. While some interactional phenomena can be anticipated, this (re)designing of video-mediated services will most likely demand cyclical iteration of observing interaction and its

problems, ideating, prototyping and testing potential solutions to those problems, and again observing how these changes play out in interaction. While the first two implications have more to do with the conduct of individual professionals, the third has more to do with organisational level.

6.5 Methodological remarks and directions for future research

Conversation analysis provides a methodology that is both detailed and robust for analysing interactants' perspective to social interaction. It is detailed in the sense that all the analyst's interpretations have to be grounded in the interactants' own interpretations as made publicly available through their actions. This grounding is achieved through the "next-turn proof procedure" (Edwards, 2004). This effort to obtain a detailed description of the interactants' own interpretations of ongoing action provides the method with robustness. As the analysis is based on the observable details in the data, it is based on concrete interactional practices instead of idealisations about interaction. I have pursued detailed and transparent descriptions of the interactants' perspectives on the action in the four research articles that are part of this dissertation. This gives the reader the ability to evaluate the validity of the findings and my interpretations presented in this integrative portion of the dissertation. Despite this abiding aim for detailed and robust analysis, three limitations of the data and analysis can be identified: *dataset size*, *high data diversity* and *older data*.

The first limitation relates to the comparatively small datasets (see Table 1), which could affect the generalisability of my findings. However, dataset size and number of instances analysed do not automatically make conversation analytic research good or bad. In that approach, even single case studies have been recognised as important because they can provide highly context-sensitive analyses of complex interactional situations (Schegloff, 1987). Furthermore, as Peräkylä (2004) has pointed out, when studying institutional interactions, the central finding is describing *possible* social actions and their sequential and institutional conditions in a specific context. Drawing from this idea, I suggest that research into technology-mediated interaction should describe the possible social actions with regard to their sequential and technological conditions, as shaped by technological affordances. Thus, by carefully describing the processes through which technology may shape interaction by its affordances, analyses with small datasets can provide findings that can be generalised as hypotheses to be tested and arguments to be specified. This is what I have aimed at throughout my analyses. That said, the limited

size of the datasets does highlight the need for more research using larger datasets to reveal both the frequency of the practices I have described and other, less frequent practices that were not captured in my data and to confirm, refine or rebut the arguments I have made.

The second limitation involves the diversity of the data, both in relation to their technological settings (traditional–dyadic, hybrid–triadic, hybrid–group) and their institutional context (homecare, tele-consultations, group health counselling). This diversity has both pros and cons. Having such diverse data enabled recognising similar practices of repairing intersubjectivity in different sequential and institutional contexts. In particular, having access to data with two simultaneous perspectives (the counselling data) was crucial for analysing the omnipresence of non-mutual realities and recognising their potential effects on interaction in the other datasets.

The downside of this diversity is the somewhat superficial analysis of institutional relevancies for the interactants presented in this integrative chapter. Two points can be made regarding this limitation. First, having these diverse data forced me to analyse more general phenomena that were available in the different datasets. As I could not have focused on a thick description of the relationship of video mediation and, say, group counselling, more general interactional phenomena such as overlaps, activity transition and repair and their relationship with the technological affordances of video mediation were my focus. Conducting research always involves a series of choices about what to include and what to exclude. Having institutionally diverse data afforded this analytical choice, while concentrating on one institutional setting would have afforded other choices with different emphases and certain phenomena being excluded. Second, this more general focus is probably more prominent in this integrative chapter, and the relationships between technological and institutional relevance are more apparent in the articles.

An unexplored possibility that these datasets offer is studying the same interactional phenomena, such as openings and closings, in all these settings. This would have enabled comparison of the dynamics of understanding repair between hybrid and traditional video-mediated interaction. On the other hand, using data with varied institutional contexts, technological settings, numbers of participants and thus interactional settings (group–individual in the counselling context, individual–individual in the homecare context and dyad–individual in the tele-consultation context) would have made it more difficult to differentiate between institutional and technological

relevancies, thus risking an incoherent analysis. With the benefit of hindsight, this kind of analysis could have been carried out on at least some issues, such as the aforementioned openings and closings. The comparison of traditional video mediation and hybrid settings offers an important avenue for future research, especially when data are being collected from similar institutional contexts.

The third possible critique of the analysis relates to the age of the data. This is most apparent with the counselling data, which was collected in 2006. It is true that practices in relation to technological affordance may change as these technologies improve and are integrated into people's everyday lives. However, Seuren et al. (2020) recognised remarkably similar practices in their dyadic patient–professional data, which were collected between 2015 and 2018. Thus, while for example distributed network architecture can reduce latency and younger generations, such as the high school students in Rusk and Pörn's (2019) study, might be more accustomed to technological mediation and its relevance for interaction, these changes could well be more subtle than popular discourse suggests.

This critique of using older data can be taken to a more theoretical level. Following Flanagin's (2020) notions about techno-centrism, one might ask whether my analysis reveals anything about technology-mediated interactions other than video mediation, a technology that is prone to change and, at least in techno-optimistic visions, will soon be replaced by more advanced technology such as social virtual reality or metaverse. I have strived to avoid techno-centrism by foregrounding interactional phenomena instead of technological features. Thus, by presenting practices of recognising and repairing ruptures of intersubjectivity – namely, verbal explications and physical demonstrations – I hope to have provided insights for those studying emerging forms of technology-mediated interaction. While the particulars of the ways in which non-mutual realities and fractured ecologies shape these interactional processes might not remain identical in other communication media (whether existing or envisioned), I believe that these basic processes of human understanding will remain relevant in the future.

It should also be noted that while in general I adopted multimodal perspective in the analysis, in Article 1 my co-authors and I concentrated solely on the participants' verbal actions. We did so for three reasons. First, as the data were not initially collected for conversation analysis, they provided very limited access to the gaze of the participants. Second, the ethical approval obtained to use the data did not allow for any visual

presentation of the data. Thus, the article would have comprised only transcriptions – but no images – of gaze patterns, making it potentially unapproachable, especially to the professional audience we aimed at reaching. Third, as the analysis of these data were the first that I carried out for the dissertation, I made a hasty decision not to include gaze in the analysis. At the time of writing that article, I considered the interplay of the transmission delay and verbal conduct to be "enough" and thus decided not to use limited and hard-to-present data. With hindsight, I would make a different decision.

In addition to these limitations of the analysis, a more epistemological matter arises regarding the compatibility between ethnomethodology's analytic approach, which foregrounds members' perspectives on action, and using data from two perspectives of action, as was the case with the counselling data. As another researcher of video-mediated interaction asked me, how exactly is it possible to reconcile ethnomethodology's premise of the observability of conduct for co-participants and analysts alike and the researcher's use of recordings from two perspectives of action, that are by definition, not available to participants? While I do not have a definitive answer to this question, I propose three ways to approach it: foregrounding participant perspectives during the analysis of multiple perspectives, treating data from multiple perspectives as contextual information and considering the possibility that technological mediation may call for redefining some basic principles of ethnomethodology.

First, I have focused in the analysis on describing the participants' orientations to the ongoing interaction from each perspective separately, thus holding on to the observability of conduct. That is, both perspectives are analysed in their own right, with a focus solely on the publicly available conduct of the interactants, without any presumed reliance on the idea that these perspectives differ from each other. The collision between ethnomethodological principles and the analysis of the data from two perspectives only arises when the analyses from these two perspectives are being compared to each other. In this situation, a new perspective emerges: "a god's eye perspective", which, while based on the perspectives of the interactants, does not represent those perspectives because of its all-encompassing nature. This god's eye perspective could be regarded as irreconcilably conflicting with ethnomethodology's analytical approach, which emphasises the members' perspective. To some extent, it is not unfair to say that while I employed conversation analysis, I diverge in this dissertation from the ethnomethodological tradition in this respect (see also Olbertz-Siitonen, 2015, pp. 211–212). However, this comparison enabled analysing non-mutual

interaction realities and grasping their importance for repairing intersubjectivity in a way that I cannot imagine being carried out without this god's eye perspective.

This brings us to the second point I want to make, which is that an analogy can be made between using data from two perspectives and using contextual information to back up the analysis of interaction. While the mainstream or strict ethnomethodological view on context is that only things that are publicly oriented to or by participants themselves should be regarded as relevant contexts of action (see e.g., Potter, 1998; Schegloff, 1997; ten Have, 2007; Tracy, 1998), it has also been suggested that information outside the encounter under study may be needed in order to understand the central phenomena and to clarify or correct the analysis (see e.g., de Kok, 2008; van Dijk, 1997; Waring et al., 2012). In the same way that understanding some speech acts may require the acquisition of extra-situational material (Fitch, 1998), analysis of non-mutual realities may also require the use of such material, which in this case means recordings of the same situation from different perspectives. Although the participants do not seem to orient themselves to the other perspective in their activities (rather, they seem to orient themselves to their perspective as the one true perspective; that is, the context in which they need to fit their actions), what happens in this other perspective certainly enables and limits the set of actions that opens up for the participants. Thus, taking this other reality into account in the analysis is, if not necessary, at least valuable for properly understanding the phenomenon I recognised and to answer the questions I wanted to ask from the data. If context and action are considered mutually constitutive, the emergence of that context becomes an essential topic. This makes the analysis of both perspectives ethnomethodologically relevant.

Third, because both the aforementioned views – creating a god's eye perspective though ethnomethodologically grounded analysis and treating data from multiple sources as contextual information – seem to at least partially contradict a strict ethnomethodological interpretation, the following question arises: do some basic principles of ethnomethodology need to be adjusted as a result of technology-mediated interaction? Many ethnomethodological researchers answer in the negative. Hutchby (2001a, 2001b) has argued that adopting the concept of affordances makes it possible to produce relevant analyses of technologised interactions that align with ethnomethodological principles of observability of action. In a similar vein, Arminen and colleagues (2016) suggest that foregrounding technology over members' perspectives runs the risk of producing yet another version of the "bucket theory of context" (Heritage, 1987), in which pre-existing technological relevancies are thought

to determine the course of interaction, thus repeating the shortcomings of techno-deterministic approaches. And while emerging communication technologies may afford new ways for people to present themselves in everyday life, "ethnomethodological respecification [still] provides an important analytic mentality" (Housley, 2021, p. 83) through which to understand these changes.

My own views are largely along the same lines as these notions: as stated above, the basic phenomena of interaction, such as reasoning based on sequentiality of actions, seem to retain their place in my data. However, the irreconcilable contradiction between non-mutual realities and ethnomethodology remains: it seems, at least to me, that the study of some aspects of technologically mediated interaction, especially regarding non-mutual interactional realities, require parting from the strict ethnomethodological approach in favour of the god's eye view, which carries an inherent risk of repeating the mistakes of the bucket theory of context and techno-deterministic approach. Thus, this dissertation will not serve as an answer to this matter, but it may help inspire and provoke future work to clarify this relationship.

During the process of completing this doctoral work, two ideas for future research arose, in addition to the need for larger datasets and research concentrated on a single institutional setting. First are some possibilities to bring the social cues approach and ethnomethodological conversation analysis into closer contact. A potential point of departure could be outlier instances that do not follow the general trends found in the statistical analyses used in the social cues approach. While in statistical models outliers are often dismissed as uninteresting, these "troublemaker cases" (Garfinkel, 2002, pp. 125–126) or deviant cases (Peräkylä, 2004, pp. 292–293; ten Have 2007, 151) are a fruitful domain for conversation analytic research, as they provide opportunities both to investigate and ideally identify the situational factors that explain their deviant nature and to provide more precise analyses of the ordinary instances. This could lead to more nuanced operationalisations of future statistical analyses.

Second, with the widely acclaimed rapid development of technological means of mediated interaction, longitudinal studies would provide a deeper understanding of both emerging ways of human–technology interaction and the processes through which these ways of interaction become mundane. As people become more acquainted with communication technologies and domesticate them into their everyday lives, they may find new ways to use the affordances of these technologies. For example, the combination of a landline telephone and video mediation equipment was immediately

adopted as a repair solution in my data from tele-consultations. Longitudinal studies could offer access to processes of innovation and routinisation which are important not only in the analysis of video-mediated interaction but also in computer-mediated communication more generally and ultimately the overarching human–technology relationship.

REFERENCES

Aaltonen, T., Arminen, I., & Raudaskoski, S. (2014). Photo sharing as a joint activity between an aphasic speaker and others. In M. Nevile, P. Haddington, T. Heinemann, & M. Rauniomaa (Eds.), *Interacting with Objects* (pp. 125–144). John Benjamins Publishing Company. https://doi.org/10.1075/z.186.06aal

Aidemark, J. (2022). Care by video consultations: why or why not? *Procedia Computer Science*, *196*, 400–408. https://doi.org/10.1016/j.procs.2021.12.029

Andel, S. A., de Vreede, T., Spector, P. E., Padmanabhan, B., Singh, V. K., & de Vreede, G.-J. (2020). Do social features help in video-centric online learning platforms? A social presence perspective. *Computers in Human Behavior*, *113*, 106505. https://doi.org/10.1016/j.chb.2020.106505

Antaki, C., & Kent, A. (2012). Telling people what to do (and, sometimes, why): Contingency, entitlement and explanation in staff requests to adults with intellectual impairments. *Journal of Pragmatics*, *44*(6–7), 876–889. https://doi.org/10.1016/j.pragma.2012.03.014

Arminen, I. (2005). *Institutional Interaction: Studies of Talk at Work*. Ashgate.

Arminen, I., Licoppe, C., & Spagnolli, A. (2016). Respecifying Mediated Interaction. *Research on Language and Social Interaction*, *49*(4), 290–309. https://doi.org/10.1080/08351813.2016.1234614

Baggs, E. (2021). All Affordances Are Social: Foundations of a Gibsonian Social Ontology. *Ecological Psychology*, *33*(3–4), 257–278. https://doi.org/10.1080/10407413.2021.1965477

Benjamin, T., & Mazeland, H. (2012). Conversation Analysis and Other-Initiated Repair. In *The Encyclopedia of Applied Linguistics*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781405198431.wbeal1310

Brauner, E., Boos, M., & Kolbe, M. (Eds.). (2018). *The Cambridge Handbook of Group Interaction Analysis*. Cambridge University Press. https://doi.org/10.1017/9781316286302

Broth, M., & Mondada, L. (2013). Walking away: The embodied achievement of activity closings in mobile interaction. *Journal of Pragmatics*, *47*(1), 41–58. https://doi.org/10.1016/j.pragma.2012.11.016

Carr, C. T. (2020). CMC Is Dead, Long Live CMC!: Situating Computer-Mediated Communication Scholarship Beyond the Digital Age. *Journal of Computer-Mediated Communication*, *25*(1), 9–22. https://doi.org/10.1093/jcmc/zmz018

Cipolletta, S., Frassoni, E., & Faccio, E. (2018). Construing a therapeutic relationship online: An analysis of videoconference sessions. *Clinical Psychologist*, *22*(2), 220–229. https://doi.org/10.1111/cp.12117

Connolly, S. L., Miller, C. J., Lindsay, J. A., & Bauer, M. S. (2020). A systematic review of providers' attitudes toward telemental health via videoconferencing. *Clinical Psychology: Science and Practice*, *27*(2). https://doi.org/10.1111/cpsp.12311

Cooper-White, P. (2014). Intersubjectivity. In *Encyclopedia of Psychology and Religion* (pp. 882–886). Springer US. https://doi.org/10.1007/978-1-4614-6086-2_9182

Daft, R. L., & Lengel, R. H. (1983). *Information Richness. A New Approach to Managerial Behavior and Organization Design*. https://doi.org/10.21236/ADA128980

Daft, R. L., & Lengel, R. H. (1986). Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, *32*(5), 554–571. https://doi.org/10.1287/mnsc.32.5.554

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. In *Source: MIS Quarterly* (Vol. 11, Issue 3).

Dalley, D., Rahman, R., & Ivaldi, A. (2021). Health Care Professionals' and Patients' Management of the Interactional Practices in Telemedicine Videoconferencing: A Conversation Analytic and Discursive Systematic Review. *Qualitative Health Research*, *31*(4), 804–814. https://doi.org/10.1177/1049732320942346

de Kok, B. C. (2008). The role of context in conversation analysis: Reviving an interest in ethno-methods. *Journal of Pragmatics*, *40*(5), 886–903. https://doi.org/10.1016/j.pragma.2007.09.007

de la Cruz Paragas, F., & Lin, T. T. (2016). Organizing and reframing technological determinism. *New Media & Society*, *18*(8), 1528–1546. https://doi.org/10.1177/1461444814562156

Depperman, A. (2015). When recipient design fails: Egocentric turn-design of instructions in driving school lessons leading to breakdowns of intersubjectivity. *Gesprächsforschung–Online-Zeitschrift Zur Verbalen Interaktion*, *16*, 36–101.

Dewing, J. (2008). Process Consent and Research with Older Persons Living with Dementia. *Research Ethics*, *4*(2), 59–64. https://doi.org/10.1177/174701610800400205

Drew, P. (2009). "Quit talking while I'm interrupting": A comparison between positions of overlap onset in conversation. In M. Haakana, M. Laakso, & J. Lindström (Eds.), *Talk in Interaction: Comparative Dimensions* (pp. 70–93). Finnish Literature Society.

Drew, P., & Heritage, J. (1992). Analyzing talk at work: an introduction. In P. Drew & J. Heritage (Eds.), *Talk at work. Interaction in institutional settings* (pp. 3–65). Cambridge University Press.

Due, B. L. (2021). RoboDoc: Semiotic resources for achieving face-to-screenface formation with a telepresence robot. *Semiotica*, *2021*(238), 253–278. https://doi.org/10.1515/sem-2018-0148

Due, B. L., & Lange, S. B. (2020). Body Part Highlighting: Exploring two types of embodied practices in two sub-types of showing sequences in video-mediated consultations. *Social Interaction. Video-Based Studies of Human Sociality*, *3*(3). https://doi.org/10.7146/si.v3i3.122250

Due, B. L., Lange, S. B., Nielsen, M. F., & Jarlskov, C. (2019). Mimicable embodied demonstration in a decomposed sequence: Two aspects of recipient design in professionals' video-mediated encounters. *Journal of Pragmatics*, *152*, 13–27. https://doi.org/10.1016/j.pragma.2019.07.015

Eason, K., & Waterson, P. (2014). Fitness for purpose when there are many different purposes: Who are electronic patient records for? *Health Informatics Journal*, *20*(3), 189–198. https://doi.org/10.1177/1460458213501096

Edwards, D. (2004). Proof Procedure. In M. S. Lewis-Beck, A. Bryman, & Tim. F. Lao (Eds.), *The SAGE Encyclopedia of Social Science Research Methods*. Sage Publications, Inc. https://doi.org/10.4135/9781412950589.n763

Fitch, K. L. (1998). Text and Context: A Problematic Distinction for Ethnography. *Research on Language & Social Interaction*, *31*(1), 91–107. https://doi.org/10.1207/s15327973rlsi3101_6

Flanagin, A. J. (2020). The Conduct and Consequence of Research on Digital Communication. *Journal of Computer-Mediated Communication*, *25*(1), 23–31. https://doi.org/10.1093/jcmc/zmz019

Fox, B. A., Benjamin, T., & Mazeland, H. (2012). Conversation Analysis and Repair Organization: Overview. In *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal1314

Gan, Y. (2021). Capturing love at a distance. *Social Interaction. Video-Based Studies of Human Sociality*, *4*(3). https://doi.org/10.7146/si.v4i3.128148

Gan, Y., Greiffenhagen, C., & Licoppe, C. (2020). Orchestrated openings in video calls: Getting young left-behind children to greet their migrant parents.

Journal of Pragmatics, *170*, 364–380. https://doi.org/10.1016/j.pragma.2020.09.022

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Prentice-Hall.

Garfinkel, H. (2002). *Ethnomethodology's program: working out Durkeim's aphorism*. Rowman & Littlefield Publishers.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin. https://doi.org/10.4324/9781315740218

Gibson, J. J. (1982). *Reasons for Realism: Selected Essays*. Lawrence Erlbaum Associates.

Gillespie, A., & Cornish, F. (2009). *Intersubjectivity: Towards a Dialogical Analysis*.

Goffman, E. (1981). *Forms of Talk*. University of Pennsylvania Press.

Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers*. Academic Press.

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), 606–633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Goodwin, C. (2007). Environmentally Coupled Gestures. In S. D. Duncan, J. Cassel, & E. T. Levy (Eds.), *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill* (pp. 195–212). https://doi.org/10.1075/gs.1.18goo

Goodwin, C., & Goodwin, M. H. (2004). Participation. In A. Duranti (Ed.), *A Companion to Linguistic Anthropology* (pp. 222–244).

Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, *62*(1–2), 51–75. https://doi.org/10.1515/semi.1986.62.1-2.51

Greene, C. J., Morland, L. A., Macdonald, A., Frueh, B. C., Grubbs, K. M., & Rosen, C. S. (2010). How does tele-mental health affect group therapy process? Secondary analysis of a noninferiority trial. *Journal of Consulting and Clinical Psychology*, *78*(5), 746–750. https://doi.org/10.1037/a0020158

Hansen, J. P. B. (2020). Invisible participants in a visual ecology: Visual space as a resource for organising video-mediated interpreting in hospital encounters. *Social Interaction. Video-Based Studies of Human Sociality*, *3*(3). https://doi.org/10.7146/si.v3i3.122609

Hansen, J. P. B., & Svennevig, J. (2021). Creating space for interpreting within extended turns at talk. *Journal of Pragmatics*, *182*, 144–162. https://doi.org/10.1016/j.pragma.2021.06.009

Hayashi, M., Raymond, G., & Sidnell, J. (Eds.). (2013). *Conversational Repair and Human Understanding*. Cambridge University Press. https://doi.org/10.1017/CBO9780511757464

Heath, C. (1986). *Body Movement and Speech in Medical Interaction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511628221

Heath, C. (2004). Analysing face to face interaction: video, the visual and material. In D. Silverman (Ed.), *Qualitative Research. Theory, Method and Practice* (pp. 266–282). Sage.

Heath, C., & Luff, P. (1992). Media Spaces and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction. *Human.-Computer Interaction*, *7*, 315–346.

Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Polity Press.

Heritage, J. (1987). Ethnomethodology. In A. Giddens & J. Turner (Eds.), *Social theory today* (pp. 224–272). Polity Press.

Heritage, J., & Sefi, S. (1992). Dilemmas of advice: Aspects of the delivery and reception of advice in interactions between Health Visitors and first-time

mothers. In P. Drew & J. Heritage (Eds.), *Talk at Work: Interaction in Institutional Settings* (pp. 359–417). Cambridge University Press.

Hjulstad, J. (2016). Practices of Organizing Built Space in Videoconference-Mediated Interactions. *Research on Language and Social Interaction*, *49*(4), 325–341. https://doi.org/10.1080/08351813.2016.1199087

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*(2), 174–196. https://doi.org/10.1145/353485.353487

Housley, W. (2021). *Society in the Digital Age. An Interactionist Perspective*. Sage.

Hutchby, I. (2001a). *Conversation and Technology. From the Telephone to the Internet*. Polity.

Hutchby, I. (2001b). Technologies, Texts and Affordances. *Sociology*, *35*(2), 441–456. https://doi.org/10.1177/S0038038501000219

Hutchby, I. (2014). Communicative affordances and participation frameworks in mediated interaction. In *Journal of Pragmatics* (Vol. 72, pp. 86–89). Elsevier. https://doi.org/10.1016/j.pragma.2014.08.012

Hyppönen, H., Hämäläinen, P., & Reponen, J. (2015). *E-health and e-welfare of Finland. Check point 2015*. National institute for health and welfare. http://urn.fi/URN:ISBN:978-952-302-563-9

Iacoboni, M. (2008). *Mirroring People*. Farrar, Strauss and Giroux.

Ignatowicz, A., Atherton, H., Bernstein, C. J., Bryce, C., Court, R., Sturt, J., & Griffiths, F. (2019). Internet videoconferencing for patient–clinician consultations in long-term conditions: A review of reviews and applications in line with guidelines and recommendations. *DIGITAL HEALTH*, *5*, 205520761984583. https://doi.org/10.1177/2055207619845831

Jakonen, T., & Jauni, H. (2021). Mediated learning materials: visibility checks in telepresence robot mediated classroom interaction. *Classroom Discourse*, *12*(1–2), 121–145. https://doi.org/10.1080/19463014.2020.1808496

Jakonen, T., & Jauni, H. (2022). Managing activity transitions in robot-mediated hybrid language classrooms. *Computer Assisted Language Learning*, 1–24. https://doi.org/10.1080/09588221.2022.2059518

Jefferson, G. (1984). Notes on some orderlinesses of overlap onset. In V. D'Urso (Ed.), *Discourse Analysis and Natural Rhetoric* (pp. 11–38). Cleup.

Jefferson, G. (1986). Notes on "latency" in overlap onset. *Human Studies*, *9*(2–3), 153–183. https://doi.org/10.1007/BF00148125

Jefferson, G. (2004a). A sketch of some orderly aspects of overlap in natural conversation. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the First Generation* (pp. 43–59). John Benjamins. https://doi.org/10.1075/pbns.125.05jef

Jefferson, G. (2004b). "At first I thought": a normalizing device for extraordinary events. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the First Generation* (pp. 131–167). https://doi.org/10.1075/pbns.125.09jef

Jefferson, G. (2004c). *Glossary of transcript symbols with an introduction* (G. H. Lerner, Ed.; pp. 13–31). John Benjamins. https://doi.org/10.1075/pbns.125.02jef

Karl, K. A., Peluchette, J. v., & Aghakhani, N. (2022). Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly. *Small Group Research*, *53*(3), 343–365. https://doi.org/10.1177/10464964211015286

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica*, *26*, 22–63.

Kevoe-Feldman, H. (2015). Working the overall structural organization of a call: How customers use third position as leverage for gaining service

representatives' assistance in dealing with service problems. *Language & Communication*, *43*, 47–57. https://doi.org/10.1016/j.langcom.2015.05.001

Kitzinger, C. (2012). Repair. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 229–256). Wiley. https://doi.org/10.1002/9781118325001.ch12

Klitmøller, A., & Lauring, J. (2013). When global virtual teams share knowledge: Media richness, cultural difference and language commonality. *Journal of World Business*, *48*(3), 398–406. https://doi.org/10.1016/j.jwb.2012.07.023

Kock, N. (2001). Compensatory adaptation to a lean medium: an action research investigation of electronic communication in process improvement groups. *IEEE Transactions on Professional Communication*, *44*(4), 267–285. https://doi.org/10.1109/47.968108

Kock, N. (2004). The Psychobiological Model: Towards a New Theory of Computer-Mediated Communication Based on Darwinian Evolution. *Organization Science*, *15*(3), 327–348. https://doi.org/10.1287/orsc.1040.0071

Koivisto, T., Ilomäki, S., Kurtti, E., Koskela, I., Weiste, E., Salo, S., Aalto, O., Husman, P., & Ruusuvuori, J. (2020). *Terveydenhuollon työntekijät digimurroksessa. Moniaineistoinen tutkimus asiantuntijuuden ja yhteistyön rakentumisesta.*

Kozar, O. (2016). Teachers' reaction to silence and teachers' wait time in video and audioconferencing English lessons: Do webcams make a difference? *System*, *62*, 53–62. https://doi.org/10.1016/j.system.2016.07.002

Kyytsönen, M., Aalto, A.-M., & Vehko, T. (2021). *Sosiaali- ja terveydenhuollon sähköinen asiointi 2020–2021: Väestön kokemukset [Social and health care online service use in 2020–2021: Experiences of the population]*. http://urn.fi/URN:ISBN:978-952-343-680-0

Laitinen, J., Korkiakangas, E., Alahuhta, M., Keinänen-Kiukaanniemi, S., Rajala, U., Timonen, O., Jokelainen, T., Keränen, A.-M., Remes, J., Ruokonen, A.,

Hedberg, P., Taanila, A., Husman, P., & Olkkonen, S. (2010). Feasibility of videoconferencing in lifestyle group counselling. *International Journal of Circumpolar Health*, *69*(5), 500–511. https://doi.org/10.3402/ijch.v69i5.17688

Latikka, R., Rubio-Hernández, R., Lohan, E. S., Rantala, J., Nieto Fernández, F., Laitinen, A., & Oksanen, A. (2021). Older Adults' Loneliness, Social Isolation, and Physical Information and Communication Technology in the Era of Ambient Assisted Living: A Systematic Literature Review. *Journal of Medical Internet Research*, *23*(12), e28022. https://doi.org/10.2196/28022

Lévi-Strauss, C. (1980). *The Savage Mind*. University of Chicago Press.

Licoppe, C. (2015). Video communication and "camera actions". The production of wide video shots in courtrooms with remote defendants. *Journal of Pragmatics*, *76*, 117–134. https://doi.org/10.1016/j.pragma.2014.11.008

Licoppe, C. (2017a). Skype appearances, multiple greetings and 'coucou': The sequential organization of video-mediated conversation openings. *Pragmatics*, *27*(3), 351–386. https://doi.org/10.1075/prag.27.3.03lic

Licoppe, C. (2017b). Showing objects in Skype video-mediated conversations: From showing gestures to showing sequences. *Journal of Pragmatics*, *110*, 63–82. https://doi.org/10.1016/j.pragma.2017.01.007

Licoppe, C., & Dumoulin, L. (2010). The "curious case" of an unspoken opening speech act: Avideo-ethnography of the use of video communication in courtroom activities. *Research on Language and Social Interaction*, *43*(3), 211–231. https://doi.org/10.1080/08351811003741319

Licoppe, C., & Morel, J. (2012). Video-in-Interaction: "Talking Heads" and the Multimodal Organization of Mobile and Skype Video Calls. *Research on Language and Social Interaction*, *45*(4), 399–429. https://doi.org/10.1080/08351813.2012.724996

Licoppe, C., & Tuncer, S. (2019). The initiation of showing sequences in video-mediated communication. In *Gesprächsforschung* (Vol. 20). www.gespraechsforschung-ozs.de

Licoppe, C., & Veyrier, C.-A. (2017). How to show the interpreter on screen? The normative organization of visual ecologies in multilingual courtrooms with video links. *Journal of Pragmatics*, *107*, 147–164. https://doi.org/10.1016/j.pragma.2016.09.012

Lindström, J., Laury, R., Peräkylä, A., & Sorjonen, M.-L. (Eds.). (2021). *Intersubjectivity in Action* (Vol. 326). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.326

Linell, P. (2017). Intersubjectivity in Dialogue. In E. Weigand (Ed.), *The Routledge Handbook of Language and Dialogue* (pp. 109–126).

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). *Fractured Ecologies: Creating Environments for Collaboration* (Vol. 18).

Luff, P., Heath, C., Yamashita, N., Kuzuoka, H., & Jirotka, M. (2016). Embedded Reference: Translocating Gestures in Video-Mediated Interaction. *Research on Language and Social Interaction*, *49*(4), 342–361. https://doi.org/10.1080/08351813.2016.1199088

Mäki-Petäjä-Leinonen, A. (2006). Dementian vaikutuksesta oikeudelliseen toimintakykyyn [The effect of dementia on legal capacity]. *Stakesin Työpapereita 21/2006*.

Marquis, N., Larivée, P., Saey, D., Dubois, M.-F., & Tousignant, M. (2015). In-Home Pulmonary Telerehabilitation for Patients with Chronic Obstructive Pulmonary Disease: A Pre-experimental Study on Effectiveness, Satisfaction, and Adherence. *Telemedicine and E-Health*, *21*(11), 870–879. https://doi.org/10.1089/tmj.2014.0198

McDonald, K. E., & Kidney, C. A. (2012). What Is Right? Ethics in Intellectual Disabilities Research. *Journal of Policy and Practice in Intellectual Disabilities*, *9*(1), 27–39. https://doi.org/10.1111/j.1741-1130.2011.00319.x

Mead, G. H. (1934). *Mind, Self and Society. From the Standpoint of a Social Behaviorist*. University of Chicago Press.

Merleau-Ponty, M. (2013). *Phenomenology of Perception*. Taylor & Francis Group.

Mlynář, J., González-Martínez, E., & Lalanne, D. (2018). Situated organization of video-mediated interaction: A review of ethnomethodological and conversation analytic studies. *Interacting with Computers*, *30*(2), 73–84. https://doi.org/10.1093/iwc/iwx019

Mondada, L. (2001). *Conventions for multimodal transcription*. Https://Www.Lorenzamondada.Net/Multimodal-Transcription.

Mondada, L. (2014). Pointing, talk and the bodies: Reference and joint attention as embodied interactional achievements. In M. Seyfeddinipur & M. Gullberg (Eds.), *From Gesture in Conversation to Visible Utterance in Action* (pp. 95–124). Benjamins.

Mondada, L. (2015). Ouverture et préouverture des réunions visiophoniques. *Réseaux*, *6*, 39–84.

Mondada, L. (2018). Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction*, *51*(1), 85–106. https://doi.org/10.1080/08351813.2018.1413878

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, *145*, 47–62. https://doi.org/10.1016/j.pragma.2019.01.016

Mondada, L., & Oloff, F. (2011). Gestures in overlap. The situated establishment of speakership. In G. Stam & M. Ishino (Eds.), *Integrating gestures. The*

interdisciplinary nature of gesture (pp. 321–338). John Benjamins. https://doi.org/10.1075/gs.4.29mon

Montes López, G. M. (1992). Is Interaction the Message? The Effect of Democratizing and Non-Democratizing Interaction in Video-Conferencing Small Groups on Social Presence and Quality of Outcome. In U. E. Gattiker (Ed.), *Technology-Mediated Communication* (pp. 187–223). Walter de Gruyter. https://doi.org/10.1515/9783110860542.187

Nevanperä, N., Keränen, A.-M., Ukkola, O., & Laitinen, J. (2015). Effects of Group Counseling Transmitted Through Videoconferencing on Changes in Eating Behaviors. *Journal of Nutrition Education and Behavior*, *47*(6), 555-559.e1. https://doi.org/10.1016/j.jneb.2015.07.004

Nevile, M. (2015). The Embodied Turn in Research on Language and Social Interaction. *Research on Language and Social Interaction*, *48*(2), 121–151. https://doi.org/10.1080/08351813.2015.1025499

Nevile, M., Haddington, P., Heinemann, T., & Rauniomaa, M. (Eds.). (2014). *Interacting with Objects*. John Benjamins Publishing Company. https://doi.org/10.1075/z.186

Nielsen, A. M. R. (2020). Co-constructing the Video Consultation-competent patient. *Social Interaction. Video-Based Studies of Human Sociality*, *3*(3). https://doi.org/10.7146/si.v3i3.122708

Nielsen, M. F. (2013). "Stepping Stones" in Opening and Closing Department Meetings. *Journal of Business Communication*, *50*(1), 34–67. https://doi.org/10.1177/0021943612465182

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI*, *5*. https://doi.org/10.3389/frobt.2018.00114

Oittinen, T. (2018). Multimodal accomplishment of alignment and affiliation in the local space of distant meetings. *Culture and Organization*, *24*(1), 31–53. https://doi.org/10.1080/14759551.2017.1386189

Oittinen, T. (2020). Noticing-prefaced recoveries of the interactional space in a video-mediated business meeting. *Social Interaction. Video-Based Studies of Human Sociality*, *3*(3). https://doi.org/10.7146/si.v3i3.122781

Oittinen, T. (2022). Material and embodied resources in the accomplishment of closings in technology-mediated business meetings. *Pragmatics*, *32*(2), 299–327. https://doi.org/10.1075/prag.19045.oit

Olbertz-Siitonen, M. (2015). Transmission delay in technology-mediated interaction at work. In *PsychNology Journal* (Vol. 13, Issue 3). www.psychnology.org.

Oloff, F. (2013). Embodied withdrawal after overlap resolution. *Journal of Pragmatics*, *46*(1), 139–156. https://doi.org/10.1016/j.pragma.2012.07.005

Olsher, D. (2004a). Talk and gesture: the embodied completion of sequential actions in spoken interaction. In R. Gardner & J. Wagner (Eds.), *Second Language Conversations* (pp. 221–245). Continuum.

Olsher, D. (2004b). The embodied completion of sequences of action. In R. Gardner & G. Stam (Eds.), *Gesture: Second Language Acquisition and Classroom Research* (pp. 436–486). Routledge.

Pappas, Y., & Seale, C. (2009). The opening phase of telemedicine consultations: An analysis of interaction. *Social Science and Medicine*, *68*(7), 1229–1237. https://doi.org/10.1016/j.socscimed.2009.01.011

Pappas, Y., & Seale, C. (2010). The physical examination in telecardiology and televascular consultations: A study using conversation analysis. *Patient Education and Counseling*, *81*(1), 113–118. https://doi.org/10.1016/j.pec.2010.01.005

Pappas, Y., Vseteckova, J., Mastellos, N., Greenfield, G., & Randhawa, G. (2019). Diagnosis and Decision-Making in Telemedicine. *Journal of Patient Experience*, *6*(4), 296–304. https://doi.org/10.1177/2374373518803617

Peräkylä, A. (2004). Reliability and validity in research based on naturally occurring social interaction. In D. Silverman (Ed.), *Qualitative research. Theory, method and practice* (pp. 283–304). Sage Publications.

Potter, J. (1998). Cognition as Context (Whose Cognition?). *Research on Language & Social Interaction*, *31*(1), 29–44. https://doi.org/10.1207/s15327973rlsi3101_2

Raymond, C. W. (2019). Intersubjectivity, Normativity, and Grammar. *Social Psychology Quarterly*, *82*(2), 182–204. https://doi.org/10.1177/0190272519850781

Raymond, G., & Zimmerman, D. H. (2016). Closing matters: Alignment and misalignment in sequence and call closings in institutional interaction. *Discourse Studies*, *18*(6), 716–736. https://doi.org/10.1177/1461445616667141

Reponen, J., Keränen, N., Ruotanen, R., Tuovinen, T., Haverinen, J., & Kangas, M. (2021). *Tieto- ja viestintäteknologian käyttö terveydenhuollossa vuonna 2020 Tilanne ja kehityksen suunta [Use of information and communication technology in healthcare in 2020. Current situation and direction of development]*.

Rintel, S. (2013a). Tech-tied or tongue-tied? Technological versus social trouble in relational video calling. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 3343–3352. https://doi.org/10.1109/HICSS.2013.512

Rintel, S. (2013b). Video Calling in Long-Distance Relationships: The Opportunistic Use of Audio/Video Distortions as a Relational Resource. *The Electronic Journal of Communication / La Revue Electronic de Communication*, *23*(1–2).

Rintel, S. (2013c). Video Calling in Long-Distance Relationships: The Opportunistic Use of Audio/Video Distortions as a Relational Resource. *The Electronic Journal of Communication / La Revue Electronic de Communication*, *23*(1–2).

Rintel, S. (2015). Omnirelevance in technologized interaction: Couples coping with video calling distortions. In *Advances in Membership Categorisation Analysis* (pp. 123–150). SAGE Publications Ltd. https://doi.org/10.4135/9781473917873

Rintel, S. (2023). Video-Mediated Communication. In *Oxford Research Encyclopedia of Communication*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.277

Robinson, D., & Stivers, T. (2001). Achieving activity transitions in physician-patient encounters: From history taking to physical examination. *Human Communication Research*, *27*(2), 253–298. https://doi.org/10.1111/j.1468-2958.2001.tb00782.x

Roos, C. A., Koudenburg, N., & Postmes, T. (2022). Dealing with disagreement: The depolarizing effects of everyday diplomatic skills face-to-face and online. *New Media & Society*, *24*(9), 2153–2176. https://doi.org/10.1177/1461444821993042

Roos, C. A., Postmes, T., & Koudenburg, N. (2020a). The microdynamics of social regulation: Comparing the navigation of disagreements in text-based online and face-to-face discussions. *Group Processes and Intergroup Relations*, *23*(6), 902–917. https://doi.org/10.1177/1368430220935989

Roos, C. A., Postmes, T., & Koudenburg, N. (2020b). The microdynamics of social regulation: Comparing the navigation of disagreements in text-based online and face-to-face discussions. *Group Processes & Intergroup Relations*, *23*(6), 902–917. https://doi.org/10.1177/1368430220935989

Rosenbaun, L., & Licoppe, C. (2019). Showing 'digital' objects in web-based video chats as a collaborative achievement. In R. Harper, R. Watson, & C. Licoppe

(Eds.), *Skyping the Family: Interpersonal video communication and domestic life* (pp. 119–146). https://doi.org/10.1075/bct.103.05ros

Rosetti, D. K., & Surynt, T. J. (1985). Video Teleconferencing and Performance. *Journal of Business Communication*, *22*(4), 25–31. https://doi.org/10.1177/002194368502200404

Ruhleder, K., & Jordan, B. (2001). Co-Constructing Non-Mutual Realities: Delay-Generated Trouble in Distributed Interaction. In *Computer Supported Cooperative Work* (Vol. 10).

Rusk, F., & Pörn, M. (2019). Delay in L2 interaction in video-mediated environments in the context of virtual tandem language learning. *Linguistics and Education*, *50*, 56–70. https://doi.org/10.1016/j.linged.2019.02.003

Sacks, H. (1984). On doing "being ordinary." In *Structures of Social Action. Studies in Conversation Analysis* (pp. 413–429). Cambridge University Press. https://doi.org/10.1017/CBO9780511665868.024

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696–735.

Savenstedt, S., Zingmark, K., Hyden, L.-C., & Brulin, C. (2005). Establishing joint attention in remote talks with the elderly about health: a study of nurses' conversation with elderly persons in teleconsultations. *Scandinavian Journal of Caring Sciences*, *19*(4), 317–324. https://doi.org/10.1111/j.1471-6712.2005.00346.x

Schegloff, E. A. (1987). Analyzing Single Episodes of Interaction: An Exercise in Conversation Analysis. *Social Psychology Quarterly*, *50*(2), 101–114. https://doi.org/10.2307/2786745

Schegloff, E. A. (1992). Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *American Journal of Sociology*, *97*(5), 1295–1345. https://doi.org/10.1086/229903

Schegloff, E. A. (1997). Whose Text? Whose Context? *Discourse & Society*, *8*(2), 165–187. https://doi.org/10.1177/0957926597008002002

Schegloff, E. A. (2007). *Sequence Organization in Interaction. A Primer in Conversation Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, *53*(2), 361–382. https://doi.org/10.2307/413107

Schegloff, E. A., & Sacks, H. (1973). Opening up Closings. *Semiotica*, *8*(4). https://doi.org/10.1515/semi.1973.8.4.289

Schegloff Emanuel A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, *29*(1), 1–63. https://doi.org/10.1017/S0047404500001019

Schmitz, A., Díaz-Martín, A. M., & Yagüe Guillén, M. J. (2022). Modifying UTAUT2 for a cross-country comparison of telemedicine adoption. *Computers in Human Behavior*, *130*, 107183. https://doi.org/10.1016/j.chb.2022.107183

Schnaars, S., & Wymbs, C. (2004). On the persistence of lackluster demand - the history of the video telephone. *Technological Forecasting and Social Change*, *71*(3), 197–216. https://doi.org/10.1016/S0040-1625(02)00410-9

Schuetz, A. (1953). Common-Sense and Scientific Interpretation of Human Action. *Philosophy and Phenomenological Research*, *14*(1), 1–38. https://doi.org/10.2307/2104013

Scollon, S. W. (2004). *Nexus Analysis: Discourse and the Emerging Internet*. Routledge.

Seuren, L. M., & Shaw, S. E. (2022). Using Linguistic Ethnography to Study Video Consultations: A Call to Action and Future Research Agenda. *Qualitative*

Health Research, *32*(5), 800–813. https://doi.org/10.1177/10497323221077297

Seuren, L. M., Wherton, J., Greenhalgh, T., Cameron, D., A'Court, C., & Shaw, S. E. (2020). Physical examinations via video for patients with heart failure: Qualitative study using conversation analysis. *Journal of Medical Internet Research*, *22*(2). https://doi.org/10.2196/16694

Seuren, L. M., Wherton, J., Greenhalgh, T., & Shaw, S. E. (2021). Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics*, *172*, 63–78. https://doi.org/10.1016/j.pragma.2020.11.005

Shaw, S. E., Seuren, L. M., Wherton, J., Cameron, D., A'Court, C., Vijayaraghavan, S., Morris, J., Bhattacharya, S., & Greenhalgh, T. (2020). Video consultations between patients and clinicians in diabetes, cancer, and heart failure services: Linguistic ethnographic study of video-mediated interaction. *Journal of Medical Internet Research*, *22*(5). https://doi.org/10.2196/18378

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Sidnell, J., & Stivers, T. (Eds.). (2012). *The Handbook of Conversation Analysis*. Wiley. https://doi.org/10.1002/9781118325001

Simona, P. D., Wagner, J., & Gonzáles-Martínez, E. (Eds.). (2018). *Longitudinal Studies on the Organization of Social Interaction.* . Palgrave-Macmillan.

Statista. (2022a). *Average selling price of desktop personal computers (PCs) worldwide from 2005 to 2015 (in U.S. dollars).*

Statista. (2022b). *Average selling price of notebooks worldwide from 2005 to 2015 (in U.S. dollars).*

Statistics Finland. (2005). *Purposes of the use of the Internet in spring 2005, per cent of Internet users.*

Statistics Finland. (2021). *Väestön tieto- ja viestintätekniikan käyttö (Use of information and communication technologies by the population).*

Stevanovic, M., & Monzoni, C. (2016). On the hierarchy of interactional resources: Embodied and verbal behavior in the management of joint activities with material objects. *Journal of Pragmatics*, *103*, 15–32. https://doi.org/10.1016/j.pragma.2016.07.004

Stivers, T. (2012). Sequence Organization. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 191–209). Wiley. https://doi.org/10.1002/9781118325001.ch10

Stommel, W. J. P., & Stommel, M. W. J. (2021). Participation of Companions in Video-Mediated Medical Consultations: A Microanalysis. In J. Meredith, D. Giles, & W. Stommel (Eds.), *Analysing Digital Interaction* (pp. 177–203). Springer International Publishing. https://doi.org/10.1007/978-3-030-64922-7_9

Stommel, W. J. P., van Goor, H., & Stommel, M. W. J. (2020). The Impact of Video-Mediated Communication on Closed Wound Assessments in Postoperative Consultations: Conversation Analytical Study. *Journal of Medical Internet Research*, *22*(5), e17791. https://doi.org/10.2196/17791

Stommel, W., Licoppe, C., & Stommel, M. (2020). "Difficult to assess in this manner": An "ineffective" showing sequence in post-surgery video consultation. *Social Interaction. Video-Based Studies of Human Sociality*, *3*(3). https://doi.org/10.7146/si.v3i3.122581

Stommel, W., van Goor, H., & Stommel, M. (2019). Other-attentiveness in video consultation openings: A conversation analysis of video-mediated versus face-to-face consultations. *Journal of Computer-Mediated Communication*, *24*(6), 275–292. https://doi.org/10.1093/jcmc/zmz015

Suh, K. S. (1999). Impact of communication medium on task performance and satisfaction: an examination of media-richness theory. *Information &*

Management, *35*(5), 295–312. https://doi.org/10.1016/S0378-7206(98)00097-4

Suoninen, E. (2001). Harold Garfinkel. Etnometodologia ja sosiaaliset järkeilytavat [Harold Garfinkel. Ethnomethodology and Social Sense-Making as Practical Action]. In V. Hänninen, J. Partanen, & O.-H. Ylijoki (Eds.), *Sosiaalipsykologian suunnannäyttäjiä [Trailblazers of social psychology]* (pp. 365–382). Vastapaino.

ten Have, P. (2007). *Doing Conversation Analysis. A Practical Guide*. Sage Publications Ltd.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Tracy, K. (1998). Analyzing Context: Framing the Discussion. *Research on Language & Social Interaction*, *31*(1), 1–28. https://doi.org/10.1207/s15327973rlsi3101_1

Twain, M. (1898). *From the "London Times" of 1904*.

van Dijk, T. A. (1997). Discourse as interaction in society. In T. A. van Dijk (Ed.), *Discourse as social interaction* (pp. 1–37). Sage.

Venkatesh, V., Thong, J., & Xu, M. J. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157. https://doi.org/10.2307/41410412

Walter, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. In *The Handbook of Interpersonal Communication* (pp. 443–479).

Walther, J. B. (2009). Theories, Boundaries, and All of the Above. *Journal of Computer-Mediated Communication*, *14*(3), 748–752. https://doi.org/10.1111/j.1083-6101.2009.01466.x

Waring, H. Z., Creider, S., Tarpey, T., & Black, R. (2012). A search for specificity in understanding CA and context. *Discourse Studies*, *14*(4), 477–492. https://doi.org/10.1177/1461445611433787

Whittaker, S. (2003). Theories and Methods in Mediated Communication. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 243–286). Lawrence Erlbaum Associates Publishers.

Wilkinson, H. (Ed.). (2001). *Perspectives of people with dementia: research methods and motivations*. Jessica Kingsley Publishers.

Wyatt, S. (2007). Technological Determinism Is Dead; Long Live Technological Determinism. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *The Handbook of Science and Technology Studies* (Third edition, pp. 165–180). The MIT Press.

Yoo, Y., & Alavi, M. (2001). Media and Group Cohesion: Relative Influences on Social Presence, Task Participation, and Group Consensus. *MIS Quarterly*, *25*(3), 371. https://doi.org/10.2307/3250922

PUBLICATIONS

PUBLICATION I

Effects of Transmission Delay on Client Participation in Video-Mediated Group Health Counseling

Ilomäki, S., Ruusuvuori, J. & Laitinen, J.

Qualitative Health Research, 31(12), 2328–2339
https://doi.org/10.1177/10497323211010726

Publication reprinted with the permission of the copyright holders.

Check for updates

Research Article

Effects of Transmission Delay on Client Participation in Video-Mediated Group Health Counseling

Qualitative Health Research
2021, Vol. 31(12) 2328–2339
© The Author(s) 2021

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10497323211010726
journals.sagepub.com/home/qhr


Sakari Ilomäki[1], Johanna Ruusuvuori[1], and Jaana Laitinen[2]

Abstract

In face-to-face group counseling, active client participation contributes to the counseling agenda by a variety of social processes, but little is known about how video mediation shapes client participation. In this article, we use conversation analysis to investigate how transmission delay affects client participation in video-mediated group counseling through shaping the resolution of overlapping talk. Data are video recordings from three video-mediated group health counseling sessions recorded simultaneously in the two participating locations. The delay changes the timing of the overlapping turns and pauses at each end of the mediated counseling, making it difficult to interpret who should take the turn after the overlap. This may pose obstacles to client participation. While mediated counseling services can increase access to services and thus improve client participation at a macro level, transmission delay can pose threats to active client participation at the micro level of interaction.

Keywords

group counseling; client participation; social processes; video-mediated interaction; conversation analysis; Finland; qualitative

Group health counseling offers practical knowledge about healthier living and enables beneficial social processes that are not available in individual counseling, such as social support (Cormack et al., 2018; Frigerio & Montali, 2016; Logren et al., 2019a, 2019b). With digital technologies having become widespread, counseling can be provided online, as through synchronous video-mediated (VM) encounters in which counseling clients and supervisors meet in a video conference. In addition to its effectiveness in, for instance, diabetes care (Laitinen et al., 2010), VM counseling has many other potential benefits, such as increased access to services, especially in peripheral areas, and service provision without physical contact during epidemics like COVID-19. Problems with communication are often pointed out as deficiencies of both text-based counseling (Dilkes-Frayne et al., 2019; Stommel & van der Houwen, 2014) and VM interaction in other health care settings (Dalley et al., 2021), but less is known about how VM shapes the interactional dynamics of participation in group counseling. In this article, we examine how interaction dynamics, possibilities for client participation, and related positive social processes in VM group health counseling for Type 2 diabetes are shaped by transmission delay caused by the technical processes of VM. Type 2 diabetes was selected as it is the most common type of diabetes and its onset can be reduced by lifestyle changes (Gomersall et al., 2011; Ingadottir & Halldorsdottir, 2008; Knutsen et al., 2017; Rosenbek Minet et al., 2011; World Health Organization, 2020).

A central aspect of client participation is engagement in various kinds of activities (Castro et al., 2016; Halabi et al., 2020). Engagement can be examined at the level of both attending to counseling programs or interventions and the social dynamics of participation in counseling interaction. Low health literacy, stigma, gaps between the content offered and client lifeworld and needs, and externalized motivation have all been recognized as barriers to participating in diabetes counseling (e.g., Harris et al., 2019; Kinnafick et al., 2014; Vega et al., 2014). It has been suggested that feelings of stigma, irrelevance of content, and externalized motivation are all particularly suitable to being alleviated by group counseling, which aims at reflection and finding solutions and strategies, together with peers, that fit clients' different life situations (Leong, 2008; Logren et al., 2017b). Furthermore, when people do engage in peer counseling, group-oriented activities and social support have been emphasized

[1]Tampere University, Tampere, Finland
[2]Finnish Institute of Occupational Health, Helsinki, Finland

Corresponding Author:
Sakari Ilomäki, Faculty of Social Sciences, Tampere University, Tampere 33014, Finland.
Email: sakari.ilomaki@tuni.fi

as benefits (Frigerio & Montali, 2016; Mendenhall et al., 2012). In sum, previous research maintains that active participation at the level of interaction dynamics is a prerequisite for effective group counseling: Such positive social dynamics can only emerge when counseling clients interact with one another and the supervisor (e.g., Hughes et al., 2017; Taggart et al., 2012; see Peräkylä & Ruusuvuori, 2007, for client participation as an interactional phenomenon in other health care contexts).

A growing strand of research on interaction processes has emerged to investigate the practices and processes that enable participation in different counseling settings (e.g., Miller & Silverman, 1995). These studies have described both ways in which professionals can encourage participation and how clients are able to initiate an active role in counseling encounters. For example, a counseling format that revolves around instructors' questions and clients' answers can limit client participation to merely answering questions in either individual or group counseling (Karhila et al., 2003; Logren et al., 2017a, 2017b; Poskiparta et al., 1998, 2001; Tiitinen et al., 2018). Furthermore, Karhila et al. (2003) have shown how direct presentation of troublesome issues in clients' lives invites discussion about their lifestyle while merely hinting at possible problems may invite only minimal participation. Logren et al. (2017a) have described how clients can ask questions to produce shifts from leader-driven to member-driven discussion, which then affords increased client participation and exchange of experiences among peers. By responding to one another's experiences, clients can provide social support, offer different perspectives on the topic at hand, and challenge other clients' thinking in constructive ways (Logren et al., 2017b, 2019a, 2019b). When group members engage in these practices, they work collaboratively toward the goals of counseling, sometimes in ways that are not available to the counselor. While these previous studies have provided important knowledge of the practices and processes through which participation is enabled in copresent counseling situations, detailed study of such practices in VM settings and how the VM setting affects them are lacking.

VM settings demand new kinds of interactional competence from professionals because of the limitations of the medium (Dalley et al., 2021). One omnipresent limitation of VM interaction is transmission delay, which is the difference between the point in time when a participant says something and when the other participants hear it, a delay caused by the technical processes of transmission (Schoenenberg et al., 2014). It has been shown that transmission delay can complicate turn-taking and therefore cause interactional dysfluency and a lack of awareness regarding the rights to speak, such as unintended pauses and overlapping talk (Ruhleder & Jordan, 2001; Rusk & Pörn, 2019; Seuren et al., 2020). In their recent study on VM consultations between individual patients and their doctors, Seuren et al. (2020) showed that transmission delay can cause problems with turn-taking and produces unintended interruptions and silences.

In this study, we explore how VM affects the interaction dynamics in group counseling contexts and how such effects shape the emergence of positive social processes of group counseling. We focus on examining how the nutritionist leading the group and the group members resolve overlapping talk (Schegloff, 2000) in VM group counseling and what kind of client participation emerges. Our overall aim is to reveal the potential risks that VM may pose to active participation in counseling and to suggest ways to overcome these problems.

Data and Method

The data are video recordings of three VM health counseling sessions for adults at increased risk for Type 2 diabetes. We conducted a secondary analysis of these data, which were originally used for studying the feasibility of VM counseling, features of behavior change, and motivational factors of health behavior (Laitinen et al., 2010; see also Alahuhta et al., 2011; Korkiakangas et al., 2010; Nevanperä et al., 2015). Nurses and doctors working in occupational and primary health care recruited the clients. The clients were eligible to participate if they were at high risk for Type 2 diabetes, did not have any ongoing serious illness, did not use medication to treat obesity, and were not on a very low-calorie diet. VM counseling was offered to people living in municipalities where face-to-face counseling was not available (from 40 to 91 km away from the regional center). At the time of the data collection (2006), secure videoconferencing systems were not widespread, so the group gathered in a local health care service unit and connected with the nutritionist via video link. There were no professionals in the room during the meetings, but local nurses helped with setting up the system, and group members could contact them if any problems arose. Despite being from the early 2000s, the data afforded a fruitful analysis (see "Discussion" section for further commentary on the limitations of the data).

The group counseling intervention aimed at promoting learning and a process of change. The overall goals of the counseling were to become aware of diabetes-related lifestyle habits and their importance in preventing diabetes, to learn new skills for a healthier life, to normalize eating behavior, and to exercise regularly. The groups met four times fortnightly and a fifth time 6 months after the fourth session. Each session lasted approximately 1½ hours and followed a designated structure of activities and topics. Group activities included mindfulness-like attuning, discussing the homework given in the last sessions (e.g., keeping a food diary) and the topic of the meeting, and playing a board game with diabetes-relevant information and scenario-based dilemmas. Further information about

Table 1. Group Characteristics.

Group Number	Meeting Number	Summary of Meeting Content	Group Members	Nutritionist	Number of Sequences Analyzed ($N = 79$)
1	1/5	Overview, methods, and rules of the group. Reflecting on current weight management and health.	8 females	A	34
2	3/5	Healthy eating.	7 females	B	22
3	5/5	Reflecting on current weight management and health. Imagining future.	4 males	A	23

the groups is presented in Table 1; for more detail, see Laitinen et al. (2010).

The counseling intervention was developed and delivered by the Finnish Institute of Occupational Health. Jaana Laitinen was part of the team that oversaw the intervention and organized the data collection. Sakari Ilomäki and Johanna Ruusuvuori conducted the analysis and did not participate in organizing the intervention. Informed consent to participate in the study was obtained from all participants. The coordinating ethics committee of the hospital district of Helsinki and Uusimaa (Document Number 50/E0/2007) approved the study.

The counseling sessions were video recorded from both the nutritionists' and the groups' perspectives; that is, each encounter was simultaneously recorded at the nutritionist's office and in the room where the group met. This enabled us to study how the interaction unfolded differently when perceived from the nutritionist's perspective in relation to the group's perspective (we refer to this difference in perspectives as *non-mutual interactional realities*; see Ruhleder & Jordan, 2001). These three sessions were selected from a larger corpus of 21 VM sessions, as they were the only ones with recordings from both perspectives. The availability of the two separates but simultaneous perspectives created a unique opportunity to explore how transmission delay can shape interaction. We transcribed the recordings using the Jeffersonian transcription system (Jefferson, 2004a: see the supplementary materials for transcription symbols). In the extracts, we present data from both perspectives and refer to the line numbering of different transcripts by using the letters G (the group's perspective) and N (the nutritionist's perspective). The original data are in Finnish, but the extracts are in English (see the supplementary material for extracts with Finnish data).

We used conversation analysis (CA), an inductive method for examining the practices and structures of interaction from the participants' viewpoint (e.g., Sidnell & Stivers, 2013). In CA, attention is paid to how turns are formulated, how participants manage speaker change, and how they make sense of the ongoing interaction turn by turn. We focused on the instances that started with the nutritionist's counseling-relevant questions and contained overlapping talk (N=79). We selected the nutritionist's questions as the starting point because of their importance in directing counseling interaction (e.g., Logren et al., 2017a; Poskiparta et al., 1998). We analyzed how the participants decide who gets to talk after overlapping talk in these segments (Schegloff, 2000). By focusing on how overlaps are resolved, we aimed to analyze situations that would be of particular importance for the direction that the interaction takes and thus more prone to the effects of technical mediation. We paid special attention to the location of the overlapping turns and the implications that these locations have on interaction (see Drew, 2009). The ways of resolving overlaps were then analyzed from the viewpoint of how the participants' different perspectives affect the overlap resolution. Finally, we reflected on how these differences shaped the clients' possibilities of participating in the counseling discourse.

Before focusing on the effect of VM on counseling interaction, we briefly show how overlap resolution takes place in face-to-face counseling. The group has been discussing a healthy diet (Extract from Logren et al., 2017a, p. 1837). N refers to the nutritionist, while A, B, and C are group members.

Extract 1: Overlap resolution in face-to-face situation

1 A: Isn't banana fa:ttening.
2 (1.2)
3 **N: No it [its**
4 **B:** **[Don't** throw it out right away **[wh(h)en ( )** hah hah hah
5 N: **[Weh heh heh**
6 N: **I[:t,**
7 A: **[One hears** all **kinds[of things.**
8 N: **[It is,**
9 C: Mm
10 N: ↑like a belief that is (.) terribly strong.

Extract 1 exemplifies four issues that are typical of overlapping talk in face-to-face or co-present interaction. First, overlapping talk occurs frequently and usually lasts for only short periods (Sacks et al., 1974). Second, the participants monitor the ongoing stretch of talk to deduce when it is appropriate to take their turn (Goodwin & Goodwin, 2004). For example, in lines 4 and 5 the nurse starts laughing immediately when B's turn is understandable or, in CA-terms, reaches a transition-relevance place (Sacks et al., 1974, p. 704), resulting in an overlap with the last word spoken by B. Third, when facing overlaps, the participants solve them beat by beat—word by word or even syllable by syllable—until only one person is talking (Schegloff, 2000). Fourth, the participants use the timing of the overlap as a hint for speaker selection (Drew, 2009; Jefferson, 2004b). Overlaps occurring after the transition relevance place may result in giving one's turn to the individual initiating the overlap (as the nutritionist interrupts her turns when group members start talking in lines 4 and 7), while overlaps occurring close to the end of another participant's turn are treated as legitimate, not resulting in such ruptures (like the nutritionist's overlapping turns near the closure of group member's turns, in lines 5 and 8). In all, resolving overlapping talk is closely tied to the timing of turns. Next, we show what happens when this timing is complicated by the delay that occurs in VM counseling.

Results

Transmission delay changes the timing of the turns and overlaps and affects the different ways of resolving overlap such as timing of the pauses within the turn. This leads to different interpretations of who should talk after the overlap, which affects the group members' possibilities of participating in three ways. First, unintended overlaps and uncertainty about the next speaker can lead to competition over the turn when activity is about to change (Extract 2). Second, perceived silences in the nutritionist's talk, which in face-to-face conversation would offer the turn to the client, may be treated as pauses within the nutritionist's turn and not as turn offerings (Extract 3). Third, to secure the speakership of the client, the nutritionists can offer the turn explicitly by asking the client to speak (Extract 4).

The Delay Produces Different Orientations to the Progression of Interaction, Which Leads to Competition Over the Turn

Changes in the appearances of turns, overlaps, and ways of resolving overlaps lead to different interpretations of the ongoing activity in each location of the VM encounter. This is exemplified in Extract 2, where delay-generated changes lead to different understandings about the relevance of a topic change and the need for overlap resolution. Before Extract 2, the group was doing an exercise in which they wrote down something they had already succeeded at, and the nutritionist then asked everybody to read their answers. Group member A said that she has switched from candy to dried fruit. The extracts illustrate the discussion during the same time slot, as it appears to the participants in the two different locations, showing first the group's perspective and then the nutritionist's perspective.

Extract 2

Group's Perspective (Group 2)

1	N:	Has it been <hard> to walk past that candy [shelf.]
2	A:	[No i-] (.) it hasn't
3		been hard when one has made that decision that °I'm not going
4	A:	there anymore [(-]-)°
5	?:	[°.hhm°]
6		(0.4)
7	N:	°↓Yes.°
8		(1.1)
9	N:	**.hh[hh] [(Eating) sweets-]**
10	A:	**[Bu]t th[at craving- (.) ↑CRAVING FO]R SWEETS** where does it
11		come from.

((Continues))

Nutritionist's Perspective

1	N:	Has it been HARD to walk past that candy shelf.
2		(0.7)
3	A:	No i- (.) ↑it hasn't been hard when one has made that decision

4 A: °(that I'm not?)°
5 (0.3)
6 N: Yes:.
7 (1.3)
8 N: **.hhhh EATING swe[ets-]**
9 A: **[(-ut that cra]ving,** (.) craving for sweets >where
10 does it< come from.

((Continues))

The extracts show how transmission delay causes the occurrence of overlapping talk to be situated differently in each location. While the overlap does not happen until in lines 9–10G/8–9N, the preceding interaction lays the ground for different interpretations of the overlap. From the *group's perspective*, it seems that following the nutritionist's question and A's answer, group members have an opportunity to take their turn (line 10G), as there is a gap of 1.1 seconds (line 8G/line 7N) and the nutritionist has not yet started her turn. But from the *nutritionist's perspective*, it seems that the interactants are ready to move on to the next activity (Heritage & Sefi, 1992), so she starts her turn (line 8N). When she hears group member A, who seems to have started later, or in CA terms in *interjacent overlap* (line 9N, Drew, 2009, 88–91), she makes the choice to abandon her turn to implicitly offer it to group member A. This temporally caused confusion in terms of the progression of the discussion results in the group member starting to compete over the turn with the nutritionist. This is shown in her continuing to talk in spite of hearing the nutritionist also starting to talk, and in speaking louder to maintain the turn (lines 9–10G/8–9N). From the nutritionist's perspective, however, as the group member starts clearly later than she does, she is not competing over the turn but gives it to the group member (who has seemingly interrupted her, line 8–9N). A's competition over the turn and the nutritionist's dropping out enable A to participate by posing her topic-relevant question, thus producing a shift from leader-driven to member-driven discussion (Logren et al., 2017a).

Due to delay-generated changes, the participants perceive the overlapping talk and its implications for interaction differently: Competition over the turn appears relevant to A, while competition is not relevant for the nutritionist. In this case, the delay and the changes it generated in the position of the overlap eventually helped A to take the turn despite the lack of shared perspective on interaction, but as the following extracts show, these changes can also prevent client participation.

Implicit Turn-Offering Is Perceived as a Within-Turn Pause

When the participants implicitly offer the turn to one another by pausing, the transmission delay can alter the location of the pause and make it appear as a pause within the speaker's turn. This kind of pause does not unambiguously afford speaker change and participation (Sacks et al., 1974). Furthermore, as was exemplified with the nutritionist abandoning her turn in the previous extract, when participants find themselves in interjacent overlap, the position of the overlap hints that there are some reasons for the other participant to continue and for the current speaker to stop (Drew, 2009, pp. 88–91). These two dynamics are at play in the following extract, where the nutritionist offers the turn to group member C by pausing (line 34N), but the delay leads C to perceive the nutritionist's turn as interjacent overlap and the turn-offering pause as occurring within the nutritionist's ongoing turn (line 34G). Due to the changes in timing, the nutritionist's turn-offering pause is apparent only in the second transcript (line 34N), not in the first (line 34G). The group is discussing their challenges regarding eating.

Extract 3

Group's Perspective (Group 2)

1 N: .hhhhhhh Well how about A?
2 (0.3)
3 N: What do you have.=
4 A: =Well I have a point exact[ly about this?] (0.2) kr kr krhm (0.8)
5 ?: [.nffft]
6 A: this evening- evening tiredness and that when I take supper so,
((16 lines omitted. A describes her difficulties in managing eating in the evenings. B responds by telling how she prepares the number of sandwiches she is about to eat and then puts the ingredients away to avoid overeating.))

22 B: .hhh Heh[heh£] [.hhh]
23 A: [°Yeah.°] So [it's like] that, (.) my problem that I eat
24 more then >↑in the evening<=I don't like during the day I don't
25 have problems and neither then still when I come from work so? .hhh
26 there is no problem with eating but the supper is kind of a, (0.3)
27 A: < stumbling [°block.°>]
28 B: [°(I have <exactly] the same.) (--)°
29 (0.4)
30 N: Ye:s.=
31 B: =°In principle.°
32 (0.4)
33 C: **Yeah and then [that:,]**
34 N: **[Rea]lly good, (0.7) really good suggestion,**

((Continues)) ((N refers to B's suggestion which is not shown.))

Nutritionist's Perspective

1 N: .hhhhhhh Well how about A?
2 (0.3)
3 N: What do you have.
4 (1.3)
5 A: Well I have a point exact[ly about this?] (0.2) kr kr krhm (0.8)
6 ?: [.nffft]
7 A: this evening- evening tiredness and that when I take supper so,
((16 lines omitted.))
22 B: .hhh heh[heh£] [.hhh]
23 A: [°Yeah.°] So [it's like] that, (.) my problem that I eat
24 more then >↑in the evening<=I don't like during the day I don't
25 have problems and neither then still when I come from work so?
26 .hhh (0.2) there is no problem with eating but the supper is kind
27 of a, (0.3) stumbling °block.°
28 (1.1)
29 N: Ye:s.
30 (1.0)
31 B: °In principle.°=
32 N: **=Reall[y good m?]**
33 C: **[Yeah (---)]**
34 **(0.7)**
35 N: **Really good suggestion,**

((Continues))

Again, what happens before the overlap (lines 33–34G/32–35N) lays the ground for the different interpretations of participation. After the nutritionist's question and A's answer, A and group member B start discussing a possible solution to A's problem. A's repetition of the central point of her answer (lines 23–27G/23–27N; see, for example, Barnes, 2007), B's statement of sharing the same experience (line 28G/not hearable by the nutritionist; Logren et al., 2019a), and brief turns by the nutritionist and B (lines 30–31G/29, 31N; Schegloff, 2007) all imply that they are finished with the previous topic and ready move on in the discussion. From the *nutritionist's perspective*, she starts to evaluate the idea, which was discussed earlier, of how to avoid evening snacking (line 32N) but stops immediately when she hears C's voice and implicitly offers the turn to her by pausing (lines 33–34N). As C does not appear to continue, the nutritionist finishes the evaluation and repeats the advice provided by the group members (line 35N). From the *group's perspective*, the implicit turn offering appears a bit differently due to the alterations that the delay causes on the location of the overlap and the nutritionist's pause. The third group member, C, starts adding a new perspective to the topic ("yeah and then that," line 33G) but finds herself in an interjacent overlap with the nutritionist (line 34G). Like the nutritionist in Extract 2, C takes this overlap to imply that the nutritionist should take the turn and stops talking. From the group's perspective, the pause (with which the nutritionist, from her

perspective, implicitly offered a turn to C) appears as a pause within the nutritionist's turn, during which speaker change is not relevant.

In Extract 3, the interactional implications of the overlapping talk and turn-offering pause appear to be different in the different locations because of the non-mutual interactional realities. From the *nutritionist's perspective*, she has provided C with a reasonable amount of time to take the turn, but from the *group's perspective*, the pause that was supposed to offer the turn occurred in a location that did not make the speaker change relevant. The comparison of the two perspectives reveals a central shortcoming of implicitly offering the turn in overlap situations during VM counseling: Technological mediation produces a situation where implicit turn offerings may appear as pauses within the leader's turn and thus be inadequate to secure client participation. Furthermore, unlike Extract 2, where A meets the nutritionist's overlapping talk in a similar interactional environment and competes over the turn, C does not engage in competition over the turn here. This might show her orientation to herself as not a ratified speaker at this point, as she is not the one to whom the question was posed (as A was in Extract 2). Whatever the reason, since C does not engage in competition and the nutritionist's implicit turn offering is ineffective to secure the turn for C, her participation in the discussion by responding to another group member's challenging experience (Logren et al., 2019a) and related positive social contribution are not actualized.

Explicit Turn Offering Secures Client Participation

When the nutritionist recognizes that her implicit turn offering has fallen short, the turn can be offered more explicitly with minimal offerings (such as "yeah" with a rising intonation and other continuers; Müller, 1996) or more direct offerings like "go ahead." As in co-present interaction, explicitly offering the turn secures client participation as it indicates that the one who offered the turn has a right—and in some sense an obligation—to take the turn (Sacks et al., 1974). This is exemplified in Extract 4. The group is playing a boardgame that involves information about healthy lifestyle, questions about diabetes, and scenario-based dilemmas to solve. The group member A has been given the task of naming three good practices of stress management and she has suggested physical exercise, handicrafts, and reading.

Extract 4

Group's Perspective (Group 2)

N: Really good tips.
N: You got three points.
N: [Does somebody want] to add something more, (.) to this issue.
?: [↑mm:?]
(4.1)
B: **Well=nothing more than [that-]**
N: **[As we know] stress affect- (.) ↑yeah?**
B: **Th[at- (.) e-] I was thinking that when=it feels really**
N: **[>Go ahead.<]**
B: >stressful,< (0.4) I at least try to come up with something
to do >that I< like.
(1.0) ((Continues))

Nutritionist's Perspective

N: Really good tips.
N: You got three points.
N: Does somebody want to [add some]thing more, (.) to this issue.
?: [(-)]
(4.7)
N: **As we kn[ow stress affect-]**
B: **[Well=nothing more than,]**
(.)
N: **↑yeah?**
(0.5)
N: >**Go ahead.**<
B: **That- (.) e- I was thinking that when=it feels really**
B: >**stressful,**< (0.4) I at least try to come up with something to do
B: >that I< like.

((Continues))

After the nutritionist has commented on A's ideas for stress management, she checks whether anybody has something to add (line 3G/3N). The question is not directed to a specific group member and not answering would be taken as the group's readiness to move on to the next activity. From the *nutritionist's perspective*, the long pause (line 5N) hints that nobody has anything to add, and she moves on to providing information about the interconnectedness of stress and eating (line 6N). From the *group's perspective*, however, group member B starts her answer after a lengthy pause (line 3G) but finds herself interrupted by the nutritionist's turn (line 4G). Like group member C in Extract 3, B ceases to talk when facing this kind of interjacent overlap and does not take the turn during the pause in the nutritionist's turn (line 7G), since, from the group's perspective, the turn-offering pause appears as a pause within the nutritionist's turn. But, unlike in Extract 3, the nutritionist explicitly asks B to continue ("yeah," lines 7G/9N). From the *group's perspective*, B restarts her answer immediately after the explicit offer to take the turn, while from the *nutritionist's perspective* the nutritionist experiences a half-second pause and further expands the offering with "go ahead" (lines 8/10–11 N). By building on the implicit turn offering with an explicit offer, the nutritionist enables B to take the turn and thus participate by a self-reflective turn (Logren et al., 2017b). In her response, B both expresses sharing the experience of stress management and reframes A's list of potential practices of stress management at a more general level, thus helping others to identify practices that could fit their lifestyle and situation. Had the nutritionist not been sensitive to B's turn and offered the turn explicitly, reflection and reframing would likely not have occurred.

Discussion

We have shown how transmission delay shapes perceived rights to take the turn and thus client participation in VM group counseling, where the group meets in a single location to connect with the nutritionist via video link. The transmission delay caused by VM produced changes in the timing of the overlaps and the ways of resolving overlapping talk resulted in confusion in terms of speaker choice. Due to the delay, the group and the nutritionist had slightly different perspectives concerning the ongoing action; they had *non-mutual interactional realities* (Ruhleder & Jordan, 2001) that led to different interpretations of the progression of the discussion and which speaker rights would be timely in the moments of delay. This limited the group members' possibility to participate. To secure their participation, the group members competed over the turn, especially in interactional intersections where the topic or broader line of action was changing and where the turn allocation was thus not explicitly determined. The nutritionists worked toward securing client participation by offering the turn implicitly by pausing or explicitly asking the client to continue. As Extract 3 shows, implicit turn offerings often proved to be inadequate to secure a speaker change and client participation. To overcome this problem, nutritionists could strive to offer turns explicitly. Compared with implicit turn offerings, explicit offerings were more likely to secure the client speakership and participation.

The analysis has revealed that, while the effect of transmission delay is apparent for the analyst and the readers of this article—who have data from both perspectives—it may have remained hidden from the participants. Technological aspects were not mentioned in meta-talk when facing overlapping talk but only when some of the participants could not produce a relevant next action (e.g., the nutritionist could not evaluate the client's inaudible answer). These cases were rare, and the delay was never explicitly mentioned as a source of trouble. In addition, as implicit turn offering recurrently failed to secure a speaker change, it is plausible to suggest that the participants were not aware that the delay caused such drastic barriers to participation.

Our findings align with earlier research that has highlighted communication difficulties as the central problem of technologically mediated counseling (Dilkes-Frayne et al., 2019; Stommel & van der Houwen, 2014), VM interaction in health care (Dalley et al., 2021), and VM interaction in general (Ruhleder & Jordan, 2001; Rusk & Pörn, 2019). We add to this knowledge by showing that these difficulties may be difficult to notice and name and therefore shape the interaction even when they remain hidden. Our findings are in line with Seuren et al. (2020) who showed that delay interferes with regular turn-taking in VM consultations. We expand on this finding by showing that these interferences have consequences for client participation and beneficial social processes of counseling, such as peer support. To our knowledge, this topic has not been studied in any kind of group setting before.

Earlier research on interaction dynamics in counseling settings has demonstrated how clients' possibilities are shaped on at least two intertwined levels: The lines of activities put forward in the interaction and how single turns of talk are designed to encourage different kinds of participation (Karhila et al., 2003; Logren et al., 2017a, 2017b, 2019a, 2019b; Poskiparta et al., 1998, 2001; Tiitinen et al., 2018). Our analysis showed that participation dynamics is also managed at the level of speakership after overlaps. Since resolving overlaps demands meticulous monitoring of the timing of overlapping turns and pauses, this level is heavily influenced by technological mediation, potentially even more than the other two. Furthermore, these ruptures, in participation, occur in

relation to activities that are central to fulfilling the positive social processes in counseling such as steering the direction of interaction to member-led discussion (Extract 2; Logren et al., 2017a) and responding to other members' turns in constructive ways (Extracts 3 and 4; Logren et al., 2019a). In all, while VM counseling services can increase access to services and thus improve client participation at the macro level, transmission delay caused by the technology involved can pose threats to active client participation and the associated positive social dynamics at the micro-level of interaction.

The existing qualitative research on patient and client participation has predominantly concentrated on experiences of participation (e.g., Gomersall et al., 2011). Considering participation as an ongoing interactional achievement (Collins et al., 2007; Goodwin & Goodwin, 2004) offers important insights into the conceptualization and study of participation by providing a more nuanced understanding of how engagement in counseling activities evolves in situ. The interactional perspective contributes to understanding how the working relationship and engagement in activities that contribute to the goals of the counseling unfold in discourse and how technological mediation shapes the possibilities to do so. These two perspectives—client experience and interaction dynamics—are not conflicting but complementary (De Jaeger et al., 2016). Connecting the study of interaction dynamics and clients' and professionals' experiences could be strengthened in the future to form a more holistic understanding of participation in VM counseling.

To understand the ways in which technological mediation shapes interaction dynamics in different health care settings more deeply, further research in at least two areas is needed. First is how the features of the specific health setting shape ways of participating: for example, is participation in VM counseling, which aims at reflection and problem solving, different from tele-consultations, which aim at diagnosis and decision-making about the treatment, and are there differences between patient groups and illnesses? Second is how participation is managed in technologically mediated interactions with different numbers of participants and sites of interaction. As troubles for client participation arose even with only two sites of interaction, we hypothesize that managing participation in settings with three or more sites would be even more complex.

The use of VM counseling and other health services is likely to continue to grow in the future. Urbanization and the decay of peripheries, the ideals of aging in place and, perhaps most pressingly, recurring pandemics that demand refraining from physical contact will increase the use of VM services. Simultaneously, Type 2 diabetes and other chronic conditions that could be alleviated with proper lifestyles are becoming more common globally. If high-quality care and counseling are to be offered through VM, it is crucial to increase our understanding of how different features of the technology affect participation, working relationships, and engagement.

Conclusion

Strengths and Limitations of the Study

The small dataset of this study was based on necessity: We used all the data that the project had gathered from two perspectives of action. This is a clear strength, as our findings regarding the non-mutual interactional realities would have been impossible to make without this kind of rich data. While the small dataset could limit the generalizability of our findings, CA of institutional encounters aims not only at finding generalizable practices but also at describing what kind of practices are possible in a specific context (Peräkylä, 2004). In this research model, generalizability stems from comparing the findings from different contexts, which in our case means earlier findings from the counseling context (Karhila et al., 2003; Logren et al., 2017a, 2017b, 2019a, 2019b; Poskiparta et al., 1998, 2001) and research on VM interaction in different contexts (Ruhleder & Jordan, 2001; Rusk & Pörn, 2019; Seuren et al., 2020). Moreover, by using the concept of participation (Castro et al., 2016; Collins et al., 2007; Halabi et al., 2020) as a basis for theoretical comparison, we have been able to show how delay, which is a general feature of any kind of VM interaction, becomes consequential specifically in counseling contexts. This has enabled us to participate in discussions beyond our empirical cases. The fact that the data were gathered approximately 15 years ago could be considered as a limitation to its validity. However, despite 15 years of technological development and improved tele-health competencies, the delay-generated problems that we described are present in more recently gathered data as well (see Seuren et al., 2020), thus justifying the secondary analysis of data.

Implications for Practice

Video mediation is a challenging environment for nutritionists and other professionals to lead peer groups. In general, we hope to have sparked reflection among health care professionals about the effects that VM has on interaction and client participation. As the analysis has shown, the methods used in face-to-face interaction for securing client participation after overlaps may be inadequate in a VM setting. To secure client participation and positive social processes, the nutritionists in our data offered explicitly the turn to participants. Even though offering the turn explicitly can, in some cases, result in new overlaps or awkward feeling, our data suggest that it is more reliable for ensuring client participation after overlaps

than implicit turn offerings. Furthermore, meta-talk about the effects of technical mediation on the interaction process and working relationship might alleviate the possible negative feelings associated with overlaps and confusion concerning speaker choice. Our findings about the fine details of interaction and their relationship with client participation might have remained unnoticed if the participants were merely interviewed about the quality of interaction. Therefore, we encourage practitioners and researchers to engage in studies that draw data from real-life health care interactions and pay attention primarily to interaction dynamics as an endogenous phenomenon.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclose receipt of the following financial support for the research, authorship, and publication of this article: The Juho Vaino Foundation, and the Strategic Research Council at the Academy of Finland (Grant Number 303430).

ORCID iD


Sakari Ilomäki https://orcid.org/0000-0003-1791-7117

Supplemental Material

Supplemental Material for this article is available online at journals.sagepub.com/home/qhr. Please enter the article's DOI, located at the top right hand corner of this article in the search bar, and click on the file folder icon to view.

References

Alahuhta, M. A., Korkiakangas, E. E., Keränen, A.-M., Kungäs, H. A., & Laitinen, J. H. (2011). Using pictures as vignettes to assess stages of change in weight management. *Scandinavian Journal of Public Health*, *39*(4), 403–409. https://doi.org/10.1177%2F1403494810395822

Barnes, R. (2007). Formulations and the facilitation of common agreement in meetings talk. *Text & Talk*, *27*(3), 273–296. https://doi.org/10.1515/TEXT.2007.011

Castro, E. M., Van Regenmortel, T., Vanhaecht, K., Sermeus, W., & Van Hecke, A. (2016). Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review. *Patient Education and Counseling*, *99*, 1923–1939. https://doi.org/10.1016/j.pec.2016.07.026

Collins, S., Britten, N., Ruusuvuori, J., & Thompson, A. (Eds.). (2007). *Patient participation in health care consultations: Qualitative perspectives*. Open University Press.

Cormack, D., Jones, F. W., & Maltby, M. (2018). A "collective effort to make yourself feel better": The group process in mindfulness-based interventions. *Qualitative Health Research*, *28*(1), 3–15. https://doi.org/10.1177/1049732317733

Dalley, D., Rahman, R., & Ivaldi, A. (2021). Health care professionals' and patients' management of the interactional practices in telemedicine videoconferencing: A conversation analytic and discursive systematic review. *Qualitative Health Research*, *31*(4), 804–814. https://doi.org/10.1177/1049732320942346

De Jaeger, H., Peräkylä, A., & Stevanovic, M. (2016). The co-creation of meaningful action: Bridging enaction and interactional sociology. *Philosophical Transactions of the Royal Society. Biological Sciences*, *371*(1693), 1–10. https://doi.org/10.1098/rstb.2015.0378

Dilkes-Frayne, E., Savic, M., Carter, A., Kokanović, R., & Lubman, D. I. (2019). Going online: The affordances of online counseling for families affected by alcohol and other drug issues. *Qualitative Health Research*, *29*(14), 2010–2022. https://doi.org/10.1177/1049732319838231

Drew, P. (2009). "Quit talking while I'm interrupting": A comparison between positions of overlap onset in conversation. In M. Haakana, M. Laakso, & J. Lindström (Eds.), *Talk in interaction: Comparative dimensions* (pp. 70–93). Finnish Literature Society.

Frigerio, A., & Montali, L. (2016). An ethnographic-discursive approach to parental self-help groups: The case of ADHD. *Qualitative Health Research*, *26*(7), 935–950. https://doi.org/10.1177/1049732315586553

Gomersall, T., Madill, A., & Summers, L. K. M. (2011). A metasynthesis of the self-management of type 2 diabetes. *Qualitative Health Research*, *21*(6), 853–871. https://doi.org/10.1177/1049732311402096

Goodwin, C., & Goodwin, M. H. (2004). Participation. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 222–243). Basil Blackwell.

Halabi, I. O., Scholtes, B., Voz, B., Gillain, N., Durieux, N., Odero, A., . . . The APPS Interreg Group. (2020). "Patient participation" and related concepts: A scoping review on their dimensional composition. *Patient Education and Counseling*, *103*, 5–14. https://doi.org/10.1016/j.pec.2019.08.001

Harris, S., Miller, A., Amiel, S., & Mulnier, H. (2019). Characterization of adults with type 1 diabetes not attending self-management education courses: The barriers to uptake of type 1 diabetes education (BUD1E) study. *Qualitative Health Research*, *29*(8), 1174–1185. https://doi.org/10.1177/104973231882371

Heritage, J., & Sefi, S. (1992). Dilemmas of advice: Aspects of the delivery and reception of advice in interactions between health visitors and first-time mothers. In P. Drew & J. Heritage (Eds.), *Talk at work* (pp. 359–417). Cambridge University Press.

Hughes, S., Lewis, S., Willis, K., Rogers, A., Wyke, S., & Smith, L. (2017). The experience of facilitators and participants of long-term condition self-management group programmes: A qualitative synthesis. *Patient Education and Counseling*, *100*, 2244–2254. https://doi.org/10.1016/j.pec.2017.06.035

Ingadottir, B., & Halldorsdottir, S. (2008). To discipline a "dog": The essential structure of mastering diabetes.

Qualitative Health Research, *18*(5), 606–609. https://doi.org/10.1177/1049732308316346

Jefferson, G. (2004a). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–23). John Benjamins.

Jefferson, G. (2004b). A sketch of some orderly aspects of overlap in natural conversation. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 43–59). John Benjamins.

Karhila, P., Kettunen, T., Poskiparta, M., & Liimatainen, L. (2003). Negotiation in type 2 diabetes counseling: From problem recognition to mutual acceptance during lifestyle counseling. *Qualitative Health Research*, *13*(9), 1205–1224. https://doi.org/10.1177/1049732303257153

Kinnafick, F.-E., Thøgersen-Ntoumani, C., & Duda, J. L. (2014). Physical activity adoption to adherence, lapse, and dropout: A self-determination theory perspective. *Qualitative Health Research*, *24*(5), 706–718. https://doi.org/10.1177/1049732314528811

Knutsen, I. R., Foss, I. C., Todorova, E., Roukova, P., Kennedy, A., Portilla, M. C., . . . Rogers, A. (2017). Negotiating diet in networks: A cross-European study of the experiences of managing type 2 diabetes. *Qualitative Health Research*, *27*(3), 299–310. https://doi.org/10.1177/1049732315610318

Korkiakangas, E., Alahuhta, M., Husman, P., Keinänen-Kiukaanniemi, S., Taanila, A., & Laitinen, J. (2010). Motivators and barriers to exercise among adults with a high risk of type 2 diabetes: A qualitative study. *Scandinavian Journal of Caring Sciences*, *25*(1), 62–69. https://doi.org/10.1111/j.1471-6712-2010-00791.x

Laitinen, J., Korkiakangas, E., Alahuhta, M., Keinänen-Kiukaanniemi, S., Rajala, U., Timonen, O., . . . Olkkonen, S. (2010). Feasibility of videoconferencing in lifestyle group counseling. *International Journal of Circumpolar Health*, *69*(5), 500–511. https://doi.org/10.3402/ijch.v69i5.17688

Leong, F. L. (2008). *Encyclopedia of counseling. Vol. 1: Changes and challenges for counseling in the 21st century*. SAGE.

Logren, A., Ruusuvuori, J., & Laitinen, J. (2017a). Group members' questions shape participation in health counselling and health education. *Patient Education and Counseling*, *100*(10), 1828–1841. https://doi.org/10.1016/j.pec.2017.05.003

Logren, A., Ruusuvuori, J., & Laitinen, J. (2017b). Self-reflective talk in group counselling. *Discourse Studies*, *19*(4), 422–440. https://doi.org/10.1177/1461445617706771

Logren, A., Ruusuvuori, J., & Laitinen, J. (2019a). Peer responses to self-disclosures in group counselling. *Text & Talk*, *39*(5), 613–647. https://doi.org/10.1515/text-2019-2042

Logren, A., Ruusuvuori, J., & Laitinen, J. (2019b). Stories of change: Comparative time-framed experience telling in health promotion group discussions. *Qualitative Health Research*, *30*(2), 279–292. https://doi.org/10.1177/1049732319877858

Mendenhall, T. J., Seal, K. L., GreenCow, B. A., LittleWalker, K. N., & BwornOwl, S. A. (2012). The Family Education Diabetes Series: Improving health in an urban-dwelling American Indian community. *Qualitative Health Research*, *22*(11), 1524–1534. https://doi.org/10.1177/1049732312457469

Miller, G., & Silverman, D. (1995). Troubles talk and counselling discourse: A comparative study. *Sociological Quarterly*, *36*(4), 725–747. https://doi.org/10.1111/j.1533-8525.1995.tb00462.x

Müller, F. E. (1996). Affiliating and disaffiliating with continuers: Prosodic aspects of recipiency. In E. Coupler-Kuhlen & M. Selting (Eds.), *Prosody in conversation: Interactional studies* (pp. 131–176). Cambridge University Press.

Nevanperä, N., Keränen, A.-M., Ukkola, O., & Laitinen, J. (2015). Effects of group counseling transmitted through videoconferencing on changes in eating behaviors. *Journal of Nutrition Education and Behavior*, *47*(6), 550–559.e1. https://doi.org/10.1016/j.jneb.2015.07.004

Peräkylä, A. (2004). Reliability and validity in research based on naturally occurring social interaction. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (pp. 283–304). SAGE.

Peräkylä, A., & Ruusuvuori, J. (2007). Components of participation in health care consultations. In S. Collins, N. Britten, J. Ruusuvuori, & A. Thompson (Eds.), *Patient participation in health care consultations. Qualitative perspectives* (pp. 167–155). Open University Press.

Poskiparta, M., Kettunen, T., & Liimatainen, L. (1998). Reflective questions in health counseling. *Qualitative Health Research*, *8*(5), 682–693. https://doi.org/10.1177/104973239800800508

Poskiparta, M., Liimatainen, L., Kettunen, T., & Karhila, P. (2001). From nurse-centered to empowermental health counseling. *Qualitative Health Research*, *45*(1), 69–79. https://doi.org/10.1016/s0738-3991(01)00140-9

Rosenbek Minet, L. K., Lønvig, E.-M., Henriksen, J. E., & Wagner, L. (2011). The experience of living with diabetes following a self-management program based on motivational interviewing. *Qualitative Health Research*, *21*(8), 1115–1126. https://doi.org/10.1177/1049732311405066

Ruhleder, K., & Jordan, B. (2001). Co-constructing non-mutual realities: Delay-generated trouble in distributed interaction. *Computer Supported Cooperative Work*, *10*, 113–138. https://doi.org/10.1023/A:1011243905593

Rusk, F., & Pörn, M. (2019). Delay in L2 interaction in video-mediated environments in the context of virtual tandem language learning. *Linguistics and Education*, *50*, 56–70. https://doi.org/10.1016/j.linged.2019.02.003

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735. https://doi.org/10.2307/412243

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, *29*, 1–63. https://doi.org/10.1017/S0047404500001019

Schegloff, E. A. (2007). *Sequence organization in interaction. Vol. 1: A primer in conversation analysis*. Cambridge University Press.

Schoenenberg, K., Raake, A., Egger, S., & Schatz, R. (2014). On interaction behaviour in telephone conversations under transmission delay. *Speech Communication*, *63–64*, 1–14. https://doi.org/10.1016/j.specom.2014.04.005

Seuren, L. M., Wherton, J., Greenhalgh, T., & Shaw, S. E. (2020). Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics*, *172*, 63–78. https://doi.org/10.1016/j.pragma.2020.11.005

Sidnell, J., & Stivers, T. (Eds.). (2013). *The handbook of conversation analysis*. Wiley-Blackwell.

Stommel, W., & van der Houwen, F. (2014). Complaining and the management of face in online counseling. *Qualitative Health Research*, *24*(2), 183–193. https://doi.org/10.1177/1049732313519706

Taggart, J., Williams, A., Dennis, S., Newall, A., Shortus, T., Zwar, N., & Harris, M. F. (2012). A systematic review of interventions in primary care to improve health literacy for chronic disease behavioral risk factors. *BMC Family Practice*, *13*, Article 49. https://doi.org/10.1186/1471-2296-13-49

Tiitinen, S., Weiste, E., Ruusuvuori, J., & Laitinen, J. (2018). Ryhmänohjauksen vuorovaikutusmekanismit vertaistuen kokemusten taustalla [Interactional mechanisms explaining the experiences of peer support in group counselling]. *Psykologia*, *53*, 345–363.

Vega, G., Vega, M., Alajarín, L. I., Portero, B., Ruíz, T., . . . Conde, F. (2014). Profile of the patient who refuses to participate in the expert patients program. *Qualitative Health Research*, *24*(6), 846–859. https://doi.org/10.1177/1049732314535850

World Health Organization. (2020). *Diabetes*. https://www.who.int/news-room/fact-sheets/detail/diabetes

Author Biographies

Sakari Ilomäki is doctoral researcer at the Tampere University. His research concentrates on social interaction in video-mediated health and social care, namely group counseling, telehomecare and tele-consultations.

Johanna Ruusuvuori is professor of Social Psychology at the Tampere University. Her research interests include social interaction in institutional settings, such as healthcare, working life, and lifestyle counseling, work-life participation, and emotion in interaction.

Jaana Laitinen is adjunct professor in Nutrition and in Public Health at the Finnish Institute of Occupational Health. She specializes in intervention studies aiming to promote healthy lifestyles and work ability.

Supplementary material 2: Data extracts with original Finnish transcripts and word-by-word glossing

Extract 1

1 A: **eikö banaani liho:ta.**
not+Q banana fatten
Isn't banana fa:ttening.
2 (1.2)
3 N: **ei se [sen**
No it [its
4 B: **[elä nyt heti tyrmää [k(h)u ( )hah hah hah**
don't+you now immediatelly knock+out when
[Don't throw it out right away [wh(h)en ( ) hah hah hah
5 N: [Weh heh heh
6 N: **s[e:,**
I[:t,
7 A: **[sitä kuulee kaiken[laista.**
it 0+hears all+kinds+of+things
[One hears all kinds[of things.
8 N: **[se on,**
[It is,
9 C: Mm
10 N: **↑semmonen hirveen vahvassa, (.) oleva usko.**
the+kind+of, terribly strongly existing belief
↑like a belief that is, (.) terribly strong.

Extract 2

Group's perspective (Group 2)

1 N: **onko ollu <vaikeeta> kävellä sen karkkihyllyn o[hi.]**

has+Q been hard to+walk that candy+shelf past

has it been <hard> to walk past that candy [shelf.]

2 A: ***[ei s-] (.) ei oo***

no not has

[no i-] (.) it hasn't

3 **ollu vaikeeta kun sen tehny päätöksen että °ei ny (ruukaa)°**

been hard when that done decision that not now (have+habit+of)

been hard when one has made that decision that °I'm not going

4 A: ***[(-]-)°***

there anymore [(-]-)°

5 ?: [°.hhm°]

6 (0.4)

7 N: **°↓*joo.*°**

yes

°↓yes.°

8 (1.1)

9 N:**.hh[hh] [makean (syöntihä-)]**

sweet's eating

.hh[hh] [(eating) sweets-]

10 A: **[mut]ta s[e makkeen- (.) ↑MAKKEEN HI]MO NI mistä se voipi**

but that sweet's sweet's graving like where it can

[bu]t th[at craving- (.) ↑CRAVING FO]R SWEETS where does it

11 **johtua.**

be+due

come from.

((Continues))

Nutritionist's perspective

1 N: **onko ollu VAIKEETA kävellä sen karkkihyllyn ohi.**

has+Q been hard to+walk that candy+shelf past

has it been HARD to walk past that candy shelf.

2 (0.7)

3 A: **ei s- (.) ↑ei oo ollu vaikeeta kun sen on tehny pä<u>ä</u>töksen**

no not has been hard when that has done decision

No i- (.) ↑it hasn't been hard when one has made that d<u>e</u>cision

4 A: °(that I'm not?)°

°(ettei?)°

that+not

5 (0.3)

6 N: **joo:.**

yes

yes:.

7 (1.3)

8 N: **.hhh MAKEAN syön[tihä-]**

sweet's eating

.hhhh EATING swe[ets-]

9 L: **[(-ta se ma]kee,) (.) makeen himo nii >mistä se<**

that sweet's sweet's graving like where it

[(-ut that cra]ving, (.) craving for sweets >wh<u>e</u>re

10 **voipi johtua.**

can be+due

does it< come from.

((Continues))

Extract 3

Group's perspective (Group 2)

1 N: **.hhhhhhh no entäs A?**

well how+about A

.hhhhhhh well how about A?

2 (0.3)

3 N: **mitäs sulla ois.=**

what you would+have

what do you have.=

4 A: **=no mullon jus[t tähä?] (0.2) kr kr krhm (0.8)**

well I+have exactly to+this

=well I have a point exact[ly to this?] (0.2) kr kr krhm (0.8)

5 ???: [.nffft]

6 A: **tämä ilta- ilta(nen) väsymys ja se et kun iltapaa otan nii,**

this evening evening tiredness and it that when supper I+take so

this evening- evening tiredness and that when I take supper so,

((16 lines omitted. A tells about her difficulties in managing eating in the evenings. B responds by telling how she prepares the amount of sandwiches she is about to eat and then puts the ingredients away to avoid over eating.))

22 B: .hhh heh[heh£] [.hhh]

23 A: **[° nii.°] et [se on niinku] se, (.) minu' ongelma että syön**

yeah so it is like it my problem that I+eat

[°yeah.°] so [it's like] that, (.) my problem that I eat

24 **enemmän >↑illalla<=mulle ei niinku päivällä mullei**

more in+evening to+me no like during+day I+don't

more then >↑in the evening<=I don't like during the day I don't

25 **oo ongelmaa eikä oo sillonkaa vielä ku mä töistä mee' niin? .hhh**

have problem neither have then+even still when I from+work go so

have problems and neither then still when I come from work so? .hhh

26 **siinä syönnis ei oo ongelmaa mut se iltapala on semmone, (0.3)**

in+that eating no is problem but that supper is that+kind+of

there is no problem with eating but the supper is kind of a, (0.3)

27 A: **< kompastus[°kivi.°>]**

stumbling+block

< stumbling [°block.°>]

28 B: **[°(mullon <aivan] samallailla.) (--)°**

I+have exactly same+way

[°(I have <exactly] the same.) (--)°

29 (0.4)

30 N: **joo:.=**

yes

ye:s.=

31 B: **=°ajatuksen tasolla.**

thought's level

=°in principle.°

32 (0.4)

33 C: **nii ja sitte [se:,]**

yeah and then that

yeah and then [that:,]

34 N: **[hirv]eän hyvä, (0.7) hirveän hyvä ehdotus,**

awfully good awfully good suggestion

[rea]lly good, (0.7) really good suggestion,

((Continues))

Nutritionist's perspective

1 N: **.hhhhhhh no entäs A?**

well how+about+s A

.hhhhhhh well how about A?

2 (0.3)

3 N: **mitäs sulla ois.**

what you would+have

what do you have.

4 (1.3)

5 A: **no mullon jus[t tähä?] (0.2) kr kr krhm (0.8)**

well I+have exactly to+this

well I have a point exact[ly to this?] (0.2) kr kr krhm (0.8)

6 ?: [.nffft]

7 A: **tämä ilta- ilta(nen) väsymys ja se et kun iltapaa otan nii,**

this evening evening tiredness and it that when supper I+take so

this evening- evening tiredness and that when I take supper so,

((16 lines omitted. A tells about her difficulties in managing eating in the evenings. B responds by telling how she prepares the amount of sandwiches she is about to eat and then puts the ingredients away to avoid over eating.))

22 B: .hhh heh[heh£] [.hhh]

23 A: **[°nii.°] et [se on niinku] se, (.) minu' ongelma että syön**

yeah so it is like it my problem that I+eat

[°yeah.°] so [it's like] that, (.) my problem that I eat

24 **enemmän sillon >↑illalla<=mulle ei niinku päivällä mullei**

more then in+evening to+me no like during+day I+don't

more then >↑in the evening<=I don't like during the day I don't

25 **oo ongelmaa eikä oo sillonkaa vielä ku mä töistä mee' niin?**

have problem neither have then+even still when I from+work go so

have problems and neither then still when I come from work so?

26 **.hhh (0.2) siinä syönnis ei oo ongelmaa mut se iltapala on**

in+that eating no is problem but that supper is

.hhh (0.2) there is no problem with eating but the supper is

27 **semmone, (0.3) kompastus°kivi.°**

that+kind+of stumbling+block

kind of a, (0.3) stumbling °block.°

28 (1.1)

29 N: **joo:.**

yes

ye:s.

30 (1.0)

31 B: **°ajatuksen tasolla.=**

thought's level

°in principle.°=

32 N: **=hirveä[n hyvä m?]**

awfully good

= reall[y good m?]

33 C: **[nii (---)]**

yeah

[yeah (---)]

34 **(0.7)**

35 N: **hirveän hyvä ehdotus,**

awfully good suggestion

really good suggestion,

((Continues))

Extract 4

Group's perspective (Group 2)

1 N: **Oikei' hyviä konsteja.**

really good tips

Really good tips.

2 N: **Sait kolme pistettä.**

you+got three points

You got three points.

3 N: **[Haluaako] joku lisätä tähä' vielä, (.) jonku asian.**

want+Q somebody add to+this still some thing

[Does somebody want] to add something more, (.) to this issue.

4 ?: **[↑mm:?]**

5 (4.1)

6 A: **No= ei muutaku [(sen-)]**

well no else+than that

Well=nothing more than [that-]

7 N: **[stres]sihän vaikut- (.) ↑ni?**

tress+CLT yeah

[As we know] stress affect- (.) ↑yeah?

8 A: **sem[most- (.) e-] ajattelin että siis sillon=kun tuntuu oikein**

that I+thought that so then when feel+0 really

Th[at- (.) e-] I thought that when=it feels really

9 N: **[>sano vaan.<]**

say PTCL

[>Go ahead.<]

10 A: **>tympeeltä nii,< (0.4) mää ainakii yritän keksiä jotakin semmosta**

dull so I at+least try invent something that+kind+of

>stressful,< (0.4) I at least try to come up with something

11 A: **tekemistä >mistä mää< tykkään.**

activity that I like

to do >that I< like.

12 (1.0) ((Continues))

Nutritionist's perspective

1 N: **Oikei' hyviä konsteja.**

really good tips

Really good tips.

2 N: **Sait kolme pistettä.**

you+got three points

You got three points.

3 N: **Haluaako joku lisä[tä tähä'] vielä, (.) jonku asian.**

want+Q somebody add to+this still some thing

Does somebody want to [add some]thing more, (.) to this issue.

4 ?: **[(-)]**

5 (4.7)

6 N: **Stressi[hän vaikut-]**

stress+CLT

As we kn[ow stress affect-]

7 B: **[No=ei muutaku,]**

well no else+than

[Well=nothing more than,]

8 (.)

9 N: ↑ni?

yeah

↑yeah?

10 (0.5)

11 N: **>Sano vaan.<=**

say PTCL

>Go ahead.<=

12 B: **=(Et se-)} (0.5) ajattelin että siis sillon=kun tuntuu oikein**

that I+thought that so then when feel+0 really

That- (.) e- I thought that when=it feels really

13 B: **>tympeeltä nii,< (0.4) mää ainakii yritän keksiä jotakin semmosta**

dull so I at+least try invent something that+kind+of

>stressful,< (0.4) I at least try to come up with something

14 B: **tekemistä >mistä mää< tykkään.**

activity that I like

to do >that I< like.

((Continues))

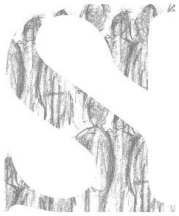
PUBLICATION II

From appearings to disengagements: Openings and closings in video-mediated tele-homecare encounters

Ilomäki, S. & Ruusuvuori, J.

Social Interaction. Video-Based Studies of Human Sociality, 3(3)
https://doi.org/10.7146/si.v3i3.122711

Publication reprinted with the permission of the copyright holders.



Social Interaction. Video-Based Studies of Human Sociality.
2020 Vol. 3, Issue 3
ISBN: 2446-3620
DOI: 10.7146/si.v3i3.122711

Social Interaction

Video-Based Studies of Human Sociality

From appearings to disengagements: Openings and closings in video-mediated tele-homecare encounters

---

Sakari Ilomäki & Johanna Ruusuvuori

Tampere University

Abstract

In this article, we examine openings and closings in video-mediated tele-homecare for older adults in Finland, using multimodal conversation analysis. We demonstrate how participants organise these boundaries sequentially and multimodally, how visual appearing and disengaging are of key importance in these processes, and how openings and closings mirror each other in this institutional setting. In the openings, the participants orient to sequential structures that resemble those from mundane telephone conversations and Skype interactions: summons–answer, appearing–noticing, greeting–greeting and the "how are you" question–answer. The participants treat appearing as an accountable part of the opening, and delay advancing to the "how are you" question until a proper visual appearing is produced. Closings are managed through stepwise transition practices that result in a terminal exchange and both participants disengaging from the encounter: the clients, by walking away; the nurses, by closing down the connection. In addition to managing visuality, time-oriented talk is present in both openings and closings. A comparison of our results with findings from other technology-mediated encounters emphasises the importance of visuality in managing closings, and shows that tele-homecare is an interesting hybrid of institutionality and informality.

Keywords: tele-homecare, openings, closings, video-mediated interaction, multimodal conversation analysis

1. Introduction

Openings and closings serve as the boundaries for institutional encounters. They set the conditions for the first topic and later make explicit the moment when the issues to be dealt with are sufficiently resolved. In face-to-face interaction, these boundaries are achieved through the close coordination of verbal and bodily practices, and the use of material surroundings and artefacts (Broth & Mondada, 2013, 2019; Harjunpää, Mondada, & Svinhufvud, 2018; Hartford & Bardovi-Harlig, 1992; Heath & Luff, 1992a; Heritage & Robinson, 2006 Robinson, 1998; 2001; Ruusuvuori, 2001). These transition practices gain meaning through their relationship to the overall structure of the institutional encounter (Drew & Heritage, 1992; Robinson & Stivers, 2001). Even though participants engage in different practices to achieve openings and closings, there are similarities between the boundaries in terms of their structural characteristics (Femø Nielsen, 2013; Schegloff & Sacks, 1973, 297). In business meetings, for example, participants engage in similar kinds of practices in reverse order when opening and closing encounters (Femø Nielsen, 2013). Encounters that take place in technologically mediated settings challenge the multimodal organisation of openings and closings (see, e.g. Arminen, Licoppe, & Spagnolli, 2016; Heath & Luff, 1992b; Luff et al., 2003; Luff, Heath, Yamashita, Kuzuoka, & Jirotka, 2016). In this article, we examine openings and closings in video-mediated (henceforth, VM) tele-homecare for older adults in Finland. We show how participants organise these boundaries sequentially and multimodally, how visual appearing and disengaging are of key importance in these processes, and how the openings and closings mirror each other in this institutional setting.

In institutional telephone interaction, the participants simplify the sequential organisation of both openings and closings. Openings in institutional telephone interaction are kept short, and consist of the call-taker answering and identifying the service, and then the caller acknowledging this and proceeding to the central issue (e.g. Kevoe-Feldman, 2015; Leydon, Ekberg, & Drew, 2013; Whalen & Zimmerman, 1987; Zimmerman, 1992). In comparison, mundane telephone calls operate through three adjacency pair structures: summons–answer, identification–recognition and the "how are you" question–answer (henceforth, HAY) (Schegloff, 1968, 1986). Closings in institutional telephone interaction are often managed through a caller's expression of gratitude or acceptance of a service, which leads to the terminal exchange (e.g. Kevoe-Feldman, 2015; Raymond & Zimmerman, 2016; Woods, Drew, & Leydon, 2015), while in mundane telephone interactions, the closings are initiated by a closing section, starting with potential pre-closing tokens and ending with a terminal exchange (Schegloff & Sacks, 1973). One recurrent practice in both institutional (see, e.g. Ekberg & Lecouteur, 2014) and mundane telephone interactions (Schegloff & Sacks, 1973, 315) is the forming of future arrangements as a closing implicative action. While institutional telephone encounters may involve the professional's

use of artefacts and technologies (Kevoe-Feldman, 2015; Zimmerman, 1992), openings and closings are achieved solely through talk, as the interaction lacks visual cues.

In the openings of VM encounters, orientation to visibility is central to determining whether the participants can proceed (e.g. Licoppe, 2017; Pappas & Seale, 2009). Analogical to the aural summons–answer adjacency pair, participants in mundane Skype interactions organise openings around the visual *appearing–noticing* adjacency pair (Licoppe, 2017). Appearings are differentiated from merely "becoming visible on the screen", and participants can refrain from advancing to greetings until the talking heads configuration, which shows the interactants' face and upper body, is established (Licoppe & Morel, 2012). Furthermore, participants often topicalise seeing and visibility in the openings (Duuly & Tudini, 2016; Licoppe & Dumoulin, 2010; Pappas & Seale, 2009; Stommel, van Goor & Stommel, 2019). Additionally, participants coordinate their verbal and bodily conduct to establish a shared orientation and readiness to proceed to the first topic of these encounters. For example, in tele-consultations, doctors' HAY questions both establish attentiveness to the patient and serve as an implicit means of checking the audio connection (Stommel et al., 2019). While openings in VM settings have gained substantial attention, to the best of our knowledge there are few published EM/CA research findings on closings in VM settings.

This study focuses on tele-homecare for older adults, which has so far remained understudied (for a review on EM/CA research on VM interaction, see Mlynár, González-Martínez, & Lalanne, 2018). Homecare encounters are institutionally managed visits, in which a home helper (in Finland, this is often a practical nurse) assists an older adult with everyday tasks and minor medical issues. While activity transitions have been studied in face-to-face homecare settings (Lindström & Heinemann, 2009), the openings and closings of the encounters have not been studied. Sävenstedt, Zingmark, Hydén and Brulin (2015) studied older adults as participants in VM interaction, emphasising the importance of gaze-direction practices and social talk in building joint attention between older adults living in residential care and nurses, using a CA-inspired method. So far, however, no rigorous EM/CA analysis of activity transitions in VM interactions involving older adults has been performed.

In this study, we examine how practical nurses (for the sake of convenience, we will henceforth refer to the professionals in our data as *nurses*) and homecare clients organise openings and closings in VM tele-homecare encounters. We will show how the participants organise openings around four adjacency pairs, and closings through stepwise transition, how these transitions are multimodal achievements in which visual appearing and disengaging are of key importance,

and how the management of institutional and technological relevancies produces these boundaries as a mirroring each other (Femø Nielsen, 2013).

2. Data and method

The data for this study comprises video recordings from 14 tele-homecare encounters, collected from a Finnish homecare unit undergoing a service pilot in which one of the daily visits was replaced by a video call. Twelve encounters were recorded in the nurses' office, and two were recorded in the clients' homes. Each encounter was recorded from one side of the interaction. The transcripts are in Finnish with English translations, accompanied by line drawings. Word-by-word translations are provided as supplementary material. We used the Jeffersonian (Jefferson, 2004) transcription system, accompanied by Mondada-style (2001) annotation for visual conduct. The data collection and analysis were part of the *Healthcare Workers in the Eye of the Digital Turbulence* research project, conducted in Tampere University and the Finnish Institute of Occupational Health, with funding from the Finnish Work Environment Fund. All clients lived alone and had some level of mild memory deficit. During the recruitment process, special attention was given to ensuring that the clients' rights were protected. An ethical statement for the study was granted by the ethical committee for the Tampere region (document number 49/2017).

Four older adults living at home used a tablet with a simple program that allowed them to answer calls. Before the encounter, the tablet was in screen-saving mode. When the nurse initiated the call, the screen changed to indicate the incoming call and display the caller identification. The client answered by tapping the caller identification icon. None of the clients used the phonebook feature to initiate calls themselves. (Figure 1a) After the client had tapped the icon, the screen changed, showing the caller's information, as well as a loading bar to indicate the establishment of the connection (1b). Subsequently, the screen turned to encounter mode, in which the client's own image is visible in the "vanity screen" at the bottom-right corner (1c). While the clients were able to close the connection if they wished, this never occurred in our data.

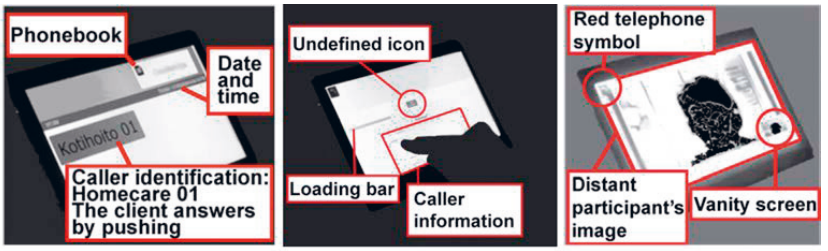


Figure 1a *Figure 1b* *Figure 1c*

Figure 1. Clients' user interface

The three nurses who participated in the study used a computer in their shared office. In order to initiate the call, the nurses first had to select the client from the contact list showing all the clients, and then click an icon to confirm that they wanted to proceed (Figure 2a). This would open a dialogue box with contact information, which the nurse then clicked to proceed (2b). The screen would then change, showing the contact information, and the nurse would click once more to initiate the call (2c). Between the nurses' initiation and the clients' answer, the nurses would see their own image both in full screen and on the "vanity screen" on the bottom-right corner (2d). When the clients answered, the clients' image would replace the nurses' image on the screen (2e).

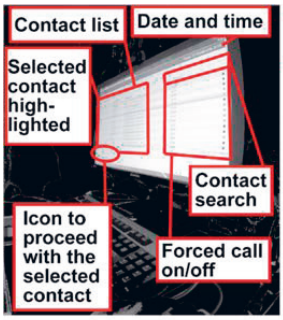


Figure 2a

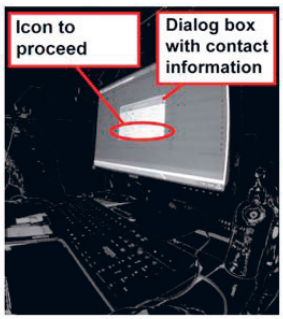


Figure 2b

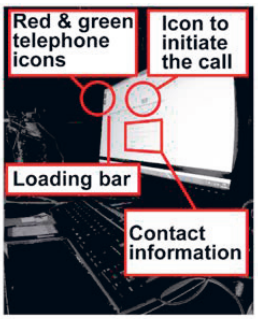


Figure 2c

Figure 2d

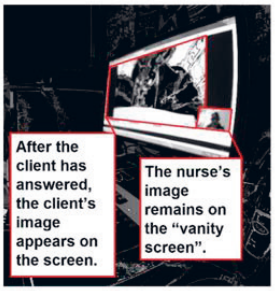


Figure 2e

Figure 2. Nurses' user interface

We employed multimodal Conversation Analysis (Mondada, 2019; Sidnell & Stivers, 2013) as our analytic approach. After a preliminary analysis of all the topic and activity transitions, we chose the openings and closings as the focus for this article, since they are fundamental for managing the encounter. We then analysed how the participants achieved the transition from opening to the institutionally relevant HAY question, and from the last topic to closing the encounter, as well as how the management of institutional and technological relevancies produces these boundaries as mirroring each other.

Before the analysis, we wish to highlight two aspects about the data. First, when the participants look at their own screens, they appear to look slightly sideways at the other interactant (Arminen et al., 2016; De Fornel, 1994), and when they look directly at the web-camera, thus appearing as gazing directly at the distant participant, they cannot reciprocally see the distant participant's gaze on the screen. Therefore, mutual gaze, in the sense of gazing directly into each others' eyes, as occurs in face-to-face encounters, is impossible. However, analogous to gaze contact in face-to-face encounters, the co-interactant's gazing at, as opposed to away from, the screen is treated as relevant when managing transitions (c.f. Satar, 2013). Second, as each encounter was recorded from only one location, we cannot analyse how the distant participant receives the turns at talk or bodily actions. What is produced at one end of the encounter differs from what is perceived at the other (see, e.g. Luff et al., 2016; Ruhleder & Jordan, 2001). Thus, both participants may interact on the basis of a slightly different understanding of the ongoing action. Therefore, to assume that recordings from only one location would capture both members' perspective would be to overlook the fundamental ways in which technical mediation becomes sequentially relevant. We will concentrate on features that are analysable with data from only one perspective, and reflect on this limitation in the discussion section.

3. Analysis

3.1 Openings in tele-homecare

In their simplest form, openings in our data are organised around a recurring sequential structure consisting of four subsequent adjacency pairs:

(1) summons(S)–answer(An)

(2) appearing(Ap)–noticing(N)

(3) greeting(G)–greeting(G) and

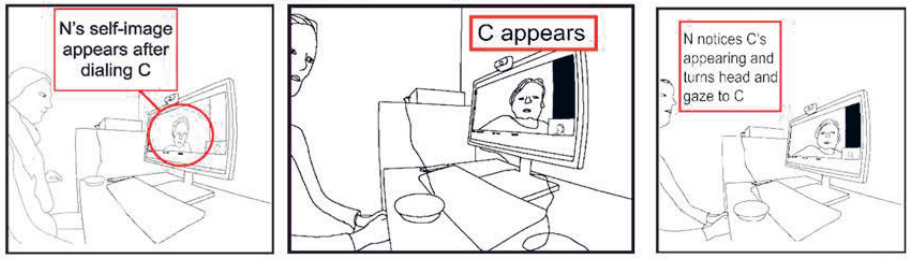
(4) HAY-question(Q)–answer(An).

This is exemplified in Extract 1. The nurse (N) has selected the client (C) from the phone book and initiated the call.

Extract 1: Orienting to four adjacency pairs in the opening

```
01          +(0.3)♦   (2.5)   #♦(10.5)+(0.4)+(3.6)+(4.5)+(0.3)+(4.5)
   n: S-> +N initiates the call by clicking the icon
   n: >>-----gaze to screen-------+.....+g. speaker +.....+gaze screen
   s:              ♦load. bar ♦N's self image
                                #IMG1.1
02          +(0.3)+(4.4)#%(0.3)+(0.9)#+(0.2)
   c: An/Ap->              %C's image appears, gazing to screen /N->
   n: N->  +.....+ gaze right +......+gaze to screen /C->
```

#IMG1.2 #IMG1.3



IMG1.1 IMG1.2 IMG1.3

```
03 N:N/G-> tervehdys.
           hello.
04          (0.9)
05 C: G-> no tervehdys.
           well hello.
06          (0.3)
07 N: Q-> m:itäs sinne.
           wh:at's up.
08          (0.5)
09 N: Q-> kuuluu.
           over there.
10          (1.1)
11 C: An-> mitäs täällä.=vanha (rouva) ja entise' kujeet.
           not much.=old (lady) with her usual habits.
           ((Continues))
```

The first adjacency pair, summons–answer, occurs in lines 1–2, when the nurse initiates the call (line 1) and the client appears on the screen (line 2, IMG 1.2, see also Licoppe, 2017). As the video mediation offers the nurse visual access to the client, there is no need for the client to produce a verbal answer, and the participants achieve this part of the opening without talk. In addition to answering the summons, the client's visual appearance also serves as the first-pair part for an appearing–noticing adjacency pair. The nurse produces noticing both bodily and verbally, by turning her head and gaze to the screen (line 2, IMG1.3) and then producing a verbal greeting (line 3). The appearing–noticing adjacency pair is thus achieved both verbally and through bodily conduct (by becoming visible on the screen and observably turning to the screen). The nurse's verbal turn also functions as the first pair part for a greeting–greeting adjacency pair (lines 3 and 5). By producing the noticing via both visual and verbal conduct, the nurse can manage the limitations that the technical mediation brings to intersubjectivity. While she cannot know for certain whether the client has noticed her, or if the client can recognise her gaze shift as noticing, verbalisation makes the noticing

salient. Furthermore, the first greeting projects a reciprocal greeting from the client, thus allowing for a testing of the connection – if the client does not produce the greeting, the nurse can imply that the client has not heard her (c.f. Stommel et al., 2019). When the client produces the forecasted greeting, the participants establish that they can both hear each other and proceed to the HAY question (lines 7 and 9), followed by an answer by the client (line 11).

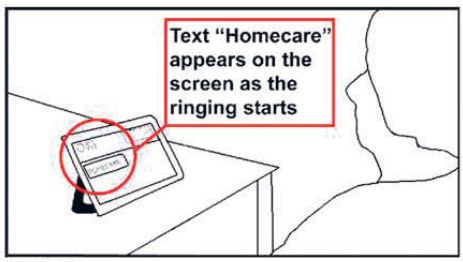
As is observable in Extract 1 (and in Licoppe, 2017), the participants manage the openings not only through talk, but also through visual conduct, monitoring each other's visibility and adjusting their conduct to that visibility. Compared to telephone conversations, in which the summons–answer adjacency pair is achieved via a combination of technology use (calling and picking up the phone) and talk (verbal answer), in our data both the summons (the nurse calling the client) and the answer (answering the call and appearing on the screen) are achieved without talk. Furthermore, the appearing–noticing adjacency pair is based on the visual appearance of the summoned participant (the client, line 1, IMG 1.2), followed by both participants establishing gaze towards the screen (line 2, IMG1.3), and the nurse verbally noticing this visual appearing when initiating the greetings–adjacency pair (line 3).

Treating visibility as essential in managing the opening is also observable when either of the participants fails to appear properly in the opening. This is exemplified in Extract 2, which shows the client answering the call in her home. The nurse has called the client twice, and the client has tried to answer. However, for some reason, the call has not connected. The client is sitting in front of the tablet in her living room. In the transcript, the letter *a* refers to the non-human sounds from the tablet, while *s* indicates screen changes.

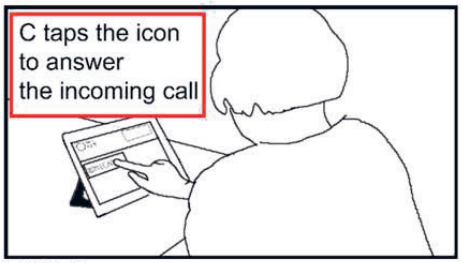
Extract 2: Postponing the HAY question until visual appearing is achieved

((The client is sitting in front of the tablet. The tablet has rung twice before, the last time, 71 seconds ago. The client has answered, but for some reason, the connection has not started.))

01 **#♦(0.3)%(0.2)%(1.0)♦(0.5)%(2.4)#♦(1.3)♦(3.2)♦(0.2)♦**
a:S-> **♦-------ring------♦ ♦ring ♦ ♦ring ♦**
c:>>g. over tablet%.....%gaze to screen->
c:An-> **%tapping icons on the screen->**
#IMG3.1 **#IMG3.2**

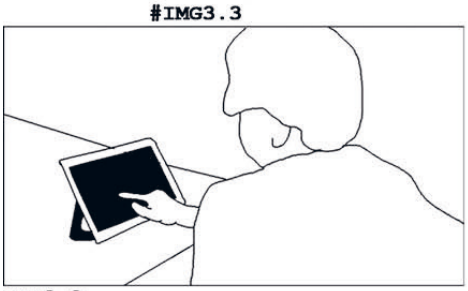


IMG3.1

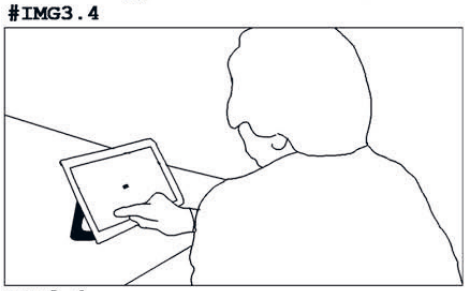


IMG3.2

02 **(0.5)#♦(2.1)♦(2.2)%(1.1)%(0.5)#♦(9.7)**
c: An-> **%.....%taps an icon on the screen, holds->**
s: **#♦black♦---white screen---♦ icon appears on the screen->**
#IMG3.3 **#IMG3.4**

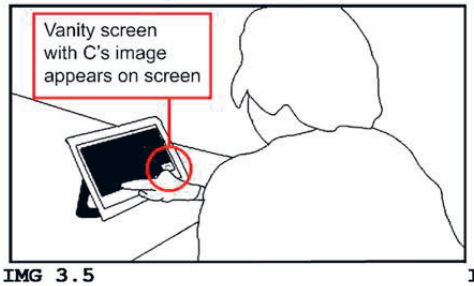


IMG3.3

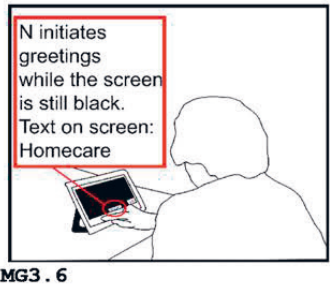


IMG3.4

03 **#♦(6.7)**
s: **->♦black screen**
#IMG3.5



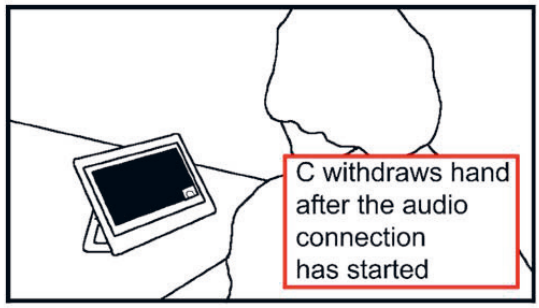
IMG 3.5



IMG3.6

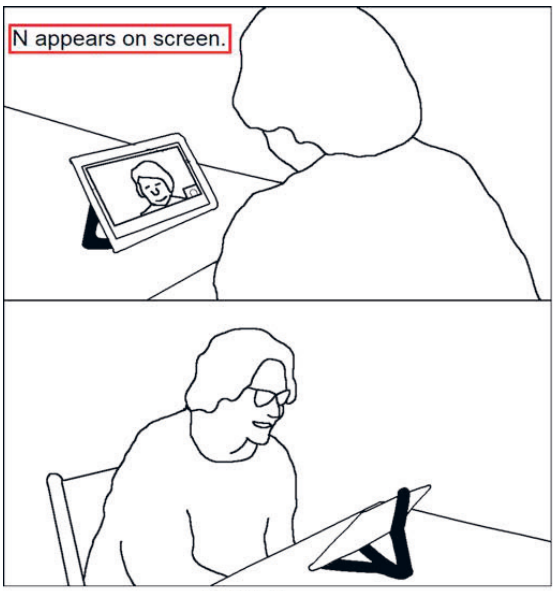
04 N:G-> **#♦iltaa Ann[a.**
good evening Ann[a.
s: **♦text on screen appears**
#IMG3.6

```
05 C:G->                          [no ehtoota %ehtoota.#=
                                  [well good evening.=
   c:                                       ->%withdraws hand->
                                                  #IMG3.7
```



IMG3.7

```
06 C:        =on%ko ongelmia?
             =is there some trouble?
   c:         ->%both hands on lap->

07 N:     (no:=o[n.)
          (well:=yes.)
08 C:           [£hehe[hahhah£                    [£hehheh£
09 N:                [no    vähän    tuntuu   ole[van >ei näy,< (.)
                     [well it feels a bit like th[at >the picture,<(.)
 (.)
10 N:N-> kuvaa   >(jaa no)   mutta   nyt<   näkyy     [kuva ja ääniki
         doesn't show >(well oh) but now< the picture [shows and sound
11 C:                                                 [jaaha?
                                                      [well then?
12 N:     [kuu(luu.)
          [(works) as well.
13 C:     [no nii?
           okay?
14        (.)
15 C:     £(h)jo#♦o.£
           £(h)yes.£
   s: Ap->       ♦Screenchange: N appears
                #IMG3.8
```



IMG3.8

```
16          (.)
17 C:       £hm hm .hh[h
18 N:                 [£no: ni?£
                      [£o:kay?£
19 N:       [kuuluuko      sinne        hy[vin.
            [does the sound work there we[ll.
20 C:N->    [.ng                          [joo kuuluu ja näkyy kuvaki
            [.ng                          [yes the audio works and also
21 C:       jo.
            the picture works now.
22          (.)
23 C:       £hoh hoh£ (0.5) .hh[h
24 N:                          [£no(h)o n[i?£       [hyvä.
                               [  £o(h)ka[y?£       [good.
25 C:                                    [£hkmhkmkh[m£
26 C:       [joo.     [joo.
            [yes.     [yes.
27 N:       [hyvä jut[tu.
            [good thi[ng.
28          (0.4)
29 N:Q->    mites päivä o menny?
            How has the day been?
            ((Continues))
```

During the third attempt to answer the call, the technical difficulties remain. Only after the client has tapped different icons on the screen for several seconds (lines 1–2) does the screen finally change to encounter mode (line 3, IMG3.5). However, at this stage, the nurse's visual appearance is still inadequate. The nurse's image remains black, and only the client's vanity screen is visible. Almost seven seconds after establishing this configuration, the nurse launches the first pair part of greetings (line 4), and the client answers in a last item overlap (line 5). So far, the nurse's visual appearance has been inadequate, and the participants have only established an aural connection. Compared to Extract 1, it

is noteworthy that the dual function of visual appearing is not present from the client's perspective. For the client, the call-taker, the nurse's appearing serves solely as the first-pair part in an appearing–noticing adjacency pair, and the first greeting solely as the first pair part in reciprocal greetings. Thus, these actions, visually appearing on the screen and verbal greeting, appear differently from various perspectives of VM interaction (c.f. Ruhleder & Jordan, 2001).

The client treats the lack of the nurse's visual appearance as relevant by, asking a polar question about the VM technology immediately after the greetings (line 6), thereby starting an insert expansion in the basic structure of the opening. The question focuses on problems in general, embedded with an assumption that some problems do exist. This turn design enables the nurse to handle both the recurrent difficulties in establishing the connection and the inappropriate visual appearing, i.e. the black screen. The nurse produces a type-confirming answer (line 7) and, partly overlapping with the client's laughter, elaborates the answer by first topicalising the inappropriate visual appearing of the client (*the picture doesn't show / ei näy kuvaa*), and then cutting off her turn and explicating the change in the client's visual appearance with *now the picture shows* (*nyt näkyy kuvaki*, lines 9–10, 12). Furthermore, in her response, the nurse treats visibility as shared by presuming that the problem has been resolved. In her turn (lines 9–10, 12), the nurse states the changes in what she can see without adding modifiers that would emphasise the difference in perspectives (such as *now <u>your</u> picture shows*). Nor does she explicate the client's perspective by, for example, asking whether the client can see her. The client receives the nurse's turn (lines 11, 13, 15), and during the last turn, the nurse's image finally appears on the client's screen (line 15, IMG3.8).

The nurse asks the client about the quality of the audio (line 19), which further shows her treating the problems with the visual connection as being resolved. However, in her answer (lines 20-21), the client retopicalises visibility by adding *and also the picture works now* (*ja näkyy kuvaki jo*) after answering the nurse's question about the audio connection. In her response, the client makes salient the dissimilarity of their visual perspectives. As the nurse visually appears only after she has explicated that she sees the client (lines10–13), it is apparent to the client that there is some level of incongruity regarding what they can visually perceive. By retopicalising visibility at this point, and explicating that she can see the nurse, the client confirms that a mutual visual appearance has been achieved, and that the participants can now proceed with the encounter. It is only after the participants have established that both have adequately visually appeared that they move on to the HAY question (line 29).

In the tele-homecare setting, the HAY question serves the institutional purpose of preparing for the transition to the actual business at hand. This is apparent from two features: the nurses routinely insert the question after the greetings (c.f.

Stommel et al., 2019), and can repeat the question if the client does not answer it in a proper way. This is exemplified in Extract 3. The nurse and the client have opened the connection and briefly talked about the connection problems.

Extract 3: Treating the HAY answer as inadequate

```
01 C:      £no [nii.£]        [£heh£]
           £  o[kay?£]        [£heh£]
02 N:Q1->      [.hhh ] mitäs [   Ar]ja.
               .hhh what's up [   Ar]ja.
03 C:      .hhh [mitä- ]
           .hhh [not-  ]
04 N:Q1->      [mite' ] päivä o'=men[ny.   ]
               [how's ] the   day=be[en.   ]
05 C:                               [(no)  ] menny oikei hyvi,=
                                    [(well)] it has been really good,=
06         ja mulla on tää, ö- ö- tää- m-=m- (.) mikä tää nyt on? (0.2)
           and I have this, eh- eh- this-m-=m-(.)what is it again? (0.2)
07         .hhh (0.6) hhhh täällä, ↓hmh (.) sa[n:=ny?       ]
           .hhh (0.6) hhhh here, ↓hmh (.) wh:a[t's that=now?]
08 N:                                         [         vähä] video-
                                              [         some] video-
09         (0.6)
10 C:      joo? (.) videoimassa.
           yeah? (.) video recording.
           ((6 lines omitted, N and C talk about the recording.))
17 C:      ↑juu  sa[atii-     ]
           ↑yeah we[ got it-  ]
18 N:Q2->        [(no) mite'] sun päivä on menny.=
                 [(well) how] has your day been.=
19 C:An->=.hhh ↑oikei hyvi.
           =.hhh ↑really good.
20         (0.9) ((Continues))
```

After the client has closed the talk about the connection problems (line 1), the nurse initiates the HAY adjacency pair (lines 2 and 4). The client answers, in a co-produced turn, that the researcher is visiting to record the encounter (lines 3–17). While the client treats this as a noteworthy event, the nurse doesn't take an answer to her previous question to have been produced, as it does not provide information about the client's wellbeing, as a recounting of daily activities would. The nurse therefore repeats the question, focusing on the client's day in particular, with the addition *your day* (line 18). Thus, as in medical consultations, where the HAY question is used for gathering information on the reason for the visit (Heritage & Robinson, 2006), here it appears to be gathering knowledge on the wellbeing of the client, rather than merely functioning as a vehicle for small talk, other attentiveness or testing the connection (c.f. Stommel et al., 2019).

Extracts 1–3 show the participants' orientations to the four adjacency pair structures (summons–answer, appearing–noticing, greeting–greeting, HAY-question–answer) when managing the openings. Parts of this structure are achieved without talk, and other parts with the co-use of talk, bodily conduct and

technology. This basic structure can be expanded when the participants need to manage technical problems to produce a proper visual appearing (Extract 2) or when the clients answer the HAY question in ways that are not institutionally relevant (Extract 3). Thus, the openings are organised around the close coordination of verbal and bodily practices and the monitoring of another participant's visibility.

3.3 Closings in tele-homecare

The closings in our data are organised in a stepwise manner (c.f. Schegloff & Sacks, 1973), often primed with the nurses' positive evaluations (E), talk about future arrangements (F), either about the prospects of the care or merely the nurse mentioning that they will leave the client to continue their evening, pre-closing tokens (P) and the client's service appreciation (A), resulting in the terminal exchange (T) accompanied by mutual visual disengagement (D). Extract 4 exemplifies this stepwise progression towards the closing. Before the extract, the nurse has asked whether the client has already taken her medicine. The client has answered that she will take it after she has eaten, and the nurse has received this answer with *okay (no nii)*.

Extract 4: Stepwise progression to closing

01 **#+(0.2)+(0.9)**

n: >>g. to phone+.....+gaze to screen /C->

c: >>gaze to screen/N->

#IMG4.1

IMG4.1

02 N:F-> **#sää [voit si]itä=sitte ruveta syömään kohta iltapalaa ja**

you [can]=then start eating soon that supper and

03 C: **[(-)] ((possibly *kyllä /yes*))**

#IMG4.2

IMG4.2

04 N:F-> **sitte ottaa lä%äkkeet%** **[ni.]**

then take the medicine [like.]

05 C: **[(-)]**

c: **->%.....%gaze to left->**

06 **(0.2)**

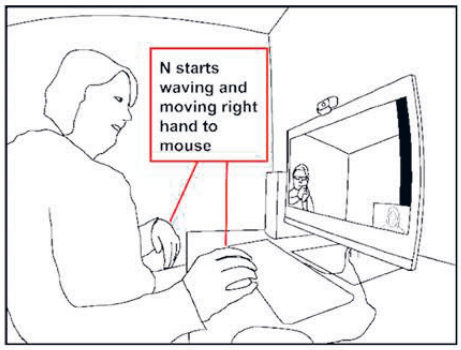
07 C:F-> **(-%-)% ja menen katsomaan telkkaria +samalla.**

(--) and I'll go watch the telly as well.

c: **->%..%gaze to screen /N->**

n: **+nods->**

```
08 N:     no: ni.+
          o:kay.
   n:          ->+nod finished

09        (.)
10 N:E-> kuulostaa hyvältä.
          sounds    good.
11        (0.9)
12 C:P-> joo: joo.
          ye:s yes.
13        (.)
14 C:P-> kyllä.
          right.
15 N:P-> selvä.
          okay.
16        (0.2)

17 N:F-> ei kuule muuta ku mukavaa illajatkoo ja taas so%itellaa.%
      nothing else then than have a nice evening and let's call again.
   c:                                                      %--nods--%

18        (0.7)%(0.2)
   c:          %nods twice->18.22

19 C:A-> kiitos,
          thank you,

20        +(0.2)
   n:     +nods->

21 N:E-> hy+v[ä.+]
           goo[d. ]
   n:     ->+nod finished
   n:            +moves right hand to mouse->
22 C:A->      [  s]a%moi ja   #+kii- (.) + soitosta.#          +
              [  t]oo    and  #tha- (.) for calling.#
   c:              ->%nods finished
   n:                          +.........+waves with left hand+
                              #IMG4.3               #IMG4.4
```



IMG4.3

IMG4.4

23 (0.2)

24 N: +KII:+TOS?#
THA:NK YOU?
n: +moves left hand to lap->
n:D-> ->+right hand on mouse, operates->
#IMG4.5

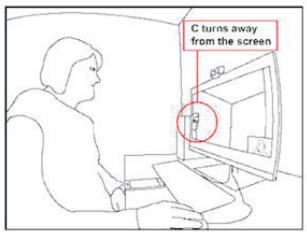
IMG4.5

25 N:T-> [↑MOI] ↑M+O+I?
[↑BYE] ↑BYE?
n: ->+left hand on lap->
n: +moves cursor->
26 C: [(-)]

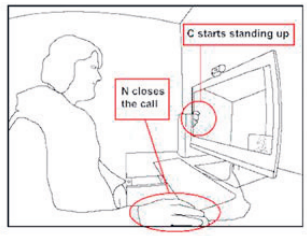
27 %(0.3)%(0.2)
c: %nods %

28 C:T-> %>moi % moi.<
>bye bye.<
c:D-> %turns%

```
29        #%(0.5)#+(1.3)+(0.5)#+♦%(0.7)+(1.7)♦
    n:D->          +click        +......+operates phone with right hand->
    n:                      +......+gaze to phone->
    c:D->  %------stands up------%
    s:                               ♦scrn freezes♦screen to client list
          #IMG4.6#IMG4.7       #IMG4.8
```



IMG4.6



IMG4.7

IMG4.8

Throughout the extract, the nurse and the client approach the closing in a stepwise fashion, through various verbal practices. First, the participants produce future-oriented talk, both when the nurse directs the client to take their medicine and the client aligns (lines 2–7), and then when the nurse wishes the client a pleasant evening and says they will be in contact in the future (line 17). Second, the client produces a service appreciation in response to the nurse's future-oriented turn (lines 19 and 22). Third, the nurse produces an evaluation (line 21) as a response to the client's service appreciation. Fourth, the participants produce various pre-closing tokens when closing the central task of the call (lines 12–15). These practices, especially the reciprocal thanks in this case, ensure that neither of the participants has any unmentioned mentionables before closing the encounter. The participants proceed to the terminal exchange (lines 25 and 28), which is accompanied by their disengagement – the client by physically withdrawing from the screen (lines 27-29, IMG4.6–4.8) and the nurse by closing the connection, thus digitally disengaging from the encounter (line 29, IMG4.3 and IMG4.7).

The ways in which the participants treat visibility as meaningful for organising the closings are apparent in how they closely coordinate their disengagement with the terminal exchange. During the evaluation and pre-closing tokens, the nurse projects the closing by waving and operating the computer to close the encounter (lines 21–22, IMG 4.3-4.4), and the client withdraws from the screen, in coordination with the terminal exchange (lines 27–29, IMG4.6). While the nurse could close the connection immediately following the terminal exchange, she postpones the closing for 0.5 seconds and waits for the client to start disengaging from the encounter by turning and standing up (line 29, IMG4.7). The client's withdrawal from the screen is therefore treated as meaningful in organising the closing.

Extract 5 further demonstrates the importance of the client's disengagement, the interplay between verbal and bodily conduct and the sequential organisation of the closings. Before the extract, the nurse has explained that another nurse will visit later in the evening, and the client has received this information with a token *yes*, *okay* (*joo selevä*). During this exchange, the client's gaze has wandered around her apartment, and she has appeared to look at the screen only briefly during the service announcement.

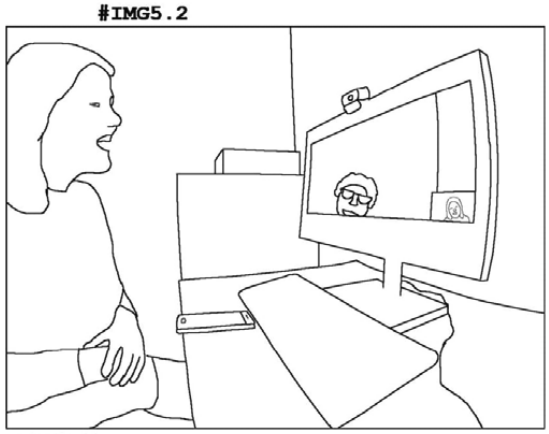
Extract 5: Orienting to disengaging as sequentially relevant next action in the closing sequence

```
01        (0.7)
   n:     >>gaze right->
   c:     >>gaze left->

02 N:     .thhhh+hhh+ ei  mu-=+sit+te muu+ta+ kuule mun +puol+esta ku
           .thhhhhhh nothing e-  else  from  my  side  then  than
   n:         ->+...+gaze down+...+screen+..+gaze phone +....+screen/C>
   n:           +eyes closed      +eyes open->


03 N:     oikee hyvää ill%a'jatkoo.%#
          have a nice evening.
   c:                  ->%.........%gaze to screen /N->
                                    #IMG5.1
```

IMG5.1

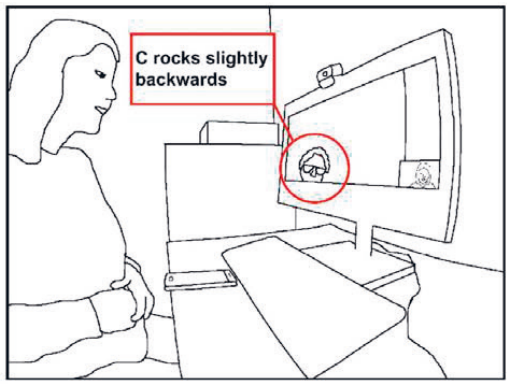
```
04        +(0.3)+(0.4)%(0.3)%(0.7)+%(0.6)
   n:     +.....+-----smile-------+neutral face
   c:                %.....%gaze down->
   c:                             %nods->
```

```
05 C:    ja työn%iloo.
         and enjoy your work.
   c:          ->%nod finished, gaze to screen /N->

06       %(0.3)+#
   n:          +smiling->
   c:    %smiling->6.9
                #IMG5.2
```



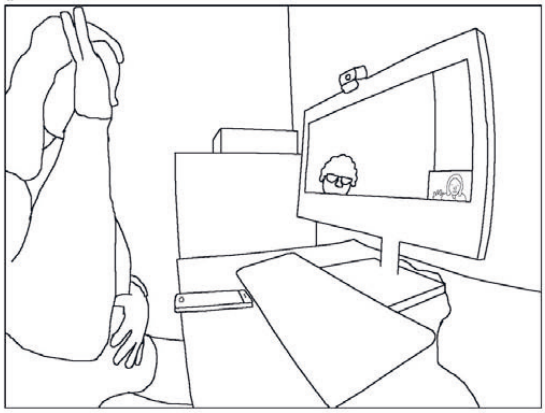
IMG5.2

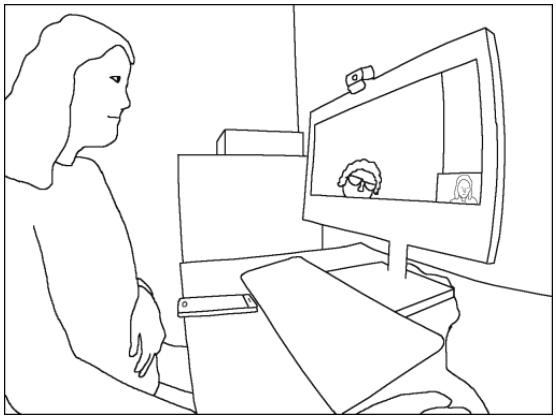
```
07 N:F-> [kiitos.   nähd]ää' +huomenna.
         [thank you. see] you tomorrow.
   n:                        +straightens posture->
08 C:    [£heh heh heh£ ]

09       (0.8)%(0.6)% (0.7)  %
   c:       ->%stops smiling
   c: potential  D->%rock bwd%  ((potential disappearing))

10 N:E-> #+hy%vä.
           good.
   n:     +lifts right hand->
   c:        %rocks fwd->
         #IMG5.3
```



IMG5.3

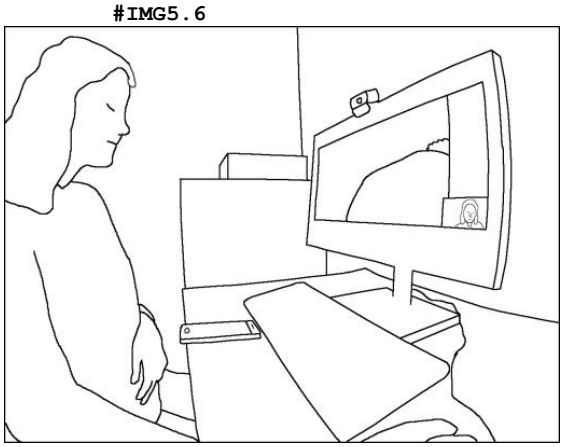
11 (.)
12 N:T-> #+moi%kka?
bye bye?
n: ->+waves->
c: ->%sits still
#IMG5.4





IMG5.5

```
15        %(0.2)+(1.1)#+%(0.7)+(0.4)+♦(0.5)+(1.9)♦
   n:D->        +click +click
   n:                           +.....+gaze to phone
   n:                                 +......+left hand operates phone->
   c:D-> %stands up----%face outside the screen
   s:                                  ♦scr freezed♦contact list
                      #IMG5.6
```



IMG5.6

The participants manage the closing through the coordination of verbal and bodily practices in various ways. First, the nurse keeps her gaze fixed on the screen throughout the sequence, and the client gazes at the screen, thus appearing to be gazing at the nurse, in a sequential location in which the participants negotiate that no unmentioned mentionables will arise (lines 1–6). First, the nurse wishes the client a pleasant evening (lines 2–3). During the nurse's turn, the client turns her gaze to the screen, thus appearing to be gazing at the nurse, and during the following pause responds bodily by nodding (line 4), and then verbally by saying *and enjoy your work* (*ja työniloo*), enabling her to produce reciprocal wishes (line 5), thus confirming that there are no unmentioned issues left to be dealt with. As the client's gaze reaches the screen, the nurse starts to smile (line 4), and the participants share a moment of mutual smiling (line 6, IMG5.2).

Second, the nurse does not treat the client's lack of response to her farewell (line 7) as accountable, and proceeds towards the closing based on the sequential features of the turns and the bodily practices of the client. The nurse's thanks and farewell (line 7) are followed by a 2.5-second pause, during which the client rocks backwards slightly (line 9), after which the nurse produces an evaluation (line 10). While the nurse's farewell makes a response a relevant next action, two features explain why the verbal response from the client is not mandated. First, from the perspective of the sequential organisation of closings, this farewell is a somewhat superfluous addition: the participants have already established the prospects of the care and that neither of them has any unmentioned mentionables. Second, given that the client's rocking (lines 9-10, IMG5.3) takes place in a sequential environment in which the terminal exchange and disengagement are relevant actions in the near future, the nurse might interpret it as preparation for standing. Thus, while not responding verbally, the client is seen to align bodily with proceeding towards the closing (see Licoppe, 2017, 382 on "noticing whatever may count as an appearance" in openings). Thus, proceeding to the closing is possible even without the client's response to the nurse's turn, as the client's bodily action does not appear as misaligning with the closing (c.f. Stommel et al., 2019, 286–287) – rather, she appears to be expecting and preparing for the closing.

Third, the participants achieve closing not only through talk, but also through reciprocally visually disengaging from the encounter. The nurse first produces a bodily farewell, waving (lines 10–12, IMG5.4), which is accompanied with a first-pair part of the terminal exchange (line 12). As the verbalisation ends, the nurse starts to move her hand towards the mouse and then operates it (line 13, IMG5.5), hence observably preparing for the closing. This, however, stays unavailable to the client due to the limited visual access afforded by the webcam (c.f. Luff et al., 2003). After 1.3 seconds, the client produces the second-pair part of the terminal exchange (line 14), which is immediately accompanied by her disengaging from the screen (line 15). The nurse carefully adjusts her digital disengagement (closing the connection) to the client's physical disengagement. While the nurse has started preparing the closing immediately after the first-pair part of the terminal exchange (line 13), it is only after the client starts withdrawing from the screen that the nurse closes the connection (line 15). The second click, which closes the connection, only appears as the client's face and gaze observably abandon the encounter (line 15, IMG5.6). While the client does not operate the connection, the nurse does not treat her merely as a passive disconnectee, but instead treats her disengaging as an integral part of organising the closing. Thus, mutual disengagement consists of the close coordination of verbal and bodily practices, which make the closing intersubjectively understandable.

4. Discussion

The participants achieved the openings and closings in VM tele-homecare through coordinating their verbal and bodily conduct with the visibility of each other in recurrent sequential structures. Openings consist of four subsequent adjacency pairs: summons–answer, appearing–noticing, greeting–greeting and HAY-question-answer. This structure resembles mundane telephone conversation openings (Schegloff, 1968, 1986), rather than the prompt advancing to the central issue of many institutional telephone encounters (see, e.g. Leydon et al., 2013; Whalen & Zimmerman, 1987; Zimmerman, 1992). The visual appearing, and the noticing of that appearing, is a central part of the opening sequence. Furthermore, the participants treat each other's visual appearance as an accountable part of the openings and proceed only after a proper visual appearing has been produced (Extracts 1 and 2). This is in line with findings from mundane Skype conversations (Licoppe, 2017; Licoppe & Morel, 2012). However, contrary to Licoppe's (2017) findings, the participants simplified the openings, and it is always the nurse who greets first. This may relate to the adjacency pair organisation of the openings and the participant's orientation to the institutionality of the interaction. From the client's perspective, both the summons–answer and appearing–noticing adjacency pairs are initiated by the nurse, thus encouraging the client to give the floor to the nurse in order to initiate the next action (see Femø Nielsen, 2013). Compared to tele-consultations, in which participants routinely proceed to the reason for the encounter after the opening and the HAY sequence is used to test the connection and display other-attentiveness (Stommel et al., 2019), in tele-homecare, a HAY question is routinely inserted after the greetings. When the answer does not provide care-relevant information, the nurses repeat the question. Thus, the HAY question serves to establish not only that both participants can see and hear each other, but also to gain institutionally relevant information.

Closings in our data are approached step by step with evaluations, future-oriented talk, service appreciations and pre-closing tokens, resulting in terminal exchange and mutual disengagement. The stepwise transition towards closings resembles mundane telephone calls (Schegloff & Sacks, 1973), compared to a straightforward closing through service appreciation (e.g. Kevoe-Feldman, 2015). Our analysis adds to the earlier body of EM/CA research on VM interaction by showing how visuality is of the utmost importance in the organisation of closings, not just openings. In closings, the participants' visual disengagement, either by withdrawing from the screen or closing the connection, is expected. Furthermore, the participants coordinate disengagements with ongoing talk, especially with terminal exchanges, and with each other's visual disengagements (see Extracts 4 and 5). One could say that there are both verbal and visual terminal exchanges in VM tele-homecare. This resembles mobile interactions, in which walking away is closely coordinated with talk and functions as both forecasting and doing

closing (Broth & Mondada 2013, 2019). Similarly to openings in which "recognizing and noticing whatever may count as an appearance … becomes a powerful resource in the process of achieving, collaboratively, a proper joint interactional frame" (Licoppe, 2017, 382), in closings the participants may interpret whatever appears in the sequential location where disengagement is relevant as doing preparing for disengagement (see Extract 5). Thus, both the openings and closings of VM tele-homecare encounters are multimodal collaborative achievements in which the participants adjust their actions to each other's conduct and the technological affordances of the medium (Hutchby, 2001).

Some practices of openings serve double functions. As mentioned, the client's appearing has two functions. In order to work through the limitations that technical mediation brings to intersubjectivity, the nurse's greeting serves as both a verbal noticing and the first greeting. These double functions relate to the different projects they achieve as parts of adjacency pairs. A summons–answer achieves the opening of the overall connection; appearing–noticing, the visual connection; and greeting–greeting, the aural connection. Managing these projects demands the use of different semiotic fields (Goodwin, 2000). In the opening phase of VM encounters, the participants need to establish both the social relationship between the interactants and the medium for the encounter, which produces these distinct yet connected interactional projects. Managing both the medium and the social relationship seems to result in different interactional projects and *multiple temporalities* in interaction (Mondada, 2018, 104). It can be said that "responsive actions can be produced [not only] during initiating actions" (Mondada, 2018, 104), but also as initiating actions in managing these different projects.

The ways of achieving openings and closings mirror each other in two ways. First, the encounter progresses from reciprocal adjacency pairs of visual appearings and greetings to the terminal exchange adjacency pair and disengagement. While adjacency pairs are a central way of organising both openings and closings (see also Schegloff & Sacks, 1973, 297), the possibility of a missing second-pair part in closings (see Extract 5) emphasises how these practices acquire importance as parts of broader sequences of action and interactional projects. Second-pair parts are mandatory when opening the connection, as they ensure that the other interactant can hear, but in closings, their absence can be tolerated if the participants have otherwise established that they can proceed towards the closing. Second, openings and closings mirror each other with regard to time. The opening ends with talk about the client's life in the past, while the closings start with talk about the client's life in the future. The fact that not everything from the client's past counts as institutionally relevant, combined with the recurrence of talk about the prospects of the care, seems to suggest that a certain kind of time-oriented talk is a fundamental aspect of managing these boundaries in the

context of tele-homecare. Through time-oriented talk, the participants enable the flow of care-relevant information from the client to the nurse and vice versa, thereby building the relationship between the nurse and the client as continuing and personal. Through these practices, the participants achieve and manage the sequential organisation of the boundaries of the encounter, and establish and dissolve a shared digital space for a certain kind of encounter.

This study has at least two clear limitations. First, due to the small dataset, some findings, especially regarding the potential disengagement, remain somewhat speculative. However, the central objective of institutional CA is to describe what practices are possible in a particular context, and the generalisability stems from comparisons between the settings (Peräkylä, 2004). We have hopefully demonstrated this. Second, we were only able to attain data from one location of each encounter. Thus, we could not study the *non-mutual interactional realities* (Ruhleder & Jordan, 2001) that technological mediation produces, nor the ways in which the distant participant received the turns. However, as pointed out by Olbertz-Siitonen (2015, 211–212), the participants do not have access to both ecologies, and adopting this "God's-eye view" might distance the analyst from the members' perspective. In our analysis, we have concentrated on phenomena that are available for the participants in their respective local ecologies. An analysis of openings and closings in VM settings with larger data sets and data from two perspectives (or more in multi-party settings) would offer important elaborations to the findings presented here.

In this article, we have described the sequential and multimodal features of openings and closings in Finnish tele-homecare encounters. These boundaries are managed through practices familiar from both mundane and institutional settings, and thus tele-homecare appears as an interesting hybrid of institutionality and informality. The centrality of visual appearings and disengagements suggests that, when available, visuality is an important element in the management of interaction and the relationship between the interactants in technologically mediated interaction. As it is envisioned that interpersonal contacts will be increasingly digitalised in the future, understanding how technology becomes sequentially consequential (Arminen et al., 2016) in the management of institutional tasks, the flow of interaction and intersubjectivity remain key questions for EM/CA.

References

Arminen, I., Licoppe, C., & Spagnolli, A. (2016). Respecifying mediated interaction. *Research on Language and Social Interaction, 49*(4), 290–309. https://doi.org/10.1080/08351813.2016.1234614

Broth, M., & Mondada, L. (2013). Walking away: The embodied achievement of activity closings in mobile interaction. *Journal of Pragmatics, 47*, 41–58. https://doi.org/ 10.1016/j.pragma.2012.11.016

Broth, M., & Mondada, L. (2019). Delaying moving away: Place, mobility, and the multimodal organization of activities. *Journal of Pragmatics, 148*, 44–70. https://doi.org/10.1016/j.pragma.2019.05.013

De Fornel, M. (1994). Le cadre interactionnel de l'échange visiophonique [International framework of the videophone exchange]. *Réseaux, 64*, 107–132. doi:10.3406/reso.1994.2473

Duuly, M., & Tudini, V. (2016). 'Now we are teachers': The role of small talk in student language teachers' telecollaborative task development. *Journal of Pragmatics, 102*, 38–53. https://doi.org/10.1016/j.pragma.2016.06.008

Drew, P., & Heritage, J. (1992). *Talk at work: Interaction in institutional settings.* Cambridge: Cambridge University Press.

Ekberg, S., & LeCouteur, A. (2014). Making arrangements in talk-in-interaction. *Text & Talk, 34*(4), 377–400.

Femø Nielsen, M. (2013). 'Stepping stones' in opening and closing department meetings. *Journal of Business Communication, 50(1)*, 34–67. https://doi.org/10.1177/0021943612465182.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics, 32(10)*, 1489–1522. https://doi.org/10.1016/S0378-2166(99)00096-X

Harjunpää, K., Mondada, L., & Svinhufvud, K. (2018). The coordinated entry into service encounters in food shops: Managing interactional space, availability, and service during openings. *Research on Language and Social Interaction, 51*(3), 271–291. https://doi.org/10.1080/08351813.2018.1485231

Hartford, B. S., & Bardovi-Harlig, K. (1992). Closing the conversation: Evidence from the academic advising session. *Discourse Processes, 15(1)*, 93–116. https://doi.org/10.1080/01638539209544803.

Heath, C., & Luff, P. (1992a). Collaboration and control. Crisis management and multimedia technology in London Underground line control rooms. *Computer Supported Cooperative Work (CSCW), 1*, 69–94.

Heath, C., & Luff, P. (1992b). Media spaces and communicative asymmetries: Preliminary observations of video-mediated interaction. *Human-Computer Interaction, 7,* 315–345.

Heritage, J., & Robinson, J. D. (2006). The structure of patients' presenting concerns: Physicians' opening questions. *Health Communication, 19*(2), 89–102. https://doi.org/10.1207/s15327027hc1902_1.

Hutchby, I. (2001). Technologies, texts and affordances. *Sociology, 35*(2), 441–456. https://doi.org/10.1177/S0038038501000219

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). Amsterdam/Philadelphia: P. A.: John Benjamins.

Kevoe-Feldman, H. (2015). Working the overall structural organization of a call: How customers use third position as leverage for gaining service representatives' assistance in dealing with service problems. *Language & Communication, 43,* 47–57. https://doi.org/10.1016/j.langcom.2015.05.001

Leydon, G. M., Ekberg, K., & Drew, P. (2013). 'How can I help?' Nurse call openings on a cancer helpline and implications for call progressivity. *Patient Education and Counselling, 92*, 23–30. https://doi.org/10.1016/j.pec.2013.02.007.

Lindström, A., & Heinemann, T. (2009). Good enough: Low-grade assessments in caregiving situations. *Research on Language and Social Interaction, 42*(4), 309–328. https://doi.org/ 10.1080/08351810903296465

Licoppe, C. (2017). Skype appearances, multiple greetings and '*coucou*'. The sequential organization of video-mediated conversation openings. *Pragmatics, 27*(3), 351–386. https://doi.org/10.1075/prag.27.3.03lic.

Licoppe, C., & Dumoulin, L. (2010). The 'curious case' of an unspoken opening speech act: A video-ethnography of the use of video communication in courtroom activities. *Research on Language & Social Interaction, 43*(3), 211–231. https://doi.org/ 10.1080/08351811003741319

Licoppe, C., & Morel, J. (2012). Video-in-interaction: 'Talking heads' and the multimodal organization of mobile and Skype video calls. *Research on Language & Social Interaction, 45*(4), 399–429. https://doi.org/10.1080/08351813.2012.724996

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). Fractured ecologies: Creating environments for collaboration. *Human-Computer Interaction, 18(*1), 51–84. https://doi.org/10.1207/S15327051HCI1812_3

Luff, P., Heath, C., Yamashita, N., Kuzuoka, H., & Jirotka, M. (2016). Embedded reference: Translocating gestures in video-mediated interaction. *Research on Language and Social Interaction, 49*(4), 342–361. https://doi.org/10.1080/08351813.2016.1199088

Mlynár, J., González-Martínez, E., & Lalanne, D. (2018). Situated organization of video-mediated interaction: A review of ethnomethodological and conversation analytic studies. *Interacting with Computers*, *30*(2), 73–84. https://doi.org/10.1093/iwc/iwx019

Mondada, L. (2001). *Conventions for multimodal transcription.* Tutorial available online: https://www.lorenzamondada.net/multimodal-transcription

Mondada, L. (2018). Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction, 51(1)*, 85–106. https://doi.org/10.1080/08351813.2018.1413878

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics, 145*, 47–62. https://doi.org/10.1016/j.pragma.2019.01.016.

Olbertz-Siitonen, M. (2015). Transmission delay in technology-mediated interaction at work. *PsychNology Journal, 13(2-3)*, 203–234.

Pappas, Y., & Seale, C. (2009). The opening phase of telemedicine consultations: An analysis of interaction. *Social Science & Medicine, 68*(7), 1229–1237. https://doi.org/10.1016/j.socscimed.2009.01.011

Peräkylä, A. (2004). Reliability and validity in research based on naturally occurring social interaction. In D. Silverman (Ed.) *Qualitative research.*

Theory, method and practice (pp. 283–304). London, U.K., Thousand Oaks, CA., & New Delhi, India: Sage Publications Ltd.

Raymond, G., & Zimmerman, D. H. (2016). Closing matters: Alignment and misalignment in sequence and call closings in institutional interaction. *Discourse Studies, 18*(6), 716–736.

Robinson, J. D. (1998). Getting down to business: Talk, gaze, and body orientation during openings of doctor–patient consultations. *Human Communication Research, 25*(1), 97–123. https://doi.org/10.1111/j.1468-2958.1998.tb00438.x

Robinson, J. D. (2001). Closing medical encounters: Two physician practices and their implications for the expression of patients' unstated concerns. *Social Science & Medicine, 53*, 639–656. https://doi.org/10.1016/s0277-9536(00)00366-x

Robinson, J. D., & Stivers, T. (2001). Achieving activity transitions in physician-patient encounters. *Human Communication Research, 27*(2), 253–298. https://doi.org/10.1111/j.1468-2958.2001.tb00782.x

Ruhleder, K., & Jordan, B. (2001). Co-constructing non-mutual realities: Delay-generated trouble in distributed interaction. *Computer Supported Cooperative Work, 10*, 113–138. https://doi.org/10.1023/A:1011243905593

Ruusuvuori, J (2001). Looking means listening: Coordinating displays of engagement in doctor-patient interaction. *Social Science & Medicine, 52(7)*, 1093–1108. https://doi.org/10.1016/s0277-9536(00)00227-6

Satar, H.M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL, 28(3)*, 305–325. https://doi.org/10.1017/S0958344012000286

Schegloff, E.A. (1968). Sequencing in conversational openings. *American Anthropologist, 70*(6), 1075–1095. https://doi.org/10.1525/aa.1968.70.6.02a00030

Schegloff, E.A. (1986). The routine as achievement. *Human Studies, 9*(2),111–151. https://doi.org/10.1007/BF00148124

Schegloff, E.A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*(4), 289–327. https://doi.org/10.1515/semi.1973.8.4.289.

Sidnell, J., & Stivers, T. (2013). *The handbook of conversation analysis.* Chichester, UK: Wiley–Blackwell.

Stommel, W., van Goor, H., & Stommel, M. (2019). Other-attentiveness in video consultation openings: A conversation analysis of video-mediated versus face-to-face consultations. *Journal of Computer-Mediated Communication, 24*, 275–292. https://doi.org/10.1093/jcmc/zmz015

Sävenstedt, S., Zingmark, K., Hydén, L., & Brulin, C. (2005). Establishing joint attention in remote talks with the elderly about health: A study of nurses' conversation with elderly persons in teleconsultations. *Scandinavian Journal of Caring Sciences 19*(4), 317–324.

Whalen, M. R., & Zimmerman, D. H. (1987). Sequential and institutional contexts in calls for help. *Social Psychology Quarterly, 50*(2), 172–185.

Woods, C. J., Drew, P., & Leydon, G. M. (2015). Closing calls to a cancer helpline: Expressions of caller satisfaction. *Patient Education and Counselling, 98*, 943–953. https://doi.org/10.1016/j.pec.2015.04.015

Zimmerman, D. H. (1992). The interactional organization of calls for emergency assistance. In P. Drew and J. Heritage (Eds.), *Talk at work. Interaction in institutional settings* (pp. 418–469). Cambridge: Cambridge University Press.

PUBLICATION III

Preserving client autonomy when guiding medicine taking in telehomecare: A conversation analytic case study

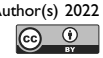
Ilomäki, S. & Ruusuvuori, J.

Nursing Ethics, 29(3), 719-732
https://doi.org/10.1177/09697330211051004

Publication reprinted with the permission of the copyright holders.

Check for updates

Original Manuscript

Nursing Ethics
2022, Vol. 29(3) 719–732
© The Author(s) 2022

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/09697330211051004
journals.sagepub.com/home/nej

Preserving client autonomy when guiding medicine taking in telehomecare: A conversation analytic case study

Sakari Ilomäki
Tampere University, Tampere, Finland

Johanna Ruusuvuori
Tampere University, Tampere, Finland

Abstract

Background: Enhancing client autonomy requires close coordination of interactional practices between nurse and client, which can cause challenges when interaction takes place in video-mediated settings. While video-mediated services have become more common, it remains unclear how they shape client autonomy in telehomecare. **Research aim:** To analyse how video mediation shapes client autonomy when nurses guide medicine taking remotely through video-mediated home care. **Research design:** This is a conversation analytic case study using video recordings of telehomecare encounters. The theoretical approach draws on ethnomethodology and empirical ethics. **Participants and research context:** Four home-dwelling older adults and three nurses participated in the data collection; data extracts include one client and two nurses. The study was conducted in Finland. **Ethical considerations:** Special attention was given to protect the rights of home care clients. An ethical statement for the study was given by the Ethics Committee of the Tampere Region **Findings:** Video mediation complicates interacting remotely with care-relevant artefacts because of nurses' limited visual access to the medicine and client's need to simultaneously engage in vocal interaction and medicine taking. This can be overcome by dividing the guidance into manageable steps which invite the client to explicate their readiness to take the medicine and situating the video-mediation equipment and medicine close together. Different interactional practices and ways of situating video-mediation equipment and medicine have consequences for client autonomy. **Discussion:** Understanding client autonomy in digitalised settings demands empirical examination that recognises the importance of different human and non-human aspects of care that shape client autonomy. **Conclusions:** To harness the benefits of video-mediated home care, communication technologies' reliance on home space and interactional practices should be recognised. Empirical ethics research is needed in order to make normative suggestions that fit a wide variety of care situation.

Corresponding author:
Sakari Ilomäki, Faculty of Social Sciences, Tampere University, 33014 Tampere, Finland.
Email: sakari.ilomaki@tuni.fi

Keywords
Autonomy, telehomecare, older adults, video-mediated interaction, qualitative research, Finland

Introduction

Digital technologies are envisioned to enhance client autonomy in home care for older adults. In Finland, the context of this study, the newest quality recommendations for services for older adults suggest that 'technology may support the initiative, independence and privacy of elderly patients' and 'help people lead healthier lives … independently and safely in their homes'.[1] One example of how digital telecare has been suggested to enhance client autonomy is through the deinstitutionalisation of care.[2] According to this concept, as the emphasis of care shifts from concrete, hands-on care to the management of services, care becomes more disembodied, resulting in greater client involvement and autonomy. However, especially privacy risks and the involuntary adoption of new devices have been recognised as risks for client autonomy.[3] As shown by these contrasting views, it remains unclear how the digitalisation of care will shape client autonomy. In this study, we approach client autonomy empirically, from the perspective of interaction dynamics and ask how video mediation (VM) shapes client autonomy in telehomecare encounters.

Here, we adopt the view that the ways in which nurses and older adults interact is central to shaping client autonomy.[4,5] For instance, the care practices that drive older adults to ask for help have been recognised as potential risks for client autonomy in residential care for older adults.[6] On the other hand, respecting others' rights to determine their own course of action has been found to enhance client autonomy.[7–12] This is achieved, for example, by using suggestions rather than straightforward orders in care work[9] and taking into consideration the guided person's ability to comply with the directions.[7,8] In sum, enhancing client autonomy requires close coordination of interactional practices between nurses and clients.

This coordination of activities can become challenging when interactions move to a VM environment. This is especially true of the limited visual access that cameras provide of the client setting, including medicines and other relevant artefacts.[12–15] In terms of vocal and non-vocal social interaction, it has been shown that health and social care professionals can overcome the limitations that VM produces and support client participation by talking casually about clients' daily lives and imitating eye contact by gazing directly into the screen,[15] by showing how the client should act through 'mimicable embodied demonstrations',[12] or by tolerating clients' missing turns that are not mandatory for maintaining shared focus and continuity of interaction.[13]

However, to our knowledge, medicine taking has not been studied in the telehomecare context, and the topic has not been connected to the broader discussion on client autonomy. Furthermore, research on geriatric nursing has generally emphasised care that takes place outside the home, leaving home care an under-researched area.[10,16,17] In this study, we aim to bridge these gaps by examining how VM shapes client autonomy in telehomecare when nurses guide medicine taking. Drawing on empirical ethics and ethnomethodology, we conduct a conversation analytic case study[18] to show that different ways of situating the VM equipment and medicine in the home space, along with the interactional practices that nurses use to guide clients in medicine taking, have an impact on clients' degrees of independent action and thus on client autonomy.

Theoretical approach: Empirical ethics and ethnomethodology

The concept and ideal of autonomy actualise differently depending on the ethical approach one takes to the issue. The *biomedical ethics perspective* emphasises the importance of formulating ethical guidelines in accordance with the principles of beneficence, non-maleficence, justice and autonomy. Autonomy is defined

negatively as the lack of hindrances for decision making, and it can be protected by ethical guidelines that diminish these hindrances.[19–21]

As a critique of the biomedical perspective's individualism, the *care ethics perspective* has emerged, emphasising the social and relational aspects of autonomy.[22,23] The care ethics perspective treats autonomy as enabled by people's interaction.[5,22–25]

Portraying itself as the successor of the care ethics perspective, the *empirical ethics perspective* expands this relationality by including not only other humans but also technologies and other artefacts, material infrastructure, various norms, values and ideals of care and different kind of practices as important parts in configuring autonomy.[26–28] In line with this perspective, several ethnographic studies on telehomecare have highlighted how digitalised care is dependent on the physical world, including artefacts like memory aids, and on social relationships.[29–31]

From the empirical ethics perspective, care and autonomy are accordingly understood as something that actualises in the constantly evolving and contextual relationships between human and non-human actors. Autonomy is not an abstract moral principle but a consequence of people's concrete actions in a given situation and thus something that varies with context.[26,28] That is, instead of asking how we can protect clients' autonomy from various threats through ethical guidelines, empirical ethics invites us to ask what good care consists of and how autonomy emerges in concrete care situations comprised of specific material, social and cultural features.[28] From the empirical ethics perspective, autonomy is thus viewed as situationally produced and context-specific.

To connect the idea of the contextual production of autonomy to our empirical data, we employ ethnomethodology as our general theoretical background. Similar to empirical ethics, ethnomethodology takes human action and the ongoing production of social context as the central foci of research.[32,33] Social organisation is regarded as built in and through ongoing actions rather than as something that pre-exists and explains those actions. Norms and rules of action always contain something that escapes explicit articulation, so social actions and situational sense-making merit analysis as they appear in everyday social interactions. Thus, instead of merely asking people what they think about autonomy, the ethnomethodological perspective examines how autonomy is 'talked into being'[33] as they are in telehomecare encounters. To study this process, we use the ethnomethodological method of conversation analysis which uses actual video-recorded encounters as data.

Data and method

Setting

The study was conducted in Finland in a home care unit undergoing a service pilot, where one of the daily home care visits was replaced with a video call from a nurse. The clients, older adults living at home, were provided with a tablet computer loaded with a simple program that allowed them to answer the calls[13]. The nurses used a computer in a shared office in the building where their social room was located. Figures 1 and 2 illustrate the two video-connected settings.

Participants

Three nurses and four home care clients participated in the data collection. We used convenience sampling to recruit both nurses and clients: all nurses who wanted to participate and all clients deemed eligible by the municipality for the service pilot and wanted to participate were recruited. Permission for research was obtained from the municipal council before beginning the recruitment process, and participants gave informed consent before joining the study.

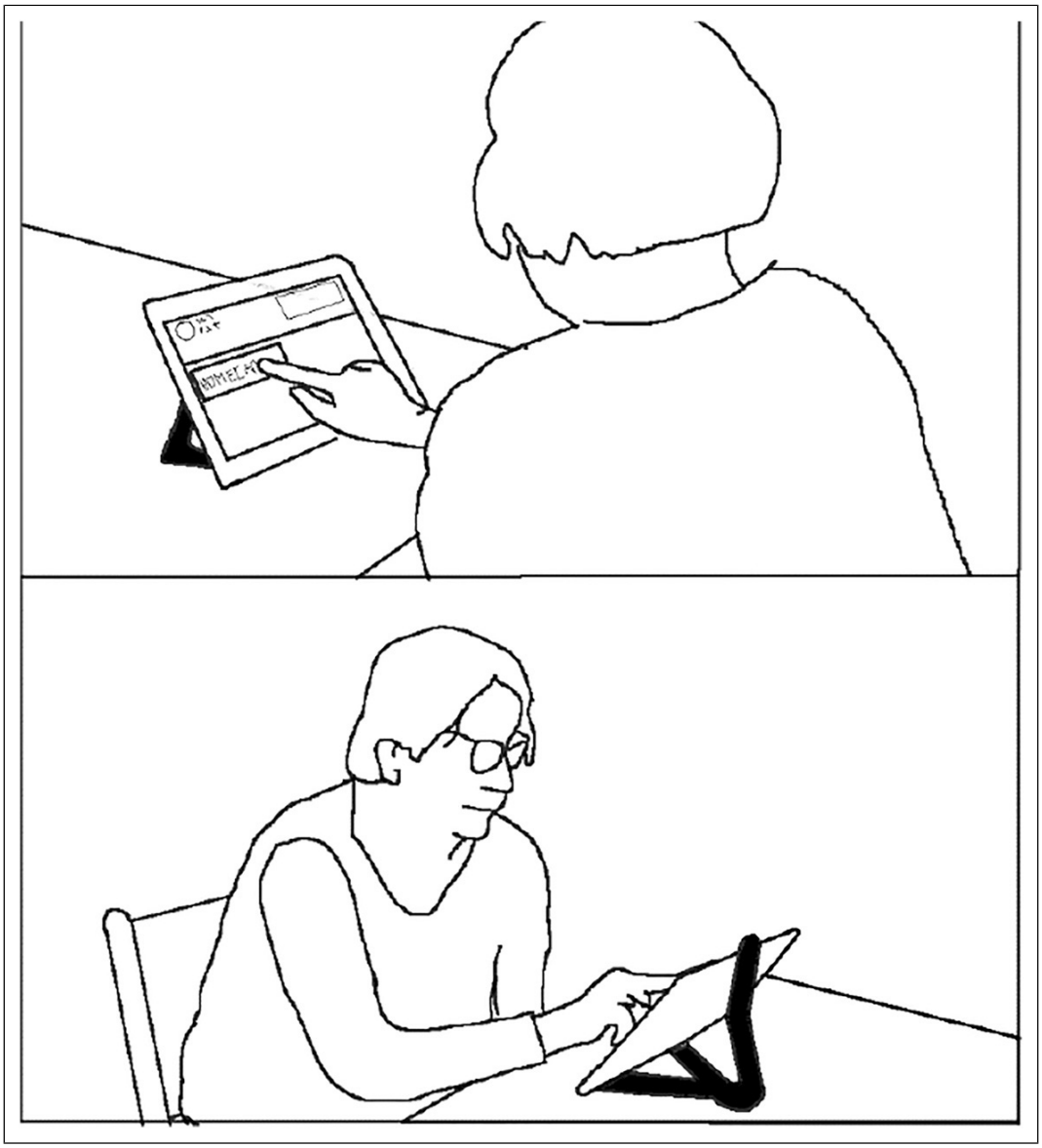


Figure 1. Client answering an incoming call in her home.

Figure 2. Nurse making a call at the office.

Ethical considerations

All clients lived alone in their own homes and had mild memory deficits. While this raises ethical dilemmas regarding informed consent,[34] it has been noted that people should be treated as capable of making decisions about their life to avoid stigmatisation[35] and that restricting their right to participate should be avoided unless their ability to decide can reasonably be doubted.[36] Thus, we carefully considered how to balance the right to participate and the right for protection[34] with a variety of procedures. *First*, older adults with only mild memory deficits were recruited by the municipality for the pilot and thus this study. *Second*, the participants were given information on the study both orally and in writing, with the opportunity to ask questions before deciding to participate. The language and typography of the information leaflet and the consent form were designed to be simple; for example, they used plain language. Meanwhile, Author 1 explicitly asked each client whether he or she understood the content. *Third*, the participants were interviewed before the recordings and individuals with obvious problems of understanding or remembering (e.g. if the client forgot who the researcher was) were excluded from the study despite their having provided consent. *Fourth*, we used a research method that foregrounds the choices of action and interpretations of the participants themselves within the situation studied. Thus, the focus shifts from studying the potential problems caused by memory

deficits within the ongoing interaction to the observable competences of the participants in successfully acting in relevant ways in challenging situations such as VM encounters.

The study was conducted according to the relevant measures for data protection[37] and the ethical guidelines for human studies by the Finnish National Board on Research Integrity TENK.[38] An ethical statement was obtained from the Ethics Committee of the Tampere Region (document number 49/2017)

Data

The data consist of video recordings of 14 VM telehomecare encounters. Twelve encounters were recorded in the nurses' office and two in clients' homes by Author 1. For the case study presented in this article, data are drawn from two encounters with one client and two different nurses. The data collection and analysis were part of the research project Healthcare workers in the eye of the digital turbulence, conducted by Tampere University and the Finnish Institute of Occupational Health, with funding from the Finnish Work Environment Fund.

Data analysis

The data were analysed using multimodal conversation analysis (CA), which is an inductive qualitative research method that studies recurring patterns and structures of interaction and aims to uncover the interlocutors' own perspective and interpretations of one another's conduct by drawing on their observable next actions.[39] Attention is focused particularly on how interlocutors link their verbal and nonverbal actions with the other participant's actions to construct broader activities, such as medicine taking.[40,41] This requires a detailed scrutiny of participant interactions in authentic video-recorded encounters and meticulous transcription of the words, gestures, and tones and pitches of voice used by all participants. The video-recorded encounters were originally transcribed according to CA conventions[42,43] and streamlined for the present article. Transcripts in Finnish with translations may be found in Supplementary Material 1, while transcription symbols are explained in Supplementary Material 2.

During transcription, we identified medicine taking as a potential phenomenon of interest regarding client autonomy as it involves both directing the client's actions and distant collaboration with the care-relevant artefacts (the medicine). We first analysed how the transition from previous activity to medicine taking is performed before analysing how the nurses guided their clients in taking the medicine. This stage included an analysis of both vocal and non-vocal behaviour and of how VM and non-mutual access to the medicine affected the coordination of the participants' actions in these sequences. Finally, this descriptive analysis was complemented by a more interpretative analysis of how autonomy is enacted in these sequences.

The present article draws on insights from the analysis of all 14 encounters, but we concentrate on a case study of two instances of medicine taking involving the same client. These instances were selected because they most clearly exemplify the dilemmas that arise in this process. In the CA research tradition, case studies are a widely employed research strategy that enables a context-sensitive analysis of the turn-by-turn/action-by-action process of performing institutional tasks.[18,41,44] Thus, the approach is suitable for analysing complex interactional processes such as guiding medicine taking in VM telehomecare.

Results

Our analysis shows two ways of guiding clients' medicine taking: a straightforward approach (shown in Extract 1 in Supplementary Material) and a stepwise transition into guiding (Extract 2 in Supplementary Material), each followed by different interactional consequences. The use and usability of these different practices are closely connected to the different ways in which VM technology is situated in the home space.

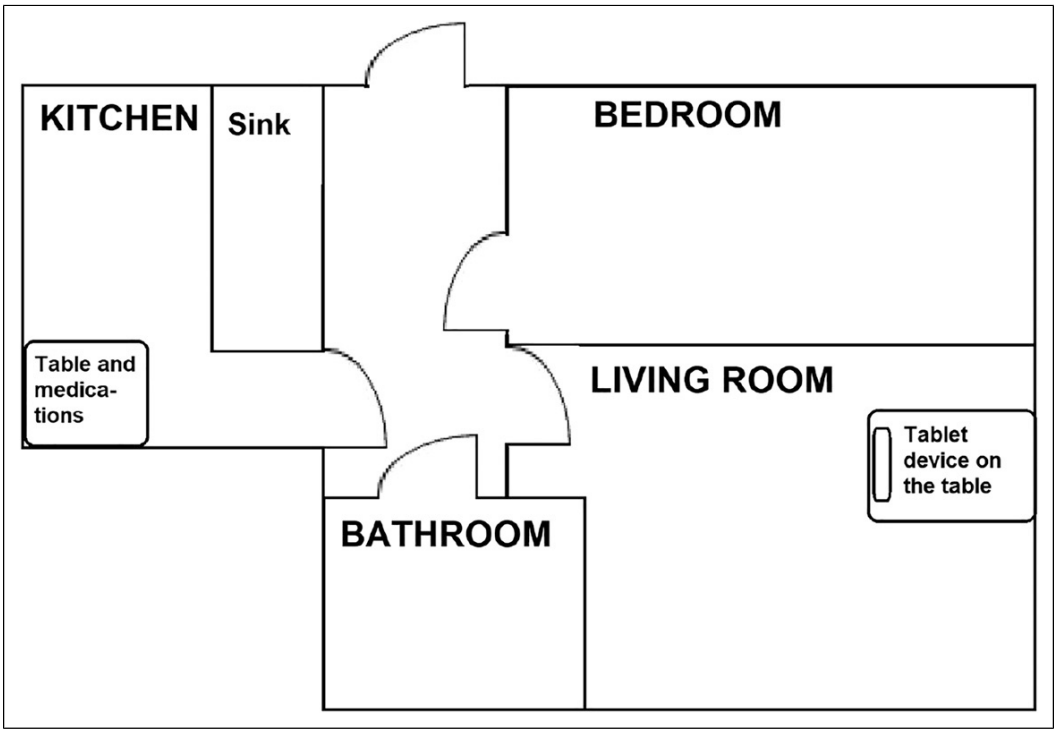


Figure 3. Plan of the client's home in the original configuration.

We also show how each practice, combined with the material conditions of the encounter, results in different possibilities for client autonomy.

Extract 1 demonstrates how situating the VM equipment and medicine some distance from each other, so that the client cannot see them both at the same time, can cause interactional dysfluency and decrease client autonomy, especially when straightforward guiding is used. In the situation in Extract 1, the tablet computer is located in the client's living room, while the medicine is in the kitchen. Figure 3 depicts the home space. Data were recorded at the client's home, and the client is shown in the images in the transcript sitting in front of the tablet computer. Each image contains two angles of the client's home, one from behind showing the tablet screen (the upper parts of the pictures) and one from the front (the lower parts). The timing of the pictures in relation to talk are marked with hash signs (#) in the transcripts. Before Extract 1, the nurse and the client have discussed how the client's day has been, have closed the topic, and are now ready to move on to the next activity, medicine taking.

As the interaction unfolds, the ways in which the spatial arrangements shape the interaction become apparent. Since the VM equipment and the medicine are far away from each other, the client cannot simultaneously direct her attention to the ongoing interaction with the nurse and to the medicine. This leads to interactional dysfluency as the interactants progress to medicine taking. Despite the fact that both participants should consequently be aware that a new topic or activity will follow next, the client appears to be surprised by the nurse's guiding turn (line 1). During the long pause after the nurse's turn (line 2), the client raises her eyebrows, frowns (IMG 1.2), a potential sign of an upcoming problem,[45] and says *jasso*

(line 3), an expression with its roots in Swedish, where it is often used in receiving new information.[46] The nurse's guiding turn *well would you go and take that evening medication of yours* (line 1) projects the client's acceptance or refusal of the suggested action, but the client's *jasso* does neither. Instead, it seems to treat the nurse's turn as new information, as if the client did not expect this activity to become topical.

The nurse partially repeats her guiding turn, omitting the object of the suggested activity (the medicine (line 5)), thus treating herself as entitled to guide and the client as able and willing to comply. Overlapping with the nurse's turn, the client asks where the medicine is, thus postponing her response to the guiding turn (lines 6–7). This turn makes salient that the client lacks the knowledge necessary to either comply with or reject the suggested activity. By preparing to stand up (line 7, IMGs 1.5 and 1.6), the client also expresses her preliminary alignment with the suggested activity and her understanding that she must exit the encounter momentarily to take the medicine. After the nurse has answered the question (line 12) and the client has the adequate knowledge to proceed, she stands up and walks to another room to take the medicine.

From the perspective of client autonomy, we see how the nurse originally treats the client as autonomous, but the physical disposition of the VM equipment and the medicine leads to a rupture of autonomy. By transitioning directly to medicine taking, the nurse treats medicine taking as a routine procedure which does not threaten the client's autonomy. As the nurse designs the initial guiding turn as a conditional question (instead of, for example, using the imperative mood), she treats the client as having some control over the task and at least ostensibly able to act differently (line 1). When the nurse only partially repeats the guiding turn, she shows she recognises that the client has heard and understood the necessary parts of the guiding turn. Furthermore, the nurse does not adjust guiding to the client's readiness to comply by asking, for example, if she *could* take the medicine. This further shows her treating herself as entitled to guide, the client as able and willing to comply and the whole question of medicine taking as non-problematic and not threatening the client's autonomy. However, the material setting, which situates the VM equipment and the medicine in the home, hinders the client's ability to engage simultaneously in the interaction and the medicine taking. During the nurse's repetition, the client starts to search for the medicine and then verbalises her search with a question (lines 6–7). So, despite the nurse's orientating herself to the client's autonomy, the combination of direct guiding to medicine taking and this specific material setting push the client to ask for help, which can jeopardise her autonomy.

Extract 2 demonstrates how moving the medicine and the VM equipment closer to each other and using a stepwise approach to medicine taking enable a smoother process of medicine taking and support client autonomy. The data are recorded from a different encounter with the same client as in Extract 1, but the nurse is different. The client's tablet computer has been moved from the living room to the kitchen table, as depicted in Figure 4. Before the extract, the nurse and the client have been talking about the client's plans for the week. Unlike in Extract 1, the data in Extract 2 were recorded at the nurse's office, and the client now appears on the screen of the computer. Due to audio problems, the nurse is holding the speaker-microphone in her right hand (The camera microphone used in data collection blocks parts of the nurse's face.)

Compared to Extract 1, the new arrangement enables the client to simultaneously engage in interaction with the nurse and locate the medicine in her physical surroundings. Furthermore, compared to the earlier extract, the nurse now initiates medicine taking in three steps. First, she topicalises the medicine, asks for the client's confirmation (line 1) and inquires whether the client has taken the medicine (line 2). The nurse delays her guiding turn at this point. Second, already during the nurse's turn, the client starts to align with the activity initiated by the nurse in her bodily interaction. The client shifts her gaze first towards the nurse (IMG 2.2) and then down to her right, the presumed location of the medicine (IMG 2.3), reaches in that direction with her right arm (IMG 2.4) and leans towards the medicine (IMG 2.5). After this, the client answers the nurse's question (lines 4–10), which the nurse receives with laughter (line 12). It is only after this establishment of

shared orientation to the medicine and thus the client's ability to take it that the nurse finally deploys the third part, the actual guiding turn: proposing that the client take the medicine (line 15). This is followed by a clarification about the night-time medicines that the client could take (lines 16 and 18).

This stepwise entry into medicine taking, consisting of (a) the nurse's questions (which invite the client's perspective display in lines 1–2), (b) the client's response (which aligns with the task of taking the medicine in lines 4, 8, and 10, and IMGs 2.2–2.5) and (c) the nurse's guiding turn (line 15), enables the participants to move to medicine taking without breaks in interaction. By dividing medicine taking into manageable steps, the nurse enables working with physical objects in a situation where the participants have unequal access to this resource for interaction. Unlike in Extract 1, the participants preserve client autonomy throughout the action sequence in Extract 2. The combination of the spatial arrangement, where the client has visual access to the medicine while engaged with the tablet device and the tele-encounter, and the nurse's practice of dividing medicine taking into manageable steps allow the client to take her medicine without difficulties and thus appear more autonomous and able. Furthermore, the nurse formulates her guiding turn as a possibility for the client to act upon, thus reinforcing her own lower entitlement to control the activity.

Discussion

Our analysis shows that the ways in which the VM equipment is situated in the home and the practices that the nurses use to guide the client shape the interaction dynamics and the client autonomy in VM telehomecare. When the VM equipment and the medicine are located apart from each other, the client cannot simultaneously engage in both vocal interaction and medicine taking, which can lead to interactional dysfluencies (Extract 1). As previous studies of other contexts of VM encounters have shown,[12,14] due to the limited visual contact and spatial arrangements that situate VM equipment and medicine far from each other, VM hinders participants' collaboration with care-relevant artefacts and the creation of a shared understanding of relevant activities involving these artefacts. This complicates nurses' opportunities to support clients' independent actions and hence client autonomy. In our analysis, this happened as the arrangement of VM equipment and guiding practices pushed the client to ask for help, which has been recognised as a potential threat for situational autonomy in residential care.[6] Situating the equipment and medicine closer to each other eliminated the problem of simultaneous engagement in interaction and medicine taking (Extract 2). Furthermore, compared to straightforward guiding, the stepwise approach portrays the client as a knowledgeable participant who thus has more time to locate the medicine. The stepwise approach also enables the participants to build a shared understanding about the client's readiness to take the medicine through perspective display, which helps coordinate medicine taking in a VM environment. Similar practices of stepwise entry to perspective display have been recognised as a way to introduce delicate topics in conversation,[47] such as determining whether a given piece of advice is appropriate.[9] When comparing our findings to this earlier research, we suggest that stepwise entry into medicine taking is one way through which both participants can collaborate remotely with care-relevant artefacts, despite the limitations of VM, and work around the delicacy of guiding another person. These features – supporting shared understanding of the ongoing action and leveraging collaboration with artefacts remotely through perspective displays – worked in our data to support independent and fluent participation of the client, that is, autonomy.

In line with earlier research that follows the empirical ethics perspective,[28–31] our findings show how the production of autonomy in VM telehomecare depends on the physical world, care-relevant artefacts and social interaction. For this kind of analysis, the empirical ethics perspective offered important viewpoints that could have been missed if the biomedical or relational perspective had been chosen as the starting point of the analysis. The general guidelines of the biomedical ethics perspective are not sufficient to reveal the intricate local negotiation of autonomy in situ, while the care ethics perspective, although importantly focusing on

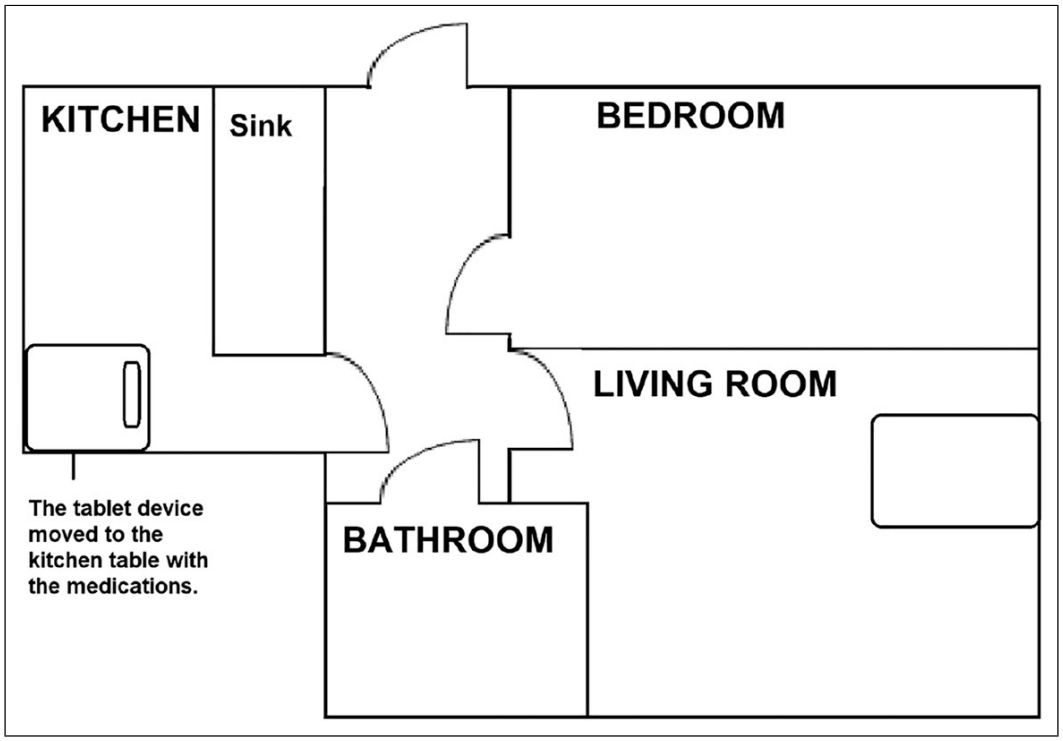


Figure 4. Plan of the client's home in the changed configuration.

social relational aspects of the situation, leaves out the material affordances available to the participants and relevant to the care situation's ongoing negotiations. Thus, the extensive context sensitivity embedded in the empirical ethics' perspective – the sensitivity to human action and social relations in situ, technologies and other artefacts, including the material infrastructure – proved to be essential for investigating questions of autonomy as they appear for participants in authentic encounters of social and health care. While other perspectives might be suitable for considering other ethical dilemmas, such as drafting guidelines for service provision at the level of policy making, we argue that empirical ethics appears to be a more appropriate alternative for analysing the daily production of client autonomy.

Uncertainty about how the digitalisation of care shapes client autonomy remains an issue. While the ideals of disembodied care[2] include the possibility of more active and autonomous clients, telecare also poses risks to client autonomy: telecare can enable independent living in the home environment but can simultaneously lead to situations where clients' autonomy is threatened at the level of interaction; there are also critical perspectives on the idea of ageing in place.[48,49] From the empirical ethics viewpoint, there is no clear relationship between new technologies and autonomy. Instead, the VM equipment, VM interactional practices, care-relevant artefacts and home spaces create complex arrangements in which client autonomy is interacted into being. In this complex setting, the changes that new technologies can bring to client autonomy are nuanced and context-specific. Thus, they call for comparison of situational practices of good care, in order to describe the processes through which different applications of new technologies can simultaneously support and hinder client autonomy. As a consequence, in addition to discussing risks regarding privacy and

involuntary use of new technologies,[3] understanding client autonomy in digitalised settings demands empirical examination of the spatial, material and interaction-related aspects of care within actual telehomecare encounters.

Methodological reflections

Conversation analysis case study design carries both limitations and advantages. The small dataset could limit the generalisability of our findings. However, CA of institutional encounters aims not only at finding generalisable practices but also at describing what kind of practices are possible in a specific context.[50] In this research model, generalisability stems from comparing the findings from different contexts, in our case research on client autonomy specifically and studies employing the empirical ethics perspective more generally. This theoretical sampling has enabled us to participate in discussions beyond our empirical cases. The empirical ethics perspective calls for context-sensitive analysis, which is a recognised strength of CA case study design.

Conclusions

Video mediation can complicate interactions with care-relevant artefacts, which further shapes client autonomy. In our data, the nurse used stepwise transition to invite the client to explicate their perspective before using the artefacts. This practice appeared to overcome some of the limitations associated with VM and could be applied as a good practice to soothe the interaction and support client autonomy. The use and usability of VM technology as part of telehomecare depends on the home space as a whole. Therefore, when implementing new devices, attention needs to be paid to situating the device in the home space and arranging other care-relevant artefacts as an ensemble that supports the central activities of care. By taking video-recorded data of everyday telehomecare encounters as our starting point, adopting the empirical ethics perspective, and analysing a sample of the interactional work that is necessary for participants in negotiating medicine taking, we have shown how this kind of analysis is potent in revealing the involvement of the participants, making visible the interactional work that is needed to 'put new normative suggestions to work'.[28] Accordingly, we encourage practitioners and researchers to engage in studies that draw data from real-life care encounters to make normative suggestions that fit a wide variety of care situations.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Finnish Work Environment Fund grant number 117151 and Jenny and Antti Wihuri Foundation.

ORCID iD

Sakari Ilomäki https://orcid.org/0000-0003-1791-7117

Supplemental Material

Supplemental material for this article is available online.

References

1. Finish Ministry of Social Affairs and Health. *Quality recommendation to guarantee a good quality of life and improved services for older persons 2020–2023: the aim is an age-friendly Finland*. Publications of the Ministry of Social Affairs and Health, 2020, Vol. 37, http://urn.fi/URN:ISBN:978-952-00-8427-1 (16 December 2020, accessed 5 March 2021).
2. Bashshur R, Reardon TG and Shannon GW. Telemedicine: a new health care delivery system. *Annu Rev Pub Heal* 2000; 21: 613–637.
3. Sundgren S, Stolt M and Suhonen R. Ethical issues related to the use of gerontechnology in older people care – a scoping review. *Nurs Ethics* 2019; 27(1): 88–103.
4. Scott PA, Välimäki M, Leino-Kilpi H, et al. Perceptions of autonomy in the care of elderly people in five European countries. *Nurs Ethics* 2003; 10(1): 28–38.
5. Jacobs G. Patient autonomy in home care: nurses' relational practices of responsibility. *Nurs Ethics* 2019; 26(6): 1638–1653.
6. Moilanen T, Kangasniemi M, Papinaho O, et al. Older peoples perceived autonomy in residential care: an integrative review. *Nurs Ethics*. 2020; 28(3): 414–434.
7. Antaki C and Kent A. Telling people what to do (and, sometimes, why): contingency, entitlement and explanation in staff requests to adults with intellectual impairments. *J Pragmatics* 2012; 44(6–7): 876–889.
8. Curl TS and Drew P. Contingency and action: a comparison of two forms of requesting. *Res Lang Soc Interac* 2008; 41(2): 129–153.
9. Heritage J and Lindström A. Advice giving – terminable and interminable: the case of British health visitors. In: Limberg H and Locher MA (eds) *Advice in discourse: pragmatics & beyond new series 221*. Philadelphia, PA: John Benjamins Publishing, 2012, pp. 169–194.
10. Kindell J, Keady J, Sage K, et al. Everyday conversation in dementia: a review of the literature to inform research and practice. *Int J Lang Comm Dis* 2017; 52(4): 392–406.
11. Lindholm C. Parallel realities: the interactional management of confabulation in dementia care encounters. *Res Lang Soc Interac* 2015; 48(2): 176–199.
12. Due BL, Lange SB, Nielsen MF, et al. Mimicable embodied demonstration in a decomposed sequence: two aspects of recipient design in professionals' video-mediated encounters. *J Pragmatics* 2019; 152: 13–27.
13. Ilomäki S and Ruusuvuori J. From appearings to disengagements: openings and closings in video-mediated tele-homecare encounters. *Social Interaction. Video-Based Studies of Human Sociality* 2020; 3(3).
14. Luff P, Heath C, Kuzukoka H, et al. Fractured ecologies: creating environments for collaboration. *Hum Comput Interact* 2003; 18(1): 51–84.
15. Sävenstedt S, Zingmark K, Hydén L-C, et al. Establishing joint attention in remote talks with the elderly about health: a study of nurses' conversation with elderly persons in teleconsultations. *Scand J Caring Sci* 2005; 19(4): 317–324.
16. Clancy A, Simonsen N, Lind J, et al. The meaning of dignity for older adults: a metasynthesis. *Nurs Ethics* 2020; 28(6): 878–889.
17. Haahr A, Norlyk A, Martinsen B, et al. Nurses experiences of ethical dilemmas: a review. *Nurs Ethics* 2020; 27(1): 258–272.
18. Schegloff EA. Analysing single episodes of interaction. An exercise in conversation analysis. *Soc Psychol Quart* 1987; 50(2): 101–114.
19. Beauchamp TL and Childress JF. *Principles of biomedical ethics*. New York: Oxford University Press, 2001.
20. Stirrat GM and Gill R. Autonomy in medical ethics after O'Neill. *J Med Ethics* 2005; 31: 127–130.
21. O'Neill O. *Autonomy and trust in bioethics*. Cambridge, UK: Cambridge University Press, 2002.
22. Verkerk MA. The care perspective and autonomy. *Med Health Care Philos* 2001; 4: 289–294.

23. Killackey T, Peter E, MacIver J, et al. Advance care planning with chronically ill patients: a relational autonomy approach. *Nurs Ethics* 2019; 27(2): 360–371.
24. MacDonald C. Nurse autonomy as relational. *Nurs Ethics* 2002; 9(2): 194–201.
25. Santos Salas A and Cameron BL. Ethical openings in palliative home care practice. *Nurs Ethics* 2010; 15(5): 655–665.
26. Musschenga AW. Empirical ethics, context sensitivity, and contextualism. *J Med Philos* 2005; 30(5): 467–490.
27. Parker M. Two concepts of empirical ethics. *Bioethics* 2009; 23(4): 202–213.
28. Pols J. Towards an empirical ethics in care: relations with technologies in health care. *Med Health Care Philos* 2015; 18: 81–90.
29. López Goméz D. Little arrangements that matter. Rethinking autonomy-enabling innovations for later life. *Technol Forecast Soc* 2015; 93: 91–101.
30. López D and Domènech M. Embodying autonomy in a Home Telecare Service. *Sociol Rev* 2008; 56(2): 181–195.
31. Sánchez Criado T and Domènech M. Older people in a connected autonomy? Promises and challenges in the technologisation of care. *Rev Esp Investig Sociol* 2015; 152: 105–120.
32. Garfinkel H. *Studies in ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall, 1967.
33. Heritage J. *Garfinkel and ethnomethodology.* Cambridge, UK: Polity Press, 1984.
34. Dewing J. Process consent and research with older persons living with dementia. *Res Ethics* 2008; 4(2): 59–64.
35. Wilkinson H (ed). *Perspectives of people with dementia: research methods and motivations*. London: Jessica Kingsley Publishers, 2001.
36. Mäki-Petäjä-Leinonen A. (2006) *Dementian vaikutuksesta oikeudelliseen toimintakykyyn*. [On the effects of dementia on legal agency]. In: Topo P (ed) Eettiset kysymykset vanhustenhuollon tutkimuksessa *[Ethical considerations in care research for older adults]. Stakesin työpapereita 2006: 21*. National Research and Development Centre of Welfare and Health. (26 January 2006, accessed 5 March 2021).
37. Office of the Data Protection Ombudsman. *Guidelines of the European Data Protection Board*. https://tietosuoja.fi/en/guidelines-of-the-european-data-protection-board (accessed 5 August 2021).
38. Finnish Advisory Board on Research Integrity. *Responsible conduct of research and procedures for handling allegations of misconduct in Finland*, 2012. https://tenk.fi/en/advice-and-materials/RCR-Guidelines-2012 (14 November 2012, accessed 5 August 2021).
39. Sidnell J and Stivers T (eds). *The handbook of conversation analysis*. Chichester, UK: Wiley-Blackwell, 2013.
40. Mondada L. Contemporary issues in conversation analysis. Embodiment and materiality, multimodality and multisensoriality in social interaction. *J Pragmatics* 2019; 145: 47–62.
41. Ten Have P. *Doing conversation analysis*. London: SAGE, 2007.
42. Jefferson G. Glossary of transcript symbols with an introduction. In: Lerner GH (ed) *Conversation analysis: studies from the first generation*. Philadelphia, PA: John Benjamins Publishing, 2004, pp. 13–31.
43. Mondada L. Conventions for multimodal transcription, https://www.lorenzamondada.net/multimodal-transcription (2001, accessed 5 March 2021).
44. Peräkylä A and Vehviläinen S. Conversation analysis and the professional stocks of interactional knowledge. *Discourse Soc* 2003; 14(6): 727–750.
45. Kaukomaa T, Peräkylä A and Ruusuvuori J. Foreshadowing a problem: turn-opening frowns in conversation. *J Pragmatics* 2014; 71: 132–147.
46. Lindholm C. Boundaries of participation in care home settings: use of the Swedish token *jaså* by a person with dementia. *Clin Linguist Phon* 2016; 30(10): 832–848.
47. Maynard DW. Perspective-display sequences in conversation. *West J Speech Commun* 1989; 53: 91–113.

48. Lowenstein A. Elder abuse and neglect: 'old phenomenon': new directions for research, legislation, and service developments. *J Elder Abuse Negl* 2009; 21: 278–287.
49. Andrews GJ, Cutchin M, McCracken K, et al. Geographical gerontology: the constitution of a discipline. *Soc Sci Med* 2007; 65: 151–168.
50. Peräkylä A. Reliability and validity in research based on naturally occurring social interaction. In: Silverman D (ed.) *Qualitative research: theory, method and practice*. London: SAGE, 2004, pp. 283–304.

Supplementary material 1: Data extracts with original Finnish data and English translations

Extract 1: Straightforward guiding to take the medicine and the lack of visual cues lead to confusion

N: **[ihan] kiva.**
[pretty] nice.
C: [mm.]
(.)
C: mm?
(0.2)
C: **niin o.**
it is.
(0.5)
N: **joo.**
yes.
(0.4)
N: ↑**joo:.**
↑ye:s.

IMG 1.1

IMG 1.2

01 N: **.hh (.) #tuota, (.) menisiks sää sen: sun %iltalääkkeen# ottaa.**
.hh (.) #well, (.) would you go and take that evening %medicine# of yours.
c: %raises eyebrows
#IMG 1.1 #IMG 1.2

IMG 1.3

02 (0.7)#%(2.4)
c: %frowns
#IMG 1.3

03 C: **.mt jasso.**

04 (1.3) ((The client stares at the screen of the tablet.))

IMG 1.4

05 N: **jo[o.] [kä]yks=s[ä ot%] t a[#massa,]**
ye[s.] [wou]ld=yo[u go%]and [#take,]
06 C: **[(n]o [m-] [mis-%] [#missä-] (.) missäs ne nyt**
[we]ll [wh-] [whe-%] [#where-] (.) where are they
c: %turns to right
#IMG 1.4

IMG 1.5

IMG 1.6

07 C: **sitten onkaan.#=**
again.#=
#IMG 1.5

08 C: **=ne on sie-**
=they are there-

09 (0.5)

10 C: .hhhh

11 (0.6)

12 N: **[#ne on siel kei]ttiön <pöydälläh.>**
[#they are there on the kit]chen <tableh.>

13 C: **[# °het-kine?°]**
[# °just a moment?°]
#IMG 1.6

14 N: (0.2)%(0.5)
c: %stands up and walks to kitchen->

IMG 1.7

15 C: #**aha?**
#oh?
#IMG 1.7

16 (0.5)

17 C: **no minäpäs kävele.**
well I'll go.

18 (0.8)

Extract 2: Stepwise entry into medicine taking affords a seamless transition.

IMG 2.1

IMG 2.2

01 N: **.hhh #sulla,(.) on siellä % ↑ilta#lääkkeet eikö vaan.=**
.hhh #you, (.) have the %↑evening# medicine there don't you.=
c: %shifts gaze to nurse and then down to right
#IMG 2.1 #IMG 2.2

IMG 2.3

IMG 2.4

02 N: **=ni? (.) #%ootko sää #ne=jo ottanu.**
= so? (.) #%have you #taken=them already.
c: %leans to right and reaches with right hand
#IMG 2.3 #IMG 2.4

03 (1.5)

IMG 2.5

04 C: **#emmää=nyt ↑iltaa vielä oo ottanu.**
#well=I haven't taken the ↑evening ones yet.
#IMG 2.5

05 (0.2)

06 N: **↑ni?**
↑yes?

07 (1.0)

08 C: **(-) ((*joo /jaha*))**
(yeah.)

09 (.)

10 C: **nyt on kaikissa purkissa jotai.**
now every jar has something.

11 (0.5)

12 N: **£o(h)ok(h)e:i?£**
£o(h)ok(h)a:y?£

13 (0.3)

14 C: **h[h] [heh heh heh] [(-)]**[((inaudible))]
15 N: **[.h]hh (.) sää [voisit mel]kein ne il[tal][ääkkeet ottaa.]**
[.h]hh (.) you [could actua]lly take t[he][evening medicine.]

16 N: **ja [jätt]ää si[tte-] (0.2) yölääkkeet:=sinne, (0.4) ennen kun menet**
and [leav]e th[en-] (0.2) night medicine:=for, (0.4) when you go to
17 C: [(-)a] [(-)]

18 N: **n[u k]kumaa.**
b[ed]
19 C: **[(ni?)]**
[(yeah?)]

20 (1.4)

Supplementary material 2: Transcription symbols in the data extracts

[word]	Brackets: onset and offset of overlapping talk
=	Equals sign: contiguous utterances, second is latched immediately onto the first
(0.2)	Timed interval within or between utterances, measured in seconds and tenths of seconds
(.)	Interval of less than 0.2 s
wo:rd	Colon: extension of the sound or syllable
.	Full stop: falling intonation
,	Comma: continuing intonation
?	Question mark: rising intonation1
↑	Upward arrow: Rising pitch
wo-	Dash: abrupt cut-off
WORD	Capital letters: louder volume
<word>	Slower-paced talk than the surrounding talk
°word°	Degree signs: quieter volume
hh	Audible aspiration.
hh	Audible inhalation
w(h)ord	Laughter
(----)	Lines in paranthesis: Unclear and unidentifiable talk
((word))	Text in parentheses: transcriber's comments
%	Percentile sign: The clients bodily actions
#	Hash sign: The timing /location of the images

PUBLICATION IV

Distributed cognition in fractured ecologies: Collaborative problem solving in video-mediated interaction

Ilomäki, S. & Stevanovic, M.

Article submitted and under review