Social Interaction

Video-Based Studies of Human Sociality

_____

Distributed Cognition in Fractured Ecologies:

Collaborative Problem-Solving in Video-Mediated Interaction

_____

Sakari Ilomäki & Melisa Stevanovic

Tampere University

Abstract

In this article we present a conversation analytic single-case study of a hybrid video-mediated teleconsultation in which the participants solve a problem with the audio connection. During the problem-solving process, the interactants engage in probing for a solution, reframing the problem and experimenting with the affordances of both video mediation technology and telephones. The hybrid configuration poses challenges to these processes, since access to other participants' conduct (e.g., talk and gestures) and physical surroundings are limited. Participants overcome this by fitting interactional practices to the communicative media available at different moments to enable directing attention and action in a co-present or mediated ecology of action. Complementing a distributed cognition outlook with a conversation analytic perspective on participation and multimodality, we propose refinements to existing theories of problem-solving in an effort to develop integrative approaches to problem-solving in the wild.

Keywords: real-world problem-solving, video-mediated interaction, computer-mediated communication, distributed cognition

1. Introduction: Cognition and Problem-Solving

Problem-solving involves working one's way from some current problem state to another more desirable state. Two main approaches to cognitive processing in problem-solving have been proposed; the *gestalt perspective* (Koffka, 1935; Wertheimer, 1982) and *information processing theory* (Gigerenzer & Todd, 2001; Öllinger & Goel, 2010). From the gestalt perspective, problem-solving is regarded as productive thinking in which the solver restructures the relations between the problem constituents in new ways, aiming to transform the problem gestalt into a good gestalt (Öllinger & Goel, 2010). By contrast, information processing theory conceptualises problem-solving as computational, with the transition from the initial problem state to a goal state advanced by the application of a series of operations that lead to intermediate states (Gigerenzer & Todd, 2001). Both perspectives have benefits and limitations, and there have been calls to integrate elements from both (Weisberg, 2015; Öllinger & Goel, 2010).

Compared to problem-solving by individuals in laboratory settings, on which research into the cognitive processes of problem-solving has predominantly concentrated (Sarathy, 2018), spontaneous collaborative problem-solving outside the laboratory context is more complicated for three basic reasons. First, while problems in laboratory settings are often clearly defined, spontaneous problem-solving often involves dealing with ill-defined problems or multiple problems simultaneously (Sarathy, 2018; Steffensen, 2013; Weisberg, 2015). Second, it requires the creative use of and interaction with the physical environment and affordances of this environment (Sarathy, 2018; Tarasmundi & Linell, 2017; Vallée-Tourangeau, 2014) – that is, the possibilities and limitations that the physical environment sets for actions (Hutchby, 2001; Gibson, 1979) – while in laboratory settings such materials are predefined and limited. Third, collaborative problem-solving requires interaction and coordination between people, which at a minimum consists of defining who should participate in the problem-solving (Stasser & Abele 2020). Thus, integrating useful elements from various approaches to problem-solving is particularly important when studying the phenomenon outside the laboratory.

One potential starting point for building an integrative approach to problem-solving is distributed cognition (DC), which portrays cognitive activities as relationships between people and their immediate surroundings rather than strictly mental processes (e.g., Cowley & Vallée-Tourangeau, 2013; Hollan et al., 2000; Hutchins, 1980, 1995; Järvilehto, 2009; Zhang & Patel, 2006). This perspective brings both theoretical and methodological changes to our conceptualisations of problem-solving. Theoretically speaking, problem-solving is understood as participants' ongoing achievement that actualises in the interaction between people and the physical environment (Hutchins, 2011; Steffensen, 2009, 2013; Vallée-Tourangeau, 2014). From this viewpoint, cognition expands 'beyond the skull' (Cowley & Vallée-Touragnant, 2013), as people coordinate human actors, artefacts and environments of action to form

distributed cognitive systems (Clark & Chalmers, 1998; Hollan et al., 2000; Vallée-Tourangeau, 2014; Zhang & Patel, 2006). These systems allow people to assign certain parts of cognitive processes to artefacts, which then enables them not only to gain more memory by writing things down or orally explaining their inner thoughts but also and more generally 'to make use of a different set of internal and external processes' (Hollan et al., 2000, p. 176). Consequently, the coordination of participants' engagement in the materiality of the situation becomes a central focus. Methodologically, studying distributed cognitive systems relies on video data to capture problem-solving in natural settings.

Despite these promising aspects, DC has been criticised for overlooking the setting in which actions take place, equating practical social actions with cognitive actions and fundamentally reducing them to mechanistic models of information processing (e.g., Aagaard, 2021; Button, 2008). To address these concerns, we integrate the general theoretical perspective of DC with analytical concepts provided by conversation analysis (CA) to emphasise the situated conduct of problem-solvers. Through a conversation analytic single case study, we examine a problem-solving episode in a hybrid video-mediated (VM) encounter involving both co-located and remote participants. We observe the use of multiple modalities, such as spoken utterances and gestures, to solve a technical problem related to a faulty audio connection which meant that one participant could not hear the others. Furthermore, we show how technological mediation as part of the overall material setting and affordances available in mediated and co-present ecology are made consequential for the problem-solving activity. By combining DC and CA, our analysis focuses on the contextual and practical sense-making of problem-solvers while contributing to the broader theoretical discourse on problem-solving. The aim of the analysis is to show that problem-solving is an emergent activity in which different affordances are made relevant and the goal state is transformed moment-by-moment as the problem is redefined.

2. Coordination of Distributed Cognitive Systems and Problem-Solving from a CA Perspective: Participation and Multimodality

In this article, we augment the discussion on the central DC notion of *coordination* with reference to three CA concepts that are related to the DC perspective: *participation* (Goodwin & Goodwin, 2004), *co-operative action* (Goodwin, 2017) and *multimodality* (Mondada, 2019). In distributed cognition, coordination refers to the processes through which internal and external representations are brought together within the cognitive system. Thus, material parts and artefacts have a representational role, as something to be juxtaposed with internal representation. This approach to coordination and materiality takes the cognitive system comprised of people and tools as its unit of analysis, emphasising the flow of information within the system. By comparison, CA emphasises sequences of action as its primary unit of analysis (Schegloff, 2007).

This action-centeredness is characteristic of the conceptualisation of both participation and multimodality.

From the CA perspective, we consider participation as a series of practical actions through which problem-solvers express and regulate their engagement with the ongoing activity and fit their actions to those of others (Goodwin & Goodwin, 2004). Through this expression and regulation, problem-solvers build on one another's actions and the material artefacts that have been made relevant in those actions, forming *co-operative action* (Goodwin, 2017). From the perspective of problem-solving, this co-operative action involves collaborative engagement with both defining the problem and probing for solutions, such as by collaboratively combining various forms of lay and expert knowledge to define the problem in a way that enables all relevant actors to participate (Arminen & Poikus, 2009). Participation is shaped by material artefacts which allow people to create new forms of engagement and perspective sharing, for example by affording not only individual sense-making but also collaborative problem problem-solving, which occurs when the interactants highlight and verbalise the relevant parts of the environment (Tarasmundi & Linell, 2017). Interacting with objects can sometimes help adopt the role of a non-present third party, as when two co-workers seeking to solve a problem with an invoice find a solution by physically simulating the situation in which the client received the physical invoice (Steffensen, 2013). Furthermore, participation involves the regulation of who will take part in the problem-solving in the first place; for example, one might recruit those considered more knowledgeable in the matter to join the effort in a coding problem or refer to analogous features of familiar programs (Bowden, 2019).

The material dimension of distributed cognitive systems can be approached through the concept of multimodality in interaction (Mondada, 2019). Multimodality calls attention to how participants build a shared understanding of an ongoing situation using multiple modalities, including spoken language, non-lexical vocalisation, a wide range of body movements from gaze shifts to walking, artefacts and other material aspects of the environment. From the point of view of problem-solving, these material aspects are not fixed; rather, the cognitive affordances embedded in artefacts are actualised in the interaction between problem-solvers and the environment. For example, manipulating artefacts can enable exploration of new possibilities without a clearly defined goal, thus representing bottom-up cognitive processing (Bjørndahl et al., 2014) or reframing the problem to overcome frustration and fixation on non-functioning solution options (Steffensen, 2013). However, as we discuss below, in the context of VM interaction, where some material aspects of the physical ecologies are more readily available to some participants than others, interactants need to take the technological affordances of the communication medium into account.

In sum, the DC concepts of coordination and materiality can be viewed as emphasising the cognitive system and goals, while the CA notions of participation, co-operative action and multimodality focus on how interactants achieve a shared understanding through inter-action (Due, 2016; Heath & Luff, 1992). By drawing on these CA concepts, our aim is to illuminate the practical interactional details of distributed cognitive systems without prioritising cognitive actions over practical ones. Next, we consider ways in which distributed cognitive systems may be especially vulnerable in VM settings.

3. Fractured Ecologies and Cognitive Systems

When examining the impact of VM on problem-solving processes, it is crucial to describe how technical mediation within a broader socio-material context becomes consequential for the action (Arminen et al., 2016). A key theoretical concept in this respect is *fractured ecologies* (Luff et al., 2003, 2016). In VM interactions, camera technologies offer limited perspectives on the local environments of distant participants, resulting in partial views that obscure the physical relationships between material objects. Thus, the notion of fractured ecologies highlights the ways in which VM detaches bodily actions from both "the environment in which [they are] produced and from the environment in which [they are] received" (Luff et al., 2003, p. 55). This detachment has consequences for the production and interpretation of various bodily actions and the use of material resources, such as pointing to artefacts (for reviews on VM interaction, see Arminen et al., 2016, Due & Licoppe, 2020, and Mlynár et al., 2018).

People can work around the limitations that VM brings to interacting with artefacts through making them visible by showing them or by camera movement (Due & Lange, 2020; Licoppe, 2017; Seuren et al., 2020; Stommel et al., 2020) and by directing one another's actions in relation to them (Due et al., 2019; Ilomäki & Ruusuvuori, 2022). In addition to embedding 'gestural showings' in the ongoing talk (Licoppe, 2017), different highlighting practices like pointing can be used (Goodwin, 1994) to make parts of what has been shown stand out from the overall picture and build a shared understanding (Due & Lange, 2020). When directing a distant participant's actions in relation to material artefacts, participants can, for example, divide activities into smaller sub-actions, such as locating the activity-relevant artefact before manipulating it (Ilomäki & Ruusuvuori, 2022), or show how one should turn one's head to find the relevant artefact through 'mimicable embodied demonstrations' (Due et al., 2019, p. 19–20).

Fractured ecologies also shape the management of participation; for example, participants may need to explicate what kind of participation is expected from others (Hansen, 2020; Seuren et al., 2020; Shaw et al., 2020; Stommel & Stommel, 2021). Furthermore, hybrid interactions can demand additional work to manage

participation between interactants in distributed participation frameworks: the digital space shared by all the participants and the physical space shared by only the co-located people (Oittinen, 2018). Shifting between physical and digital spaces, as occurs when there are transmission distortions to be solved (Oittinen, 2020), needs to be managed through various multimodal resources like gaze shifts accompanied with talk (Oittinen, 2018; Saatçi et al., 2020).

In VM meetings, interactants make fractured ecologies salient through recognising and solving routine technological ruptures (Oittinen, 2020; Rintel, 2013a, 2013b, 2015). However, the practices of noticing and remedying these troubles have not thus far been studied from the DC perspective. Thus, from this starting point, we examine how collaborative problem-solving happens in a hybrid VM setting where technical mediation challenges the management of participation and access to others' local ecologies. We ask how interactants employ various modalities to manage participation despite the absence of some communication channels and a lack of access to other interlocutors' local ecologies.

4. Data and Method

We draw data from a video recording of one hybrid VM teleconsultation where a general practitioner and a patient met in the general practitioner's office to consult a specialist from another location. The data are drawn from a larger corpus of five teleconsultations comprising a total of 255 minutes of data that was collected during a service pilot in a private clinic in an urban area in Finland in 2019. The data are in Finnish and were recorded with two video cameras. Figure 1 depicts the setting in the general practitioner's office, with the two integrated frames showing the action from different camera angles. These data were collected as part of the *Healthcare Workers in the Eye of Digital Turbulence* project conducted by Tampere University and the Finnish Institute of Occupational Health, with funding from the Finnish Work Environment Fund (grant number 117151). Additional funding for the analysis and writing came from the Jenny and Antti Wihuri Foundation and the Strategic Research Council at the Academy of Finland (grant numbers 335288 and 336277).

Figure 1. *The physical setting and participants*
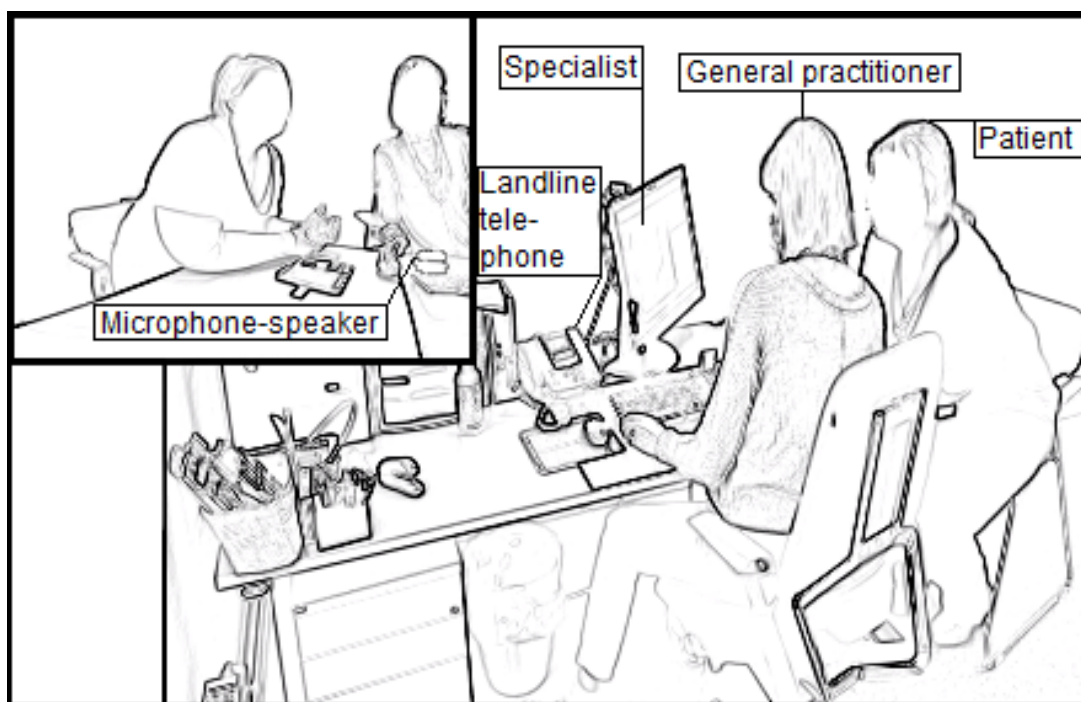
We used multimodal CA as our method; CA helps identifying recurring patterns and structures of interaction to discover interlocutors' moment-by-moment interpretations of the ongoing action (Mondada, 2019; Sidnell & Stivers, 2013). The data were transcribed according to CA conventions (Jefferson, 2004; Mondada, 2019; see Appendix 1 for symbols).

We focused on the problematic and prolonged opening of the first teleconsultation in the pilot, in which the participants tried to begin the consultation but realised that the specialist could not hear the other two participants via the VM equipment. This instance provides an interesting example of spontaneous problem-solving. The solution found by the participants immediately became routine and was applied in subsequent session when the problem persisted. Thus, it was the only case of this kind of problem-solving in the data corpus (see Mlynář & Arminen, 2023 on practices becoming obsolete). Conversation Analytic single-case analysis (Schegloff, 1987) enabled a context-sensitive examination of the action-by-action process of unplanned, naturally occurring problem-solving in a complex socio-technical setting with hybrid participation (see Discussion for methodological reflections).

This case is particularly interesting because the problem lies with the audio channel that would usually be employed as part of problem-solving. Our analysis reveals that the same technical artefact that is part of an ordinary communication system, the landline telephone, can become a critical part of a distributed cognitive system. Furthermore, as the interaction is not only VM but hybrid, the

management of participation might demand extra work than in solely face-to-face or VM settings.

We first transcribed the data following CA conventions and conducted a sequential analysis of the problem-solving processes. Partially overlapping this more descriptive analysis, we examined how the participants use different modalities to manage participation. Below, we discuss and explicate the connection between our analysis of participation and multimodality and the notion of coordination, thus linking our findings to the broader DC context.

5. Results

In line with existing research, our analysis shows how problem-solving in a VM setting demands coordinating a variety of resources, including knowledge and technological artefacts and affordances of these technologies, to form a distributed cognitive system. We add to this established understanding by showing how despite the limitations that VM places on different participants to access physical parts of the cognitive system, participants manage to build and coordinate a shared distributed cognitive system. To enable participation with remote interactants and facilitate the coordination of the overall cognitive system, participants fit their conduct to the media they considered available to others. We now demonstrate these processes through four data excerpts.

5.1 Solution probing through directing attention and action in co-present and mediated ecologies

After acknowledging the problem (that the general practitioner and the patient can hear the specialist, but the specialist cannot hear them), the participants move on to testing possible solutions. They direct one another's attention to certain parts of the VM equipment and software as potential sources of the trouble and suggest how to act on them. Directing both attention and action requires fitting one's actions to the media available to others through an appropriate communication channel.

The participants' process of identifying and addressing potential trouble sources in the co-present setting is shown in Extract 1, in which the patient (P) uses both talk and pointing to direct the general practitioner's (G) attention and action in a shared physical space. Throughout the segment, the general practitioner and the patient hear the specialist's voice through the VM system, but the specialist cannot hear the other two participants. We join in as the participants have just figured out and explicated the problem.

Extract 1. *Testing a possible solution in the co-located ecology*

```
    g: >>gaze down->
    p: >>gaze to computer screen->
01 G: @hmh?
       @PIC 1.1
```


PIC 1.1

```
02    (0.5)+(.)
   p:      +...

03 G: .m[t]
04 P:    [m]itäs jos%-@ (0.3)%(.)+(0.6) @%+pai%naa:, (0.3) +mitä siitä
           what   if                      one clicks          then what
   p: ........................+points to screen +,,,,,,+ H under chin->
   g:            ->%........%gaze to screen->
   g:                      %  leans right %leaning right->
                           %..........%left hand on mouse->
                      @PIC1.2            @PIC1.3
```
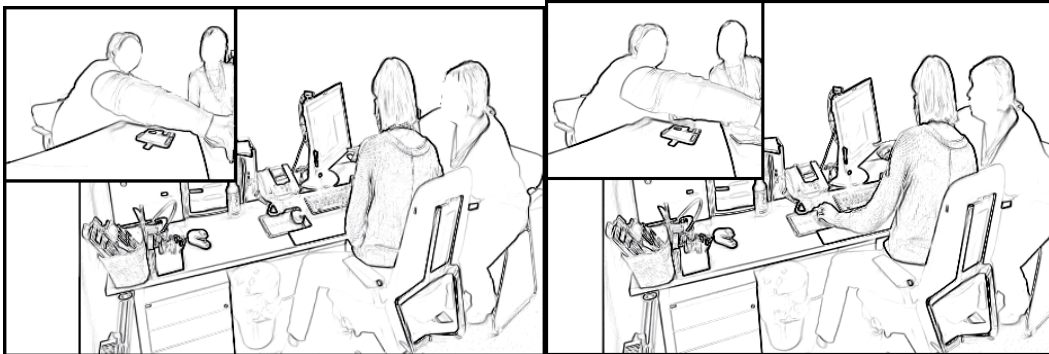

PIC 1.2   PIC 1.3

```
05 P: <tapahtu.>
      happens
06    (0.3)%(1.0)
   g:      %mouseclick

07 P: +[e::i.  ]
         no
08 G: +[se mene]e vaan sitte mikki +pois.
       it only closes the microphone then
   p: +,,,,,,,,...................+gaze to microphone->

09    (0.4)%(0.5)+(0.3)
   g:      %mouseclick
   g:    ->%lean right
   p:          +reaches microphone->
```

Testing a possible solution comprises initiating testing (line 4–5), the test itself (lines 4–6), feedback from the system (line 6) and the human actor's evaluation of the test (lines 7–8). Starting at line 4, the patient formulates a possible solution by first pointing towards the screen and verbalises the solution while stretching out her arm. Although the general practitioner has visual access to the monitor, at this point she is gazing down at the keyboard after typing a chat message to the specialist (the message content was not captured on camera) and is therefore not fully engaged in the interaction with the patient or able to see where the patient is about to point. Thus, the patient institutes a brief pause to establish a shared focus between her and the general practitioner. She pauses after producing the first part of the *if X then Y*-type question (Grigorov & Snoeck Henkemans, 2019; Speer, 2012), making it relevant for her to produce another part in the turn as a possible solution, and completes the suggestion only after the general practitioner focuses her attention on the software feature on the screen to which the patient is pointing: *what if (1.0) one clicks (0.3) then what happens* (lines 4–5; Goodwin, 1980; see also Stivers & Rossano, 2010). This turn design highlights a certain part of the VM equipment as a potential source of a solution and the patient's epistemic stance towards it: the solution could be found in a specific feature of the software, but the patient is not sure if the solution will work. As she is referring to an artefact to which both relevant participants have immediate visual access, she can use a pointing gesture to highlight it and omit verbalising the object while still producing an understandable proposal (at least once a shared focus is established). Furthermore, by producing her initiation as a *what if X*-type proposal (instead of, say, a directive), the patient builds the proposal as a seamless part of the broader problem-solving activity that began with the problem being recognised (cf. Stivers & Sidnell, 2016 on *how about* proposals) and the entire activity as shared, while producing a lower deontic stance, potentially linked to her role as a patient (Couper-Kuhlen, 2014; Stevanovic & Peräkylä, 2012).

Even as the patient produces the proposal, the general practitioner starts orienting herself to testing the proposed solution by preparing to operate the mouse. By doing so, she expresses a shared understanding of a potential source of trouble and identifying herself as the actor who should carry out the testing. Through her actions, the general practitioner aligns with the patient's unknowing but curious stance, displaying interest in the potential trouble source while being unsure of the outcome before testing. The general practitioner tests the proposed solution by clicking the suggested part of the software, which is visually available to both participants on the screen (line 6), the participants wait in silence for the feedback from the system (line 6), and as it becomes apparent that the action has not yielded the hoped-for solution; they both evaluate the test by verbalising the absence of the preferred outcome (lines 7–8). These actions are observable to the patient who initiated the testing, so the general practitioner does not need to work to produce her actions as understandable for the patient, who can simply observe them unfold. By coordinating their actions

with reference to what each has access, the participants manage to form a cognitive system in which ideation through manipulating the features of the VM software and its outcome are mutually intelligible. Throughout this sequence, the specialist has only partial access to the problem-solving activity: while he can (probably) see that the patient is pointing at something, he lacks an audio connection, so his resources for interpreting what that pointing contributes to are very limited. Furthermore, the other interactants do not work in any observable way to include him as a recipient of these actions at this point.

In the following extract, which begins 41 seconds after the end of Extract 1, the participants demonstrate remote coordination of the distributed cognitive system. When interacting in this kind of technologically mediated setting, the participants need to fit their actions to the affordances of the media, both when speculating about the potential trouble sources and when communicating their understanding and knowledge of those sources. The distant specialist proposes a solution through talk, but the general practitioner uses a showing gesture and highlighting through pointing, both behaviours designed to be observable for the specialist. In this way, the general practitioner indicates that the suggested solution is not workable in this situation. Prior to Extract 2, the patient and general practitioner have examined the speaker-microphone as a physical artefact. At the beginning of Extract 2, the specialist informs the others that they can hear him despite not wearing a headset, while he is unable to hear them.

Extract 2. *Testing a possible solution remotely*

```
   g: >>gaze to keyboard->
   g: >>typing->
   p: >>gaze to screen->
01    %(0.5)%(0.9)%(2,5)
   p: %g. GP%,,,,,%gaze to screen-> (1.3)

02 S: >eihän siell<ä:, (.) e+hh+ teillä, (.) oo mikään #ä ähhhh# @(0.7)
       you don't happen to  ehh you don't    have  any  e ehhhh
   g:                      ->+typing finished
   g:                      ->+,,+gaze to screen->
```
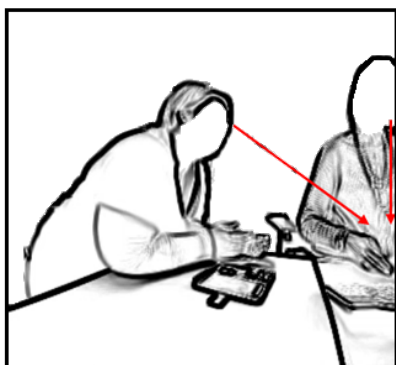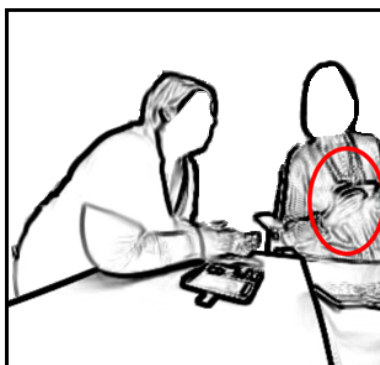
                                                    @PIC 2.1

PIC 2.1

```
03 S: #ö# (.) eihän siellä teill+ä ole +mi+kään% tota,% (0.4)+(0.4) m, (.)
       e         you  don't happen to have  any    like
   g:                               ->+......+gaze to system unit->
   g:                                        +.................+manipulates
                                                   cords of the microphone->
   p:                                      ->%......%gaze to G's hand->

04 S: #ee# se  kuu- kaiutin, +(0.3)+(0.3)    pois päältä tai,+
      ee  that head- speaker                 turned off  or
   g:                      ->+.....+gaze to speaker->
   g:                         ->+moves both hands to spkr+lifts speaker>
```
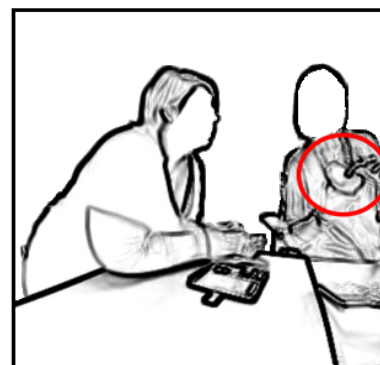

PIC 2.2   PIC 2.3   PIC 2.4

```
05      @(1.3)+@(0.3)+(0.4)+(0.6)+(0.2)%+@(0.2)+(.)
   g:       ->+turns spkr + shows speaker to S+puts speaker on table, right
   g:             +...........+points+moves hand to keyboard->
   g:                       ->+......+gaze to screen->
   p:                          ->%directs gaze to screen->
       @PIC2.2@PIC 2.3                  @PIC 2.4

06 S: %ne::,+ (.) ää+nen+voimak%kuu+det,%
      those       volume levels
   g: ,,,,,,+left hand above keyboard->
   g:        ->+speaker on table
   g:               +..........+right hand above keyboard
   s: %----gaze to screen-----%gaze to keyboard->
```

The specialist first asks whether the general practitioner's speaker-microphone is on, thus implying that one potential solution would be simply turning it on (lines 2–4). Different forms of asking invoke different assumptions of knowledge in relation to the object being queried (Heritage & Raymond, 2012; Raymond & Heritage, 2021); in this case, the question's turn-design features invoke the specialist's knowledge in two ways. First, it depicts him as knowing that such a thing as not having the microphone on is a common occurrence that could well be the problem (see also VISK §1681, for the use of the Finnish clitic particle -*hän* as a way of marking something as commonly known)[1]. Second, it demonstrates that he has less information about this specific case: since he lacks visual access to the microphone-speaker in the general practitioner's

---

[1] VISK is the online version of the comprehensive Finnish grammar, edited by the Finnish Literature Society.

office, he needs to access information about the state of the microphone by collaborating with the other participants. (By contrast, the patient can see the computer screen to identify potential trouble sources and thus points to them in Extract 1). Moreover, as the VM system does not afford the specialist to see where the hardware is situated in the office, he can only use talk to initiate the test. Thus, the specialist produces his initiating turn verbally through a polar question. Furthermore, by using the *hän*-clitic and the word *mikään* (*any*), the specialist projects a negative answer as the more expected answer, thus treating the turned-off microphone as an unlikely solution. By doing so, he potentially works around the delicacy associated with the solution: if the microphone-speaker was turned off, that might characterise the general practitioner as a non-competent user who has not adequately prepared for the encounter.

While the specialist could have written the question in the chat, at this point he employs the partially functioning audio-connection between the two ecologies of action. This type of action design has two advantages: the production of a spoken turn is faster than typing, and the specialist's action is now more readily accessible to all co-participants, compared to a small chat screen that would have been more readily accessible only for the general practitioner. Now, both the general practitioner and the patient are made relevant as recipients of the action.

During the specialist's question, the general practitioner starts to orient herself to different parts of the VM equipment as potential trouble sources, anticipating what the specialist is about to say. Since the specialist can only use talk to produce his initiating turn, the general practitioner cannot conclude from the specialist's bodily actions (such as gaze) where she should direct her attention but must infer this as the specialist's verbal turn unfolds. As the participants have already tested the software features (Extract 1), the general practitioner turns to the hardware as a possible solution, first checking to see whether the USB cords are connected correctly (lines 3–4) and, as the specialist stipulates the *speaker* as the potential trouble source, reorients her attention (line 4), still focusing on the speaker as a physical object instead of a software feature. Reciprocally to the specialist's turn, the general practitioner chooses not to use the chat function that would allow her to verbalise her answer. Instead, she produces a showing gesture accompanied with highlighting by pointing, taking advantage of the visual affordances of VM. During the specialist's suggestion, the general practitioner starts to answer by lifting the microphone-speaker so that the specialist can see it on the camera, turns it so that its user interface is visible and points to an icon indicating that the speaker is on, thus highlighting the relevant aspect of her showing gesture (lines 4–5; cf. Due & Lange, 2020). By doing so, the general practitioner works to answer the specialist's question while simultaneously maintaining all the participants as potentially relevant.

The general practitioner's response to this suggestion (lines 3–6) highlights two important aspects of the distributed cognitive system of problem-solving in this

context. First, the general practitioner seeks the solution from her physical surroundings, thus outlining the software features from the sphere of potential trouble sources based on the earlier interactional problem-solving (cf. Arminen & Poikus, 2009). Second, through the showing gesture, the general practitioner fits her actions to the affordances of the technological medium to render her actions and sense-making behind them accessible to the specialist (lines 5–6). The specialist's polar question design, which affords a yes/no-answer that could be achieved without talk – for example by nodding or a headshake – and ability to show the microphone-speaker enable the participants to build a shared understanding despite the audio connection. Furthermore, the general practitioner's answer to this question is non-minimal and non-type-conforming: it includes not only a yes/no-answer but also the justification for this answer. By highlighting through pointing, the general practitioner both works towards letting the specialist know that his suggestion has not worked and demonstrates the reason for the negative outcome (on type non-conforming answers, see also Raymond, 2003). Unlike the situation in which the patient suggests a solution and the general practitioner can both show and say why a solution is not adequate, only the limited visual medium is accessible to both the specialist and the general practitioner.

When prompting for a solution, the participants need to accomplish two tasks: direct each other's attention to certain features of the VM equipment and provide some sort of instruction on how the others should act in relation to those features. The possibilities of employing multimodal resources to achieve these tasks differ in co-located and VM ecologies. In Extract 1, the affordances of the local ecology enabled the patient to gather information about potential troubles from her immediate physical environment and suggest them to the general practitioner by highlighting the features of the software to which they both had immediate access. The actions of both the general practitioner and the patient were mutually accessible. By contrast, as a distant participant, the specialist had to work towards both assessing the relevance of the specific parts of the VM equipment and directing the others' attention to them, all without having visual access to those resources. The specialist used talk to complete both tasks: by asking a polar question about the artefacts in the others' physical environment, the specialist was able to gather information and direct the others to act by implying the solution. This work to coordinate a distributed cognitive system in a mediated setting was reciprocated by the general practitioner, who strove to establish an answer to the specialist's suggestion by showing, thus adjusting her conduct to the visual affordances that enabled building a shared understanding with the specialist. In sum, to test possible solutions, the participants needed to fit their practices of directing attention and action to the affordances of the co-present or mediated ecology, thus enabling the manipulation of the relevant artefacts in the fractured cognitive system.

5.2 Reframing the problem and dividing audio and video to different technological artefacts

After several solution candidates are tested or shown to be inadequate and fetching the researcher to help (with no success), the participants move on to formulate a plan B: they have to figure out some way to definitively cease problem-solving and move on to the central business of the patient's case. Extract 3 shows how by formulating this plan B, the participants not only establish a solution that allows them to move on to the consultation but also discuss and redefine the roles of different communication technologies. The general practitioner has called the specialist on the landline to talk about the solution. Throughout the extract, both the patient and general practitioner can hear the specialist through the computer's speaker, and the specialist can converse with the general practitioner via telephone.

Extract 3. *Reframing the problem and what counts as a solution*

```
01 G: nii ni[i (-)         ]
      so like (-)
02 S:      [yritämmekö ke]rran vielä vai?
           shall we try  once  more  or
03    (1.2)
04 G: .mhh .mt mitähän me nyt yritettäs sitte.
      .mhh .mt well what should we try now then
      ((5 lines omitted. G tells about problems before the
        teleconsultation.))
10 G: >↑mut ↑toki ↑me< ↑voidaan nyt mennä vielä ↑ulos tästä vastaanotosta
       but of course we can   now   still log out from this consultation
11 G: ja yrittää uudesta[an.]
      and  try     again
12 S:                   [ jo]o.
                          yes
13    (0.3)
14 S: ja se'=jälkee me voidaan tehdä sillä=lailla että me voidaan,(.)
      and after that we  can   do    it so     that we can    via
15    puhelimen kautta >mä voi' vaikka se?< (0.8) #ööö# potilaan
      telephone       I can like with the        u:hm  patient
16 S: k[anssa keskus]tella ja käydä läpi tätä keissiä ja? (.) sit mä teen
        talk about   and go through this case   and     then I'll make
17 G: [↑nii::::?    ]
        yeah
18 S: ehdotuksia.
      suggestions
19 G: no=ni?
      okay
20    (0.4)
21 G: tehään [nii. ]
      let's do that
22 S:        [jos=e]i toimi.
              if it doesn't work
23    (0.3)
```

Following several failed attempts to solve the problem, the specialist proposes that the participants give the VM system one more try (line 2). After a long pause (line 3), the general practitioner answers this proposal with a long complaint about the VM system (lines 4–9, omitted from the transcript). The pause and

complaint embody the general practitioner's frustration: all relevant lines of action appear to have been tested, and there is no clear path forward (see VISK §837, on *s*-clitic in questions as a way of taking a stand when asking questions, and VISK §1681, on *hän*-clitic as a way of marking a question as lacking an answer). After the complaint, the general practitioner produces a change in position with *but of course* (line 10) and proposes a more concrete solution: the participants will turn the software off and turn it on again (lines 10–11). Overlapping with the general practitioner's turn, the specialist appears to align with this proposal (line 12; see Ruhleder & Jordan (2001) on transmission delay and timing of turns in VM) and expands on it: if the suggested solution does not work, the participants will change from a VM consultation to a telephone consultation (lines 14–16, 18). Finally, the general practitioner agrees with that plan (lines 17, 21).

When formulating plan B, the participants start to engage in two processes crucial for solving the overall problem. First, they implicitly reframe the problem from 'how do we get the audio in the VM system to work?' to 'how can we proceed to the patient's case?' After the specialist's unspecified solution proposition and the general practitioner's frustration, the participants reframe the scope of what can be considered a successful outcome in the situation: instead of fixing the audio in the VM system, the participants can simply refrain from using it. This hypothetical scenario enables the participants to enter a foreseeable future where they can proceed to their central task: working on the patient's case. In line with the new problem framing, the participants recognise what is routine and will definitely work and thus identify a guaranteed way to the new desired state.

Second, as part of this process, the participants topicalise the telephone. This has two important consequences for the solution that they ultimately implement. It expands the potential material base of the solution to cover not only the hardware and software of the VM system but also the other technical artefacts that are available. Perhaps even more importantly, it separates the visual medium from the aural and situates them in different technical artefacts: the computer system (including the screen) and the landline telephone, respectively. Thus, the role of the telephone in the cognitive system shifts from a mere channel of communicating *about* the solution to being a proposed *part* of the solution. As we see from the following extract, this division of media and associating them with specific technological artefacts turns out to be the key to formulating the solution that is ultimately chosen.

5.3 Experimentation based on the separation of media to find the solution

The participants then close the connection, make a new videocall and test plan B. However, the suggested solution proves ineffective, and the problem remains.

Extract 4 depicts how instead of progressing to the telephone consultation as agreed, the specialist reorients to problem-solving as a still relevant activity by engaging in experimentation with the affordances of the telephone and the VM system. Throughout the extract, we see how the participants collaboratively manipulate the audio connection through different technological artefacts: first, they implicitly negotiate abandoning plan B (lines 8–14), followed by experimenting with different sources of audio connection, telephone and the VM system (lines 14–41) and finally combining the audio connection from the telephone with the visual connection through VM (42–54). During these three phases, the different features of the technological artefacts are emphasized, especially turning the audio connection on and off from different devices, as well as interactional practices that enable all interactants to participate in the same actions; namely, visual conduct through gesturing and nodding and spoken interaction when suitable. Between lines 1–26 and 34–44, the co-present participants hear the specialist via the VM system, and the specialist hears the general practitioner (and potentially parts of the patient's talk) via the telephone; between lines 27–33 and lines 45–48, only the professionals hear each other on the telephone; finally, from line 49 onwards, all participants hear one another through the landline telephone's conference mode.

Right before Extract 4 begins, the participants have started a new videocall and carried out several audio checks while maintaining the audio connection via the landline. We join the action as the general practitioner initiates the final audio check of this testing.

Extract 4. *Elaborating on plan B to find the solution*

```
   g: >>gaze to screen->
   p: >>gaze to screen->
01 G: @kuuletko     [mei°tä?° ]
       do you hear   us
   g:                                    ->+.....................+phone on
                                                                   ear->
02 S:             [mutta nyt] jostain s+yystä edelleenkään=ei,+ @(0.4)
                   but   now  for some reason   still     I can't
      @PIC 4.1                                                @PIC 4.2
```

PIC 4.1   PIC 4.2

```
03 G: ei.
      no
04 S: sieltä päin kuulu <ääni.>
      hear your voice from there
05    (0.2)
06 G: ei kuulu.
      can't hear
07    (1.1)



08 G: .hhh >no tehäänkö me nii että mä an+nan< tämän luur+in@
      .hhh well shall we do it like that that I give the telephone
   g:                                    ->+...............+grabs phone with
                                                              right hand
                                                           @PIC 4.3
```

PIC 4.3

```
09 G: [Ai+nol+le-          ]
       to Aino
   g:    +,,,+L hand off phone  +............+left hand on phone->
```

```
10 S: [mul=on tää pu&helin]=edel+leen päällä.+
        but I have this telephone still on
   s:                 &brings telephone back to screen and ear

11    (0.2)+&(0.3)
   g:      +......
   s:        &moves telephone away from ear

12 G: ni?
      yes?
   g: ...

13 G: >tehäänkö     [@+nii että,<]
      shall we do it  like that
   g: ...............+right hand on table->
14 P:              [@+mut me kuu]llan sua ilman puhelin£taki.£
                     but we hear you even without the telephone
                 @PIC 4.4
```
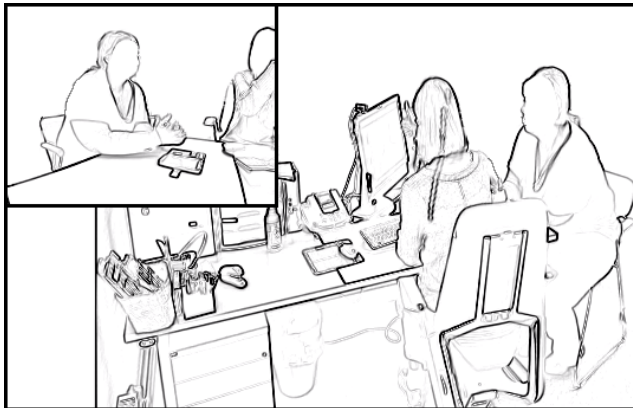


```
     PIC 4.4
```

```
15    (0.8)
16 G: jo[o.]
      yes
17 S:   [ n]ii.
          yeah
18 S: te kuulette mut, ö- #mm# nyt >juuri mä laitoin< puhelimen
      you   hear   me  ö   mm   now exactly I put   the telephone
19 S: tänne pöydän alle niin te kuulette täältä, (0.3)
      here under the table so you hear   here

20 S: +läppärin:, (.) k+au+tta
       through the laptop
   g:>+...............+TU+lifts hand->   ((TU=Thumbs up gesture))
```

```
21 S: mutta?+@
      but
      g:  ->+thumbs up on camera->
           @PIC 4.5
```



```
      PIC 4.5
```

```
22     (1.3)+(0.5)+
   g:      ->+.....+Right hand on table->
```

```
23 G: .hhhh
```

```
24 S: ↑hyvä?
      good
```

```
25     (0.8)
```

```
26 S: >mutta katotaas ny &siis,<
       but let's see   now like
      s:                 &brings telephone back to ear
```

```
27     (5.1)%(0.6)
   p:       %headshake->
```

```
28 G: no=nyt sulta hävis mikrofoonistaki=äänet.
      well now you lost the voice from the microphone as well
```

```
29 G: nyt me ei kuulla% sua.
      now we can't hear you
   p:              ->%stop headshake
```

```
30     (0.2)
31 G: >mihin se nyt katos.<
       where did it go now
32     (1.4)
33 G: no mä kuulen nyt mutta nyt Aino sitten että
      well I hear  now  but  now  Aino then  like
```

```
34 G: [nyt mi]tä ta-
      [now wh]at ha-
35 S: [entä- ]
       how about
36     (.)&(0.2)
   s:    &moves telephone away from ear and screen
```

```
37 S: mutta nyt.
      but   now
```

```
38    (0.5)%(.)
  p:        %nodding->

39 P: ny[t  kuul]uu.
      now we hear
40 G:    [ nii? ]
            yeah?
41    (0.3)

42 S: .th%hh mut kato mä- (.) mietin sitä että, (0.4) jotta, (0.3)
      .thhh  but look I      was thinking that     in order to
  p:  ->%stops nodding

43 S: #ääääääää ä yyyy# nyt mä voin puhua, (.) mutta tavallaan, (0.5)
       eeeeeeeee e yyyy  now I  can  speak     but   in a way
44 S: puhutteks=te sitten, (0.3) e- (.) >et jos mä,< (.) suljen tää-
      do you talk then           i-         if   I   turn off from h-

45    (1.2)&(6.1)%(0.2)%(0.2)
  s:        &brings telephone to screen, manipulates and gestures
  p:              %.....%gaze to screen
  p:                    %.....

46 P: py[styy%ks% sii%tä@ sun puh]+el+imes[ta+  lait]taan +kaiut[timelle.]@
      is it possible to turn the speaker on from your telephone
  p:        ->%..%gaze to G->
  p:  ...............%points to telephone->
47 G:    [      ↑nii:::?       ]     [>joo.< ]           [↑°joo.° ]
          ↑yea:::h?                   yes.               ↑°yes.°
  g:                                +..+gaze to phone->
  g:                                +..........+right hand on headset
  g:                                                    +L hand to
                                                          telephone->
```
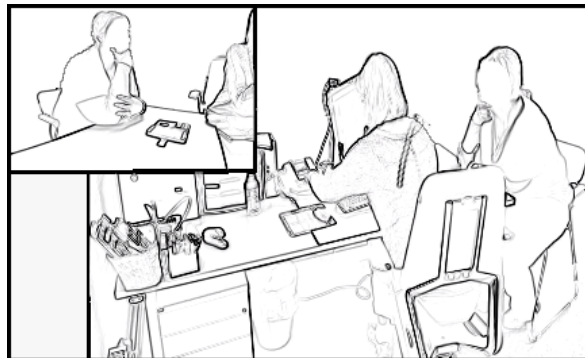
PIC 4.6   PIC4.7

```
48    (0.2)+(0.4)
  g:        +pushes a button on phone
  g:     ->+withdraws left hand from telephone

49 S: kaiutin=o  pää+llä  nyt   [+  mo]lem[mil+la. ]
      speaker is  on      now       on both of us
50 G:                            [+no, ]  [nyt+  o]n, (.)
                                  well      now    it is
  g:            ->+L hand on headset
  g:                            +,,,,,,,,,,,,,+R hand on table->

51 G: £ ka(hh[hhh)   heh    heh    he?£ ]               [.hh ]
       on the spe(hhhhh heh   heh   he                     .hh
```

21

```
52 S:        [>tavallaan   nyt   on<  ku]vayhteys, (0.3) [mutt]a ääni
                in a way     now  we have visual connection but   the voice
53 S: tulee kaiutti[mesta.]
      comes from the speaker
54 G:            [£joo.£]
                  yes
```

After the unsuccessful audio check of the testing sequence (lines 1–6), the general practitioner prepares to give the earpiece to the patient (line 8), with which the specialist's next turn overlaps (line 10). While the specialist expresses that he has the telephone, which could be understood as implying a readiness to proceed with plan B, the general practitioner treats this turn as misaligning with it: she withdraws her right hand (with which she could offer the earpiece to the patient) and puts it on the table as she repeats the question (line 13). Overlapping with the general practitioner's question, the patient explicates that while the audio connection does not fully work, they can still hear the specialist through it (lines 13–14). Neither the specialist nor the patient explicitly answers the general practitioner's question, and the patient does not bodily prepare to receive the earpiece. The participants have now collaboratively abandoned the previously agreed plan B. During this negotiation, the professionals maintain an auditory connection via telephone.

After abandoning that plan, the participants progress to experimenting with the available technological resources; that is, the audio connection via the telephones and the VM system. Unlike the previous extracts, where the conditions for participation (the quality of audio and video connections) remain the same throughout the segment, these conditions vary here as the participants manipulate the audio connection. When interacting in this kind of multimedia environment, it becomes crucial to produce actions in ways that are available and understandable to others in order to build a foundation for relevant participation and action (Heath & Luff, 1992). However, as both the visual and the aural connections are limited in this setting, the participants adjust their conduct from moment to moment to the technological affordances available at a given time.

When the specialist engages in experimentation, he ensures with the explication *I put the telephone here under the table* (lines 18–19) that the others have seen his bodily action and continues by adding what he considers this to mean for the interaction (lines 19–21). The general practitioner designs her responsive turn according to this changed situation: since the specialist can no longer hear her voice, she refrains from verbal confirmation and instead uses an emblematic thumbs-up gesture (line 20), which she further modifies by moving her hand towards the camera, thus making her acknowledgement of the specialist's turn and bodily action visually available to him (lines 20–22). After achieving this shared understanding that the general practitioner and patient can hear the specialist on the VM system, the specialist starts to experiment with the technology by marking the transition to a new activity and potentially upcoming

changes in the patient's possibility to participate in verbal interaction with *but let's see* (line 26). While neither we as analysts nor the patient can hear what the specialist does or says, we can draw certain inferences from the general practitioner's turn. It is apparent from the general practitioner's turns (lines 28, 29, 31), the cut-off in the specialist's turn (line 26) and continuing verbal exchange between the professionals (lines 28–33) that the specialist has shut off the VM system's microphone but is maintaining an auditory connection through the telephone. This momentarily excludes the patient from verbal interaction, as she cannot hear the specialist (lines 27–33). However, the patient still engages in the activity by shaking her head, not only expressing that she cannot hear but also doing this in a way that is accessible to the specialist (lines 27–29). Thus, she produces an action that shows her understanding of her exclusion from the aural participation in a way that is accessible to both professionals as they engage in dyadic verbal conversation via telephone. The general practitioner explicates the situation (line 28) and expands by formulating the upshot regarding the patient's possibilities of participating (33), after which the specialist marks the transition back to the shared verbal interaction between all three participants with *how about (0.3) but now* (line 37), and the patient confirms this change both verbally and by nodding (lines 38–42).

As the specialist and general practitioner maintain their verbal exchange without the audio connection from the VM system (lines 27–33), the potential solution becomes apparent: an aural connection can be maintained via the telephone without using the VM system audio while the visual connection is delivered by the VM system. Through this experimentation with the audio connection, the telephone's role in the cognitive system has again changed. A slight reframing of the problem occurs yet again: instead of a solution that will allow the participants to move on to the patient's case, they are now figuring out whether it is possible to converse on the telephone so that all participants can contribute while maintaining the visual connection via the VM. As the telephone conversation can be maintained with the VM system audio turned off to prevent echo, the only problem that remains at this stage is that the general practitioner and the patient cannot simultaneously hear what the specialist says, as only one person can hold the telephone's earpiece. Thus, the problem has been transformed. The new problem is whether the telephone connection could afford both the general practitioner and the patient to hear the specialist's voice simultaneously.

Building on the division of aural and visual media into different technological artefacts, the participants move on to the final stage, which is (re)combining the two. After the participants have re-established a shared aural connection and thus included the patient in the problem-solving activity (lines 34–40), the specialist proceeds to suggest the solution. The specialist marks the transition from testing whether the audio connection can be isolated to the telephone to figuring out how all participants can converse over the phone with both the contrastive conjunction *but* and the attention-getter *look* (line 42), which serves

to initiate and redirect courses of action and an explanatory function (e.g., Hakulinen & Seppänen 1992). However, instead of explaining the plan, the specialist engages in executing it; in the middle of his turn, he turns off the VM system's microphone (line 44) as he continues to explain the next steps to the general practitioner over the telephone. Simultaneously, he brings his telephone to the screen, taps it and gestures. By combining talk on the telephone and gesturing, the specialist can both communicate the suggestion in detail to the general practitioner and engage the patient in the activity through solely visual conduct. While the screen is not completely visible and what the specialist is doing or saying is not fully available to the analyst, it is not unclear to the patient; despite not hearing what the specialist says, the patient starts pointing towards the telephone on the table (line 45) and, overlapping with the general practitioner's change-of-state token (Heritage, 1984; line 47), utters the possible solution of turning on the telephone speaker (line 46). By doing so, the patient expresses her shared understanding with the specialist about the potential solution. Here, the patient and the specialist jointly highlight a feature of the technological environment that can lead to the solution: using the landline telephone's conference mode. Thus, the participants manage to orient themselves to the same artefact and its detail, despite the lack of a shared physical ecology.

The same four-part structure as in earlier testing (Extract 1) takes place here. The initiation is made by the patient and possibly simultaneously by the specialist on the phone (lines 45–46), testing is achieved when the general practitioner turns the speaker on (line 48), the feedback from the system occurs as the specialist's voice is heard (line 49), and user evaluation takes the form of a burst of laughter from the general practitioner (line 51; on laughter and celebration following problem-solving, see Bowden, 2019) and the specialist explicating the positive situation (line 49, 52–53). The participants conduct final audio checks (omitted from the transcript) and move on to the teleconsultation. Throughout this segment, the participants enable one another to participate in shared action multimodally, using talk, gesturing and the manipulation of artefacts to build a shared understanding about what should be done next, despite the momentarily varying aural connection. Through these practices, they are able to coordinate a distributed cognitive system where a solution can be probed and tested, despite the fractured ecologies of action.

6. Discussion

Problem-solving in the case presented above demanded coordinating the actions and knowledge of the problem-solvers and the affordances of the VM equipment and the telephones, forming a distributed cognitive system where potential solutions could be proposed. This coordination was shaped by non-mutual access to different parts of this system caused by the VM technology and by an insufficient auditory connection. The participants thus adjusted their

conduct to the media and affordances available to them moment-by-moment, for example, by producing binary questions that could be answered by gesture. In so doing, the participants combined the affordances of different technologies as a collection of devices, using the VM equipment to transmit visual information and the telephone to convey aural messages. During the problem-solving, the participants redefined the problem several times, which shaped which parts of the material surroundings were made relevant as aspects of the distributed cognitive system and how they were relevant. By building on actions of other participants and the material artefacts made relevant in those actions, the problem was solved co-operatively (Goodwin, 2017).

How then was this coordination shaped by the VM? Coordination as understood in DC concentrates on the distributed cognitive system and how the parts of the cognitive system are brought together (Hollan et al., 2000), while the Goodwinian view of participation emphasises the actions through which interactants make observable that they are taking part in the same actions to foster co-operative action (Goodwin, 2017; Goodwin & Goodwin, 2004). Thus, these concepts can be seen as somewhat hierarchical, with *coordination* a more abstract or general theoretical concept that enables treating cognition as distributed and connected to the physical context of the activity at hand, and *participation* enabling observers to pinpoint the practical ways of carrying out coordination. Accordingly, what emerged as a central aspect of coordination of the cognitive system was the participants' constant orientation to one another as relevant to problem-solving. This orientation became apparent when prompting a possible solution, which was comprised of directing the others' attention to certain characteristics of the technological artefacts, suggesting a course of action in relation to those features and producing recognisable departures from triadic participation structure that involved all participants. As in recent research on VM interaction (Due & Lange, 2020; Ilomäki & Ruusuvuori, 2022; Saatçi et al., 2020; see also Oittinen, 2018, on audio-mediated interaction with document sharing), VM presence and physical presence afforded different grounds for participation and coordination of material parts of the cognitive system. To enable participation with remote interactants and the coordination of the distributed cognitive system as a whole, the problem-solvers adapted their conduct to the media that were available to others on a moment-by-moment basis. The practices that the participants employed – using talk to build a shared understanding of the ecology to which another participant did not have access and deploying showing gestures to the participants who could not hear the others – align with existing research on VM interaction (Due & Lange, 2020; Due et al., 2019; Ilomäki & Ruusuvuori, 2022; Licoppe, 2017). We expand on this established knowledge by showing how these practices can be used not only to make distant participants do something themselves but also to remotely collaborate with those artefacts. Furthermore, expanding on previous research that shows how a lot of interactional work goes into recognising technological disturbances among co-present participants in hybrid meetings (on audio-

mediated interaction, see Oittinen, 2018, and Olbertz-Siitonen, 2015), our analysis showed that in the context of problem-solving, the practice of highlighting relevant parts of technological artefacts served to direct attention and action between the two ecologies of action (Goodwin, 1994). It was not only important that the participants saw (figuratively or literally) the other participants' ecologies, but also that they saw them in particular ways that enabled meaningful actions.

The successful coordination of human and non-human parts of the cognitive system laid the ground for different affordances to arise as relevant at particular points of the problem-solving activity; namely, probing for a solution, reframing the problem and experimenting. As in the existing literature (Arminen & Poikus, 2009; Steffensen, 2013), earlier attempts at problem-solving served as a foundation for both identifying potential trouble sources and redefining the problem: they were used as a substrate for further actions (Goodwin, 2013). This was most salient when the frustration with earlier attempts contributed to redefining the problem, which in turn led the participants to make different affordances of the telephone relevant: first to enable discussion about a possible solution among the professionals, then to enable conversation between the specialist and the general practitioner and finally to enable a group discussion in conference mode (cf. Steffensen, 2013, on frustration leading to reframing in problem-solving). The actualisation of the particular affordances of the telephone and VM system in relation to specific activities highlights the need for interactional practices to integrate these affordances into the cognitive system.

The successful coordination of the human and material parts of the cognitive system were central to the changes in the problem-solving process. The solution did not arise solely from thinking about the solution but through carrying out actions that were supposed to lead to that solution. Similar to physically simulating the actions and perspective of the invoice receiver to imagine a solution to the problem with the invoice (Steffensen, 2013) and exploring new possibilities by manipulating artefacts without clearly defined goals (Bjørndahl et al., 2014), this experimentation with technological artefacts allowed the participants to explore what could be possible with the materials at hand and what would need to be done to solve the problem as it was redefined in the process.

This study has two notable limitations. First, the use of data from only one location limits the analysis of the distant participant's perspective (Ruhleder & Jordan, 2001). Thus, we have focused on phenomena observable within the recorded ecology and avoided speculating on the distant participant's viewpoint. This approach proved fruitful, as it aligns with the goal of staying close to participants' perspectives (Olbertz-Siitonen, 2015, pp. 211–212). Using data from a VM setting made the distinction between directing attention and action salient because that appeared to be a problem for the participants. This

distinction might offer potential starting points for analyses of problem-solving in collocated and other technologically mediated settings.

Second, due to the single case nature of our data, we cannot make precise, generalised claims about the recurrence of these particular problem-solving practices, whether within medical consultations or elsewhere. However, some interactional phenomena, especially the four-part solution-probing sequence and the need to direct other participants' attention and actions, did recur in the data. It can be hypothesised that similar phenomena occur in problem-solving activities more broadly. Furthermore, while a hybrid presence and shared digital documents, for example, can afford different ways of doing directing, they still need to be achieved through interaction (e.g., Balaman, 2021; Oittinen, 2023). In addition, as we have shown, managing participation can rise as the participants' practical problem in new ways when interaction is a hybrid of tele-presence and co-presence (cf. Oittinen, 2018, 2020; Saatçi et al., 2020). Therefore, further research exploring emergent problem-solving in various settings (co-present, technologically mediated and hybrid) is needed to better understand the relationship between communication media and the underlying interactional processes of problem-solving (Carr, 2020; Flanagin, 2020).

The intertwining of coordination and materiality with the notion of cognition expands cognitive processes 'beyond the skull' (Cowley & Vallée-Touragnant, 2013), as observed in our case. The solution emerged not from cognitive models but from the physical manipulation of the telephone, which integrated it into the cognitive system rather than just using it as a communication tool. Thus, when studying spontaneous and collaborative problem-solving outside the laboratory, it is necessary to expand both the idea of good and problem gestalt and what should be considered part of the problem-solving process. Good and problem gestalt are determined not only in relation to cognitive models but also in relation to the activity, to its limits and possibilities and to the goals of the interactants as they redefine the problem on a moment-by-moment basis: that is, what is considered good gestalt can change as some possible lines of actions are opened and closed. Similarly, problem-solvers do not simply process information between the initial problem state and the goal state. Instead, what is essential information at each stage is determined as problem-solving progresses. In the analysis presented above, problem-solving progressed through several intermediate stages. However, they were not so much mental as practical, such as the concrete assignment of the visual and the aural to different devices by manipulating the audio connection. Actions of problem-solving are both practical and cognitive. Ultimately, in spontaneous problem-solving, problems are not merely multiple and ill defined: the whole problem, and thus both the good gestalt and the goal state, can be redefined as problem-solving progresses. That is, what could be considered a good gestalt or goal state in our example varied depending on how the participants defined the problem. Thus, we conclude by noting that research on problem-solving in the wild demands approaches in which the social and the material and the human and non-human

are considered potentially equally important parts of problem-solving situations in which certain parts of the cognitive process are distributed to different actors. Our analysis, based on the ideas of distributed cognition and the principles of conversation analysis, contributes to developing such an integrative perspective for studying problem-solving while avoiding the reduction of situated practical actions into mere cognitive processing.

References

Aagaard, J. (2021). 4E cognition and the dogma of harmony. *Philosophical Psychology*, *34*(2), 165–181. https://doi.org/10.1080/09515089.2020.1845640

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

Arminen, I., & Poikus, P. (2009). Diagnostic reasoning in the use of travel management system. *Computer Supported Cooperative Work*, *18*, 251–276. https://doi.org/10.1007/s10606-008-9086-3

Balaman, U. (2021). The interactional organization of video-mediated collaborative writing: Focus on repair practices. *TESOL Quarterly*, *5*(3), 979–993. https://doi.org/10.1002/tesq.3034

Bowden, H. M. (2019). Problem-solving in collaborative game design practices: Epistemic stance, affect, and engagement. *Learning, Media and Technology*, *44*(2), 124–143. https://doi.org/10.1080/17439884.2018.1563106

Bjørndahl, J. S., Fusaroli, R., Østergaard, S., & Tylén, K. (2014). Thinking together with material representations: Joint epistemic actions in creative problem solving. *Cognitive Semiotics*, *7*(1), 103–123. https://doi.org/10.1515/cogsem-2014-0006

Button, G. (2008) Against 'Distributed Cognition'. *Theory, Culture and Society*, *25*(2), https://doi.org/10.1177/0263276407086792

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

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Couper-Kuhlen, E. (2014). What does grammar tell us about action? *Pragmatics, 24*(3), 623–647. https://doi.org/:10.1075/prag.24.3.08cou

Cowley, S. J., & Vallée-Tourangeau, F. (Eds.). (2013). *Cognition beyond the brain: Computation, interactivity and human artifice*. Springer.

Due, B. L. (2016). Fælles orientering som ressource for idéudvikling: En single case-analyse baseret på distributed cognition (DC) conversation analysis (CA) [Shared orientation as a resource for idea development: A single case analysis based on distributed cognition (DC) conversation analysis (CA)]. Nydanske Sprogstudier NyS, *50*, 86–119. https://doi.org/10.7146/nys.v1i50.23799

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., Femø Nielsen, M., & Jalskov, 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

Due, B. L., & Licoppe, C. (2020). Video-mediated interaction (VMI): Introduction to a special issue on the multimodal accomplishment of VMI institutional activities. *Social Interaction: Video-Based Studies of Human Sociality*, *3*(3). https://doi.org/10.7146/si.v3i3.123836

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

Gibson, J. J. (1979). *The ecological approach to visual perception*. Psychology Press.

Gigerenzer, G., & Todd, P. M. (2001). *Simple heuristics that make us smart*. Oxford University Press.

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological Inquiry*, *50*(3–4), 272–302. https://doi.org/10.1111/j.1475-682X.1980.tb00023.x

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

Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics*, *46*(1), 823. https://doi.org/10.1016/j.pragma.2012.09.003

Goodwin, C. (2017). *Co-operative action*. Cambridge University Press.

Goodwin, C., & Goodwin, M. H. (2004). Participation. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 222–244). Blackwell. https://doi.org/10.1002/9780470996522.ch10

Grigorov, D. N., & Snoeck Henkemans, A. F. (2019). Hypothetical questions as strategic devices in negotiation. *Negotiation Journal*, *35*, 363–385. https://doi.org/10.1111/nejo.12297

Hakulinen, A., & Seppänen, E. L. (1992). Finnish kato: From verb to particle. *Journal of Pragmatics*, 18(6), 527–549. https://doi.org/10.1016/0378-2166(92)90118-U

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

Heath, C., & Luff, P. (1992). Collaboration and Control Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work, 1*, 69–94. https://doi.org/10.1007/BF00752451

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 299–345). Cambridge University Press. https://doi.org/10.1017/CBO9780511665868.020

Heritage, J. (2012). Epistemics in conversation. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 370–394). Wiley-Blackwell. https://doi.org/10.1002/9781118325001.ch18

Heritage, J., & Raymond, G. (2012). Navigating epistemic landscapes: Acquiescence, agency and resistance in responses to polar questions. In J. De Ruiter (Ed.), *Questions: Formal, functional and interactional perspectives* (pp. 179–192). Cambridge University Press. https://doi.org/10.1017/CBO9781139045414.013

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

Hutchby, I. (2001). *Conversation and technology: From the telephone to the internet*. Polity Press.

Hutchins, E. (1980). *Culture and inference*. Harvard University Press.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Hutchins, E. (2011). Enculturating the supersized mind. *Philosophical Studies*, *152*, 437–466. https://doi.org/10.1007/s11098-010-9599-8

Ilomäki, S., & Ruusuvuori, J., (2022). Preserving client autonomy when guiding medicine taking in telehomecare: A conversation analytic case study. *Nursing Ethics*. https://doi.org/10.1177%2F09697330211051004

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–23). John Benjamins. https://doi.org/10.1075/pbns.125.02jef

Järvilehto, T. (2009). The theory of the organism-environment system as a basis of experimental work in psychology. *Ecological Psychology*, *21*(2), 112–120. https://doi.org/10.1080/10407410902877066

Koffka, K. (1935). *Principles of Gestalt psychology*. Harcourt, Brace and Company.

Luff, P., Heath, C., Kuzukoka, 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

Mlynář, J., & Arminen, I. (2023). Respecifying social change: the obsolescence of practices and the transience of technology. *Frontiers in Sociology, 8*(1222734). https://doi.org/10.3389/fsoc.2023.1222734

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. (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

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://10.7146/si.v3i3.122781.

Oittinen, T. (2023). Including Written Turns in Spoken Interaction: Chat as an Organizational and Participatory Resource in Video-Mediated Activities. *Research on Language and Social Interaction, 56*(4), 269–290. https://doi.org/10.1080/08351813.2023.2272524.

Olbertz-Siitonen, M. (2015). Transmission delay in technology-mediated interaction at work. *PsychNology Journal*, *13*(2–3), 203–234. https://jyx.jyu.fi/handle/123456789/51325

Raymond, C. W., & Heritage, J. (2021). Probability and valence: Two preferences in the design of polar questions and their management. *Research on Language and Social Interaction*, *54*(1), 60–79. https://doi.org/10.1080/08351813.2020.1864156

Raymond, G. (2003). Grammar and social organization: Yes/no interrogatives and the structure of responding. *American Sociological Review*, *68*(6), 939–967. https://doi.org/10.2307/1519752

Rintel, S. (2013a). Tech-tied or tongue-tied? Technological versus social trouble in relational video calling. In Ralph E. Sprague (Ed.), *Proceedings of the Forty-Sixth Hawaii International Conference on System Sciences*, 3343–3352. IEEE Computer Society. http://dx.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 Électronique de Communication*, *23*(1–2). https://cios.org/EJCPUBLIC/023/1/023123.HTML

Rintel, S. (2015). Omnirelevance in technologized interaction: Couples coping with video calling distortions. In R. Fitzgerald & W. Housley (Eds.), *Advances in Membership Categorisation Analysis* (pp. 123–150). SAGE.

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

Saatçi, B., Akyüz, K., Rintel, S., & Nylandsted Klokmose, C. (2020).
(Re)Configuring hybrid meetings: Moving from user-centered design to
meeting-centered design. *Computer Supported Cooperative Work*, *29*,
769–794. https://doi.org/10.1007/s10606-020-09385-x

Sarathy, V. (2018). Real world problem-solving. *Frontiers in Human
Neuroscience*, *12*(261). https://doi.org/10.3389/fnhum.2018.00261

Schegloff, E. A. (1987). Analysing 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. (2007). *Sequence organization in interaction: A primer in
conversation analysis*. Cambridge University Press.

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

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

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

Speer, S. A. (2012). Hypothetical questions: A comparative analysis and
implications for "applied" vs. "basic" conversation analysis. *Research on
Language and Social Interaction*, *45*(4), 352–374.
https://doi.org/10.1080/08351813.2012.724987

Stasser, G., & Abele, S. (2020). Collective choice, collaboration, and
communication. *Annual Review of Psychology*, *71*, 589–612.
https://doi.org/10.1146/annurev-psych-010418-103211

Steffensen, S. V. (2009). Language, languaging and the extended mind
hypothesis. *Pragmatics & Cognition*, *17*, 677–697.
https://doi.org/10.1075/pc.17.3.10ste

Steffensen, S. V. (2013). Human interactivity: Problem-solving, solution-probing and verbal patterns in the wild. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition beyond the brain: Computation, interactivity and human artifice* (pp. 195–221). Springer. https://doi.org/10.1007/978-1-4471-5125-8_11

Stevanovic, M., & Peräkylä, A. (2012). Deontic authority in interaction: The right to announce, propose, and decide. *Research on Language & Social Interaction*, *45*(3), 297–321. https://doi.org/10.1080/08351813.2012.699260

Stivers, T. & Rossano, F. (2010). Mobilizing Response. *Research on Language and Social Interaction*, *43*(1), 3–31. https://doi.org/10.1080/08351810903471258

Stivers, T. & Sidnell, J. (2016). Proposals for activity collaboration. *Research on Language and Social Interaction*, *49*(2), 148–166. https://doi.org/10.1080/08351813.2016.1164409

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. 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. https://doi.org/10.1007/978-3-030-64922-7_9

Tarasmundi, S. B., & Linell, P. (2017). Insights and their emergence in everyday practices: The interplay between problems and solutions in emergency medicine. *Pragmatics & Cognition*, *24*(1), 62–90. https://doi.org/10.1075/pc.17002.tra

Vallée-Tourangeau, F. (2014). Insight, interactivity and materiality. *Pragmatics & Cognition*, *22*(1), 27–44. https://doi.org/10.1075/pc.22.1.02val

VISK. (2004). *Iso suomen kielioppi* [The comprehensive Finnish grammar]. Hakulinen, A., Vilkuna, M., Korhonen, R., Koivisto, V., Heinonen, T.-R., & Alho, I. (Eds). Finnish Literature Society. http://scripta.kotus.fi/visk

Wertheimer, M. (1982). *Productive thinking* (Enlarged ed.). University of Chicago Press.

Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Thinking & Reasoning*, *21*(1), 5–39. https://doi.org/10.1080/13546783.2014.886625

Zhang, J., & Patel, V. L. (2006). Distributed cognition, representation, and affordance. *Pragmatics & Cognition*, *14*(2), 333–341. https://doi.org/10.1075/pc.14.2.12zha

Öllinger, M., & Goel, V. (2010). Problem solving. In B. M. Galtzeder, V. Goel, & A. von Müller (Eds.), *Towards a theory of thinking: Building blocks for a conceptual framework* (pp. 3–21). Springer. https://doi.org/10.1007/978-3-642-03129-8_1