Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The here-and-now of "real-time visual assistance". Assembling gestalt contexture in blind-sighted navigation

Damien Rudaz

daru@hum.ku.dk

University of Copenhagen

Brian L. Due

University of Copenhagen

Barbara Patricia Nino Carreras

University of Copenhagen

Research Article

Keywords: Gestalt contexture, Road crossing, Scene descriptions, Multimodal AI, Descriptive Accuracy, Ethnomethodology and Conversation Analysis

Posted Date: July 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10494575/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

The here-and-now of “real-time visual assistance”

Assembling gestalt contexture in blind-sighted navigation

Name and affiliation redacted for review

Abstract

This study contributes to ethnomethodological research on gestalt contexture by extending it from relatively stable and symmetrically available perceptual objects to dynamic settings marked by practical urgency and asymmetric perceptual access. Using an artificial agent’s situated conduct as a point of comparison, we explore how an experienced sighted guide’s actions methodically segmented observable and hearable properties of participants’ surroundings, and indexed various temporal and normative orders—such as those of the ongoing traffic—as constituent details of the practical activity of “crossing the street”. Our empirical observations help specify the set of practices glossed by the terms “real-time visual assistance” and “real-time descriptions” in recent research on human and human–agent interaction. We document how a human guide’s grammatically declarative statements do not merely describe a pre-existing scene (immediately experienced as coherent) but, rather, progressively make selected features of the surroundings into constituent details of a shared here-and-now.

Keywords

Gestalt contexture; Road crossing; Scene descriptions; Multimodal AI; Descriptive Accuracy; Ethnomethodology and Conversation Analysis;

Word count

8978 words

1. Introduction

City streets constitute a material arrangement to which the notion of gestalt contexture has been leveraged extensively by ethnomethodological approaches (Laurier et al., 2020; Liberman, 2013; Lloyd, 2020; Smith, 2017, 2024). Notably, among urban activities, "crossing the street" (Cromdal et al., 2026; McIlvenny, 2019; Merlino & Mondada, 2019; Relieu, 2023a) emerges as a particularly fruitful case for providing knowledge about the practical accomplishment of perception (Coulter & Parsons, 1990). As a practice of "personal mobility in public space" (Mondada & Tekin, 2024)—and alongside, for example, queuing (Ayaß, 2020; Garfinkel & Livingston, 2003)—crossing provides a heuristic glimpse into the indexical relations (and their accomplishment) at play in "generating recognizable actions" (Sacks, 1995) without formulating them in so-many-words (Mondada & Tekin, 2024). For instance, a pedestrian is only understandable as "projecting to cross a road" in relation to an assembled setting of other pedestrians, cars, bikes, and other visuospatial features of the environment such as roads and parallel white lines on the ground—which themselves become recognizable as a "zebra crossing" by virtue of their being constituents of a produced gestalt contexture that they simultaneously contribute to establishing, and so on ([Author B, redacted for review]).

The present article contributes to this line of EMCA research on perception and mobility in urban environments by documenting how features of the urban setting emerge as constituent details of "street crossings" in which verbal accounts of the environment are produced by a sighted participant navigating alongside a blind participant. We study how declaratives such as "the road appears to be clear", "there is a car behind us", or "they are waiting" emerge as resources for action. More specifically, the practical adequacy of such utterances depends not on an (exogenously defined) propositional accuracy or response speed, but on whether they index and produce a shared temporally, spatially, and normatively coherent gestalt contexture. In sum, this paper explores how street crossings are encountered by participants as practical fields of action whose relevant features are assembled over time, rather than as static scenes to be described.

We begin by studying how the spoken contribution of the multimodal voice agent (Meta smart glasses using Meta AI to generate descriptions of the visual environment) came to be treated as "detached" (Dreyfus, 1990) from the rapidly shifting temporal order of human participants' conduct and of the traffic. In being accountably "delayed"—relative to the practical expectancies of the unfolding activity—this voice agent's account of the traffic was decoupled from the "vivid present" (Au-Yeung & Fitzgerald, 2023; Garfinkel, 2006) of the co-present humans. We specify what those prospective crossers treated as a "clear road" in relation to the practical urgency of an imminent crossing—an orientation that may differ, for instance, from the one displayed by museum visitors using the same expression, "clear road", when assessing the painting of a city street. In doing so, we document the taken-for-granted system of relations as part of which such a statement about a road is rendered intelligible by participants.

Then, using the artificial agent's situated conduct as a point of comparison, we explore how, in assisting a visually impaired person, an experienced sighted guide's actions methodically segmented observable and hearable properties of participants' surroundings, and indexed various temporal and normative orders—such as those of the ongoing traffic—as constituent details of the practical activity of "crossing the street". In doing so, this guide's contributions co-constructed and indexed a contexture together with other participants. We highlight just how this sighted guide's directives constituted relevant resources for the blind individual (i.e., they were "procedurally consequential"; Schegloff, 1992) only insofar as they indexed and produced a common "here-and-now".

Finally, we connect these insights—framed in the vocabulary of gestalt contexture—with recent challenges pertaining to the design of assistive multimodal voice agents producing "real-time" descriptions (Chang et al., 2024) or "real-time visual assistance" (Xie et al., 2022) for blind participants. That is, we specify some situated practices glossed by these terms: those practices in and through which "real-time" is locally accomplished.

2. Two cases for "revealing the ('hidden') transparent work of achieved coherence"

2.1. The perspicuousness of interactions involving multimodal voice agents

By drawing on a comparison of two different situations involving two different assistive agents (human and artificial), this article employs one of countless methods for "render[ing] the phenomenon available again" (Eisenmann & Rawls, 2023), or for making "an invisible world observable" (Garfinkel, 1965). By producing contributions demonstrably discordant with human participants' ongoing courses of action, by publicly creating "split[s] in patterns of salience and obligation" (Hutchinson & Chikwira, 2026), the artificial agent's conduct rendered visible *some* praxeological foundations of human participants' gestalt-coherence. Such cases allow us—the "professional analysts" studying strips of interaction after the fact (Schegloff, 1996)—to better grasp the situated methods through which gestalt contextures are assembled[1]. In a different vocabulary, we treat these cases as "'Heideggerian'" examples, in Garfinkel's (1996) sense of being "perspicuous in revealing the ('hidden') transparent work of achieved coherence".

2.2. What street crossings involving blind and sighted participants can teach us about locally assembled gestalt contexture

Early research on gestalt principles and gestalt contexture has largely focused on two kinds of objects:

1. **"Stable visual objects"** (Meyer, 2022). Canonical Gurwitschian demonstrations, and much of the classical Gestalt tradition, were built around paper-and-pencil figures (Garfinkel, 2021) or reversible drawings (Lynch & Eisenmann, 2022). Although Gurwitsch also extended the analysis to sequential auditory objects such as melodies (Gurwitsch, 1964), those were still studied from the perspective of a "disengaged" (Meyer, 2022) perceiver rather than as locally produced phenomenon—the focus was on the listening rather than on "playing the melody" (Lynch & Garfinkel, 2022).

   In relation to such objects, participants' perception is not explored in its intertwinement with any form of practical urgency, where "urgency" is understood in two senses: (1) as threats to one's physical safety, as may be the case when crossing a road near speeding cars; (2) as concurrent temporal orders that must be managed by the perceiving agent, as in multiactivities

[1] We do not, however, take principled issue with the idea that the frictions presented in this document might disappear as these technologies are developed in the future. We merely study these technologies in their current state and as they are presently being used and embedded in the world by humans.

such as conversing while navigating an urban environment (Haddington et al., 2013, 2014). These objects are those of the phenomenologist involved in producing a second-order reconstruction of their own meaning-making practices, or of the museum visitor contemplating a work of art. They allow the perceiver to experience their own activity as that of a subject purposedly attending to features of the environment *as* objects (Dreyfus, 2014).

2. **Perceptually shared objects.** That is, objects to which participants are assumed to share symmetrical perceptual access—or, at least, regarding which asymmetries of perceptual access do not surface as practically relevant. In such cases, the gestalt organization is treated as immediately available (Gurwitsch, 1964; Kanizsa, 1979; Koffka, 1935/2013) rather than progressively assembled ([Author B, redacted for review]).

   Notably, studies involving some degree of attention to the gestalt coherence of everyday life activities (whether the technical nomenclature of "gestalt" was explicitly used or not) have extensively documented the *visual component* of gestalt contextures—most recently, for instance, in Carlin et al. (2025), and Moutinho et al. (2025). More broadly, as noted by Popova (2018), a traditional ethnomethodological approach has understood witnessability in terms of observability. Hence, while "visually available mundane orders" (Hester & Francis, 2003) have been a prominent focus of research, haptically and sonically available orders have received comparatively less attention.

In light of these two characteristics of the objects traditionally studied in research on gestalt contexture, street crossings involving sighted and blind participants constitute heuristic settings for extending our understanding of how such contextures are locally assembled. Indeed, these settings are assembled by participants witnessably…

1) …dealing with the practical urgency of various temporal orders—in the strong sense of "urgency" defined above. They allow us to study participants responding to urban space as space structured *for action*: that is, as a space of rights, obligations (Deppermann, 2019; Merlino & Mondada, 2019; Smith, 2017), and opportunities for action around rapidly moving vehicular units. These participants are not merely engaged in "seeing that" (Hanson, 1979; Wittgenstein, 1968) there is a street, a car, or a sidewalk; rather, their conduct segments and indexes features of the environment tied to they own course of action, in relation to various normative orders.
2) …engaged in the production and interpretation of grammatically declarative contributions. Notably, sighted "guides" ordinarily produce what are, in strictly grammatical terms, declarative statements—that is, utterances understood merely to report an existing state of affairs rather than to perform one (Austin, 1975). This activity, therefore, involves segmenting and reporting, in so many words, features of the environment (accomplished as features precisely in being so formulated) for the blind participant.

These blind–sighted street crossings are especially fruitful for developing an understanding of gestalt coherence as a *process* rather than as the immediate experience of a whole (for a more comprehensive exploration of a process-philosophical ethnomethodology, see ([Author B, redacted for review]). Sighted members of a setting do not ordinarily begin with an inventory of discrete environmental properties and then assemble a coherent situation ([Author B, redacted for review]): they act within an already organized phenomenal field, in which details have their sense by virtue of their place in the ongoing activity (Garfinkel, 1967; Gurwitsch, 1964). This immediacy is less evident for blind participants (Rickard et al., 2019), who must often construct a sense of the whole from elements

progressively discovered haptically, tactually, and/or by managing accounts provided by sighted co-participants ([Author B, redacted for review]). In this respect, situations of sighted "assistance" are particularly perspicuous in that they allow us to document gestalt contextures as *processually* assembled. In particular, because they involve participants without sight being reported perceptually unavailable features of their surroundings, these situations provide a particularly revealing vantage point from which to examine the linguistic practices through which gestalt contextures are built. They render available sighted participants' analysis of features of the observable surroundings as practically relevant—or, more accurately, they allow us to document how a setting is constructed, in part, through sighted participants' formulation, segmentation, categorization, of (what becomes) the surrounding environment.

3. "Real-time" descriptions and the problem of relevance

A long-standing problem for designers of multimodal AIs—and, more recently, of multimodal voice agents—is to identify which features of a setting such a system should perceive, discretize, and respond to as relevant properties of this setting. Numerous works have documented how the "problem of relevance" (Dreyfus, 2007)[2] or "frame problem" (McCarthy & Hayes, 1969) emerges in relation to image, object, or scene descriptions produced by AIs for blind or low-vision individuals (Chang et al., 2024; Stangl et al., 2021)—that is, in regard to the adequate length, level of detail, timing, thematic focus, and so on, of said descriptions. Accordingly, we do not attempt the (always indefinitely elaboratable—Garfinkel, 1967; Lynch & Eisenmann, 2022; Waismann, 1951) inventory of a settings' practically relevant features. Rather, we focus on the much humbler task of understanding how declaratives produced by a sighted co-participant become unproblematic constituents of a blind participant's gestalt contexture. Indeed, the aforementionned body of work in human-computer interaction or in computer science points to the need to better specify the situated *practices* glossed by the term "real-time descriptions"—whether in human or artificial assistance. That is, to document the methods through which spoken turns by human or artificial assistants emerge as timely and meaningful rather than "delayed" statements about a blind individual's surroundings—methods with significant, and perhaps underappreciated, implications for the design of voice agents intended to provide "real-time" assistance. As we hope to demonstrate, accounting for the situated relevance of a "guide"'s contribution requires attending to how a shared unity of space and time is built and ongoingly displayed by participants.

4. Data and methodological approach

The two excerpts analysed below belong to a corpus of interactions involving blind participants navigating various environments while relying on verbal assistance provided either by (1) multimodal voice agents, or (2) sighted human co-participants (whether physically co-present or accessing the scene remotely through the blind participant's smartphone or glasses). This corpus was collected as part of two different research projects in which we were engaged [projects names and dates redacted for

[2] A phenomenological antecedent of this "problem of relevance" can, of course, be found in Schutz (1970). Although Schutz did not engage with its implications for artificial intelligence, he discusses how certain features come to be selected from "the totality of the world which is pregiven to life" (Schutz, 1996).

review]. These particular excerpts were selected because they involve two contrasting street crossings; they are transcribed using Jeffersonian conventions for talk (Jefferson, 2004) and Mondadian conventions for embodied conduct (Mondada, 2018). Data collection took place in Denmark and was approved by the researchers' university ethics committee. The data were collected with participants' informed written consent.

Following Wiener et al.'s (2009) taxonomy, the following excerpts are situations of navigation involving "aided" wayfinding: here, through various resources, such as a cane, pedestrian traffic-light sounds, a voice assistant, or a sighted human guide. The first excerpt features a blind participant (Laura) trying out the Meta Glasses' capacity to assist her while navigating familiar streets. Laura was recorded by two researchers during an afternoon of ethnographic fieldwork focusing on her discovery and everyday use of multimodal voice agents. One of these researchers also acted as Laura's sighted assistant that day. The second excerpt involves a blind participant (Sophia) during her regular riding session with her usual sighted instructor (Trine). Both riders were filmed by a researcher (Frank), who was also riding a horse, at the rear of the single-file formation. The first author of the present paper was involved in collecting the data in both cases.

This research draws on a process-philosophical version of ethnomethodological conversation analysis ([Author B, redacted for review]) that understands the world as in becoming in each moment for the first time (Garfinkel, 1967). To unpack the role of practices glossed as "scene descriptions" in constructing gestalt contextures, we rely on the technical terminology of conversation analysis to describe the sequential unfolding of interaction (Sacks et al., 1974).

5. Analysis

5.1. Excerpt 1 – Assessing whether a road is clear with a multimodal voice agent

The following excerpt was recorded during an ethnographic study on Laura's (LAU) use of various assistive technologies (see[Author A, redacted for review]), during which she was filmed navigating various environments and accomplishing her daily activities by two researchers over the course of an afternoon. LAU was born blind and regularly uses multimodal voice agents in her personal life but had never previously tried any form of smart glasses.

The voice agent used in Excerpt 1 is the combination of the Meta Glasses (as the embodied technical device) and Meta AI (the artificial assistant generating the spoken responses). This system belongs to a range of currently widely available commercial solutions, such as ChatGPT's multimodal 'voice mode' (OpenAI, 2025). It generates responses based on both hearable information (mainly, human speech) and observable data (the data streams provided by the smart glasses' camera, which faces the front of the glasses). Whether or not such systems are designed for the assistance tasks they are required to perform, they are increasingly used by blind individuals—as one resource among others—in their daily lives (Gonzalez Penuela et al., 2024, 2025).

This excerpt begins as LAU stands facing a busy crossroad in Copenhagen (see Figure 1). LAU is wearing and interacting verbally with the Meta glasses (MET), which she has been testing as part of her journey through this urban environment for several minutes now. Cars hearably passing on the road directly in front of LAU are identified as "car1", "car2", and "car3". Sandra (SAN), one of the researchers, is standing to LAU's right, filming her. LAU and SAN are themselves being filmed from a

wider angle by another researcher positioned in the background. LAU has previously informed SAN that she may try to cross the road, and to stop her if she attempts something dangerous. The data are in English.



Figure 1. Laura (LAU), the blind participant, facing a busy crossroad next to Sandra (SAN), the researcher, while wearing the Meta Glasses (MET).

```
01 LAU    +[hey] meta? (0.4)#* tell me if the road is clear or not
   lau  >>+faces crosswalk, leaning forward->
          ((red light sound indicator is ticking))>>
   car1                       *passes on the road facing LAU->
   fig                       #figure 2
02         (0.2)*(1.2)
   car1       ->*
03 MET     #click ((photo capture sound))#
   fig     #figure 3                      #figure 4
```



Figure 2. Laura standing next to the pedestrian traffic light pole, her head and body oriented toward the crosswalk, while her cane remains close to her body.



Figure 3. Laura faces the empty road immediately in front of her at the exact moment when the meta glasses take a photo.



Figure 4. The exact photo taken by the meta glasses. This photo is used as a reference by the meta glasses for generating their response to Laura's question during the rest of the fragment.

```
04          (3.0)*(0.5)
   car2          *passes on the road facing LAU->
05 MET      tududu ((processing sound))
06          (0.2)
07 MET      the# road appears* to be* clear
   car2                    ->*
   car3                              *passes on the road facing LAU->
   fig         #figure 5
```



Figure 5. Cars passing by on the road in front of Laura, while the meta glasses state that the road appears to be clear.

```
08 MET    [with] no] v+isible obstructions# or tr*[affic congestion]
   lau              ->+straightens up, transfers cane to right hand->
   car3                                        ->*
   fig                                    #figure 6
```



Figure 6. LAU straightens up before producing a negative assessment of MET's response

```
09 LAU    [uh::    ]                          [definitely% not. .h[uhuhu]]
10 SAN                                                        [no.hoh]o:%
   san                                                    %---shakes head----%
11        (0.2)+(0.3)
   San       ->+transfers cane to left hand>>
12 LAU    yeah. ok.+              so no- definitely:             +
   lau             +touches the pedestrian crossing pole's button+
13        (0.4)
14 LAU    not (.) not passing that test .huhuh
15        (0.2)
16 SAN    no.
17 LAU    =.huhuh
18        (0.1)
19 SAN    also becau[se] (.) i- don't you (.) think there is a bit of a
20                [huhuh]
21        delay: (.) when it tells you maybe [things] have changed already
22                                           [yeah]
23        (0.2)
24 LAU    yeah. maybe- yeah things could have changed too (maybe soon after)
```

This fragment starts as LAU is displaying her readiness to cross the street (Figures 2 and 3). This readiness as a "prospective crosser" (Relieu, 2023b; Sudnow, 1972) is in part recognizable (and made recognizable) in LAU's spatial and embodied positioning within the urban environment: she is facing the road, slightly bent forward, feet a few centimeters from the end of the curb, in front of the zebra crossing—itself so configured as a "visible inscription" (Merlino & Mondada, 2019) rather than as mere consecutive white lines—and next to a pedestrian traffic light, as the sound of the red pedestrian signal is hearable. Additionally, she is standing on the textured pavement ordinarily used in Danish urban

environments to tactually signal the presence of a pedestrian crossing to blind individuals. LAU's embodied conduct and positioning relative to these features of the urban environment render her stance recognizable as that of a (blind) pedestrian projecting an imminent crossing of the road. Significantly, her position in space and stance are achieved as part of a sequential context configured by her request, a few moments earlier, that the researcher standing to her left (SAN) stop her should she attempt to cross dangerously. All these visuospatial and sequential features "hang together" (Gurwitsch, 1964, p. 128): they are constituents of a gestalt contexture whose relations render LAU's projected action immediately meaningful without being put into words (see Mondada and Tekin, 2024, on the unarticulated intelligibility of navigation in public spaces). SAN contributes to the accomplishment of this contexture by positioning herself beside LAU rather than in front of her: she thereby leaves open the space in front of LAU as unobstructed and "walkable", reinforcing the projectability of LAU's trajectory. The zebra crossing in front of LAU is thus constituted as the space into which her next movement may be launched.

In short, this assembled setting—which participants' practices both develop and index—can be summarized using Liberman's (2013) formulation: LAU's prospective crossing practices are "not merely actions—they are demonstrations" rendering publicly observable "the action one is taking" and involving "methods for exhibiting". Participants' spatial positioning, their stance, the sequential history, and the material arrangement of the crosswalk constitute salient features of the "ecology of action" (Goodwin, 2018) through which crossing is made recognizable and accountable as a relevant next action for LAU. These elements "implicate each other" as an "ensemble of mutually relevant functional characters" (Garfinkel, 1965) that constitute LAU as a "potential crosser" (Merlino & Mondada, 2019) whose projected action is identifiable "at a glance" (Sudnow, 1972).

As part of this setting, LAU triggers MET's listening state by using the preset vocal command "hey Meta" (l. 1). She then asks MET to tell her "if the road is clear or not" (l. 1). Significantly, LAU's question is produced in relation to—and as a new constituent of—the accomplished features of the situation described in the previous paragraph. Rather than merely requesting a detached description of the state of traffic, LAU's question reinforces the "scenic intelligibility" (Jayyusi, 1988) of her orientation to walk forward on the zebra crossing. This request to MET enacts assumptions regarding what counts as a "clear" road in relation to a prospective crossing: what is requested from MET is not merely "seeing that" there is a road, but seeing the road "as" the space of a projected action of walking forward. As she finishes uttering this turn, a car passes by (l. 1). Although the passing car is not commented upon by LAU, it nevertheless provides hearable evidence that the road is not "clear" at the moment her turn comes to completion.

After a pause of 1.4 s (l. 2), MET emits a sound indicating that it is taking a photo using its forward-facing camera—currently oriented toward the crossroad (l. 3). From an exogenous analytic perspective, with retrospective access to the photo taken by MET, this photo displays a clear road immediately in front of LAU, a red pedestrian signal in the distance, and zebra lines extending across the lane toward the opposite pavement (Figure 4). Still from a technical point of view, this picture contains the visual information from which the MET glasses will generate their response to LAU's spoken request during the remainder of this excerpt. After another pause of 3.5 seconds (l. 4), the MET glasses emit a sound (l. 5), conventionally indicating that they are processing the photo that was just taken. Just prior to MET's statement that "the road appears to be clear" (l. 7), an approaching car becomes audible (l. 6) in the lane directly in front of LAU. This car then passes in front of LAU exactly as MET utters "appears to be clear" (l. 7, Figure 5). A second car then passes in front of LAU as MET utters the second part of its turn, namely that there are "no visible obstructions or traffic congestion" (l. 8, Figure 6).

In overlap with the end of MET's turn, and immediately after its description of the road as "clear", LAU produces the elongated response token "uh::" (l. 9). This token here functions as a trouble marker (Schegloff, 2005): it displays an incipient disaffiliation with MET's ongoing account of the traffic, subsequently made explicit by LAU's negative assessment of MET's turn-so-far ("definitely not", l. 9) and laughter—here positioned as a display of disagreement (Osvaldsson, 2004). However, at no point does LAU explicitly surface the perceptually available features of the setting (whether the "audible order"—Relieu, 2023a—of the traffic, the sound of the two cars passing by right in front of LAU, the ticking of the red light, etc.) in relation to which MET's statement is being challenged. In rejecting MET's statement that the road is "clear"—as cars are passing by—LAU indexes her surroundings as "an ordered as well as a temporally defined ecology of noise" (Atkinson, 2007); however, these dissonant features of the setting remain unformulated, and unproblematically so. In sum, LAU's turn does not segment specific properties of a locally assembled "traffic gestalt" (Laurier et al., 2020) that may be experienced as incoherent with MET's description-in-so-many-words.

Significantly, LAU's negative assessment co-occurs with a shift in her posture: she straightens up and transfers her cane to her left hand (l. 8, Figure 6). This bodily reorientation interrupts what had been, until that moment, her "ready" stance. It is interpretable either as interrupting the projection of her imminent crossing of the street and/or as ending a sequence of examination of the road, as part of which LAU had provided MET with a stable point of view on the traffic. LAU then extends her account to MET's overall performance by stating that it (i.e., MET) is "not passing that test" (l. 14). Through this formulation, LAU retrospectively displays that she had so far been engaged in layered courses of action, i.e., in crossing the road and, simultaneously, in a form of "testing" (Tuncer et al., 2024) of the Meta glasses' capabilities. SAN participates in accounting for MET's performance by affiliating verbally with LAU's negative assessments ("no", l. 10 and l. 16) and by shaking her head (l. 10).

In the end, SAN formulates a possible explanation for MET's inadequate description of the traffic, namely that there may be "a bit of a delay" (l. 19 and l. 21) between MET's accounts of the situation and what is treated as the current situation. In doing so, SAN thematizes the possibility of trouble with the "now" of the interaction, i.e., a temporal discrepancy between MET's talk and the ongoing "vivid present" (Au-Yeung & Fitzgerald, 2023; Schutz, 1944) of participants as they engage in various courses of action. SAN's next turn clarifies this hypothetical dissonance between MET's talk and a rapidly shifting temporal and material order: she states that "when it tells you maybe things have changed already" (l. 21). LAU produces a token of agreement ("yeah", l. 22) in overlap with SAN's explanation, then ultimately affiliates with it (l. 24), although her affiliation is epistemically downgraded through the modal adverb and epistemic stance marker "maybe" (Stivers et al., 2011).

SAN's suggestion of a latency with MET's responses (and LAU's candidate affiliation, l. 24) further displays and enacts what counts as a response produced "on time". SAN's turn manifests an ephemeral duration of the "now" relative to which the description of the road as "clear" was valid or useful before becoming "delayed". It calls into question the adequacy—so far taken for granted in this excerpt—between participants' talk (including MET's spoken contributions) and co-occurring temporal orders relevant to the task at hand (here, the unfolding traffic). This co-produced "now", whose duration is contingent on the multiple activities being accomplished, was immanent to the contexture against which MET's description was originally resisted by LAU (l. 9 and l. 12). Among other layered activities, it is the "now" of a course of action projecting an imminent forward movement from the blind user, as soon as an opportunity for safely crossing the street becomes salient.

5.2. Excerpt 2 - Crossing the road as group of horse riders

In the previous excerpt, the artificial voice agent's account of the traffic is negatively assessed and is followed by a manifest suspension of the projected road crossing. By contrast, the following excerpt documents a "successful" guided road crossing in two respects: the participants do cross the road and, more importantly, a sighted participant's accounts of the traffic are not problematized by the co-participants. This excerpt comes from a video recorded during a horseback ride involving an experienced blind rider (Sophia – SOP) on her horse (SH), her instructor (Trine – TRI) on her own horse (TH), and a researcher responsible for collecting the data (Frank – FRA) also on a horse (FH). The three cars passing on the road near the riders are identified as "car1", "car2", and "car3". The ride was recorded using two GoPros: one attached to Sophia's chest, and the other to the researcher's chest. The instructor was positioned at the front of the group, followed by Sophia, and then Frank (see Figure 7). Trine and Sophia are Danish; in this fragment, their talk alternates between Danish and English. Frank neither speaks nor understands Danish, a fact known to all participants.

For the sake of readability, the transcript attributes movements in space to the horses (SH, TH, and FH). Nevertheless, horses and their riders are (often) analyzable as forming assemblages whose movements result from finely tuned coordination that cannot be unpacked on the basis of this recording alone.

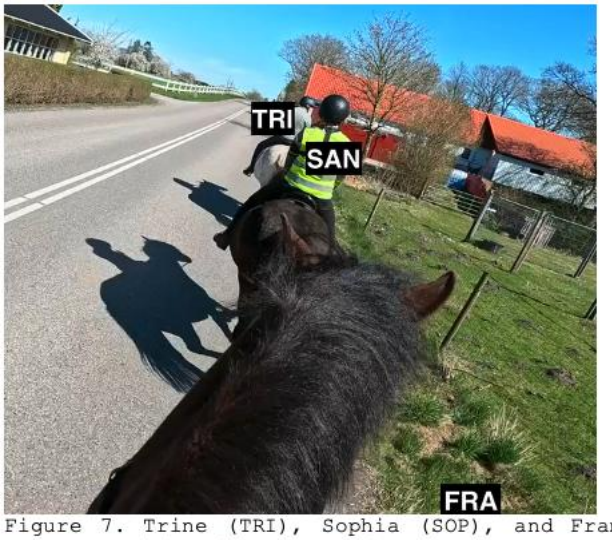


Figure 7. Trine (TRI), Sophia (SOP), and Frank (FRA – out of frame, holding the camera), riding in single file along the edge of a paved road on their respective horses.

```
01 SOP     it depends if you have been born uh: (0.2) blind from birth (.) or:
02         if you a:re
03         (0.4)
04 TRI     oh ye[ah+ yeah (...)]
   tri             +raises left arm, whip extended horizontally->
05 SOP          [have become >blind] if you're blind< from birth i think its
06         (0.7)
07 SOP     *yo+#u- (0.2) you*                  + have to  +eh:m
   car1    *---passes by----*
   tri      ->+palm of the hand facing the car+lowers arm+
   fig         #figure 8
```



Figure 8.

```
08          (0.7)
09 SOP      maybe it's more difficult because you have to learn to accept
10          that you have a balan- a- (0.9) >what to call it that you
11          can< balance yourself
12          (0.6)
13 TRI      w[e +#will] stop here!
   tri          +raises arm, palm up front->
   fig           #figure 9.1
14 SOP       [like-]
15          (0.2)%(0.4)+(0.2)#
   tho           %rotates right, neck across the lane->
   tri                 +torques towards back->
   fig                       #figure 9.2
```



Figure 9.1.



Figure 9.2.

```
16 TRI      °wo:%°
   tho         %stops->
17          (0.2)
18 SOP      °stop°
19          (1.1)$(0.3)
   sho           $stops->
20 TRI      der kommer:: >car behindΔ us.<
            there is a car coming behind us
   fho                              Δstops->
21          (1.1)
22 TRI      *and (.) one in front.
   car2     *stops behind horse riders, in their lane>>
23          (6.8)
24 SOP      dja:
25          (3.7)+*# (0.4)  *          (0.2)             +#(0.5)
   tri           +gazes twd  car2, points to hiking trail+points to self->
   car3           *passes by*
   fig             #figure 10.1                            #figure 10.2
26          +#                  (1.5)                  +         (0.2)        +
   tri   ->+points to hiking trail, fully extended arm+partially retracts arm+
   fig       #figure 10.3
```


Figure 10.1.


Figure 10.2.


Figure 10.3.

27 TRI **+vi% ridder over de +holder tilbage (.) ko[m]-%#**
we are crossing the street because they (car) are waiting come
tri +gives squeezes with heels, alternates gaze between lanes->
tho %,,%walks forward->
fig #figure 11.1
28 SOP **tsk [ok]$ay**
sho ->$walks behind THO->
29 **(0.2)+#(0.4)**
tri ->+
fig #figure 11.2
30 SOP **weΔ are+#(pa- eh-) (.) cross**
fho ->Δwalks behind SHO->
tri +fully raises left arm, palm up front, gazes towards car behind->
fig #figure 11.3


Figure 11.1.


Figure 11.2.


Figure 11.3.

31 **(0.9)+ (1.1) +(0.2)%#(1.7)$(2.9)Δ#(0.3)**
tri ->+partially retracts left arm+both hands on reins>>
tho %reaches hiking trail>>
sho $reaches hiking trail>>
cho Δreaches trail>>
fig #figure 12.1 #figure 12.2


Figure 12.1.


Figure 12.2.

32 SOP **og så rider vi til venstre snart ikke?**
and then we'll turn left soon right?
33 **(0.2)**
34 TRI **ja**
yes
35 **(0.2)**
36 SOP **yes. (0.2) then we will go to the left (.) soon**

The excerpt begins as SOP, FRA, and TRI have been engaged for several seconds in a conversation about balancing oneself on a horse without vision (l. 1 to l. 9). However, after a pause following SOP's turn (l. 12), TRI directs the group ("we") to stop. The togetherness of the riders as a "mobile formation" (McIlvenny et al., 2014) had so far been accomplished and rendered recognizable for other vehicular units (Goffman, 1971)—here, cars—partly through the coordinated single-file positioning of the horses, which follow one another closely and maintain a similar pace. In referring to the three riders and their horses as a "we", TRI formulates, for the group itself, their status as such a vehicular unit.

TRI's directive to "stop here" (l. 13) enacts an abrupt shift in participants' common course of action: it is followed by an orientation of the riders toward *features of the surrounding environment* as now relevant to the task at hand. SOP participates in producing this pragmatic shift, as she does not pursue her turn-initial "like-", started in overlap with TRI (l. 14). SOP's cut-off turn is analyzable as launching a post-gap self-continuation (Rönnqvist & Lindström, 2021) that ties it to the previous talk: in abandoning it, she aligns with the new course of action initiated by TRI. All participants then remain silent (l. 15), further enacting a shift away from the alternation of turns-at-talk characteristic of ordinary conversation (Sacks et al., 1974). The two verbal contributions subsequently uttered by SOP and TRI ("wo" and "stop") are hearably addressed to their horses: these are terms routinely used by both riders to instruct their horses to stop, here produced at low volume. Altogether, 3 seconds elapse before TRI produces a new turn addressed to the group of riders (l. 20). Thus, unlike, for instance, in "problematic crossings" (Merlino & Mondada, 2019), these participants' suspension of their previously alternating turn-by-turn talk does not constitute a "slight discontinuity" (Merlino & Mondada, 2019) in the verbal exchange or its topical trajectory; rather, it initiates a sequence of coordination accomplished through several turns—both within the group of riders and with other mobile units. Moreover, in this specific ecology, where cars can be heard before they are seen, silence not only displays a shift in attention toward the surrounding environment but also constitutes a condition for this shift. By suspending talk, participants produce optimal hearing conditions for detecting incoming cars, while also providing TRI with a slot in which to produce further instructions or calls for action.

As she requests the group to stop, TRI raises her arm, palm up (l. 13, Figure 9.1). This gesture, which accompanies her instruction to stop, is not visually available to SOP but only to FRA and to potential drivers. TRI then turns her horse so that its neck extends across the closest lane (and toward the path across the street) while she torques toward the back (l. 15, Figure 9.2). This embodied conduct by TRI and her horse not only projects an endeavour to cross the street but also physically occupies the space from which this crossing would begin. After both TRI's and SOP's horses have fully stopped (l. 16 and l. 19), TRI states that there is a car coming "behind" them (l. 19). After a pause of one second, she adds that another car is coming from the "front" (l. 22). In doing so, she engages in ordinary "structuring" work (Mlynář, 2026) as she establishes a "front" and a "back" in the space surrounding participants.

The specific segmentation of space produced by TRI is activity-relevant: As Mlynář (2026) remarks, "structure is 'given' only in and through the demonstrable practical orientations that are constitutive of particular activities"[3]. In this excerpt, rather than being intrinsic properties of the urban environment, "front" and "behind" are practical accomplishments whose relevance emerges in relation to, among other things, the location of the hiking trail, the riders' and their horses' projected trajectory within the material arrangement recognizable as a crossroad, locally significant properties of their bodies' front/back asymmetry (Todes et al., 2001) such as the fact that horses and humans move forward

[3] Although this remark is formulated by Mlynář (2026) with respect to structured objects understood as artifacts, we posit that it remains equally valid when the object being structured is the surrounding environment.

more readily than backward, etc.[4] In a different vocabulary, "front" and "behind" emerge as such through their "functional significations" (Garfinkel, 2021), that is, in relation to each other and to other constituents of a gestalt-contexture—much like the pairs of dots used as a paradigmatic example in the literature on gestalt contextures (Gurwitsch, 2010; Lynch & Eisenmann, 2022; Meyer, 2024).

Independently of TRI's experience of the environment in the traditional phenomenological understanding of gestalt (i.e., how the world "shows up" to her as a whole), this fragment thus highlights how the "here and now" of the situation is accomplished as a world in common partly through TRI's methods of guiding (and, reflexively, through SOP's and FRA's mutually coordinated conduct). In and through TRI's witnessable orientation to the environment, the participants' surroundings are produced as a field of possible action for embodied subjects with particular physical capacities. Akin to Goodwin's (1994) canonical archaeology example, TRI emerges as a guide or an expert insofar as she organizes another participant's seeing and acting within what is thereby constituted as the environment—she does not merely "describe" an already pre-assembled setting (Zimmerman & Pollner, 2013). Rather than being an arbitrary selection of features among the indefinite ways in which these surroundings could be discretized and labeled from an exogenous perspective, her segmentation of space is relevant for and as the practical accomplishment of the crossing.

Using the interaction with MET (Excerpt 1) as a contrastive foil brings into relief a property of TRI's contributions: their embeddedness within multiple local orders. For instance, while participants' talk had so far been designed as responsive to the normative organization of conversation (i.e., to other participants' talk), TRI's successive reports of the cars' position (l. 20, l. 22, and their update in l. 27) introduce into the interaction another temporal and normative order, that of traffic. Significantly, as part of this account, TRI *describes the car behind the group as "waiting"* (rather than as merely "immobile", for instance), thereby indexing an expectancy for the horse riders to cross without delay: the "car" is thematized as an entity embedded within a local organization. TRI's successive reports thus produce specific norms and temporal constraints as relevant to the activity of coordinating a crossing. It is in relation to these (emergent) temporal and normative features of the situation that the car stopped behind the horse riders can be recognized as "waiting". In uttering the factual declarative (Rossi, 2018) "they are waiting", TRI does more than reporting properties of the surrounding environnement, she formulates features of the setting in their relation to a moral order—that is, she indexes (and, in doing so, enacts) the local organization as part of which her turn-final explicit call for action ("come") emerges as warranted.

Another facet of this embeddedness of TRI's contributions in multiple local orders is their adjustment to *selected shifts* in the observable environment (which emerge as shifts for the blind participants through these contributions from TRI). For example, the car "behind" the riders (l. 20) has its position updated. TRI's report that this car is now "waiting" behind the riders (l. 27) is designed as responsive to changes in the position and speed of this very car. The other car—described as "in front" (l. 22)—can be seen and heard passing in front of the horse riders (l. 25, image 4.1); yet, these changes in its position are not described by TRI. Although this goes beyond what can be analytically demonstrated from these data, it may display TRI's orientation to avoiding the formulation of shifting

[4] A two-way road is not a "fronted object" (Abkarian, 1982), that is, an object whose own features define an intrinsic front–back axis, such as chairs or vehicles. TRI's structuring of the environment may therefore not index a "fronted" (intended-) design of the road itself. Yet, from a praxeological perspective, even the parts of a "fronted object" derive their functional meaning from the situated activities within which they are mobilized and which they shape. As documented in Smith's (2021) ethnomethodological analysis of mountain-rescue stretcher work, the environment is not a neutral container; rather, it is ordered as a meaningful whole through participants' embodied and categorial work. That is, at the risk of stating the obvious, as spatial categories, "front" and "behind" are, by definition, activity-relative and locally occasioned.

features of the environment that are already perceptually available to SOP. This possibility is supported by several ethnographic observations extending beyond this fragment: as an experienced guide who developed routinized practices with SOP, TRI regularly orients to the availability of the "audible order" (Relieu, 2023a) of the traffic to SOP.

In sum, TRI's talk is designed for a vehicular unit embodied in space and time. Yet this space-time is neither geometrical space nor "clock" time: that is, neither the "abstract space, devoid of landmarks or any privileged centre" of the map (Bourdieu, 1977), as opposed to the "practical space of journeys actually made" (Bourdieu, 1977), nor impersonal mathematical time, as opposed to time as kairos—that is, the "praxeological time" of opportunity (Mondada, 2018; see Au-Yeung & Fitzgerald, 2023, for a discussion). Rather, TRI's accounts ongoingly index and construct (for the entire group of horse riders) a space-time structured in relation to several normative orders and relevancies.

6. Concluding remarks: The accomplished "here-and-now" of crossing a road

6.1. Constituting an urban setting as "crossable"

The analysis of the two previous excerpts sheds light on the methods through which, and the local orders as part of which, verbal accounts of a sighted co-participant can emerge as resources in crossing a street. Notably, our second excerpt allows us to foreground how the horse-riding guide contributed to accomplishing the urban surroundings as a structured and actionable whole rather than as a mere juxtaposition of independent features.

A first way of making sense of this accomplishment is through the respecification of the notion of "scene description" in human remote sighted assistance and AI assistance (Gonzalez Penuela et al., 2024). *TRI does not merely describe a pre-given environment:* she constructs it in relation to, and as a manifestation of, a shared course of action. As Schegloff (1972) notes in his classic discussion of formulations, any formulation of location both orients to and displays a particular (spatial, temporal, pragmatic, membership-categorial, etc.) here-and-now. Congruently, TRI's use of the terms "front" and "behind" to account for the cars' location participates in producing the situation—among countless possible ways of framing it—as one involving a vehicular unit, facing a particular direction, as part of a projected road crossing, while exhibiting a front/back asymmetry (Todes, 2001), etc., for whom the relative position of cars on the road becomes relevant only in certain respects. For instance, the precise number of centimeters separating the cars from the group of riders is not treated as practically relevant, and so on. In sum, it is, in part, because TRI's referential practices are produced so as to be recognizable in relation to the group of horse riders' ongoing course of action (here, most notably, crossing the road) that these practices emerge as unproblematic members of the contexture they contribute to constituting.

In comparison, the Meta Glasses did not utter any action-bound deictics in their description of the traffic: their description of the street was "detached" in the sense of (being designed as) removed from any form of practical urgency. Because of its timing, length, grammatical construction, and absence of ongoing adjustment to the evolving traffic, the Meta Glasses' contribution can be characterized as recipient-designed for a distant observer treating the traffic as an object of theoretical study (Dreyfus, 1990)—rather than for a prospective crosser for whom the rapidly shifting state of the traffic is directly relevant to the organization of the task at hand. By being timed and designed independently from the

practical urgency of the ongoing activity (and notably, *because its description of the road as "clear" is not modified as a car is passing by in front of participants*), the agent's contribution was configured, in James's (1878) words, as "a mirror floating with no foothold anywhere, and passively reflecting an order that he comes upon and finds simply existing". Responsiveness to human talk alone was not sufficient for the artificial agent to provide "real-time assistance", as it failed to orient to the "structures providing for the organization of the endogenous activity systems within which strips of talk are embedded" (Goodwin, 1996).

6.2. Assembling contexture in guided navigation

This study extends an ethnomethodological line of inquiry that moves the study of gestalt contexture beyond relatively stable and symmetrically available perceptual objects toward dynamic settings marked by practical urgency and asymmetric perceptual access. We have explored how a human or artificial agent's structuring of the environment contributes constituent details to the contexture indexed by a blind participant's conduct. The comparison between these two excerpts reminds us of the long-documented contextualist dictum (Andler, 2000) that the accuracy of a description is not merely a matter of occasion-independent correspondence between words and world. A human guide's grammatically declarative statements do not merely report a pre-existing setting to a blind participant; they *make* features of the surroundings into constituent details of a shared here-and-now. The contrast with an "assistive" multimodal agent's contributions reveals the practical work produced by a human guide through which temporal, spatial, and normative features are brought together as a gestalt contexture. Although it would be entirely accurate to characterize this guide as merely "inform[ing] the other [i.e., the blind participant] of the event's relevant implications for their mobility" (Haddington, 2013), we have attempted here to *foreground the complexity of the taken-for-granted system of relations as part of which an informing becomes "relevant"* (and, indeed, an "informing") for all practical purposes. Rather than treating gestalt coherence as a necessary pre-given visual whole available to an individual perceiver, we have shown how the relevant "whole" of a traffic environment is progressively produced in and as the linguistic practices of this guide. Because practices glossed as "scene descriptions" are always part of the setting which they contribute to describing (Sacks, 1963; Wooffitt, 1992), they do not merely occur "at the right time" and "with the right degree of detail": they must be understood to participate in accomplishing "what is going on", that is, to index and enact features of a situation as members of an organized whole.

On a more fundamental level, this research then argues that the dualistic distinction between "foregrounded" action and a passive "background" environment, often implicit in social research, is empirically difficult to sustain. Rather than treating environments, technologies, bodies, materials, and the rest as contextual conditions for action, our analyses have shown how they participate in the ongoing accomplishment of situations. From a post-praxeological perspective ([Author B, redacted for review]), this also points to a limitation in contemporary AI systems. By segmenting the world into computational representations and subsequently into linguistic units, they presuppose a separation between entities and their relations that sits uneasily with a one-world process ontology (Whitehead, 1985), where agents, actions, environments, and methods of knowing and the rest continuously emerge together through the unfolding of practice.

Acknowledgements

[Redacted for review]

References

Abkarian, G. G. (1982). Comprehension of deictic locatives: The object "Behind" it. *Journal of Psycholinguistic Research*, *11*(3), 229–245. https://doi.org/10.1007/BF01067566

Andler. (2000). Context and Background. Dreyfus and Cognitive Science. In W. Wrathall (Ed.), *Heidegger, Coping and Cognitive Science, Cambridge*. The MIT Press.

Atkinson, R. (2007). Ecology of Sound: The Sonic Order of Urban Space. *Urban Studies*, *44*(10), 1905–1917. https://doi.org/10.1080/00420980701471901

Austin, J. L. (1975). *How to do things with words*. Harvard university press.

Au-Yeung, T. S. H., & Fitzgerald, R. (2023). Time structures in ethnomethodological and conversation analysis studies of practical activity. *The Sociological Review*, *71*(1), 221–242. https://doi.org/10.1177/00380261221103018

Ayaß, R. (2020). Doing Waiting: An Ethnomethodological Analysis. *Journal of Contemporary Ethnography*, *49*(4), 419–455. https://doi.org/10.1177/0891241619897413

Bourdieu, P. (1977). *Outline of a Theory of Practice* (R. Nice, Trans.; 1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511812507

Carlin, A., Varanda Marques, J. B., & Watson, R. (2025). Encountered astronomy: Aesthetics and authenticity in the public communication of science. *Ethnographic Studies*, *20*. https://doi.org/10.26034/lu.ethns.2024.6912

Chang, R.-C., Liu, Y., & Guo, A. (2024). WorldScribe: Towards Context-Aware Live Visual Descriptions. *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, 1–18. https://doi.org/10.1145/3654777.3676375

Coulter, J., & Parsons, E. D. (1990). The praxiology of perception: Visual orientations and practical action. *Inquiry*, *33*(3), 251–272. https://doi.org/10.1080/00201749008602223

Cromdal, J., Broth, M., & Willemsen, A. (2026). Corrections on the kerb: How preschool groups prepare for crossing the street. *Linguistics and Education*, *93*, 101509. https://doi.org/10.1016/j.linged.2026.101509

Deppermann, A. (2019). Intersubjectivity and other grounds for action-coordination in an environment of restricted interaction: Coordinating with oncoming traffic when passing an obstacle. *Language & Communication*, *65*, 22–40. https://doi.org/10.1016/j.langcom.2018.04.005

Dreyfus, H. L. (1990). *Being-in-the-World: A Commentary on Heidegger's Being in Time, Division I* (Vol. 102, Issue 2, pp. 290–293). Bradford.

Dreyfus, H. L. (2007). Why Heideggerian Ai Failed and How Fixing It Would Require Making It More Heideggerian. *Philosophical Psychology*, *20*(2), 247–268. https://doi.org/10.1080/09515080701239510

Dreyfus, H. L. (2014). *Skillful coping: Essays on the phenomenology of everyday perception and action*. OUP Oxford.

Eisenmann, C., & Rawls, A. W. (2023). The Continuity of Garfinkel's Approach: Seeking Ways of "Making the Phenomenon Available Again" Through the Experience and Usefulness of "Trouble." In P. Sormani & D. vom Lehn (Eds.), *The Anthem Companion to Harold Garfinkel*

(pp. 19–42). Anthem Press. Cambridge Core. https://www.cambridge.org/core/product/A67C7283F3130116E23A88DC2F6D6CC2

Garfinkel, H. (1965). Sign functions: Organized activities as methods for making an invisible world observable. *Paper Delivered to the Social Psychology Section at the Annual Meetings of the American Sociological Associations, Chicago, IL (30 August)*.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Polity Press.

Garfinkel, H. (1996). Ethnomethodology's Program. *Social Psychology Quarterly*, *59*(1), 5. https://doi.org/10.2307/2787116

Garfinkel, H. (2006). *Seeing sociologically: The routine grounds of social action*. Paradigm.

Garfinkel, H. (2021). Ethnomethodological Misreading of Aron Gurwitsch on the Phenomenal Field. *Human Studies*, *44*(1), 19–42. https://doi.org/10.1007/s10746-020-09566-z

Garfinkel, H., & Livingston, E. (2003). Phenomenal field properties of order in formatted queues and their neglected standing in the current situation of inquiry. *Visual Studies*, *18*(1), 21–28. https://doi.org/10.1080/14725860320001002 9

Goffman, E. (1971). *Relations in public: Microstudies of the public order*. Basic Books.

Gonzalez Penuela, R. E., Collins, J., Bennett, C., & Azenkot, S. (2024). Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24*. https://doi.org/10.1145/3613904.3642211

Gonzalez Penuela, R. E., Hu, R., Lin, S., Shende, T., & Azenkot, S. (2025). Towards Understanding the Use of MLLM-Enabled Applications for Visual Interpretation by Blind and Low Vision People. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '25*. https://doi.org/10.1145/3706599.3719714

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

Goodwin, C. (1996). Transparent vision. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 370–404). Cambridge University Press. Cambridge Core. https://doi.org/10.1017/CBO9780511620874.008

Goodwin, C. (2018). Why Multimodality? Why Co-Operative Action? (transcribed by J. Philipsen). *Social Interaction. Video-Based Studies of Human Sociality*, *1*(2). https://doi.org/10.7146/si.v1i2.110039

Gurwitsch, A. (1964). *The Field of Consciousness*. Duquesne University Press.

Gurwitsch, A. (2010). The Phenomenological and the Psychological Approach to Consciousness. In A. Gurwitsch, *The Collected Works of Aron Gurwitsch (1901–1973)* (Vol. 193, pp. 99–118). Springer Netherlands. https://doi.org/10.1007/978-90-481-2942-3_4

Haddington, P. (2013). Projecting mobility: Passengers directing drivers at junctions. In P. Haddington, L. Mondada, & M. Nevile (Eds.), *Language and the Body in Motion* (pp. 179–209). De Gruyter. https://doi.org/doi:10.1515/9783110291278.179

Haddington, P., Keisanen, T., Mondada, L., & Nevile, M. (2014). Towards multiactivity as a social and interactional phenomenon. In P. Haddington, T. Keisanen, L. Mondada, & M. Nevile (Eds.), *Multiactivity in Social Interaction* (pp. 3–32). John Benjamins Publishing Company. https://doi.org/10.1075/z.187.01had

Haddington, P., Mondada, L., & Nevile, M. (Eds.). (2013). *Interaction and Mobility: Language and the Body in Motion*. De Gruyter, Inc.

Hanson, N. R. (1979). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cup archive.

Hester, S., & Francis, D. (2003). Analysing visually available mundane order: A walk to the supermarket. *Visual Studies*, *18*(1), 36–46. https://doi.org/10.1080/14725860320001000056

Hutchinson, P., & Chikwira, L. (2026). Praxeological analysis (PA/CPA) for stigma, health inequalities, and coercion in women's services. *Frontiers in Psychiatry*, *17*, 1766029. https://doi.org/10.3389/fpsyt.2026.1766029

James, W. (1878). Remarks on Spencer's Definition of Mind as Correspondence. *The Journal of Speculative Philosophy*, *12*(1), 1–18. JSTOR.

Jayyusi, L. (1988). Towards a socio-logic of the film text. *Semiotica*, *68*(3–4), 271–296. https://doi.org/10.1515/semi.1988.68.3-4.271

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In *Gene H. Lerner (Ed.) Conversation Analysis: Studies from the first generation* (pp. 13–31). John Benjamins Publishing Co.

Kanizsa, G. (1979). *Organization in Vision: Essays on Gestalt Perception*. Praeger.

Koffka, K. (2013). *Principles of Gestalt psychology* (First issued in paperback). Routledge. (Original work published 1935, Harcourt, Brace)

Laurier, E., Muñoz, D., Miller, R., & Brown, B. (2020). A Bip, a Beeeep, and a Beep Beep: How Horns Are Sounded in Chennai Traffic. *Research on Language and Social Interaction*, *53*(3), 341–356. https://doi.org/10.1080/08351813.2020.1785775

Liberman, K. (2013). The local orderliness of crossing Kincaid. *More Studies in Ethnomethodology*, 11–30.

Lloyd, M. (2020). Getting by: The ethnomethods of everyday cycling navigation. *New Zealand Geographer*, *76*(3), 207–220. https://doi.org/10.1111/nzg.12274

Lynch, M., & Eisenmann, C. (2022). Transposing Gestalt Phenomena from Visual Fields to Practical and Interactional Work: Garfinkel's and Sacks' Social Praxeology. *Philosophia Scientiæ*, *26–3*(3), 95–122. Cairn.info. https://doi.org/10.4000/philosophiascien

Lynch, M., & Garfinkel, H. (2022). *Harold Garfinkel: Studies of Work in the Sciences* (1st ed.). Routledge. https://doi.org/10.4324/9781003172611

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463--502). Edinburgh University Press.

McIlvenny, P. (2019). How did the mobility scooter cross the road? Coordinating with co-movers and other movers in traffic. *Language & Communication*, *65*, 105–130. https://doi.org/https://doi.org/10.1016/j.langcom.2018.04.002

McIlvenny, P., Broth, M., & Haddington, P. (2014). Moving Together: Mobile Formations in Interaction. *Space and Culture*, *17*(2), 104–106. https://doi.org/10.1177/1206331213508679

Merlino, S., & Mondada, L. (2019). Crossing the street: How pedestrians interact with cars. *Language & Communication*, *65*, 131–147. https://doi.org/https://doi.org/10.1016/j.langcom.2018.04.004

Meyer, C. (2022). The Phenomenological Foundations of Ethnomethodology's Conceptions of Sequentiality and Indexicality. Harold Garfinkel's References to Aron Gurwitsch's "Field of Consciousness." *Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion*, *23*, 111–144.

Meyer, C. (2024). Context-sensitivity and context-productivity: Notions of "practice" and "practicality" in ethnomethodology and conversation analysis. *Frontiers in Sociology*, *9*, 1221026. https://doi.org/10.3389/fsoc.2024.1221026

Mlynář, J. (2026). Structured objects: The organized relationships of parts and wholes. *Przegląd Socjologiczny*, *75*(1), 101–120. https://doi.org/10.26485/PS/2026/75.1/6

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., & Tekin, B. S. (2024). The intelligibility of mobile trajectories: Walking in public space. *Mobilities*, *19*(6), 1076–1098. https://doi.org/10.1080/17450101.2024.2337256

Moutinho, R., Carlin, A. P., & Marques, J. B. (2025). Visually informed accounts: Instructed achievements during planetarium visits and sky observations. *Visual Communication*, *24*(2), 447–468. https://doi.org/10.1177/14703572221130441

Osvaldsson, K. (2004). On laughter and disagreement in multiparty assessment talk. *Text & Talk*, *24*(4), 517–545. https://doi.org/doi:10.1515/text.2004.24.4.517

Popova, K. (Author). (2018). *Ethnomethodological Studies of Visuality*. https://doi.org/10.5281/ZENODO.1475767

Relieu, M. (2023a). Teaching and learning how to identify an audible order in traffic. In S. Keel, *Medical and Healthcare Interactions* (1st ed., pp. 152–175). Routledge. https://doi.org/10.4324/9781003312345-8

Relieu, M. (2023b). The production and reception of assistance proposals between pedestrians and visually impaired persons during a course in orientation and mobility. In B. L. Due, *The Practical Accomplishment of Everyday Activities Without Sight* (1st ed., pp. 26–48). Routledge. https://doi.org/10.4324/9781003156819-2

Rickard, C., Strother, M., Fox, B. A., & Raymond, C. W. (2019). Scaffolding Embodied Access for Categorization in Interactions between a Blind Child and Her Mother. *Languages*, *4*(1), 2. https://doi.org/10.3390/languages4010002

Rönnqvist, S., & Lindström, J. (2021). Turn Continuations and Gesture: "And Then"-Prefacing in Multi-Party Conversations. *Frontiers in Communication*, *6*, 670173. https://doi.org/10.3389/fcomm.2021.670173

Rossi, G. (2018). Composite Social Actions: The Case of Factual Declaratives in Everyday Interaction. *Research on Language and Social Interaction*, *51*(4), 379–397. https://doi.org/10.1080/08351813.2018.1524562

Sacks, H. (1963). Sociological description. *Berkeley Journal of Sociology*, *8*, 1–16.

Sacks, H. (1995). *Lectures on conversation: Volumes I & II* (G. Jefferson, Ed.; 1. publ. in one paperback volume). Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696. https://doi.org/10.2307/412243

Schegloff, E. A. (1972). Notes on a conversational practice: Formulating place. In D. Sudnow (Ed.), *Studies in Social Interaction* (pp. 75–119).

Schegloff, E. A. (1992). On talk and its institutional occasions. *Talk at Work: Interaction in Institutional Settings*, *101*, 101–134.

Schegloff, E. A. (1996). Confirming Allusions: Toward an Empirical Account of Action. *American Journal of Sociology*, *102*(1), 161–216. https://doi.org/10.1086/230911

Schegloff, E. A. (2005). On Complainability. *Social Problems*, *52*(4), 449–476. https://doi.org/10.1525/sp.2005.52.4.449

Schutz, A. (1944). On multiple realities. *Philosophy and Phenomenological Research*, *5*(4), 533–576.

Schutz, A. (1970). *Reflections on the Problem of Relevance*.

Schutz, A. (1996). Outline of a Theory of Relevance. In *Collected Papers: Volume IV* (pp. 3–5). Springer.

Smith, R. J. (2017). Membership categorisation, category-relevant spaces, and perception-in-action: The case of disputes between cyclists and drivers. *Journal of Pragmatics*, *118*, 120–133. https://doi.org/10.1016/j.pragma.2017.05.007

Smith, R. J. (2021). Categorisation practices, instructed actions, and teamwork as occasioned phenomena: Structuring the "carry off" in mountain rescue work. *Gesprächsforschung - Online-Zeitschrift Zur Verbalen Interaktion*, *2021*(22), 648–669.

Smith, R. J. (2024). Fieldwork, participation, and unique-adequacy-in-action. *Qualitative Research*, *24*(1), 60–80. https://doi.org/10.1177/14687941221132955

Stangl, A., Verma, N., Fleischmann, K. R., Morris, M. R., & Gurari, D. (2021). Going Beyond One-Size-Fits-All Image Descriptions to Satisfy the Information Wants of People Who are Blind or Have Low Vision. *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '21*. https://doi.org/10.1145/3441852.3471233

Stivers, T., Mondada, L., & Steensig, J. (2011). Knowledge, morality and affiliation in social interaction. In J. Steensig, L. Mondada, & T. Stivers (Eds.), *The Morality of Knowledge in Conversation* (pp. 3–24). Cambridge University Press. Cambridge Core. https://doi.org/10.1017/CBO9780511921674.002

Sudnow, D. (1972). Temporal parameters of interpersonal observation. In *Studies in Social Interaction*. Free Press.

Todes, S. (2001). *Body and world* (H. L. Dreyfus & P. Hoffman, Eds.). The MIT Press.

Tuncer, S., Licoppe, C., Luff, P., & Heath, C. (2024). Recipient design in human–robot interaction: The emergent assessment of a robot's competence. *AI & SOCIETY*, *39*(4), 1795–1810. https://doi.org/10.1007/s00146-022-01608-7

Waismann, F. (1951). Verifiability. In G. Ryle & A. Flew (Eds.), *Logic And Language* (pp. 35--68). Blackwell.

Whitehead, A. N. (1985). *Process and reality: An essay in cosmology* (D. R. Griffin, Ed.; Corr. ed). Free Press.

Wiener, J. M., Büchner, S. J., & Hölscher, C. (2009). Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach. *Spatial Cognition & Computation*, *9*(2), 152–165. https://doi.org/10.1080/13875860902906496

Wittgenstein, L. (1968). *Philosophical Investigations*. Blackwell.

Wooffitt, R. (1992). *Telling tales of the unexpected: The organization of factual discourse.* (pp. xii, 217). Harvester Wheatsheaf.

Xie, J., Yu, R., Lee, S., Lyu, Y., Billah, S. M., & Carroll, J. M. (2022). Helping Helpers: Supporting Volunteers in Remote Sighted Assistance with Augmented Reality Maps. *Proceedings of the 2022 ACM Designing Interactive Systems Conference, DIS '22*, 881–897. https://doi.org/10.1145/3532106.3533560

Zimmerman, D. H., & Pollner, M. (2013). The Everyday World as a Phenomenon'. *People and Information: Pergamon General Psychology Series*, *6*, 33.