Check for updates

Gaze alternation predicts inclusive next-speaker selection: evidence from eyetracking

Christoph Rühlemann*

Deutsches Seminar - Germanistische Linguistik, Albert-Ludwigs-University, Freiburg, Germany

Next-speaker selection refers to the practices conversationalists rely on to designate who should speak next. Speakers have various methods available to them to select a next speaker. Certain actions, however, systematically co-select more than one particular participant to respond. These actions include asking "open-floor" questions, which are addressed to more than one recipient and that more than one recipient are eligible to answer. Here, next-speaker selection is inclusive. How are these questions multimodally designed? How does their multimodal design differ from the design of "closed-floor" questions, in which just one participant is selected as next speaker and where next-speaker selection is exclusive? Based on eyetracking data collected in naturalistic conversation, this study demonstrates that unlike closed-floor questions, open-floor questions can be predicted based on the speaker's *gaze alternation* during the question. The discussion highlights cases of gaze alternation in open-floor questions and exhaustively explores deviant cases in closed-floor questions. It also addresses the functional relation of gaze alternation and gaze selection, arguing that the two selection techniques may collide, creating disorderly turntaking due to a fundamental change in participation framework from focally dyadic to inclusive. Data are in British and American English.

KEYWORDS

next-speaker selection, conversation, gaze alternation, inclusion—exclusion, eyetracking, participation framework

1 Introduction

Gaze is key to talk-in-interaction. We cannot not gaze, and gazing away from co-participants or "merely" blinking has communicative value (Hömke et al., 2017). One function gaze is widely assumed to perform during the final part of the current speaker's turn is to select the next speaker (Sacks et al., 1974; Kalma, 1992; Lerner, 2003; Tiitinen and Ruusuvuori, 2012; Auer, 2021a,b; Auer et al., in preparation).

Consider extract (1).[1] Before the excerpt, participant A has used the term "steam punk." The excerpt begins when Participant B says *wait* to momentarily pause the course of action

---

1   In all extracts, gaze and, sometimes, gesture are indicated below the simultaneous verbal transcript; the letters A, B, and C indicate the looked-at participant, whereas gaze shifts from one participant to the other, or other forms of gaze aversion are marked by *; this annotation is implemented only where relevant to the argument. In some excerpts, screenshots from both the room camera video and the participants' eyetracking videos are implemented as figures (with lower-case *f*).

and then initiates a repair sequence by asking *what did you say? you said steam!punk!?* = (line 1) to ascertain she heard the expression correctly. Participant A's gaze is fixating on the answerer, participant B, throughout the question. Similarly, when B asks the follow-up question *what does steam punk mean [°exactly°]?* (line 11), B's gaze rests on A throughout the question:

```
(1) [F07]
1     B:  →   =wait what did you say? you said steam !punk!?=
      B-gaze  ------------------------A--------------------|
2             (0.085)
3     A:      =ye:ah=
4             (0.050)
5     B:      =paintings?=
6             (0.027)
7     A:      =steam punk yeah=
8             (0.175)
9     B:      =<okay>
10            (0.018)
11    B:  →   what does steam punk mean [ °exactly° ]?
      B-gaze  -------------------A------------------|
12    A:                                [steam punk ] is:: a
13            (.) >a sort of< Art !sty!le but it's also (.)
14            it has its own scene like a medieval scene,=
```

To claim, however, that it is B's *gaze* that selects A as the next speaker is not without problems. For A is already (pre-)selected exclusively as the next speaker on sequential grounds. B's first question *what did you say? you said steam!punk!?* = (line 1) serves a repair function related to a co-participant's prior utterance. Given the preference organization of repair sequences, whereby the one who produces the trouble source has first rights to repair that trouble source (see Schegloff et al., 1977; Schegloff 1992a), the question is preferentially addressed to that co-participant who introduced the repairable—that is, to A. Similarly, B's second question *what does steam punk mean [°exactly°]?* (line 11) is a *follow-up* question building on the back of the prior sequence and extending it. Participant A is therefore, again, selected sequentially. In both question cases, then, it can be argued that the exclusive next-speaker selection is effected by the questions' sequential position the questioner's gaze is redundant. In the following, I will refer to this type of question, which picks out one particular addressee to answer and give exclusive rights to respond to that specific recipient, as *closed-floor questions* and the type of next-speaker selection they facilitate as *exclusive*.

Contrast extract (1) with extract (2). In extract (2), participants B and C are brothers—hence, a party, i.e., a unit of social organization "which can be claimed to have a persistence and reality quite apart from the interaction" (Schegloff, 1995, p. 33). Before the extract, participant A has just introduced himself saying his first name. Participant C informs him that *we have a brother* of the same name (line 1). The informing is confirmed by B's = °*yeah*° = (line 3). Based on C's use of the pronoun *we* and B's confirmation, A infers that B and C must be brothers and asks accordingly = *what you guys brothers¿* = (line 5). The question is multiply answered by the brothers almost concurrently.

```
(2) [F16]
1     C:      we have a brother called ((anonymized name)) too=
2             (0.200)
3     B:      =°yeah°=
4             (0.060)
5     A:  →   =what you guys brothers¿=
      A-gaze  -*-|---B--|-*-|----C----|
6             (0.017)
7     C:      =mhm=
8             (0.016)
9     B:      =yeah=
```

As shown in the gaze annotation in extract (2), speaker A's question-final gaze is on C, that participant who is the first to answer. In that sense, C may be said to be selected as the next speaker. However,

the question = *what you guys brothers¿* = is fundamentally different from the questions in excerpt (1). In excerpt (1), given their sequential dependence on prior talk, the two questions were addressed exclusively to one participant. The question in excerpt (2), conversely, is addressed collectively to both participants: both are epistemically eligible to answer it, both are licensed to answer it—and, indeed, they do (cf. lines 7–9). The focal participation framework is not dyadic, as in extract (1), but triadic, including all co-participants. Here, then, the questioner is "ask[ing] a question without 'asking *somebody* a question'" (Schegloff, 1992b, p. 122; emphasis in original). In the following, I will term this type of question, which does not *per se* address one particular recipient but opens up the floor to more than one participant, *open-floor questions* and will refer to the kind of next-speaker selection as *inclusive*.

The argument this study seeks to make is that the multimodal design of these two types of questions differs in that the questioner's gaze fixates only on one addressee in exclusive, closed-floor questions, whereas it characteristically alternates between the co-selected addressees in inclusive, open-floor questions. To illustrate, re-consider extract (2): upon asking = *what you guys brothers¿* = (line 5), speaker A first looks at B and then shifts to C. I will demonstrate that gaze alternation is characteristic of inclusive open-floor questions in the sense that it statistically predicts such questions. Its occurrence in exclusive, closed-floor questions, by contrast, is merely occasional, and due to other reasons than indexing that the floor is open to both or either question recipient.

In the next section, Section 2, I will elaborate on the background to this research. Section 3 will outline the data and methods used. Section 4 will present the results of the statistical analysis. In Section 5, I will discuss typical and untypical cases and comment on the functional relationship between gaze alternation and gaze selection before Section 6 concludes this study.

2 Background

In this background section, I will address the issue of next-speaker selection and outline the role of gaze in turn-taking as well as previous research on gaze alternation.

2.1 Next-speaker selection

Speakers in conversation take turns. In dyadic conversation, the order in which they take turns is fixed: first you speak, then I speak, then you speak, etc. In conversation with more than two speakers, turn order varies: first you speak, then I (may or may not) speak, then she may (or may not) speak, then you (may or may not) speak, etc. How do participants to multi-party interaction *decide* whose turn it is? If turns were allocated at random, speakers would mostly talk over each other or not talk at all. However, this is not what we observe; instead, one speaker speaks at a time, and phases of overlap are brief (Sacks et al., 1974). Underlying this orderliness is a set of three rules governing turn allocation: the current speaker selects the next speaker to respond, or a current non-speaker self-selects, or the current speaker self-selects when no other participant has self-selected (Sacks et al., 1974). The three rules are ordered: the first rule, current-selects-next, applies first, and has priority over the two other rules (Sacks et al., 1974, p. 709).

How do current speakers select next speakers? Next-speaker selection is a "gradual dimension" (Auer, 2021a, p. 125) in the sense that turns/actions differ by the extent to which they make a response turn/action and, hence, next-speaker selection relevant. Auer (2021a) proposes a continuum ranging from "maximally projecting first actions such as adjacency pairs, to actions that do not project any next action, such as the ones that are typically found at sequence closure" (Auer, 2021a, p. 126; cf. also Stivers and Rossano, 2010). Moreover, to answer the above question how speakers select next speakers, it is instructive to examine turns that exert maximal projection strength. Most first-pair parts of adjacency pairs fall into this category. For example, an information-seeking question has maximal projection strength on prospective next speakers, making an answer sequentially relevant (Stivers and Rossano, 2010; Auer, 2021a).

However, as noted, questions can be used in ways that do not select "somebody" as next speaker. Cases in point are collective and distributive questions (cf. Link, 1991), while collective questions address more than one recipient as a collectivity and are answerable by one member of the collectivity for the collectivity as a whole, distributive questions address each participant as an individual and are answerable serially—first by one, then, potentially, by the other participant(s).

For illustration, consider extract (3). The three participants have only just met in the recording situation. They are thus three unrelated individuals, who have neither social ties beyond the recording situation nor yet formed a momentary party structure. Therefore, speaker C's pronoun form *you guys* refers to participants A and B as individuals. The question in extract (1) is, hence, distributive; given its answerability by either recipient, it is not surprising that both recipients produce an answer (in mutual overlap):[2]

(3) [F18]

```
1    C:    have you guys seen the movie American Psycho?
2          (0.244)
3    A:    [mhm,]
4    B:    [mhm,]
```

An example showing that distributive and collective questions need not be mutually exclusive *in situ*, is extract (4). Here, speaker A inquires about the work of speakers B and C. As the recording session is speaker A's first meeting with speakers B and C, she cannot know that B and C in fact share the same work place (the Fraunhofer Institute). Observed from A's perspective, her question is a distributive one, answerable first by one, and then by the other recipient. Once B and C, both using the pronoun *we*, reveal to her that they work at the same institute, C's answer is sufficient for both of them:

(4) [F08]

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

Questions, then, in and of themselves do not necessarily pick out one particular addressee to answer—but they often can and, in fact, do single out and give exclusive rights to a specific recipient to respond. Next, speakers can be selected based on context- or sequence-specific particulars. A question that incorporates context-specific particulars can *tacitly* select a specific participant to respond "[w]hen the requirements for responding to a sequence-initiating action limit eligible responders to a single participant" (Lerner, 2003, p. 190). These "requirements" include all conceivable kinds, of which epistemics (what Lerner refers to as "shared knowledge and experience") is arguably the most important:[3] a question may presuppose particular knowledge and/or experience and thus select the person "in the know." For illustration, consider (5):

(5) [F07]

```
1    B:    remember that (0.284) gi:rl tha- she was Swedish
2          who played at the Irish pub¿ (0.460) who sang °at
3          the Irish° pub?=
4          (0.075)
5    A:    =!oh! [(Amanda)]
```

Here, participants B and A actually went to the pub, but participant C did not, so C cannot possibly remember the girl there. The only participant epistemically eligible and thus entitled to answer the question is participant A.

Tacit addressing based on context-specific particulars as in extract (5) seems rather rare (at least in the present data). More commonly, next-speaker selection relies on sequence-specific particulars to narrow the field of possible answerers to just one. A common practice is, for example, extending a sequence occasioned by previous turns (cf. Hayano, 2013, p. 172). Extract (1) in the Introduction illustrated two such scenarios; both questions in the extract considered follow-up questions facilitating exclusive next-speaker selection.

It is worth noting that not all follow-up questions need to select just one next speaker. Questions can be clearly sequentially dependent but still invite responses from more than one participant.

Extract (6) is a case in point. The participants are two males from the US, speakers B and C, and one female from Britain, speaker A. Immediately prior to the extract, the participants have been talking about music and what instruments they play. It has turned out that the only one not to have learned an instrument is participant C. Instead, he notes, he took *Music Appreciation* (line 1) in school, which prompts laughter from participant B (line 3). Speaker C goes on to say that *it was act[ually a really fun class]* (line 5). Speaker A incredulously interrupts him by asking *[!what! £there's a!thing! called] Music App[reciation?]£*. The question is not directly answered; rather speaker B assesses the class as *a joke > of a cla:ss< (.) in the St[ates]* (lines 8–9). He thereby reveals that he, too, apparently has some knowledge of Music Appreciation classes. Participant A, still in pursuit of an answer to her previously unanswered question, reformulates the question by enquiring *[↑what is↑ Music] Appreci- th[at's]* (line 11). That question

is responded to by both recipients, largely in overlap, with B claiming to have only hearsay knowledge (>[*I do not] know* < *(.) but I've heard of it,* line 11) and C providing the sought information [*it it just*] *like gives you a survey of music and like the different styles of it* (lines 13–14). Here, then, the fact that both B and C answer the question (largely simultaneously) is an indication that they both treat the question as addressed to them based on the fact that each of them is epistemically eligible to answer the question.

```
(6) [F08]
1    C:     like (.) I took Music Appreciation=
2           (0.090)
3    B:     =((v: laughs)) [((v: laughs))]
4    C:                    [   it (.)   ]
5           it was act[ually   a   really   fun    class]
6    A:               [ !what! £there's a !thing! called]
7           Music App[reciation?   ]£
8    B:              [but that is a] joke >of a cla:ss< (.)
9           in the St[ates]
10   C:              [yeah] but I [=I           ]
11   A: →                         [↑what is↑ Music] Appreci-
12           th[at's    ]
13   B:        [°°yeah°°]
14   B:     >[I don't ] know< (.)
15        [but I've heard of it]
16   C:   [  it    it   just ] like gives you a survey of music and
17        like the different styles of it
```

Selection techniques based on contextual and sequential particulars have been referred to as *implicit* next-speaker selection techniques (Hayano, 2013). These implicit techniques are intertwined in intricate ways with *explicit* selection techniques. Most of them are *embodied* methods.[4]

To start with, embodied methods of next-speaker selection include forms of pointing. Consider extract (7). Participants C and A are talking about a non-present person speaker B is also vaguely familiar with. C and A, respectively, describe that person as a *!zen! nature man* (line 1) and *[he is a tree]* (line 2). B inquires about his name (line 3), which is given as *Helge* (line 5). The name occasions comment by all three participants: first, B qualifies it as a >*[perfect] name (.) great name* < = (line 10), then B connects it with = <*a Ge:rman £myth > fable [(…)]£* (line 12), and A finally extends that associative line with *he's the yeah he's the mythical forest m[an,]* (line 15). At this point, C poses the question *[did you] read the Hobbit?* (line 20). The question is a *pre*-question testing the ground for the provision of the informing *he's uh Tom Bombadil* = (a character from J.R.R. Tolkien's work) later in the sequence (line 23). The question could in principle be addressed to both A and B. However, C not only turns his head to B and fixates his gaze on him (which would already have the force of selecting him to answer; see below) but also points at him using his index finger, keeping that pointing gesture in place during the whole question [cf. upper right tile in the embedded figures 3, 4 in extract (8)]. C's attention is thus fully on recipient B, excluding participant A from the focal dyadic framework and unambiguously selecting B by multiple means as next speaker. B answers in the positive (*yeah yeah [°sure°]*, line 22), so C knows he can rely on common ground when comparing Helge with Tom Bombardil:

---

4   An explicit and *verbal* next-speaker selection method is using an address term. Sacks et al. (1974, p. 717) assume that this method is "an important, perhaps the central, general technique" for next-speaker selection. However, affiliating address terms are rare and, when they occur, they serves more functions than just next-speaker selection (Auer, 2021b).

```
(7) [F04]
1    C:     [but like] !zen! nature man [°yeah°]
2    A:     [he is a tree]
3    B:     >[ what's  his] name again?< Hel-
4           (0.298)
5    C:     <Helge>=
6           (0.025)
7    B:     =°Helge°
8           (0.299)
9    A:     Hel[  ge  ]
10   C:        >[perfect] name (.) great name<=
11          (0.095)
12   B:     =<a Ge:rman £myth> fable
13          [   (…)   ]£
14   C:     [yeah exactly yeah]
15   A:     [    yea:h    ] he's the yeah he's the mythical forest m[an],
16   B:                                                           [y ]eah=
17          (0.048)
18   C:     #=he !is! >he's like mythical f- >i- i- %d-<
```

fig


```
        figure 1                    figure 2
```

```
19   A:     [<and he>]
20   C: →   #[did you] read the Hobbit?%
```

fig


```
        figure 3                    figure 4
```

```
21          (0.232)
22   B:     yeah yeah [°sure°]
23   C:               [ he ]'s he's uh Tom Bombadil=
24          (0.093)
25   B:     =°okay [okay (.) yeah]°
```

Pointing as in extract (7), where a single participant is singled out, is part of the multimodal package of a closed-floor question and serves to exclusively select the pointed-to participant as the next speaker. If, by contrast, speakers use pointing in open-floor questions to co-select participants, they will either distributively point to different participants (the left hand pointing to the participant sitting on the questioner's left and the right hand pointing to the other participant) or will move both hands with extended index fingers from left to right in a "wind-screen wiper" motion. Pointing does occur in the present data as a next-speaker selection technique but it is far less frequent than gaze-directional addressing, undoubtedly the most ubiquitous embodied method, to which we turn now.

2.2 Gaze and gaze alternation

Unlike other forms of embodied conduct, gaze is omnipresent in face-to-face interaction. Considering its omnipresence, it is no surprise that research on gaze in social interaction has been productive [for example, Kendon (1967) on gaze as an index of social accessibility and attention, Onishi and Baillargeon (2005) on gaze and theory of mind, Carpenter and Liebal (2011) on gaze and joint attention, Brône and Oben (2021) on gaze and action ascription, Holler and Kendrick (2015) on gaze and recipiency, Goodwin and Goodwin (1986) and Auer and Zima (2021) on gaze and cognitive duress, Noris et al. (2012), Falck-Ytter et al. (2015), and Rigby et al. (2016) on gaze in autism, as well as Mondémé (2023) on gaze in human-pet interaction].

Moreover, given its omnipresence, gaze is always potentially available for next-speaker selection. And unlike, for example, gestures, which are rarely gaze-fixated (Beattie, 2016), the speaker's gaze behavior is closely monitored in environments of possible turn transition, which is why gaze is often assumed to provide a method of next-speaker selection (Sacks et al., 1974; Kalma, 1992; Lerner, 2003; Tiitinen and Ruusuvuori, 2012; Auer, 2021a,b). This assumption

builds on the observation that speakers often avert their gaze from the interlocutor(s) as a turn-holding strategy (e.g., Kendrick et al., 2023; Moore and Robinson, 2024). Indeed, recent quantitative research suggests that there is a robust correlation between the speaker's last look to a recipient and that recipient taking the next turn: in 74% of all cases, the recipient that was gazed at last by the current speaker during the question was the one who answered it (Auer, 2021b).

The strength of the correlation is arguably owed to no small part to the fact that gaze is inherently dyadic (Auer, 2021a,b): you cannot look at two or more people *at the same time*. What if speakers wish to "ask a question without 'asking *somebody* a question'"(Schegloff, 1992b, p. 122; emphasis in original), that is, if the selection pattern is inclusive rather than exclusive? A practiced solution to gaze's inherent dyadicity is *gaze alternation*: looking at two or more people *serially*, that is, one after another (Auer, 2018, 2021a; Stivers, 2021).

For illustration, consider extract (8). Participant B is asking a complex question referring to participant A's and C's shared apartment *[but] =°but° w- if you say it's a Dachgeschoss top floor is it like (0.493) slanted? and can you actually walk?*. During the *but*-clause and the beginning of the question *is it like* as well as a pause of almost half a second, he is not looking either to A or C but somewhere in-between them, as shown in the gaze transcription underneath line 1. Only as he produces *slanted* does he fixate on participant A (line 2). He expands the question by adding *and can;* during that expansion, he quickly shifts his gaze toward participant C (line 3). Upon completing the question with *you actually walk?*, his gaze fixates on participant C (line 4).

figure 1   figure 2

figure 3   figure 4

figure 5   figure 6

figure 7   figure 8
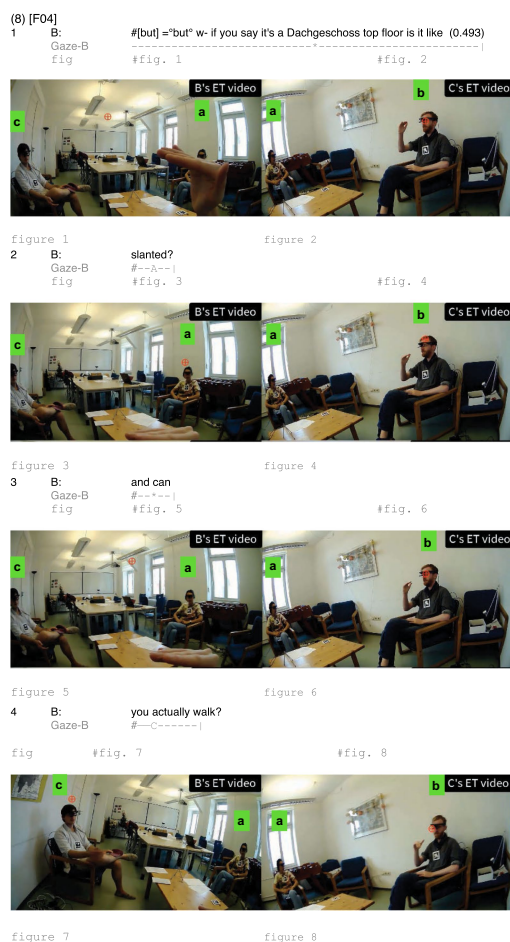
Research on *gaze alternation* in multi-party conversation is still in short supply. The phenomenon has, arguably, first been discussed in Auer (2018), who makes a distinction between gaze used for next-speaker selection and gaze for addressee involvement to prevent schism and marginalization. Gaze alternation also plays a role in reenactments, where "reenactors can mark role switches by gazing alternately at different participants of the reenacting event in order to assign different roles to them" (Pfeiffer and Weiß, 2022, p. 34). In research on storytelling interaction in triads (Rühlemann et al., 2019), gaze alternation *accelerates* as the storyteller progresses toward the climax, where the recipients' display of affiliation with the teller's stance becomes key (cf. Stivers, 2008). Moreover, Waring and Carpenter (2019) examine gaze shifts as a resource for managing attention and participation in teaching contexts.

In both storytelling and teaching contexts—but, arguably, not in reenactments—the foundational function of gaze alternation is to provide a solution to the problem posed by gaze's inherent dyadicity: "By alternatingly looking at the addressees, current speakers actively keep a three-party conversation from turning into a two-party conversation by gaze-addressing and gaze-engaging both co-participants" (Auer, 2021a, p. 124; cf. also Auer, 2018). In other words, in-turn gaze alternation serves an inclusive function, working against the dyadic focal participation framework underlying conversation (Stivers, 2021).

As noted in the Introduction, this study will demonstrate that, in questions, the presence of gaze alternation is predictive of inclusive, open-floor questions, and its absence, conversely, predicts exclusive, closed-floor questions.

3 Data and methods

3.1 FreMIC

The data for the analyses in this study come from the Freiburg Multimodal Interaction Corpus (FreMIC), a multimodal corpus, consisting of transcripts of unscripted conversation in dyads and triads prepared using transcription software ELAN (Wittenburg et al., 2006) as well as large streams of non-linguistic data; the latter include *inter alia* eyetracking data (for a more detailed description of the corpus, cf. Rühlemann and Ptak, 2023).

Recordings were made in dyadic and triadic settings using one room camera and one centrally placed scene microphone. Participants wore Ergoneers eyetracking devices (Dikablis Glasses 3), which capture the participants' foveal (focal) vision; the direction of their foveal vision is indicated by crosshairs. Participants were informed that the eyetrackers were used to record their gaze behavior during conversation.

Participants were mainly students of Albert-Ludwigs-University Freiburg, as well as their friends and relatives. All participants' first language was English. Participants gave their informed consent about the use of the recorded data, stating their individual choices as to which of their data can be used and for what specific purposes. They received a compensation of € 15 for each recording.

During the recordings, participants were requested to sit in an F-formation (Kendon, 1967) enabling them to establish eye contact, hear each other clearly, and engage in nonverbal cues. Participants in triads were seated in an equilateral triangle, with the room camera frontally capturing one of the participants and the other two from the side (cf. Figure 1). The participants were instructed to talk about whatever they liked for 30–45 min, at which point the recording would be stopped.

FIGURE 1
Still taken from a split-screen video of a tradic conversation in the FreMIC corpus. Top left, top right, and bottom left tiles show the participants' eye-tracking (ET) video and bottom right tile shows the room camera perspective. Red crosshairs in participants' eye-tracking videos indicate participants' foveal vision.

The data selected from the corpus for this study include information-seeking question-answer (QA) sequences occurring in 12 triadic conversations[5] that have to date been fully transcribed and annotated, covering a run time of 8.6 h. QA sequences were targeted for this analysis as the action of asking an information-seeking question has maximal response relevance (Stivers and Rossano, 2010) or projection strength (Auer, 2021a), i.e., it makes the provision of the sought information normatively maximally relevant.

QA sequences in triadic files in FreMIC were rigorously annotated by several researchers as part of an ongoing research project funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, or DFG),[6] consulting video-recordings in ELAN and extensively discussing cases until consensus was achieved. QA sequences were filtered for "valid" sequences, with validity requiring that a genuine question is asked (rather than a rhetorical question, a self-directed question, an ironic question, or a leading question), the answer to the question is intelligible and verbal (rather than, for example, mere laughter) as well as type-fitting (rather than a comment on the question, as in A: *How much did you have to pay?*—B: *Aw, we are really grilling him!*), the question is a request for information (rather than a request for clarification or correction as in a repair question or a confirmation question, as in *That was tough wasn't it?*). Finally, cases in which excessive overlap in questions and/or responses made an unambiguous identification of the answerer impossible were excluded.

This original pool of valid QA sequences consisted of 357 sequences.

3.2 Data pre-processing

3.2.1 Annotation of dependent variable *Selection Type*

QA sequences were examined and annotated for *Selection Type;* that is, each question was coded for whether it was (i) a closed-floor question specifically addressed to one of the two question recipients and selecting that participant as next speaker or (ii) an open-floor question addressed to both question recipients not specifically selecting one of them as next speaker. Crucially, the questioner's gaze behavior was not considered for this annotation. Annotation was binary:

- YES: question is a closed-floor question picking out one particular participant as next speaker.
- NO: question is an open-floor question that does not *per se* select one particular participant as next speaker.

The variable *Selection Type* was coded by several researchers based on inspection of the respective contexts in ELAN; again, cases that were coded differently between them were extensively discussed until the consensus was achieved.

Annotation of *Selection Type* was based on the following set of criteria, applied comprehensively:

- Lexical means: the (rare) use of a singular address form suggested the presence of a closed-floor question, whereas the occurrence

---

of "you all," "you two," or "you guys" as in extract (2) made the presence of an open-floor question likely.

- Embodied behavior (categorically excluding, however, gaze behavior): A questioner's pointing to an addressee, as in extract (6), made the question most likely a closed-floor question.
- Sequential "pre-selection": Questions were examined closely, often as far as 5–10 min into the preceding transcription and/ or using multiple keyword searches, whether they were follow-up questions, as in extract (1). The decision on whether a follow-up question should be classified as open-floor or closed-floor was contingent upon psychosocial and/or sociodemographic information.
- Epistemics: We gathered from what was revealed during the recorded conversations, important psychosocial information related to, for example, past experiences, skills, interests, habits, and plans for the future, as well as sociodemographic information related to, for example, the participants' mutual relationships, their origin, age, and work situation to determine cases of (exclusive) tacit addressing, as in extract (5), but also to disambiguate the (frequent) get-to-know questions as addressed to either a single participant or to both participants. For example, the question "what's your work?", as in extract (4), would not be inclusively addressed to both participants in the co-presence of a spouse or close friend.

Based on the set of criteria, the distinction between open- and closed-floor question was overwhelmingly clear-cut. Extract (13), discussed as one of the deviant cases in Section 5.2, was, one of two examples in the present data in which the boundaries between the two types were somewhat fluid. In both cases, the classification was conservative, classifying them contrary to the hypothesis (i.e., as deviant *Selection Type* YES cases). *Selection Type* is the dependent variable in the statistical model (cf. Section 3.3).

3.2.2 Annotation of questioner's gaze and coding of the independent variable gaze alternation

Using ELAN, two researchers with expertise in gaze annotation manually annotated the question turn in each QA sequence for the questioner's gaze behavior. Gazes to recipients' faces were coded A, B, or C depending on the recipient's ID suffix; gazes were labeled * if the questioner's focal vision as indicated by the crosshairs was not on, or in immediate vicinity of, a participant's face.[7] The key variable coded was *Alternation*, a binary yes/no variable indicating whether gaze alternation did, or did not, occur during the question.

For example, in extract (8), discussed above in Section 2.2, participant B's question *[but]=°but° w- if you say it's a Dachgeschoss top floor is it like (0.493) slanted? and can you actually walk?* is delivered with one gaze alternation from A to C. The example was,

then, coded "yes" for *Gaze Alternation. Gaze Alternation* is the independent variable in the statistical model (cf. Section 2.2).

3.2.3 Correction of gaze annotation In question turns

The hypothesis tested in this study relates to the questioner's gaze behavior during the action of asking an information-seeking question. Speakers can perform more than one action in a turn (e.g., Levinson, 2013), for example, combining an assessment (e.g., *Great! Where did you find it?*) or an apology with a question (e.g., *Sorry, what did you say?*). Distinct actions may be coordinated with different gaze directions, resulting in a coding (e.g., A*B) that may erroneously suggest the use of gaze alternation *during the question part*. Consider for illustration extract (9):[8]

```
(9) [F23]
1   C:        [it's like] eve[ryone's] studying
    C-gaze    ------------------A---------|-*-|
2             /((name speaker B)) do you like have exams coming up?/↑=
    C-gaze    ----------------------B------------------------|
3             (0.012)
4   B:        =yea:h >I have one on Thursday and one more next Thursday< (.)
5             and then I'm done=
```

Speaker C gazes to participant A while uttering the first TCU *[it's like] eve[ryone's] studying* (likely a noticing), at the end of which she shifts her gaze away from A. When doing the second TCU/((*name speaker B*)) *do you like have exams coming up?/↑*, which is the question TCU of her turn, she fixates B (whom she also specifically addresses, and thus selects, by using her name). There is, then, no gaze alternation during the question. The gaze coding for this QA sequence as well as any other such sequence with false positives was corrected in that the variable *Alternation* was set to "no" to adequately reflect the fact that there was no gaze alternation during the question part of the turn as a whole.

3.2.4 Removal of QA sequences unrelated to the hypothesis

QA sequences were discarded for this analysis if the questioner's gaze did not fixate on any recipient at all. For example, in (10), the questioner's gaze is consistently an away-gaze, indicated by *:[9]

```
(10) [F23]
1   B:        >you< have very interesting tattoos
2             (0.921)
3   A:        yeah
4             (1.907)
5   A:        °thanks°
6             (0.294)
7   A:        ((v: laughs))
8             (0.925)
9   B: →      is there a story behind it? or
    B-gaze    ----------------*-------------|
10            (0.916)
11  A:        !NO:!: I mean,
12            (1.641)
13  A:        °I guess I just think it looks ↑coo:l↑°
```

---

7   Coding gazes to the immediate *vicinity* of a recipient's face reflects two facts: (i) the calibration of the eyetrackers may deteriorate the longer the conversation lasts and (ii) the eye can perform involuntary micro-motions (so-called fixation tremor) which is so brief the human observer can normally not perceive them but which may be shown by the crosshairs.

8   In this excerpt and all the following excerpts, gaze annotation is only included where relevant for the examination of gaze behavior during questions.

9   Note that the symbol * is used (i) when the crosshairs in the eyetracked person's video is not on or in the immediate vicinity of a co-participant's face but also (ii) when there is no crosshairs visible at all; the absence of the crosshairs may be due a deterioration in the eyetracker's calibration (e.g., when a participant has inadvertently touched the device) or extreme out-of-center rotation of the eyeball.

TABLE 1 Descriptive statistics.

Selection type	N	Gaze alternation = *no*	Gaze alternation = *yes*	% yes
YES (Closed floor)	251	246	5	1.99
NO (Open floor)	86	42	44	51.16
Total	337	288	49	14.54

The final set of QA sequences on which the analysis was based included 337 QA sequences drawn from 12 triadic conversations. The questions were asked by 20 distinct participants.

3.3 Statistical analysis

The statistical analysis was conducted in the statistical analysis software R (R Core Team, 2023). First, a descriptive statistical analysis was carried out to establish how *Gaze Alternation* was distributed across *Selection Type*.

To statistically examine whether gaze alternation in question-answer sequences predicts inclusive next-speaker selection, a Conditional Inference Tree model was calculated. This method was first implemented in linguistics by Tagliamonte and Baayen (2012); it has since gained popularity specifically in corpus-linguistic studies (for the underlying algorithm, see Hothorn et al., 2006; for a readable introduction see Levshina, 2015; and for a hands-on web-based introduction see Schweinberger, 2023).

A Conditional Inference Tree (CIT) is a machine learning method that uses binary recursive partitioning to analyze and model the relationships between variables. This method is particularly useful for handling sparse data, as is the case with the present data. The underlying null hypothesis in CITs posits that there is no association between the dependent variable (*Selection Type* in this context) and the independent variable(s) (*Gaze Alternation* in this context).

The hypothesis is tested through a permutation test. This involves randomly reshuffling the labels of *Selection Type* to break any association with *Gaze Alternation*, thereby generating a null distribution directly from the data without making parametric assumptions. In the first step, a test statistic is calculated to measure the association between *Selection Type* and *Gaze Alternation*. After permuting *Selection Type*, the test statistic is recalculated. The difference in the test statistics before and after permutation indicates the strength of the association. The larger this difference, the stronger the evidence against the null hypothesis. This process is repeated multiple times to obtain a distribution of the test statistic under the null hypothesis. Based on this distribution, a *p*-value is computed for each potential split. The algorithm selects the split with the lowest *p*-value, ensuring that *Selection Type* is divided into two subsets at the point where the association with *Gaze Alternation* is strongest.

The resulting model is visualized as a tree, where each branch represents the split in the data determined by the recursion with the smallest *p*-value. The Conditional Inference Tree was implemented and visualized in R using the *ctree*-function, which is part of the *party*-library.

4 Results

The results of the descriptive statistical analysis are shown in Table 1.

Gaze alternation was observed in 49 questions vs. 288 questions without gaze alternation. In closed-floor questions (*Selection Type* YES), only five cases (accounting for 1.99%) exhibited gaze alternation. That is, except for a handful of cases, *all* YES-type questions have zero gaze alternation. By contrast, in open-floor questions (*Selection Type* NO), in as many as 44 cases (accounting for 51.16%), the questioner's gaze did alternate.

The Conditional Inference Tree is shown in Figure 2.

It splits the data along the yes/no values of *Gaze Alternation*; the split is significant at *p* < 0.001. Adding up the cases that were correctly predicted by the model (246 *Selection Type* YES cases without gaze alternation and 44 *Selection Type* NO cases with gaze alternation) and dividing the sum by the total number of cases (337), the overall accuracy rate is 86.05% (error = 13.95%). If there was no association between *Gaze Alternation* and *Selection Type*, the accuracy rate would be 74%.[10] The model therefore represents an improvement over that baseline. Hence, in triadic question-answer sequences, the questioner's gaze alternation during the question significantly predicts that the question does not exclusively select one specific recipient as the next speaker but inclusively co-selects both recipients as potential next speakers.
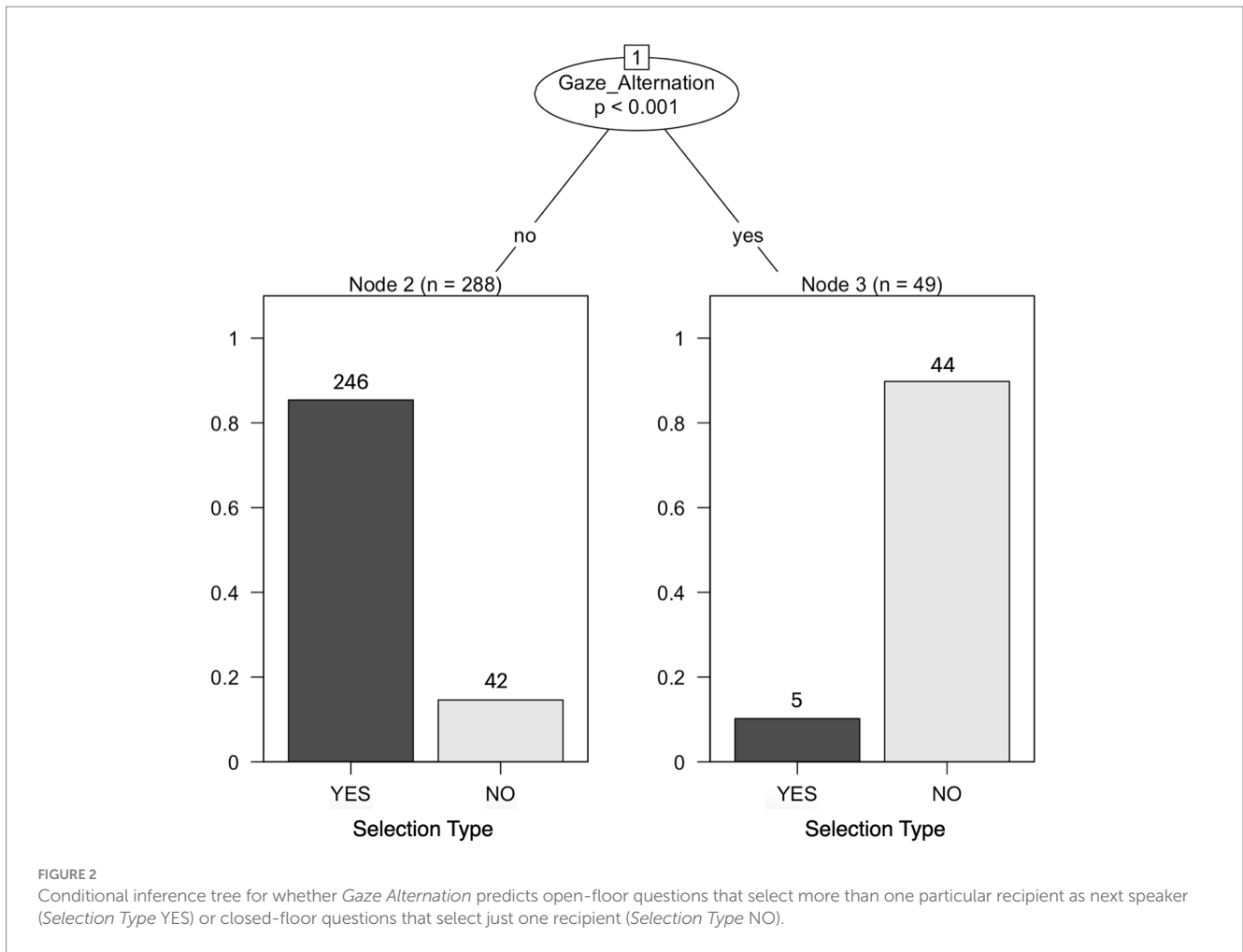
5 Discussion

The main finding of this analysis is that the use of gaze alternation by a speaker asking a question in triadic conversation in English predicts the use of an open-floor question, i.e., a question that does not *per se* select a particular participant as next speaker but licenses both question recipients to respond to the question.

The discussion of this finding falls into four sections. First, I will present an example illustrating gaze alternation in open-floor questions. Second, I will discuss all five deviant cases—that is, closed-floor questions where the speaker uses gaze alternation—in order to demonstrate that gaze alternation in these sequential contexts does not execute the function of co-selecting both recipients to respond. Third, I will assess the functional relation of gaze alternation and gaze selection. Finally, I will

---

10 The baseline is the maximum proportion obtained if the two levels (YES/NO) of the dependent variable *Selection Type* are tabulated, which is 0.7448071 or 74% (for *Selection Type* YES).

FIGURE 2
Conditional inference tree for whether *Gaze Alternation* predicts open-floor questions that select more than one particular recipient as next speaker (*Selection Type* YES) or closed-floor questions that select just one recipient (*Selection Type* NO).

consider the interactional costs incurred by the transition from the exclusive to the inclusive participation framework.

5.1 Typical cases

To illustrate the main point of this study, that gaze alternation is predictive of open-floor questions while steady gaze fixation is typical of closed-floor questions, I discuss one case of each.

In excerpt (11), speaker A inquires ≥*what do you guys study?<* (line 1). In response to this (distributive, open-floor) question, speaker C gives an extended answer detailing her study history (first studying archeology, then design). After she has concluded the telling (88 lines later) speaker A acknowledges and concludes it with = °*cool*° (line 89). Following a lapse of more than half a second (line 90) A turns to speaker B asking her *what about you¿* (line 91), a question design entirely dependent on the sequence-initiating question in line 1. A's gaze is on participant B consistently through the question. Participant B does not even reciprocate his gaze, neither at question onset or offset [see embedded figures 1, 2 in (8)]—she need not, for it is sequentially abundantly clear that, after speaker C's turn at answering the sequence-initiating question, it is now B's turn to answer it. Here, then, the question *what about you¿* is exclusively addressed to B; to mark and index that exclusive selection, speaker A's gaze is fixated on her.

```
(11) [F23]
1    A:      =>what do you guys study?<

((88 lines omitted; Speaker C is relating how she switched from studying
archeology to studying design))

89   A:      =°cool°
90           (0.565)
91   A:      #what about you¿%
     A-gaze  #--------B------%|

fig          #figure 1              %figure 2
```



```
             figure 1              figure 2

92           (0.265)
93   B:      I'm studying medicine=
94           (0.183)
95   A:      =!med!ici:ne woah
```

To illustrate cases where gaze alternation predicts the use of open-floor questions, consider extract (12). The participants have been talking about their country of origin. All three are Americans, but two of them, B and C, grew up in Germany. B and C are brothers. Before

the excerpt, they have just related that their father was employed as a lawyer by the US army at the airbase in Ramstein, a few minutes drive from the mid-sized city Kaiserlautern. At this point, speaker A asks *so d- y- you guys grew up in in>what they call<K-town (.) °right¿°* = (lines 1–2), where the term *K-town* is US military jargon for Kaiserlautern. Since B and C, as brothers, grew up together, this question is addressed to both collectively licensing them to respond. Speaker A's gaze behavior reflects this co-selection by alternating his gaze between the two potential next speakers as many as four times. As is frequently the case with collective questions, the two addressees both provide answers almost in complete overlap (lines 4–5). While B's = °*exactly*° (line 4) aligns directly with the yes/no design of the polar question underlying A's tag question °*right¿*°=, C's answer is "double-barreled" (cf. Levinson, 2013): it expresses his emphatic appreciation of the fact that A knows the jargon term and, in so doing, indirectly answers the yes/no question in the positive (note also his knocking on the arm rest twice as another sign of appreciation).

(12) [F16]

```
1     A:    →    so d- y- you guys grew up in in
      A-gaze      -B-|-*-|---C---|-*|------B-----
2           →    >what they call< K-town (.) °right¿°=
      A-gaze      -B-|-*-|---C----|-*|---------B--------
3                (0.042)
4     B:         =°exactly°
5     B:         [£↑you know it?↑ yeah  =yeah  =yeah£ ]#
5     C:         =d[        a:      #n       g      ]#
      C-gesture:                    #knocks 2x on arm rest#
```

Given that speaker A uses as many as four gaze alternations, example (12) may be considered unusual (the largest number of gaze alternations in the data is five).

We now turn to deviant cases. As noted, not all open-floor questions exhibit gaze-alternation; slightly less than half of them are without gaze alternation. However, these cases are well embedded within the distribution so they cannot be considered deviant cases. By contrast, gaze-alternation in closed-floor questions is quite untypical; they are hence considered the deviant cases. There are, as noted, five of them. I discuss each of them in the following.

5.2 Deviant cases

The first extract representing a deviant case was already briefly mentioned in Section 3.2.1 as one of the two instances in the present data where the distinction between open- and closed-floor selection was challenging.

Before the excerpt, speaker A has been questioning how the recording situation influences and possibly compromises the participants' behavior. After he has made his point, he invites the other two participants to share their view on the matter by asking what is clearly an open-floor question >°*what do y'all think*°< (line 1). After a fragmented question TCU (line 2) and a half-second pause (line 3), he goes on to narrow the scope of the inquiry to whether the co-participants *feel at °ea:se°* (line 4) and *feel comfortable* (line 5) before he wraps up the question turn asking *how do you feel* (line 5). While A's gaze is on C in line 1, he looks away from both participants during the hesitation (lines 2–3), returns his gaze very briefly to C during the launch of the more specific question *do you feel at °ea:se°* (line 4) before he settles his gaze on B for the rest of the utterance. During the concluding question TCU *how do you feel* (line 6), speaker

A not only turns his head to B but extends his left arm and hand toward B, an interactional gesture that "offers" the response turn exclusively to B.

(13) [F27]

```
1     A:          >°what do y'all think°<
      A-gaze      ------------C---------->
2                 how do you uhm
      A-gaze      ---C----|--*-*-->
3                 (0.590)
      A-gaze      -------*------>
4                 do you feel at °ea:se°
      A-gaze →    -*-|--C--|-*-|---B---->
5                 °°do you feel comfortable
      A-gaze      -----------B---------->
6                 #how          do# you feel°
      A-gaze      -----------------B-----------------|
      A-gesture   #extends hand palm-up to B#
      fig         #fig. 1
```


figure 1

```
7     B:         ((v: laughs))
8                (0.310)
9     B:         £well [for o:]ne I think >this is like my fourth time being
                        here,<((answer continues))
```

In other words, the (long) question turn starts off as an inclusive open-floor question addressed to *y'all* but is both semantically and gesturally narrowed down, across its multiple question TCUs, to a closed-floor question addressed to one co-participant only.

In the second extract representing a deviant case of gaze alternation in closed-floor questions, extract (14), C is telling how he wrote an *email last night* (line 1) to a company to which he had submitted a job application. The telling, which has been going on for a while, is directed to participant B, as participant A is C's wife and therefore most likely in the know about C's application. When C relates that *they have not responded so I think it's probably a≈no≈* (line 3) and adds an iconic sad sniffing to convey his disappointment (line 5), his wife A, who has largely abstained from contributing to the telling, also responds non-verbally by curling her lip, thus expressing sympathy with the negative outcome of the application (underneath line 6). Here, then, we observe the occurrence of a shift from a dyadic participation framework including speakers C and B but excluding speaker A (during the bulk of the telling) to one that includes speakers C and A, engaging in an intimate exchange of non-verbal actions that index their shared knowledge and shared stance toward the failed application but excludes speaker B.[11]

The gaze behavior by all participants during A's silent gesture is critical here: As she performs the facial expression, A fixates on teller (and partner in life) C, addressing her sympathy to him. Her facial gesture is noticed by C, who shifts his gaze toward A before gazing back to B. Moreover, B, whose gaze is first on C, follows C's gaze to also fixate on A. B thus perceives the silent alignment between the couple. By asking *did you get a feeling that it was: like lots of people were applying?* (lines 7–8), B

---

11   Extract (14) is the second QA sequence which we decided conservatively to classify as *Selection Type* YES (based on the relative uncertainty that A, C's wife), was in the know about the main content of the question, namely, whether many people applied for the job.

connects to the couple's dyadic framework acknowledging it and entering into it by briefly gazing to A during the matrix clause of the question *did you get a feeling that* (line 7), but then, his gaze shifts away from her as he produces the first part of the subordinate clause *that it was: like lots of* (lines 7–8), to finally fixate on C on *people were applying?* (line 8). B's gaze switch from A to C thus works as an "outside-in resource" (Stivers, 2021) to re-insert himself into the participation framework. At this point, then, the dyadic participation framework including himself and participant C but excluding participant A is re-established. The question asked is in line with that framework: whether many people were applying requires specific information that may not necessarily be available to C's wife. Therefore, the question tacitly selects C as the next-speaker, a selection that is underscored by B's final gaze to C.

```
(14) [F04]
1    C:          and I wrote him an email last night,
2               (0.580)
3    C:          °but they haven't responded so I think it's probably a ≈no≈°
4               (1.062)
5    C:          ((sniffles sadly))
6               (0.341) + @--------(0.656)--------@
     A-facial:          @curls lower lip--------@
     A-gaze:            @----------C-----------@
     C-gaze             @-*-|----A-----|-----*-|--B--@
7    B: →        did you get a feeling that it was: like
     B-gaze      ---*----|----A-----|----------*------>
8    B: →        lots of people were applying?
     B-gaze      -----*----|--------C--------|
9               (0.782)
10   C:          no: I felt like I was °the only one°
```

Here, then, the questioner's (B) gaze alternation is finely attuned to and synchronized with perceived changes in participation framework.

Before the next excerpt (15), speaker B has been telling his co-participants that he is being charged too much for the water he consumes in his apartment. At the start of the question by speaker C (line 1), C's gaze first fixates on the table between the three participants. As he launches the second TCU *ask how much the quote was* (lines 3–5), he first fixates on B but briefly shifts his gaze to A before he returns it to B. While C is repeating *and* between the first (line 1) and second TCU (line 3), participant A has started turning his gaze away from the questioner to fixate on his knee. With a swift movement, he next touches his knee apparently to remove a small object from there. Questioner C's shifting gaze to A closely follows A's swift hand movement to his knee, likely to monitor "what is going on" there.

```
(15) [F12]
1    C: →        [can't you] get back in touch with the company?
     C-gaze      ---------------------*--------------------->
2               the water company and and
     C-gaze      -------------*----------|
3               (0.523)
     C-gaze      --*--|-B->
4    C: →        ask how much
     C-gaze      ------B---->
5               (0.518)
     C-gaze      -----B---->
6    C: →        the quote was?
     C-gaze      -------B----|
7               (0.570)
     C-gaze      -*-|#-A-|
```

fig   #fig. 1   #fig. 2

figure 1   figure 2

```
8    C: →        and then reduce your rent by that much:
     C-gaze      ---*----|------------B--------------|
9               (0.401)
10   B:          I !wa:s! thinking about just telling them
11              ~hey I'll pay for it just fix it~
12              and then I'll subtract that from my rent
```

Thus, in (15), the gaze shift is likely occasioned by a momentary distraction of C's attention due to the third participant's manipulation of an object.

In the following extract (16), the participants are talking about how they came to participate in the recording. B relates that she told C about the opportunity (line 1). A infers that C is a friend of B's (line 3) and asks *are you [guys friends or:]* (line 3). B answers in the negative (lines 4–5) saying *[uh:m we] actually [£!do not! (even know each other)£ ((v: laughs))]*, which is confirmed by C saying largely in overlap with B *[(((v: laughs)) £we do not really know each other]!THAt! we(h)ll£ [((v: laughs))]* (lines 6–7). Participant A finally inquires *[>how did you< how d'you!t]ell! her then¿* (lines 9). During this question TCU, his gaze is first on participant C but then shifts away to the table (from where he picks up some object) before it finally takes into focus participant B (line 12). The fixation on B co-occurs with the TCU extension *o:r=* (line 11).

```
(16) [F23]
1    B:          [a:n]d !then! I told her
2               (0.621)
3    A:          !that's! your friends=are you [guys friends or:]
4    B:                                         [  uh:m    we ] actually
5               [£!don't! (even know each other)£ ((v: laughs))]
6    C:          [(((v: laughs)) £we don't really know each other]
7               !THAt! we(h)ll£ [        ((v: laughs))        ]
8    A: →                        [>how did you< how d'you !t]ell! her then¿
     A-gaze                      -------C-----|--------------*-----------|
9    →           o:r=
     A-gaze      --B-|
10              (0.005)
11   C:          =u::hm well technically: I go to: a Korean class with a
12              friend of !hers!,
```

Here, it is noticeable that the gaze alternation takes place during a question turn that is exclusively addressed to one participant (C, in this case) using the pronoun *you* in subject function and that includes a pronoun (*her*) in object function referring to the other participant (B). In other words, the question addresses a reciprocal relation that holds between the two recipients.

In excerpt (17), speakers A and C are an American couple, and B is a student from the UK funding his studies by giving English language classes. Speaker B has just bemoaned at great length the difficulties of becoming a *Beamter* (civil servant) in the German school system. At this point, participant C attempts to explore an alternative career path for him *if that does not work out* (with *it* referring to a permanent position as a Beamter; line 1), namely, working at private language schools (line 2). Specifically, he inquires whether Cambridge or other private language schools *offer more like full-time positions [and stuff?]* (lines 2–4). During this long and complex question turn, his gaze alternates from B, the recipient exclusively selected to answer the question, to participant A and back to B. Note that the gaze shifts neatly coincide with re-starts, i.e., breaks in the syntax of the question: the first gaze shifts from B to A co-occurs with the first re-start where the particular language school reference *Cambridge* is replaced by the more general reference to *those private schoo:ls*, the second shift is aligned with the switch from the full nominal reference *those private schoo:ls* to the pronominal reference *they*. In other words, the gaze alternation in extract (17) may be explained by its conspicuous synchrony with the multiple re-starts.

```
(17) [F04]
1    C:          if that doesn't work out
     C-gaze      -----------B---------->
2    →           does like Cambridge like those private schoo:ls
     C-gaze      ---------B---------|-----*----|-------A--------|
3    →           do they offer more like full-time positions
     C-gaze      ------*---|--------------B--------------->
4    →           [  and    stuff?  ]
     C-gaze      ---------B-------|
5    B:          [ah well the gu:y] s- seems very as from (.) a
6               (0.265)
              ((answer continues))
```

To sum up this section, gaze alternation can come about even in questions designed for one particular participant. Five such instances were examined. It transpired that the questioner's gaze can alternate between the question-addressee and the non-addressed third participant (i) if the question turn involves numerous question TCUs and develops across these TCUs from being inclusive to exclusive (extract 13), (ii) if the questioner uses the question as an "outside-in resource" (Stivers, 2021) to re-enter into a dyadic participation framework that includes them (extract 14), (iii) if the questioner's attention is momentarily distracted by the third participant performing some unusual embodied conduct (extract 15), (iv) if the question is about the addressed participant's relation with the non-addressed participant (extract 16), and (v) if a question contains multiple re-starts (example 17). These circumstances are doubtlessly not the only ones that can result in the occurrence of gaze alternation in single-addressee questions. Analysis of more instances will surely unearth more contingencies. The one function *any* case of gaze alternation in closed-floor questions does, and will, not serve is to flag a question that is not asking *somebody* a question but opens the floor to more than one question recipient.

5.3 The functional relation of gaze alternation and gaze selection

How is gaze alternation related to *gaze selection*? If gaze alternation has an inclusive function co-selecting addressees as potential next speakers, gaze selection—especially the speaker's last gaze (Auer, 2018, 2021a,b)—has the opposite effect, picking out exclusively one particular addressee. This might suggest that whatever is signaled by gaze alternation is canceled out by gaze selection because the latter is temporally later.

This assumption, however, does not stand up to scrutiny. First, as was observed in most of the extracts featuring open-floor questions discussed earlier, many open-floor questions are answered by both recipients, either in overlap or in quick succession of the respective answers (see examples 1–4, 7, 12–13),[12] thereby violating both the norm for single speakers (Sacks et al., 1974) the norm for single responses (Stivers, 2021); for the preference for single responses in multi-party storytelling, see Rühlemann and Gries (2015). In all of the closed-floor questions examined in this study, by contrast, there was a single one that was answered by both recipients. In these cases, then, the speaker's last gaze has no selecting effect at all.

Second, Auer et al. (in preparation) found that the correlation between speaker's last-gaze and next-turn by that last-gazed-at participant is very high in closed-floor questions with 98% (cf. Stivers, 2021 for similarly high percentages). In open-floor questions, however, that percentage falls almost by a third to 67%. In open-floor questions with gaze alternation, the percentage falls even further to 56%, not far above the chance threshold. One of the reasons for this substantial drop is the fact that gaze alternation

shortens the time the speaker can fixate their gaze on any one participant: in the study by Auer et al. (in preparation), the mean length of the speaker's last gaze to a participant is 1,586 ms, whereas in open-floor questions, it is 1,029 ms. That is, if the speaker's last gaze works as a selection method, the time during which its effects can unfold in open-floor questions is substantially reduced.

Consider extract (18). Speakers A and B are collaboratively designing a mental map for a joint meeting. Speaker B mentions the *Mundenhof* (line 1), a location in the Freiburg area, as a reference point that is close to their intended meeting point. Speaker A confirms that reference (line 4) and goes on to give an estimate of how long it would take to get to the Mundenhof with rollerblades (lines 7–8). Both speakers seem done with their mental map at this point acknowledging it with *mhm*s (lines 9–10). Speaker C has so far not explicitly contributed to the mental map. Now he extends the sequence by asking *[isn't the] Mundenhof by the Dreisam?* (line 11), referring to the river Dreisam in the Freiburg area and thus adding yet another locational reference to the map. Both A and B are residents of Freiburg and familiar with the reference. The question is thus an open-floor question co-selecting both A and B as potential answerers. As is characteristic for open-floor questions, C's gaze co-addresses A and B, first fixating on A on *[isn't the] Mundenhof* and then fixating on B turn-finally on *the Dreisam?*. Almost latching on to the question (after a gap of merely 0.05 s), recipient A, who was not last gazed at, answers *it's not [fa:r from the Dreisam]* = (line 13), while recipient B, who *was* last gazed at, remains silent.[13] In this example, then, the questioner's last gaze has no selecting effect. The sole selection effect that can be observed is the co-selection instigated by the open-floor question and the questioner's gaze alternation.

```
(18) [F07]
1    B:        =but you ['re !not!] you're not far from the Mundenhof right¿=
2    A:                 [    but  ]
3              (0.080)
4    A:        =!no! not at all=
5              (0.045)
6    B:        =exactly [ yeah °we  should probably° ]
7    A:                 [yeah so it's like ten to fif]t- yeah °ten minutes to
8              Mundenhof maybe fifteen with rollerblades [(my guess)°]
9    B:        °mhm°
10   B:        [  (mhm)  ]
11   C:    →   [isn't the] Mundenhof by the Dreisam?
     C-gaze    ----------A-------#|-*-|------B-----%|
     fig       #figure 1              %figure 2
```


figure 1   figure 2

```
12             (0.050)
13   A:        it's not [fa:r from the Dreisam]=
14   C:                 [    fairly °close°    ]
15             (0.090)
16   C:        =cos we're gonna be going to the Dreisam as !well!=
```

This is *not* to suggest that gaze selection becomes irrelevant. After all, as noted, in the study by Auer et al. (in preparation), the last gazed-at recipient *is* the answerer in roughly two-thirds of all open-floor questions. However, the fact that open-floor questions

---

are much more often answered by multiple recipients or are answered more often by the recipient which was not last gazed at than closed-floor questions does suggest that gaze alternation and gaze selection enter into a potentially complicated relationship.

In the two-thirds of cases where the last-gazed-at recipient is the answerer, the two selection types work hand-in-hand: first, gaze alternation signals that the question is open for both recipients to answer, and then, the speaker's last gaze functions as a "You go first" cue. In one-third of all cases, however, the inclusive selection inherent in the open-floor question overpowers the last-gaze selection, potentially resulting in disorderly turntaking: either the answer is given by the recipient that was not last looked at or multiple answers are produced, mostly in overlap, thus violating the single speaker norm and the single-answer norm.

5.4 Interactional costs of the inclusive participation framework

This relative amount of disorderliness is, however, to be expected considering that open-floor questions facilitate a fundamental change in participation framework, a framework that is overwhelmingly dyadic even in multi-party interaction, so much so that conversation may be "built for two" (Stivers, 2021). Unlike closed-floor questions, which are inherently dyadic, open-floor questions are an (rare instance of an) inside-out resource partitioning the focal dyadic framework into an inclusive framework that allows all participants free entry and re-entry onto the floor. As such, they also work against the last-as-next bias (Sacks et al., 1974), whereby the speaker prior to the current speaker is the next speaker.

If the turn-taking system is structurally dyadic and, hence, exclusive, the transition to an inclusive participation framework exposes the limits of the system. Interacting at these limits poses a challenge, as the expansion of the participation framework from dyadic/exclusive to triadic/inclusive incurs additional interactional costs: co-selected question recipients are tasked to collectively advance the sequence *and* to do so while minimizing or avoiding the disruption caused by multiple answers and overlap. This double tasking requires that recipients *coordinate* their behavior multimodally. To illustrate, consider extract (19), a sequence with an open-floor question and gaze alternation on the part of the answering participant.

Speaker A has not met speakers B and C before the recording. His inquiry about their work *and what do you guys do?* (line 1) co-selects both B and C to respond. After a long gap of more than a half second (line 2), participant B, who was last gazed at by the questioner, starts up answering with a click sound (line 3). While producing the click, he starts laying his right hand on his chest (cf. embedded figure 1), turns around, and gazes to participant C, who returns his gaze, sitting still (cf. figure 2). Only after that "mutual-gaze check," the candidate answer *[stud]ying¿* (line 5) proffered by questioner A, and another delay of more than half a second (line 6) does he turn back to questioner A (see figure 3) to finally launch the full answer (lines 7–8):

```
(19) [F12]
1    A:          and what do you guys do?
     A-gaze      -----*-----|------B----->
2                (0.625)
3    B:          !t!#
fig  #figure 1
```
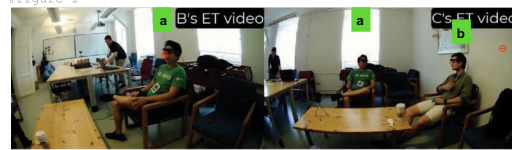

figure 1

```
4    B:          [I'm:]+
     B-gaze      --A--|-*-|--C--|+
fig  +figure    2
```


figure 2

```
5    A:          [stud]ying¿
6                (0.610)
     B-gaze      ---*---|-------A-----> ≈
7    B:          ≈I'm doing a little mix of stuff so I'm I'm ki- !I!
8                so I had actually wanted to like do uh liberal arts studies,
fig  ≈figure 3
```


figure 3

Participant B's conduct in this extract showcases the interactional effort made relevant by the inclusive participation format: B already delays even producing the initial click, thus creating space for C to take the turn; his my-side gesture to his chest indicates that his answer will provide just one of potentially two answers, thus preparing the ground for C's answer later in the sequence; he puts the answer further on hold to coordinate with C, his momentary partner in this task, by alternating his gaze from the questioner to him; and he further delays, again allowing space for C to go ahead, before he finally decides to launch the answer. Extraordinary interactional effort is invested, then, into coordinating with his momentary companion to facilitate smooth frictionless turn transition in the inclusive participation format.[14]

6 Concluding remarks

This study concerned itself with the multimodal design of next-speaker selection. The focus was on how Sacks et al.'s (1974) first rule "current-selects-next" applies to information-seeking questions in triadic conversations. The study introduced the conceptual distinction between inclusive and exclusive next-speaker selection:

---

14   Extract (19) also illustrates that gaze-selection may be subordinate to the co-selection effected by the inclusive focal participation framework: the considerable effort expended by participant B, who is the only recipient looked at by the questioner in order to coordinate with the other recipient cannot satisfactorily be explained if selection was determined by the current speaker's turn-final gaze only.

it was argued that next-speaker selection is inclusive when speakers ask a question that can be answered legitimately by either or both question recipients, whereas next-speaker selection is exclusive when speakers ask a question that selects one particular question recipient as next speaker.

The study aimed to shed light on how speakers design questions multimodally such that their embodied conduct distinguishes closed-floor questions that exclusively select a single participant as next speaker from open-floor questions where next-speaker selection is inclusive, inviting more than one participant to answer the question.

Based on eyetracking data, it was shown that a key component of the speakers' embodied conduct is their gaze behavior: speakers asking an open-floor question characteristically alternate their gaze from one participant to the other, a gaze behavior that is virtually absent from the multimodal design of closed-floor questions.

The discussion addressed deviant cases as well as the functional relation of gaze alternation and gaze selection. It was argued that gaze alternation and gaze selection enter into a complicated relationship, where, in the majority of cases (two thirds), gaze alternation is (temporally) first signaling that the floor is open to all participants, and gaze selection is (temporally) second signaling to one specific participant "You go first." In the minority case (one third), however, gaze alternation trumps gaze selection, such that, as a result, answers are produced by the non-gaze-selected participant or multiple, often choral, responses occur in violation of the norms for single responses and single speakers.

The discussion concluded by relating this element of disorder to the fact that open-floor questions represent an inside-out resource for speakers to include multiple participants in a focal participation framework (Stivers, 2021). The use of such resources in interaction initiates a diversion from the dyadic focal participation framework, which is the default in conversation. If indeed conversation is fundamentally "built for two," it will be "orderly" (Sacks, 1984) as long as the focal participation framework is dyadic; its orderliness will be challenged as soon as the focal participation framework strays away from the dyad and becomes inclusive. That challenge translates into either norm-violating turn transitions (through, mostly choral, multiple answers) or into additional interactional work expended by question recipients in order to coordinate around the *collective* task of managing turn transition smoothly, without such norm violations.

In sum, the study of gaze alternation in question-answer sequences guides us directly to social interaction where speakers co-select next speakers without selecting *somebody* as next speaker (cf. Schegloff, 1992b). In these contexts, gaze alternation is a multimodal practice that counteracts the structural tilt of the turn-taking system toward a dyadic focal participation framework and, instead, facilitates an inclusive triadic participation framework (Stivers, 2021). As conversation is built for dyadic but not triadic interaction, the challenge for the co-selected next speakers is to coordinate around the task of accomplishing smooth turn transition—a task that requires more interactional work than accomplishing turn transition in a dyadic focal participation framework (cf. Stivers, 2021).

The extra interactional work required to coordinate will, presumably, be done multimodally most of the time. Very little is known about what specific multimodal practices co-selected next speakers employ to achieve coordination, and the present study has just touched upon them. Elucidating these practices in more detail and more rigorously therefore represents an intriguing avenue for future research. A thorough examination of the practices involved in the recipients' coordination might also deepen our understanding of how turns are allocated. Sacks et al.'s (1974) set of rules makes no (explicit) provision for co-selection in inclusive participation formats. Here, the decision to take a turn is neither the result of compliance with rule 1a "currrent selects next" (where that "next" is understood to be a single participant) nor is the decision a participant's single-handed decision, as suggested by rule 1b (self-selection). The decision may instead be the outcome of intricate mutual monitoring and collaboration *among co-selected next speakers*. If this could be shown to be a systematic option, the set of rules governing turn allocation would have to be expanded to include a fourth rule, which could *tentatively* be stated as follows: If the turn-so-far is so constructed as to involve the use of a "current speaker co-selects next speakers" technique, the parties so selected share the right and the obligation to coordinate around the task of taking the next turn to speak.

The present study merely points to this fourth technique as a possibility. Needless to say, to explore it, let alone corroborate it, much further study is needed.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies involving humans because the subjects involved in the study all provided written informed consent to how different types of their data could be used. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

CR: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2024.1396925/full#supplementary-material

References

Auer, P. (2018). "Gaze, addressee selection and turn-taking in three-party interaction" in Eye-tracking in interaction. Studies on the role of eye gaze in dialogue. eds. G. Brône and B. Oben (Amsterdam: Benjamins), 197–231.

Auer, P. (2021a). Turn-allocation and gaze: a multimodal revision of the "current-speaker-selects-next" rule of the turn-taking system of conversation analysis. *Discourse Stud.* 23, 117–140. doi: 10.1177/1461445620966922

Auer, P. (2021b). Gaze selects the next speaker in answers to questions pronominally addressed to more than one co-participant. *Interact Linguist* 1, 154–182. doi: 10.1075/il.21002.aue

Auer, P., Rühlemann, C., and Gries, S. T. (in preparation). Gaze and next-speaker selection in question-answer sequences in triads.

Auer, P., and Zima, E. (2021). On word searches, gaze, and co-participation. *Gesprächsforschung* 22, 390–425.

Beattie, G. (2016). Rethinking body language: How hand movements reveal hidden thought. Abingdon: Routledge.

Brône, G., and Oben, B. (2021). Monitoring the pretence. Intersubjective grounding, gaze and irony. *Cognitive Sociolinguistics Revisited*, 48, 544–556. doi: 10.1515/9783110733945-044

Carpenter, M., and Liebal, K. (2011). "Joint attention, communication, and knowing together in infancy" in Joint attention: New developments in psychology philosophy of mind, and social neuroscience. ed. A. Seemann (Cambridge/MA: MIT Press), 159–182.

Falck-Ytter, T., Fernell, E., Hedvall, A., Von Hofsten, C., and Gillberg, C. (2015). Gaze performance in children with autism spectrum disorder when observing communicative actions. *J. Autism Dev. Disord.* 42, 2236–2245. doi: 10.1007/s10803-012-1471-6

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

Hayano, K. (2013). "Question design in conversation" in *The handbook of conversation analysis*. eds. J. Sidnell and T. Stivers, 395–414.

Holler, J., and Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Front. Psychol.* 6:98. doi: 10.3389/fpsyg.2015.00098

Hömke, P., Holler, J., and Levinson, S. C. (2017). Eye Blinking as Addressee Feedback in Face-To-Face Conversation. *Research on Language and Social Interaction.* doi: 10.1080/08351813.2017.1262143

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 15, 651–674. doi: 10.1198/106186006X133933

Jefferson, G. (2004). "Glossary of transcript symbols with an introduction" in Conversation analysis: Studies from the first generation. ed. G. H. Lerner (Amsterdam: John Benjamins), 13–31.

Kalma, A. (1992). Gazing in triads: a powerful signal in floor apportionment. *Br. J. Soc. Psychol.* 31, 21–39. doi: 10.1111/j.2044-8309.1992.tb00953.x

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4

Kendrick, K. H., Holler, J., and Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: gaze direction and manual gestures aid the coordination of turn transitions. *Philos. Trans. R. Soc. B* 378:20210473. doi: 10.1098/rstb.2021.0473

Lerner, G. H. (2003). Selecting next speaker: the context sensitive operation of a context-free organization. *Lang. Soc.* 32, 177–201. doi: 10.1017/S004740450332202X

Levinson, S. C. (2013). Action formation and ascription. In: J. Sidnell and T. Stivers (eds) *The Hand book of Conversation Analysis*, pp. 103–130.

Levshina, N. (2015). How to do linguistics with R. Data exploration and statistical analysis. Amsterdam/Philadelphia: John Benjamins.

Link, G. (1991). Plurals. In StechowA. von and D. Wunderlich (Eds.), Semantik: Ein internationales Handbuch der zeitgenössischen Forschung (pp. 418–440). Berlin: De Gruyter.

Mondémé, C. (2023). Gaze in interspecies human–pet interaction: some exploratory analyses. *Res. Lang. Soc. Interact.* 56, 291–310. doi: 10.1080/08351813.2023.2272527

Moore, C. R., and Robinson, J. D. (2024). An action-specific examination of the role of answerers' gaze orientation in managing transition relevance. *Discourse Process.* 61, 142–163. doi: 10.1080/0163853X.2024.2305528

Noris, B., Nadel, J., Barker, M., Hadjikhani, N., and Billard, A. (2012). Investigating gaze of children with ASD in naturalistic settings. *PLoS One* 7:e44144. doi: 10.1371/journal.pone.0044144

Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258. doi: 10.1126/science.1107621

Pfeiffer, M., and Weiß, C. (2022). Reenactments during tellings: using gaze for initiating reenactments, switching roles, and representing events. *J. Pragmat.* 189, 92–113. doi: 10.1016/j.pragma.2021.11.017

R Core Team. (2023). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available at: http://www.R-project.org/

Rigby, S. N., Stoesz, B. M., and Jakobson, L. S. (2016). Gaze patterns during scene processing in typical adults and adults with autism spectrum disorders. *Res. Autism Spectr. Disord.* 25, 24–36. doi: 10.1016/j.rasd.2016.01.012

Rühlemann, C., Gee, M., and Ptak, A. (2019). Alternating gaze in multi-party storytelling. *J. Pragmat.* 149, 91–113. doi: 10.1016/j.pragma.2019.06.001

Rühlemann, C., and Gries, S. T. (2015). Turn order and turn distribution in multi-party storytelling. *J. Pragmat.* 87, 171–191. doi: 10.1016/j.pragma.2015.08.003

Rühlemann, C., and Ptak, A. (2023). Reaching below the tip of the iceberg: a guide to the Freiburg multimodal interaction Corpus (FreMIC). *Open Linguist* 9:1. doi: 10.1515/opli-2022-0245

Sacks, H. (1984). "Notes on methodology" in Structures of social action. eds. J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press).

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735. doi: 10.1353/lan.1974.0010

Schegloff, E. A. (1992a). Repair after next turn: the last structurally provided defense of intersubjectivity in conversation. *Am. J. Sociol.* 97, 1295–1345. doi: 10.1086/229903

Schegloff, E. A. (1992b). "To Searle on conversation: a note in return" in Searle on conversation. eds. H. Parret and J. Verschueren (Amsterdam: Benjamins), 113–128.

Schegloff, E. A. (1995). "Parties and talking together: two ways in which numbers are significant for talk-in-interaction" in Situated order: Studies in the social Organization of Talk and Embodied Activities. eds. P. Ten and H. G. Psathas (Washington, DC: University Press of America), 31–42.

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organisation of repair in conversation. *Language* 53, 361–382. doi: 10.1353/lan.1977.0041

Schweinberger, M. (2023). An introduction to conditional inference trees in R. Bonn: Rheinische Friedrich-Wilhelms-Universität Bonn. Available at: https://MartinSchweinberger.github.io/TreesUBonn/index.html (Version 2023.01.19)

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: when nodding is a token of affiliation. *Res. Lang. Soc. Interact.* 41, 31–57. doi: 10.1080/08351810701691123

Stivers, T. (2021). Is conversation built for two? The partitioning of social interaction. *Res. Lang. Soc. Interact.* 54, 1–19. doi: 10.1080/08351813.2020.1864158

Stivers, T., and Rossano, F. (2010). Mobilizing response. *Res. Lang. Soc. Interact.* 43, 3–31. doi: 10.1080/08351810903471258

Tagliamonte, S., and Baayen, R. H. (2012). Models, forests and trees of York English: was/were variation as a case study for statistical practice. *Lang. Var. Chang.* 24, 135–178. doi: 10.1017/S0954394512000129

Tiitinen, S., and Ruusuvuori, J. (2012). Engaging parents through gaze: speaker selection in three-party interactions in maternity clinics. *Patient Educ. Couns.* 89, 38–43. doi: 10.1016/j.pec.2012.04.009

Waring, H. Z., and Carpenter, L. B. (2019). "Gaze shifts as a resource for managing attention and participation" in The embodied work of teaching. eds. J. K. Hall and S. D. Looney (Bristol, UK: Multilingual Matters), 122–141.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "Elan: a professional framework for multimodality research," In *Proceedings of LREC*, 2006 (Genoa)

Appendix

Transcription conventions underlying the (extended) Jeffersonian transcription (cf. Jefferson, 2004) in the Freiburg Multimodal Interaction Corpus (FreMIC).

Discursive:
      constructed dialogue: *she's like* ~ …~.
Sequential:
      overlap: […]
      latching: =… or …=.
Temporal:
      decelerated speech: <…>.
      accelerated speech: >…<.
      minimal pauses: (.)
Phonological:
      intensity:
            loud voice: caps, e.g., *HEY.*
            quiet voice: °…°.
      emphasis:!…*!*
      vowel stretching: colon, e.g., *dra:ft.*
      truncation: dash, e.g., *springt-.*
      intonation:
            full rise:?
            half rise: ¿.
            sentence-like drop: .
            continued intonation:,
            high pitch: ↑…↑.
            low pitch: ↓…↓.
            'scale' upward: /…/↑.
            'scale' downward: /…/↓.
      voice quality:
            tremulous voice: ≈…≈.
            creaky voice: ¥…¥.
            smiley voice: £…£.
Laughter:
      within-word laughter: laughter pulses, e.g., *ok(h)ay.*
      freestanding laughter: as event, e.g., *((v: laughs)).*