Eyebrow movements as signals of communicative problems in human face-to-face interaction

Paul Hömke [1], Stephen C. Levinson [1,2,] Judith Holler [1,3*],

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.
[2] Centre for Language Studies, Radboud University Nijmegen, The Netherlands.
[3] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands.

***** Corresponding author
Email: Judith.holler@mpi.nl

Abstract

Repair is a core building block of human communication, allowing us to address problems of understanding in conversation. Past research has uncovered the basic mechanisms by which interactants signal and solve such problems. However, the focus has been on verbal interaction, neglecting the fact that human communication is inherently multimodal. Here, we focus on a visual signal particularly prevalent in signaling problems of understanding: eyebrow frowns and raises. We present a corpus study showing that verbal repair initiations with eyebrow furrows are more likely to be responded to with clarifications as repair solutions, repair initiations that were preceded by eyebrow actions as preliminaries get repaired faster (around 230 ms), and eyebrow furrows alone can be sufficient to occasion clarification. We also present an experiment based on virtual reality technology, revealing that addressees' eyebrow frowns have a striking effect on speakers' speech, leading them to produce answers to questions several seconds longer than when not perceiving addressee eyebrow furrows. Together, the findings demonstrate that eyebrow movements play a communicative role in initiating repair in spoken language rather than being merely epiphenomenal. Thus, they should be considered as core coordination devices in human conversational interaction.

Keywords: repair, eyebrow movements, frowns, conversation, facial signals

Introduction

Unlike most other animals, humans tend to face each other in everyday communication. This allows humans to rely not only on vocal but also on various visual bodily behaviors when communicating (1). While the language sciences have made substantial progress in the study of hand gestures (e.g., 2–6), there is one part of the body that has received comparatively little attention in this field, despite its omnipresence in and intuitive relevance for everyday face-to-face communication: the face.

It is well known that the face plays an important role in expressing emotions (7,8). *Facial expressions* are often considered to be rather involuntary manifestations of an individual's emotion (e.g., fear upon seeing a spider) and they have been distinguished from more voluntary *facial gestures* (5,9,10) (see also "conversational facial signals" and "facial displays", 11,12). Rather than being part of an individual-emotional process, facial gestures are considered to be part of a social-interactive process, not so much related to an individual's inner emotions but rather to the structure and content of conversation (13).

More recently, researchers are increasingly turning towards considering facial movements as *communicative* signals, for example in the context of depictions, where facial gestures can serve to "stage a scene" (14), such as when impersonating a particular character when telling a story (see also 15–17). And of course, gaze direction has long been acknowledged to play a fundamental role in signaling communicative intentions (e.g., 18,19).

In addition to blinks and gaze shifts, some of the most prevalent facial movements in conversation are eyebrow movements, such as eyebrow raises and furrows. According to the Facial Action Coding System (FACS) (20,21)—a system that allows for coding of visually distinguishable facial movements (termed Action Units [AU])—eyebrow raises are realized by the Inner Brow Raiser (Central Frontalis; AU1) together with the Outer Brow Raiser (Lateral Frontalis; AU 2) while eyebrow furrows are realized by the Brow Lowerer (Corrugator, Depressor Supercilli, Depressor Glabellae; AU4; see Figure 1).

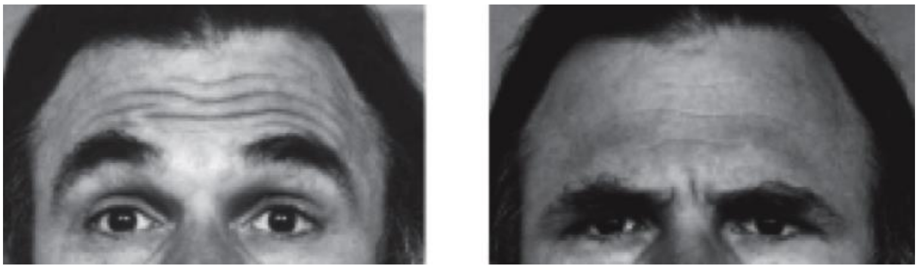

Figure 1. Example images of an eyebrow raise (AU 1+2) and an eyebrow furrow (AU 4) (Source: (21), copyright obtained by first author).

In the emotion domain, eyebrow raises have been associated with positive emotions like surprise, and eyebrow furrows with negative emotions like anger (8). In terms of non-emotional signaling, eyebrow movements have been thought to occur in requests for information from a conversational partner (7,11,22–24). Indeed, eyebrow position is a grammaticalized facial question marker in many sign languages (25–28). Specifically, eyebrow movements have been shown to fulfill an important conventionalized signaling function in signed languages in a particular type of question context—so-called 'other-initiated repair' (OIR) (29)—which is core to the process of "grounding" (30), i.e., the process of establishing the mutual belief that communicative acts have been understood well enough for current purposes (31,32). OIR is a brief exchange that momentarily interrupts the progress of a conversation to solve a communicative problem (33). An OIR sequence consist of a *repair initiation*, a signal from the addressee indicating a problem in perceiving or understanding what the speaker just said, and a *repair solution*, involving the speaker repeating part or all of the trouble source turn, clarifying certain parts of it, or confirming or disconfirming a candidate understanding offered by the addressee (34).

Judging from their linguistic functions in sign languages, eyebrow raises and furrows may also be normative practices in spoken OIR. While repair can be initiated and resolved with spoken language in the absence of the visual channel (think of speaking on the phone), in spoken face-to-face conversation eyebrow raises and furrows have also been observed in OIR contexts (35). An open question is whether eyebrow movements play a communicative role in initiating repair in spoken languages, or whether they might be epiphenomenal, that is, mere correlates or "ornaments" of verbal initiations without a signaling function in their own right.

In the present article, we address this issue, presenting two studies investigating the use and communicative consequences of eyebrow movements as potential signals of communicative problems in face-to-face interaction. In Study 1, we collected data from two corpora of face-to-face Dutch spoken conversations, coded the co-occurrence of eyebrow movements with different types of verbal signals of problems in hearing or understanding, the temporal relationship between the visual and verbal component in these multimodal signals, the type of solutions provided in response, as well as eyebrow movements alone that were treated as signals of problems in hearing or understanding. Study 2 tests the claim based on the correlational evidence from Study 1 that addressee brow furrowing alone may serve a communicative function in conversation, silently signaling “I have not received sufficient information for current purposes”. It does so by experimentally pinpointing communicative consequences that result from the perception of addressees’ brow furrows. Taken together, the two studies present converging evidence from a conversational corpus study and an experimental study suggesting that eyebrow movements indeed play a functional role in signaling communicative problems in spoken face-to-face interaction, and that speakers are sensitive to addressee brow furrows, apparently interpreting them as requests for clarification or elaboration in their own right.

Study 1

Eyebrow movements as signals of communicative problems in face-to-face conversation: A corpus study.

A few studies provide initial clues that eyebrow movements may not be epiphenomenal in spoken OIR (36–38). Comparing OIR sequences in unrelated spoken- and signed languages (Northern Italian, Cha’palaa, Argentine Sign Language), Floyd and colleagues (38) showed that if a repair initiation was accompanied by a bodily “hold”, that is, if body movements like eyebrow movements (but also, e.g., hand gestures or head movements) were “temporarily and meaningfully held static” (ibid., p. 1), this hold was often associated with communication problems and not disengaged from until the communication problem was solved. Floyd et al. interpreted these holds as displaying that a repair solution is still expected, whereas disengaging from a hold displays that a repair solution is no longer expected because one has been provided. This raises the possibility that disengaging from a brow position may play a communicative role in signaling successful grounding (i.e., through accomplishing closure of an OIR sequence), which may suggest that brow movements during repair *initiation* may be communicative too. Note that Floyd et al. (38) did not distinguish between different types of brow movements such as furrows versus raises, though. Furthermore, two individual descriptive examples—one from English (“raises her eyebrows, pulls down the corner of the mouth”; (37), p. 11) and one from Siwu (“puzzled look: furrowing of eyebrows”, (36), p. 238)—suggest that facial signals including eyebrow raises or furrows can be treated as repair initiations without relying on accompanying verbal material. While these studies suggest that eyebrow movements may serve a communicative role in initiating repair both in signed as well as spoken language, more systematic evidence for spoken language is needed. Further, little is known about the different compositions of repair initiations used in spoken language (e.g., verbal signal with versus without eyebrow movement) and about the functions of different types of eyebrow movements, such as brow raises and furrows.

Darwin (7) proposed in his principle of antithesis that two opposed movements are likely to develop distinct communicative functions. Eyebrow raises and furrows are formally opposed, constituting two maximally contrastive extremes of how eyebrows can move. They have distinct effects on vision (seeing more versus seeing less; (7,39)), and, as mentioned above, they have been associated with emotions of opposed valence (8). Assuming that eyebrow movements have a signaling function, this raises the question of whether eyebrow raises and furrows may also serve distinct communicative functions in signaling problems of perceiving or understanding in spoken face-to-face conversation. In Dutch Sign Language

(Sign Language of the Netherlands [NGT], (26)), eyebrow raises mark polar questions (e.g. 'you mean John?') and eyebrow furrows mark content questions (e.g., 'who?'). If the non-obligatory use of eyebrow actions in information requests in spoken Dutch is akin to the grammatically obligatory use of eyebrow actions in requests for information in Dutch Sign Language, one may expect that in spoken Dutch, eyebrow raises may be more often involved in repair initiations that make confirmation or disconfirmation relevant (e.g., 'you mean John?') and eyebrow furrows more often in repair initiations that make clarification relevant (often including content question words, e.g., 'who?'). While the type of brow movement involved in a multimodal repair initiation may affect which type of repair solution is provided in response, the mere presence of the brow movement and the timing of the brow movement relative to the verbal signal in repair initiations may affect the speed by which a repair solution is provided. If the brow movement is initiated before the verbal signal, it may "forewarn" the speaker about a communicative problem, providing the speaker with a timing advantage when planning an appropriate response (see also 40 on how turn-opening frowns can anticipate utterances involving some kind of trouble, e.g., epistemic challenge).

In Study 1, we hypothesize that eyebrow actions contribute to signaling problems of perceiving or understanding in spoken languages just as they do in sign languages, on the grounds that spoken languages also strongly rely on the visual channel, at least in face-to-face contexts (e.g., 3,5,41–43). We also hypothesize that eyebrow raises and furrows may serve different functions in signaling problems of perceiving or understanding. Specifically, we predict

1) the type of eyebrow action used with verbal repair initiations to be associated with the type of repair solution provided in response (e.g., confirmation vs. clarification),
2) repair time to be reduced by the presence of an eyebrow action or by an eyebrow action produced as a preliminary to verbal repair initiations, and
3) addressee eyebrow actions alone, that is, silently produced in the absence of on-record verbal repair initiations, to also occasion repair.

To address these issues, we used two corpora of dyadic Dutch face-to-face conversations, which were specifically designed for detailed analyses of facial behavior[1]. We identified OIR sequences in conversations and coded the compositionality of repair initiations, focusing on eyebrow raises and furrows. We then quantified the co-occurrence of different linguistic formats of verbal repair initiations with eyebrow raises and furrows, the temporal relationship between the visual and the verbal component in the multimodal repair initiations, and investigated whether the presence of eyebrow actions in general and early eyebrow actions in particular (produced as preliminaries to verbal repair initiations) speeded up the repair process. Finally, we identified silently produced addressee eyebrow raises and furrows that were treated as making relevant repair despite the absence of on-record verbal repair initiations.

Methods

[1] One limitation of previous studies on eyebrow movements was that the corpora they used were not suitable for detailed analyses and quantification of facial signals. As Kendrick (37) notes, "In some cases, the relevant participant is off-camera or his or her face cannot be seen due to the angle of the camera" (p.11). Floyd et al. (38) pointed out that "speakers' faces were not always clearly visible in the video" (p.190).

Participants and corpora

We used two corpora of spontaneous, dyadic Dutch face-to-face conversations: the IFA Dialog Video Corpus (IFADV, (44)) and the purpose built corpus of Dutch Face-to-Face (DF2F) conversation (see also 45). Both corpora were specifically designed to allow for detailed analyses of communicative facial behavior.

The IFADV Corpus consists of 23 dyads, all native Dutch speakers (12-72 years) who knew each other well prior to the recording. Nine of the dyads consisted of a female and a male participant, 11 were all female, and three were all male. Five of the participants participated in two dyads each. The dyads were engaged in spontaneous Dutch face-to-face conversations for 15 minutes each. Conversations were recorded in a soundproof room and participants were seated at a table, facing each other, positioned approximately 1 m from each other (see Supplementary Material 1). Two video cameras (JVC TK-C1480B, 720x576, 25 fps) were used to record frontal views of each participant and audio was recorded using head-mounted microphones (Samson QV).

The DF2F corpus consists of 10 dyads, all native Dutch speakers (18–68 years) who knew each other well prior to the recording. Four of the dyads consisted of a female and a male participant, four were all female, and two were all male. Each participant participated only in one dyad. The dyads were engaged in casual Dutch face-to-face conversations for 1 hour each. The recordings took place at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands, in a soundproof room. Participants were positioned approximately 1 m from each other at a 45-degree angle. Three HD video cameras (JVC GY-HM100) were used to record frontal views of each participant (see Supplementary Material 2) and a scene view. Audio was recorded using lightweight head-mounted microphones (DPA-d:fine-88) and an audio recorder (Roland R-44) recorded the two audio tracks in synchrony. Each recording session resulted in three videos and two audio files, which were then synchronized and exported in Adobe Premier Pro CS6 (MP4, 24 fps). Each recording session consisted of three 20-minute phases: During one 20-minute phase, participants did not wear the head-mounted microphones and audio was only recorded using a ceiling microphone. During a second 20-minute phase, audio was recorded using the head-mounted microphones, and during a third 20-minute phase, audio was recorded using the head-mounted microphones and, in addition, participants wore eye-tracking glasses. To achieve the highest audio quality and to allow for detailed analyses of facial behavior (without potential occlusion of or interference with facial behavior related to wearing eye-tracking glasses), only the 20-minute phase in which participants wore head-mounted microphones was used for this study. Each participant was paid 16 euros for the whole session which lasted about 90 minutes. The study was approved by the Social Sciences Faculty Ethics Committee, Radboud University Nijmegen, and informed consent was obtained before and after filming.

Analysis

We identified occurrences of other-initiated repair and eyebrow raises and furrows, sampling from randomly selected 10-minute segments in the IFADV corpus (one segment per dyad, resulting in 230 minutes) and from naturally occurring tellings (46) in the DF2F corpus (all tellings in all dyads, resulting in 68 minutes), resulting in a total of 298 minutes of conversation. The choice to sample from randomly selected segments in the IFADV corpus and tellings in the DF2F corpus was based on practical considerations. OIR cases were already partially coded in the IFADV corpus (by PH) and brow movements were already partially coded in tellings of the DF2F corpus (by PH) [and we had no reason to assume systematic differences in the use of eyebrow raises and furrows between these two types of selected conversational materials]. All annotations were created in ELAN 4.8.1 (47).

Verbal other-initiated repair. We first focused the analysis on verbal cases of other-initiated repair, i.e., sequences "in which a turn T0 signals some trouble in a prior turn T-1 and is treated as making relevant the provision or ratification of a repair solution in a next turn T+1" (48, p. 99). For each OIR case, the linguistic format of the verbal repair initiation as well as non-mutually exclusive characteristics of the verbal repair solution were coded. Three basic formats of repair initiations were distinguished (48). A repair initiation was coded as (1) *open request* if it targeted the prior turn as a whole (e.g., *huh?*, *what?*), typically making repetition relevant but sometimes also clarification, (2) as a *restricted request* if it targeted a specific aspect of the prior turn (e.g., *who?*), making clarification of this aspect relevant, and (3) as *restricted offer* if it targeted a specific aspect of the prior turn by offering a candidate understanding (e.g., *you mean John?*), making confirmation or disconfirmation relevant. For each repair solution, it was coded whether any material from the trouble source turn was (1) repeated, (2) clarified, or whether (3) it included a confirmation or disconfirmation (non-mutually exclusive options). A repair solution was coded as 'repeating' if some or all material from the trouble source turn was repeated (49), not taking into account whether 'dispensable' items such as a turn-initial *but* or *oh* (50) were omitted or not. A repair solution was coded as 'clarifying' if it involved modification or specification of the trouble source (51), that is, if some or all material from the trouble source was rephrased, replaced, or if something was added. A repair solution was only coded as '(dis)confirming' if it included a "'yes/no/indeed' type item, a head nod/shake, or a repetition (+/- negation)" (34, p. 42), often produced in response to an offered candidate understanding as part of the repair initiation (33,50). Note that a repair solution was not coded as '(dis)confirming' if it included an indirect (dis)confirmation, for example, by offering an alternative.

Criteria for identifying and classifying OIR cases were based on a coding scheme developed by (34). All repair sequences were identified by the first author (PH) experienced in the application of this coding scheme and resulted in a total of 116 OIR cases. Thus, there was a repair initiation about once every 2.5 minutes. This frequency is lower than the frequency of once every 1.6 minutes previously reported based on a large-scale cross-linguistic study of OIR (52). While both studies focused on maximally informal conversations suggesting a similar amount of shared knowledge among participants, this difference in frequency may be due to the fact that participants in one of the studies (52) were often engaged in parallel activities such as preparing food, eating, or playing games, potentially leading to more problems in hearing or understanding due to background noises and distractions. In contrast, both corpora used for the current study were recorded in soundproof laboratories with little to no background noises or visual interference, let alone opportunities for potentially distracting parallel activities.

Eyebrow actions. We identified eyebrow raises and furrows (see Facial Action Units 1+2 and 4, respectively; (20)), annotated from the first to the last visible movement of the eyebrows. Eyebrow actions were identified by two independent coders (KK and MK) who were blind to the hypotheses. Twelve minutes were coded for training and 59 randomly selected minutes (approx. 20% of the total data) were coded for measuring inter-rater reliability. The inter-rater reliability was 76.5 % for brow action occurrence and a Cohen's Kappa (53,54) of .88 was achieved for agreement about the brow action type (brow furrow versus brow raise) indicating substantial agreement.

Compositionality of repair initiations. For each repair sequence, firstly, we assessed whether the verbal repair initiation co-occurred with eyebrow actions or not. Secondly, if verbal repair initiations co-occurred with eyebrow actions, we assessed the temporal relationship between the visual and the verbal component. Eyebrow actions were considered to be "co-occurring" if the eyebrow action temporally overlapped with a verbal repair initiation. Eyebrow actions were also considered to be "co-occurring" if the offset of the eyebrow action immediately preceded the onset of the verbal repair initiation without perceptible interruption or if the onset of the eyebrow action immediately followed the offset of the verbal repair initiation without perceptible interruption, such that the behaviors together formed a multimodal *Gestalt* (55). More precisely, if the onset of the verbal repair initiation and the onset of the eyebrow action coincided precisely or if the onset of one preceded the onset of the other by less than 200 ms (up to which it is likely perceived as synchronous, as has been established for visible lip movements and articulatory sound (56), this was coded as "initiated simultaneously". If the onset of the eyebrow action preceded the onset of the verbal repair initiation by more than 200 ms this was coded as 'initiated visually first' (or 'verbal OIR with visual preliminary', see Results section below), and if the onset of the verbal repair initiation preceded the onset of the eyebrow action by more than 200 ms, it was coded as 'initiated verbally first'.

Eyebrow actions occasioning repair in the absence of vocalization. Finally, when eyebrow actions alone were sufficient to occasion repair, that is, without any 'on-record' verbal repair initiation (e.g., 37), they were coded as 'eyebrow actions only occasioning repair'.

Statistical Analysis. First, to statistically test whether the *presence of eyebrow action* (verbal-only repair initiation, verbal repair initiation with eyebrow action) in repair initiations is associated with the linguistic *format of the verbal repair initiation* (open request, restricted request, restricted offer), a mixed effects logistic regression analysis was performed (including random intercepts for participants). An intercept-only model was compared to a model in which 'presence of eyebrow action' was added as a predictor variable, using a Likelihood Ratio Test.

Second, to test whether the *type of eyebrow action* (verbal repair initiation with eyebrow raise, verbal repair initiation with eyebrow furrow) in repair initiations is associated with the linguistic *format of the verbal repair initiation* (open request, restricted request, restricted offer), an additional mixed effects logistic regression analysis was performed (including random intercepts for participants). Again, an intercept-only model was compared to a model in which 'type of eyebrow action' was added as a predictor variable, using a Likelihood Ratio Test.

Third, to test whether the *presence or type of eyebrow action* in repair initiation predict the *type of solution* provided, we correlated the composition of the repair initiation (verbal-only repair initiation, verbal repair initiation with eyebrow raise, verbal repair initiation with eyebrow furrow) with different non-mutually exclusive characteristics of the subsequent repair solution, namely whether any material from the trouble source turn was repeated, clarified, or whether it included a confirmation or disconfirmation. Note that this analysis could not be applied to six OIR cases in which the T+1 was absent.

Fourth, we used a mixed effects logistic regression analysis (including random intercepts for participants) to test whether the *composition of the repair initiation* (verbal-only repair initiation, verbal repair initiation with eyebrow raise, verbal repair initiation with eyebrow furrow) predicts whether the *repair solution* included a clarification or not (clarification, no clarification), while taking into account variability in the linguistic format of the verbal repair initiation (open request, restricted request, restricted offer; see Figure 1) by adding it as a predictor variable to the statistical model. This model was compared to a reduced model without the predictor variable of 'composition of repair initiation' using a Likelihood Ratio Test.

Fifth, we tested in a mixed-effects model whether *repair time* differed between verbal repair initiations *with versus without a brow action*, while taking into account variability in the linguistic format of the verbal repair initiation by adding it as a predictor variable to the statistical model. We entered 'linguistic format' (open request, restricted request, restricted offer) and 'presence of brow action' (yes, no) as fixed effects and intercepts for participants as a random effect into the model. This model was compared to a reduced model without 'presence of brow action' as a fixed effect using a Likelihood Ratio Test.

Finally, we tested in a mixed-effects model whether *repair time* differed between verbal-only repair initiations versus repair initiations with a concurrent eyebrow versus an early eyebrow action, while taking into account variability in the linguistic format of the verbal repair initiation by adding it as a predictor variable to the statistical model. We entered 'linguistic format' (open request, restricted request, restricted offer) and 'brow action' (verbal-only repair initiation, repair initiation with early brow action, repair initiation with concurrent brow action) as fixed effects and intercepts for participants as a random effect into the model. This model was compared to a reduced model without 'brow action' as a fixed effect using a Likelihood Ratio Test.

Results

We focus the analysis first on the compositionality of repair initiations, assessing the co-occurrence of different eyebrow actions with different linguistic formats of the verbal repair initiation. We then explore the corpus-based plausibility of whether eyebrow actions might merely be epiphenomena of verbal repair initiations or whether they may contribute to signaling problems in hearing or understanding by examining (1) whether the type of eyebrow action accompanying repair initiation predicts certain types of repair solutions, even after taking into account variability in the co-occurring verbal repair initiation format, (2) whether the presence of an eyebrow action as a preliminary to repair initiation speeds up the repair process, and finally, (3) whether eyebrow actions alone are sufficient to occasion repair.

Initiating repair with words and brows

Out of all identified verbal repair initiations (N=116), a substantial number co-occurred with eyebrow actions (40% [n=46]). Out of those co-occurring with eyebrow actions, about half co-occurred with eyebrow raises (46% [n=22]) and the other half with eyebrow furrows (54% [n=25]).

Which composition (verbal-only repair initiation, verbal repair initiation with eyebrow raise, or verbal repair initiation with eyebrow furrow) co-occurred with which linguistic format of the verbal repair initiation (open request, restricted request, restricted offer)? As can be see in Figure 2, restricted offer was the overall most frequent linguistic format of the verbal repair initiation (68% [n=72]), followed by restricted request (24% [n=25]) and open request

(8% [n=9]). While the distribution of linguistic formats of the verbal repair initiation is almost identical when considering just verbal-only repair initiations (restricted offer: 72% [n=44]; restricted request: 18% [n=11]; open request: 10% [n=6]) and just verbal repair initiations with eyebrow raises (restricted offer: 71% [n=15]; restricted request: 19% [n=4]; open request: 10% [n=2]), verbal repair initiations with eyebrow furrows show a lower proportion of restricted offers (54% [n=13]) and open requests (4% [n=1]), but a substantially higher proportion of restricted requests (42% [n=10]), relative to verbal-only repair initiations and verbal repair initiations with eyebrow raises[2].

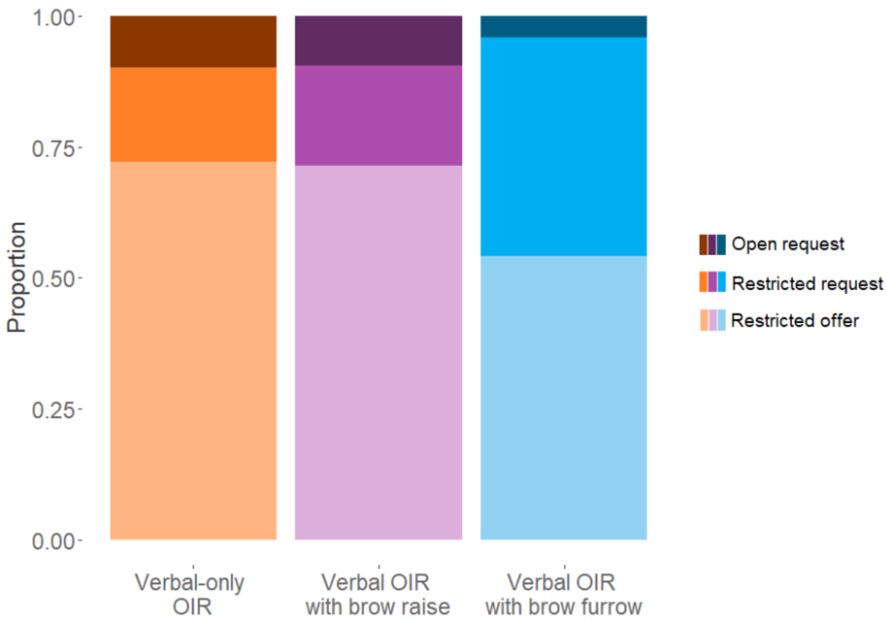

Figure 2. Compositionality of repair initiations (verbal-only OIR, verbal OIR with eyebrow raise, verbal OIR with eyebrow furrow) by linguistic format of the verbal repair initiation (open request, restricted request, restricted offer; N=106[2]).

To test whether the *presence* of eyebrow action in repair initiations (verbal-only repair initiation, verbal repair initiation with eyebrow action) is statistically associated with the linguistic format of the verbal repair initiation (open request, restricted request, restricted offer), a mixed effects logistic regression analysis was performed (including random intercepts for participants). An intercept-only model was compared to a model in which 'presence of eyebrow action' was added as a predictor variable, using a Likelihood Ratio Test. Including 'presence of eyebrow action' did not improve the model fit significantly ($\chi^2(2) = 2.56$, $p = .276$), indicating that the presence of eyebrow action *per se* did not reliably distinguish between the linguistic format of the verbal repair initiation. To test whether the *type* of eyebrow action (verbal repair initiation with eyebrow raise, verbal repair initiation

[2] Note that two rare linguistic formats of restricted OIR were excluded: alternative questions (invites a selection from among alternatives; n(with raise)=0, n(with furrow)=0, n(verbal without eyebrow action)=2) and external repair initiations (address problems about unexpressed elements of T-1; n(with raise)=1, n(with furrow)=1, n(verbal without eyebrow action)=6), resulting in a total of 106 OIR cases.

with eyebrow furrow) in repair initiations is associated with the linguistic format of the verbal repair initiation (open request, restricted request, restricted offer), an additional mixed effects logistic regression analysis was performed (including random intercepts for participants). An intercept-only model was compared to a model in which 'type of eyebrow action' was added as a predictor variable, using a Likelihood Ratio Test. Including 'type of eyebrow action' did not improve the model fit significantly ($\chi^2(2) = 1.88$, $p = .389$), indicating that the type of eyebrow action did not reliably distinguish between the linguistic format of the verbal repair initiation.

The results reveal that a substantial number of verbal repair initiations are accompanied by eyebrow actions—about as often by eyebrow raises as by eyebrow furrows. Furthermore, the results numerically mirror the hypothesized pattern based on question marking in Dutch sign language that eyebrow raises may be more often involved in repair initiations that make confirmation or disconfirmation relevant (restricted offers like 'You mean John Smith?') and eyebrow furrows more often in repair initiations that make clarification relevant (like restricted requests such as 'Who?', (26)): relative to eyebrow furrows, a larger proportion of eyebrow raises accompanied restricted offers, and relative to eyebrow raises, a larger proportion of eyebrow furrows accompanied restricted requests. However, these differences were not statistically significant.

The example below (Example 1) illustrates how an eyebrow raise may be used with a restricted offer, which is subsequently confirmed through a head nod:

```
Example 1: DF2F corpus_19_266591

1 A:  Hij heeft nu de vriendin van Boris
      he has now the girlfriend of Boris
      He now has Boris' girlfriend

2 B:  Ja m- (.) ((raises brows, see Figure 3)) Jeanette?
      yeah m-                                  Jeanette?
      yeah m-                                  Jeanette?

3 A:  ((nods))
```

Figure 3. Eyebrow raise produced with a restricted offer as linguistic format of the verbal repair initiation (‘Jeanette?’, line 2 in the example above).

By contrast, Example 2 illustrates how an eyebrow furrow may be used with a restricted request, in this case for clarification of an underspecified person reference, which is subsequently provided[3].

```
Example 2: IFADV_17_588780

1 A: Ik ben dus achternichtje met Marieke
     I  am  thus second cousin with Marieke
     It turns out I’m second cousin of Marieke

2 B: ((furrows brows, see Figure 4)) Marieke, wie is Marieke?
                                     Marieke, who is Marieke?
                                     Marieke, who is Marieke?

3 A: Ja, die ene van de Kleinkunst
     Yeah, the one from the cabaret
     Yeah, the one from cabaret
```

Figure 4. Eyebrow furrow produced with a restricted request as linguistic format of the verbal repair initiation (‘who is Marieke?’, line 2 in the example above).

Do eyebrow actions contribute to initiating repair in spoken face-to-face conversation? On the one hand, the co-occurrence of eyebrow actions and verbal repair initiations suggests they may serve similar functions, that is, they may be co-expressive in signaling problems in hearing or understanding. On the other hand, it raises the question whether the eyebrow actions in these cases might be epiphenomenal, and thus merely correlates but not functionally involved in signaling problems in hearing or understanding. Below, we present three pieces of evidence suggesting that eyebrow actions are indeed effective in signaling problems in hearing or understanding: First, we show that verbal repair initiations with eyebrow furrows are more likely to get clarifications as repair solutions compared to either

[3] See (36), Extract 5, for a strikingly similar example including eyebrow furrowing in Siwu, an African language spoken in a small community in eastern Ghana.

verbal repair initiations with eyebrow raises or verbal-only repair initiations, even after taking into account variance explained by the linguistic format of the co-occurring verbal repair initiation. This points to a potentially unique contribution of eyebrow furrows in signaling a need for clarification (relative to either verbal repair initiations with eyebrow raises or verbal-only repair initiations). Secondly, we show that, relative to repair initiations without eyebrow actions, repair initiations that were immediately preceded by eyebrow actions as preliminaries get repaired faster. Finally, and most importantly, we show that eyebrow furrows alone can be sufficient to occasion clarification. We take up these three lines of evidence in order:

(1) Does the presence or type of eyebrow action in repair initiation predict the type of solution provided? To address this question, we correlated the composition of the repair initiation (verbal-only repair initiation, verbal repair initiation with eyebrow raise, verbal repair initiation with eyebrow furrow) with different non-mutually exclusive characteristics of the subsequent repair solution, namely whether any material from the trouble source turn was repeated, clarified, or whether it included a confirmation or disconfirmation. Note that this analysis could not be applied to six OIR cases in which the T+1 was absent (N=100)[4]. Repair solutions in response to verbal-only repair initiations were slightly more likely to include repetitions (38% [n=25]) than repair solutions in response to verbal repair initiations with eyebrow raises (33% [n=7]), and slightly less likely than repair solutions in response to verbal repair initiations with eyebrow furrows (42% [n=10]). Repair solutions in response to verbal-only repair initiations were more likely to include (dis)confirmation (71% [n=46]) than repair solutions in response to verbal repair initiations with eyebrow raises (57% [n=12]), and slightly less likely than repair solutions in response to verbal repair initiations with eyebrow furrows (50% [n=12]). As one can see in Figure 4, repair solutions in response to repair initiations with eyebrow furrows were more than twice as likely to include clarification (65% [n=15]) relative to repair solutions in response to verbal-only repair initiations (31% [n=18]) and repair initiations with eyebrow raises (25% [n=5]).

[4] Out of the six OIR cases in which the T+1 was absent, four were verbal-only repair initiations, one with a brow furrow, and one with a brow raise. That is, multimodal repair initiations were ignored less often, potentially suggesting that repair initiations are "weaker" if they are not produced with eyebrow actions (but bear in mind the extremely small number of cases).

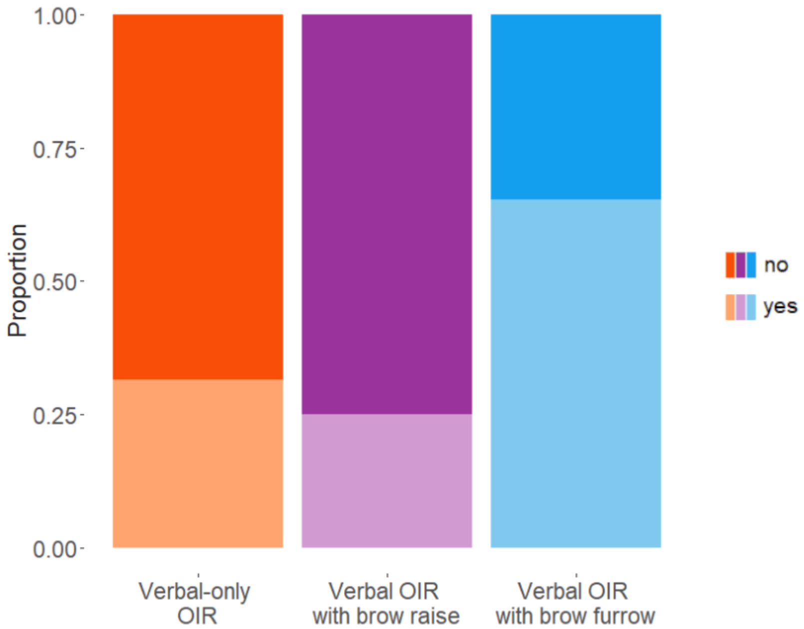

Figure 5. Compositionality of OIR (verbal-only OIR, verbal OIR with eyebrow raise, verbal OIR with eyebrow furrow) by repair solution (YES = with clarification, NO = without clarification), N=100.

Note, however, that this is not necessarily a unique contribution of eyebrow furrows. One might argue that given that eyebrow furrows are more frequent in restricted requests (see Figure 1), it is not surprising that repair solutions in response to repair initiations with furrows are more likely to include clarification. The linguistic format of the verbal repair initiation, in this case 'restricted request', rather than the accompanying eyebrow furrow, may thus underlie the increased likelihood for repair solutions to include clarifications. To explore this possibility, we used a mixed effects logistic regression analysis (including random intercepts for participants) to test whether the composition of the repair initiation (verbal-only repair initiation, verbal repair initiation with eyebrow raise, verbal repair initiation with eyebrow furrow) predicts whether the repair solution included a clarification or not (clarification, no clarification), while taking into account variability in the linguistic format of the verbal repair initiation (open request, restricted request, restricted offer; see Figure 1) by adding it as a predictor variable to the statistical model. This model was compared to a reduced model without the predictor variable of 'composition of repair initiation' using a Likelihood Ratio Test. Including 'composition of repair initiation' improved the model fit significantly ($\chi^2(2) = 7.85$, $p<.05$), revealing that, relative to repair initiations with eyebrow raises, repair initiations with eyebrow furrow changes the log odds of a subsequent repair solution including clarification by 1.66 ± 0.71 (standard error); relative to repair initiations without eyebrow actions, repair initiations with eyebrow furrow changes the log odds of a subsequent repair solution including clarification by 1.38 ± 0.55 (standard error). These results indicate that, independently of the linguistic format of the repair initiation, the presence of an eyebrow furrow increased the likelihood of a repair initiation to be treated as a request for clarification.

(2) Do eyebrow actions speed up the repair process? If eyebrow actions were merely a correlate of verbal repair initiation—say a *symptom* of cognitive effort—rather than a communicative *signal* of a problem in hearing or understanding, one should expect the repair time, measured from the end of the repair initiation to the start of the repair solution, to be unaffected by whether the repair initiation was produced with or without an eyebrow action. Alternatively, if eyebrow actions can indeed function as a communicative *signal* of a problem in hearing or understanding, one may expect that—by increasing redundancy—the presence of an eyebrow action per se may reduce potential ambiguity and express a stronger sense of urgency, which may reduce the repair time. To address this issue, we compared the repair time between verbal repair initiations without versus with a brow action (see Figure 6).

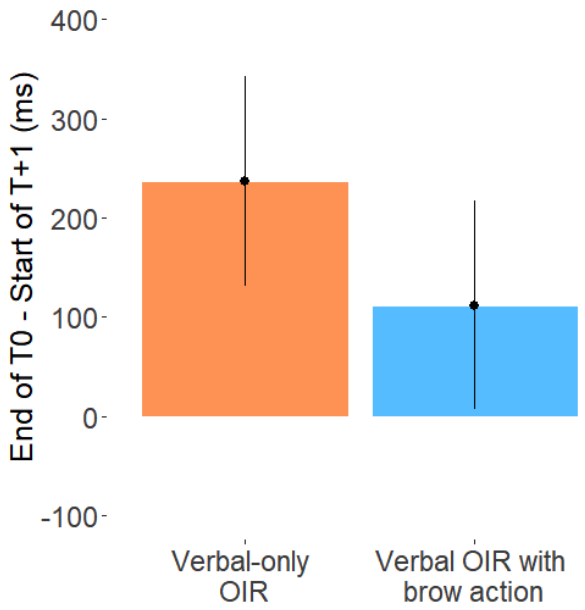

Figure 6. Repair time, measured from the end of the repair initiation (T0) until the start of the repair solution (T+1), by repair initiation *without* an eyebrow action ('Verbal-only OIR') versus *with* an eyebrow action ('Verbal OIR with brow action'). Error bars represent Standard Error.

We used R and *lme4* (57) to test in a mixed-effects model whether repair time differed between verbal repair initiations with versus without a brow action, while taking into account variability in the linguistic format of the verbal repair initiation by adding it as a predictor variable to the statistical model. We entered 'linguistic format' (open request, restricted request, restricted offer) and 'presence of brow action' (yes, no) as fixed effects and intercepts for participants as a random effect into the model. This model was compared to a reduced model without 'presence of brow action' as a fixed effect using a Likelihood Ratio Test. Including 'presence of brow action' as a fixed effect improved the model fit marginally ($\chi^2(1) = 336$, $p = .066$), revealing that—relative to repair initiations *without* a brow action (376.11 ms [mean] ± 142.48 [standard error])—the repair time for repair initiations *with* a brow action was shorter by about 135.73 ms (mean) ± 75.13 (standard error), but not reliably so ($t = -1.806$, $p = .074$). Note that adding 'brow action type' (brow raise, brow furrow) as a predictor to the statistical model did not improve the fit ($\chi^2(1) = 0.60$, $p = .436$), indicating that repair time was unaffected by the type of brow action involved.

Timing of eyebrow action: Verbal-only repair initiation versus repair initiation with concurrent versus early eyebrow action. If eyebrow actions can indeed function as a communicative *signal* of a problem in hearing or understanding, one may expect that an early eyebrow action produced as a visual preliminary, a potential visual "forewarning", may facilitate a timely response, thus reducing the repair time. To address this issue, we examined the temporal relationship between the visual and the verbal component in multimodal repair initiations (see Methods section) and then compared the repair time between *verbal-only* repair initiations, verbal repair initiations with a *concurrent* eyebrow action versus verbal repair initiations with an *early* eyebrow action (produced as a visual preliminary to the verbal repair initiation; see Figure 7).

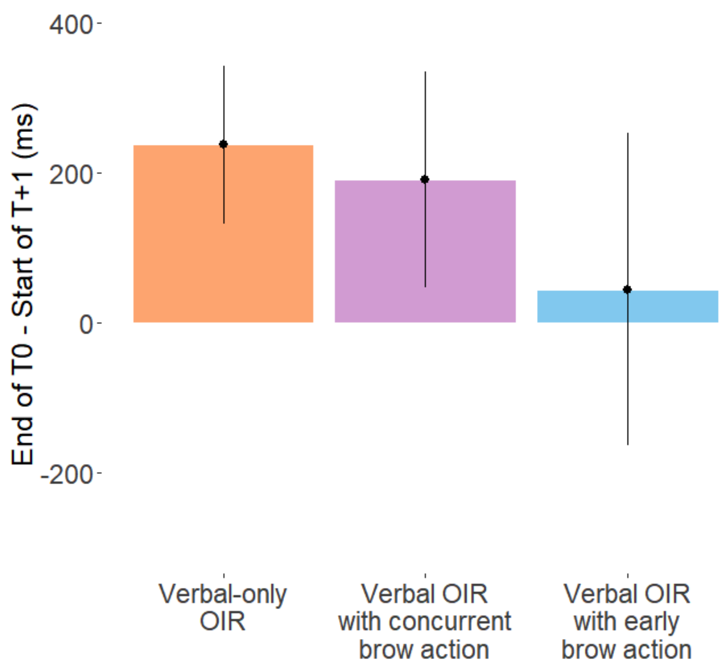

Figure 7. Repair time, measured from the end of the repair initiation (T0) until the start of the repair solution (T+1), by *verbal-only* repair initiations ('Verbal-only OIR') versus repair initiations with a *concurrent* eyebrow action ('Verbal OIR with concurrent brow action') versus an *early* eyebrow action (produced as a visual preliminary to the verbal repair initiation, 'Verbal OIR with early brow action'). Error bars represent Standard Error.

We tested in a mixed-effects model whether repair time differed between verbal-only repair initiations versus repair initiations with a concurrent eyebrow versus an early eyebrow action, while taking into account variability in the linguistic format of the verbal repair initiation by adding it as a predictor variable to the statistical model. We entered 'linguistic format' (open request, restricted request, restricted offer) and 'brow action' (verbal-only repair initiation, repair initiation with early brow action, repair initiation with concurrent brow action) as fixed effects and intercepts for participants as a random effect into the model. This model was compared to a reduced model without 'brow action' as a fixed effect using a Likelihood Ratio Test. Including 'brow action' as a fixed effect improved the model fit marginally ($\chi^2(3)$

= 6.49, p = .089). This revealed that the repair time for verbal repair initiations with a *concurrent* brow action was shorter by about 42.08 ms (mean) ± 92.95 (standard error), compared to *verbal-only* repair initiations (413.05 ms [mean] ± 141.79 [standard error]), but that this difference was not reliable (t = -0.453, p = .651). Also, the repair time for verbal repair initiations with an *early* brow action was shorter by about 189.58 ms (mean) ± 124.45 (standard error), compared to verbal repair initiations with a *concurrent* brow action (370.97 ms [mean] ± 165.56 [standard error])—but, again, not reliably so (t = -1.523, p = .131). However, relative to *verbal-only* repair initiations, the repair time for verbal repair initiations with an *early* brow action was significantly reduced by about 231.66 ms (mean) ± 107.32 (standard error), t = -2.159, p = .033. Note that adding 'brow action type' (brow raise, brow furrow) as a predictor to the statistical model did not improve the fit ($\chi^2(1) = 0.54$, p = .462).

We have seen that, while the mere presence of a brow action did not reliably speed up the repair process, the presence of an early brow action produced as a visual preliminary reduced the repair time significantly, compared to verbal-only repair initiations, again suggesting that eyebrow actions are effective in signaling problems of hearing or understanding.

(3) Can eyebrow actions alone signal problems in hearing or understanding? To address this question, we identified all silently produced eyebrow actions that occasioned repair. This resulted in eleven identified eyebrow furrows and zero eyebrow raises. None of these eyebrow furrows were treated as making (dis)confirmation relevant but *all* of them were treated as making clarification relevant (while three of these were also treated as making partial repetition relevant). Despite these observations resting on a small number of cases, the result quite convincingly suggests that eyebrow furrows alone can be sufficient as signaling a need for clarification, even in the absence of verbal repair initiations. Example 3 illustrates how an eyebrow furrow alone can occasion repair, as if it was a restricted verbal request for clarification.

```
Example 3: ETC13_151369

1 B: Ik heb het in mijn telefoon staan
     I have it  in my telephone stand
     I have it on my phone

2    in een vroeger bericht [van Floortje
     in an  earlier message from Floortje
     in an  earlier message from Floortje

3 A:                        [((furrows brows, see Figure 8))

4 B: hoe ze heet
     how she called
     what her name is

5 A: Ja ((unfurrows brows))
     yes
     yes
```

Figure 8. Eyebrow furrow alone occasioning clarification (‘what her name is’, line 4 in the example above; see video in Supplementary Material 4)

In the example above, B targets *it* (‘het’, line 1) as the trouble source by clarifying what *it* referred to through a repair *what her name is* (‘hoe ze heet’, line 4). As such, without any on-record verbal prompting, A’s eyebrow furrow was treated as if A had produced a verbal restricted request like *What do you have on your phone?* (“Wat heb je in je telefoon staan?”).

Discussion: Study 1

Do eyebrow movements serve a communicative function in signaling problems of hearing or understanding in spoken conversation? The present findings suggest they do indeed. The results are incompatible with an epiphenomenal interpretation of eyebrow movements, because (1) in addition to the linguistic format of the verbal repair initiation, the type of co-occurring eyebrow movement independently predicted the type of repair solution, (2) the presence of an eyebrow movement as a visual preliminary to verbal repair initiations enhanced repair speed, and (3) eyebrow movements alone were sufficient to occasion clarification.

First, we have seen that eyebrow raises and furrows were both used with all three basic linguistic formats of repair initiation, whether the co-occurring repair initiation targeted the prior turn as a whole (open request), a specific aspect of it (restricted request), or whether the repair initiation offered a candidate understanding (restricted offer). A higher proportion of eyebrow furrows co-occurred with restricted requests (‘who?’) relative to repair initiations with eyebrow raises, and a higher proportion of eyebrow raises co-occurred with restricted offers (‘John Smith?’) relative to repair initiations with eyebrow furrows—a numerical pattern that parallels the linguistic function of eyebrow position in Dutch Sign Language, where eyebrow furrows serve as non-manual grammatical markers of content questions and eyebrow raises as non-manual grammatical markers of polar questions (26). Bear in mind, however, that these numerical differences were not statistically significant. Repair initiations without eyebrow actions and repair initiations with eyebrow raises showed an almost identical distribution regarding restricted offers, potentially pointing to a higher optionality of the use of eyebrow raises in polar questions as repair initiations in spoken face-to-face conversation.

Second, we have also seen that the type of eyebrow movement co-occurring with repair initiations predicted differences in how these multimodal signals of problems were treated as making relevant different solutions. The presence of an eyebrow furrow uniquely—i.e., independently of the linguistic format of the verbal repair initiation—increased the likelihood of a repair initiation to get a repair solution including clarification. This suggests that the visual component is not merely a correlate of the verbal component, but that the visual and the verbal complement on another in multimodal repair initiations. More generally, while eyebrow movements are not always necessary for initiating repair in spoken conversation, this result suggests that they can nevertheless serve a communicative function, contributing to signaling the type of communicative problem and how it can best be fixed.

Third, we have seen that the presence of eyebrow movements as visual preliminaries to repair initiations reduced repair time, relative to repair initiations without eyebrow movements as visual preliminaries (i.e., verbal-only repair initiations)[5]. The eyebrow movement as a preliminary to a repair initiation here seems to serve a similar signaling function for a speaker as the orange light as a preliminary to the red light in traffic lights for a driver. While the speaker can speak through the addressee's eyebrow movement as the driver can drive through the orange light—both without being sanctioned—these preliminaries seem to facilitate a timely response, serving as 'forewarnings' of an upcoming disruption of progress. As such, this result is in line with findings from other domains of human joint action in which 'making oneself predictable' facilitates coordination (e.g., 58). More generally, it again suggests that eyebrow movements are effective in the context of initiating repair and it illustrates how visual signals may enhance communicative efficiency in spoken languages (see also 59, on questions getting faster responses if accompanied by gesture).

Fourth, we have seen that eyebrow furrows alone can silently signal insufficient understanding. This result suggests that while off-record facial action like the eyebrow furrow is usually not considered to be part of turn-constructional units in the human turn-taking system (60), it can serve sequentially equivalent functions as verbal repair initiations. As Levinson (61) noted, "Words and deeds are the same kind of interactional currency" (p. 74). The eyebrow furrow could be considered an implicit or off-record type of other-initiation of repair. While an eyebrow furrow seems slightly more accountable than a 'freeze look' (29), it still does not explicitly encode the intention to initiate repair—potentially in an effort to minimize any possible "face-threatening" consequences (62)—"just as "It's cold in here" does not explicitly encode the intention to get somebody to shut the window" (63, p. 11).

Moreover, purely visible bodily behaviors used to initiate repair have previously been classified as open requests (equivalent to e.g., *huh?*) as they do not explicitly target specific aspects of the trouble source but the trouble source as a whole, which is typically treated as making repetition relevant (e.g., 52). Accordingly, if eyebrow furrows served as open requests, one would have expected them to be typically treated as making repetition relevant. However, eyebrow furrows were treated specifically as making clarification relevant—even when not combined with a verbal repair initiation—suggesting that they may specifically target certain aspects of the prior turn as in need of clarification. How could eyebrow furrows

[5] Note that some of these results on repair time were based on a relatively small sample size (e.g., n[Verbal OIR with early brow action]=18). Further research validating its generalizability would be desirable.

possibly target specific aspects of a prior turn to be clarified? One possibility is that if the eyebrow furrow as such signals a need for clarification, it may be easy to guess for the speaker, based on estimates of shared knowledge, which aspect of the prior turn needs clarification (e.g., an underspecified person reference). Alternatively, as with visual addressee signals more generally, since they do not interfere as much with the spoken turn as verbal addressee signals, a specific troublesome aspect of a turn cannot only be targeted through explicit verbal means (e.g., *Who?*) but also through timing. That is, producing the visual signal immediately after the troublesome part (e.g., ambiguous person reference) of the ongoing turn may already signal what part of the trouble source turn needs clarification. In-depth future examinations of the precise temporal relationship between brow movement and trouble source may shed light on this issue.

Note that in the present study we did not find any eyebrow *raises* that occasioned clarification without relying on a verbal signal, which may in part be explained by the close association of eyebrow raising and speaking (64,65). While eyebrow furrows may intrinsically signal some kind of communicative trouble or puzzlement, eyebrow raises might be associated with verbal repair initiations, at least to some extent, because verbal repair initiations often have questioning prosody (e.g. Huh? ,35) and eyebrow raises can co-express questioning prosody (10). This does not mean that eyebrow raises can never occasion repair without relying on a vocal signal in spoken Dutch. At least anecdotal evidence suggests that also in spoken Dutch, eyebrow raises—especially when combined with a downward movement of the corners of the mouth—can also occasion clarification without a vocal signal, especially after a try-marked person or place references. This facial gesture combining brow and mouth actions has been termed a “facial shrug” (13,66)—a signal of “not knowing” (13, p.15, see also 37, p. 10-11).

Taken together, the results of Study 1 already provide correlational evidence regarding the hypothesized causal involvement of eyebrow movements in signaling problems of hearing or understanding in spoken face-to-face communication. However, because in conversational corpora a multitude of behaviours happen at any given time, controlled experimental work (Study 2) is required to provide conclusive evidence about potential causal relations between eyebrow movements and conversational repair.

Study 2

The cooperative eyebrow furrow: A facial signal of insufficient understanding in face-to-face conversation

Study 1 has provided correlational corpus-based evidence suggesting that addressee eyebrow furrows can indeed serve an interactional function in face-to-face communication, specifically to signal non-understanding. The results showed that verbal addressee signals of non-understanding intended to elicit repair (such as *Huh?*, *You mean John?*) accompanied by eyebrow *furrows* were more likely to prompt clarification by the speaker, compared to verbal signals accompanied by eyebrow *raises* or no eyebrow movement at all. Crucially, it was also found that eyebrow furrows alone, i.e., without words, were sufficient to occasion clarification by the speaker. Taken together, these results suggest a communicative function of addressee eyebrow furrows in signaling “I’ve *not* received enough information for current purposes” (31,32). Based on these correlational findings we ask: Is there a causal influence of addressee eyebrow furrows on speakers’ communicative behavior in face-to-face interaction?

To address these questions, we developed a novel experimental paradigm using Virtual Reality technology enabling us to selectively manipulate visual feedback in virtual addressees (see also 67). This selective manipulation allowed us to address questions regarding the causal role of eyebrow furrows in interactive face-to-face communication—questions that have previously been impossible to address with such a high degree of experimental control. Participants were asked to have a conversation with different avatars and to answer open questions (e.g., *How was your weekend, what did you do?*). During the participant's answers, the avatar produced different types of visual feedback responses, which were secretly triggered by a confederate. In one condition, the confederate responses were translated by a script to always trigger nods in the avatar (baseline 'nod' condition). In a second condition, the confederate responses triggered nods but, crucially, occasionally an eyebrow furrow instead (experimental 'nod/brow furrow' condition). A control condition was identical to the experimental 'nod/brow furrow' condition except that the eyebrow furrows were replaced with no response at all while the nods were retained (control 'nod/non-response' condition). This condition was included to control for the fact that there would be fewer nods in the 'nod/brow furrow' than in the 'nod' condition. Thus, we would be able to tease apart whether any differences between these latter two conditions were due to the reduction of nods or the presence of eyebrow frowns.

If addressees' eyebrow furrowing is irrelevant for the speaker's speaking behavior, one would not expect any differences between the nod condition and the nod/brow furrow condition. However, if addressees' eyebrow furrowing can indeed signal "I've *not* received enough information for current purposes" providing evidence for unsuccessful grounding (11) speakers should provide extra information; that is, they should provide longer answers in the nod/brow furrow condition than in the nod baseline condition. This is the main hypothesis Study 2 is testing, based on the logic that if eyebrow furrows are perceived by speakers as a request for more information, then longer answers should indicate that speakers provide additional semantic information to respond to this request. However, there is the possibility that furrowed brows throw speakers off course a little, thus leading to more hesitations than in the other conditions—unfilled, silent pauses and filled pauses like *uh* and *uhm*—which may alternatively explain any differences in overall answer length. To be able to rule out this possibility, we also measured the frequency and duration of filled pauses and unfilled pauses within each answer. Finally, if addressees' eyebrow furrowing signals a need for further information, then their presence should lead to longer answers also in comparison to the control condition, where eyebrow frowns were replaced with no feedback at all (i.e. no nod, no eyebrow frown).

Speaking behavior, like any other social behavior, varies from individual to individual (68). In this experiment, two particular individual differences measures of dispositional social sensitivity—the Empathy Quotient (69) and the Fear of Negative Evaluation scale (henceforth 'FoNE'; ,70)—were hypothesized to modulate the perception of eyebrow movements. Sensitivity to addressees' eyebrow furrows may depend on the speaker's degree of empathy, which is the "drive or ability to attribute mental states to another person/animal, and entails an appropriate affective response in the observer to the other person's mental state" (69, p.168). It has been observed that "to drive your point home in a discussion for far longer than is sensitive to your addressee" constitutes low-empathy behavior (69, p.170), suggesting that low-empathy speakers may be less sensitive to addressee feedback than high-empathy speakers. To address this issue, participants were asked to complete the Empathy Quotient questionnaire after the experiment. Sensitivity to addresees' eyebrow furrows may also depend on the speaker's degree of 'fear of negative evaluation' (FoNE). In contrast with low-FoNE individuals, high-FoNE individuals are highly concerned with seeking social approval (70). High-FoNE individuals have been shown to exhibit more pro-social behavior (71), and to try harder making a good impression during face-to-face conversations (72). According to Leary (72, p. 371), "People who are highly concerned about being perceived and evaluated negatively would be more likely to behave in ways that avoid the possibility of unfavorable evaluations and, thus, be more responsive to situational factors relevant to such concerns than individuals who are less apprehensive about others' evaluations of them". One such relevant situational factor may be others' facial expressions. Indeed, high-FoNE individuals have been shown to pay more attention to faces (73), particularly to faces expressing negative emotions due to their potentially socially devaluating meaning (74,75). Since eyebrow furrowing is associated with expressions of negative emotions like anger (8), one might expect high-FoNE individuals to be especially sensitive to addressee eyebrow furrows as they occur in the present study. Finally, high-FoNE individuals have also been shown to judge their own communicative effectiveness more accurately, that is, in a way that is more consistent with addressee's actual understanding, which might be due to their increased sensitivity to addressee feedback (76).

If addresses' eyebrow furrows are not a semiotic, conventional signal but, e.g., a symptom of the their cognitive effort, one may expect only high-empathy or high-FoNE speakers to be responsive to addressees' eyebrow furrows in the messages they design, due to their stronger social sensitivity. However, if addressees' eyebrow furrows are indeed a semiotic, conventional signal, one may expect all speakers to be sensitive to addressees' eyebrow furrows (although high-empathy or high-FoNE speakers even more so).

The overall aim of Study 2 was to experimentally test the claims based on correlational evidence suggesting that addressee eyebrow furrows may serve a communicative function in conversation (Study 1). The main hypothesis was that addressees' eyebrow furrows can function as a communicative signal of insufficient understanding, that speakers would produce longer answers in the nod/brow furrow condition than in the nod baseline condition, while individual differences in speakers' social sensitivity may modulate this effect.

Methods – Study 2

Participants

We recruited 36 native Dutch speakers through the Max Planck Institute for Psycholinguistics subject database for participation in the experiment. The data of one participant were excluded from all analyses because he provided such long answers to the avatar's questions that we had to interrupt him and end the experiment prematurely in order to be able to test the remainder of the scheduled participants. The data of one additional participant were excluded from all analyses because he excessively looked away from the screen (more often than 2.5 SD above the mean) and therefore could not have been influenced by differences in avatar addressee responses. Another participant did not complete the Empathy Quotient questionnaire and was therefore excluded from any analyses including the Empathy Quotient. This resulted in a final sample of 34 participants (18-33 years; mean age = 22.47; 18 females, 16 males), or 33 participants for analyses including the Empathy Quotient (18-33 years; mean age = 22.54; 18 females, 15 males). Each participant was paid €10 and the whole session lasted about one hour.

Design

We used a within-subject design with avatar addressee feedback (nod, eyebrow furrow, non-response) as independent variable and mean answer length as the main dependent variable. Additional variables consisted of the Empathy Quotient (69), the Fear of Negative Evaluation score (72), hesitations (frequency and duration of filled and unfilled pauses), as well as avatar evaluation questionnaire scores assessing perceived humanness, ease of understanding by the avatar of the participant, and likability of each avatar (see below for details on the questionnaires).

The experiment consisted of three blocks, one block per addressee feedback condition (i.e. one per avatar). The set of 18 spoken question stimuli were split up into three sets of 6 questions and each set was assigned to one of the three avatars, meaning each participant heard each question only once. The order of addressee feedback conditions as well as the assignment of avatars (and thus the 6 questions that were paired with the respective avatars) to the addressee feedback conditions was counterbalanced across participants. The order of items within each block was randomized.

Apparatus and Materials

Laboratory set-up and equipment. Participants were invited to the Virtual Reality laboratory at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. They were seated in front of a computer screen (HP Compaq LA2405WG) with speakers (Hercules XPS 2.010) wearing a lightweight, head-mounted microphone (DPA-d:fine-88). Audio was recorded using Adobe Audition CS6 and video was recorded using three synchronized video cameras (Sony 3CCD Megapixel) to capture the participant (1) frontally, and (2) laterally, as well as to record a separate computer screen showing exactly what the participant was seeing on their screen (i.e., the avatar). This setup allowed us to link participant and avatar behavior in a time-aligned manner. For each recording session, we synchronized the three videos and the audio file based on audible and visible markers (produced at the beginning of each block) and exported them in Adobe Premier Pro CS6 (MP4, 25 fps). The confederate was seated in the control room next to the experiment room, in front of a keyboard (Apple MB110LL/B) and a computer screen (Acer AL732). The computer screen showed the participant in real time from a frontal view. Audio from the participant's microphone was also transmitted to the control room and played via speakers (Alesis M1Active 520) in real time (see also Procedure). The confederate was thus responding to an interaction partner (i.e. the participant) who they saw and heard, but instead of producing visual and vocal addressee behaviour, the confederate was asked to press a button whenever they felt feedback should naturally occur. These button press responses were translated into different forms of avatar feedback behaviour (see below). Importantly, based on this manipulation, the avatar's feedback appeared naturally timed. Moreover, to make sure that the confederate behaved consistently, we compared the confederate's button press behaviour across conditions: predicting the confederate's feedback button press frequency (number of button presses per answer divided by the length of the same answer in minutes; $M = 10.74$; $SD = 3.52$) by feedback condition (nod, nod/brow furrow, nod/non-response), including random intercepts for participants and items, confirmed that button press frequency was consistent across conditions (nod vs. nod/brow furrow: $\beta = -0.05$, $SE = 0.25$, $t = -0.255$, $p = 0.822$; nod vs. nod/non-response: $\beta = 0.231$, $SE = 0.253$, $t = 0.913$, $p = 0.362$; nod/brow furrow vs. nod/non-response: $\beta = 0.288$, $SE = 0.253$, $t = 1.138$, $p = 0.256$).

Avatar characteristics and behaviors. The experiment was programmed in WorldViz's Vizard 5.5 and three different female avatars were created based on a stock avatar produced by WorldViz. Three different female Dutch native speakers were used to pre-record the avatars' speech, which was played at appropriate times during the experiment (one per condition). The avatars' lip movements were programmed to match the amplitude of the pre-recorded speech files (i.e., the higher the amplitude, the wider the avatar opened her mouth), creating an illusion of synchronization. The speech materials consisted of a general introduction (e.g., *Hoi, Ik ben Julia, leuk je te ontmoeten!*; 'Hi, I'm Julia, nice to meet you!' and *Ik heb een aantal vraagen aan jou*; 'I have a couple of questions for you') and a set of 18 open-ended questions (e.g., *Hoe was je weekend, wat heb je allemaal gedaan?*; 'How was your weekend, what did you do?'). The avatar also responded to the participant's answer (e.g., *Oh ja, wat interessant!*; 'Oh, how interesting!') before moving on to the next open question (e.g., *Ik heb nog een vraag aan jou;* 'I have another question for you'), or before closing the interaction (*Hartelijk bedankt voor dit gesprek, ik vond het gezellig!*; 'Thank you very much for this conversation, I enjoyed it!').

All visual feedback responses of all three avatars were triggered secretly by a confederate, a Dutch native speaker who could see and hear the participant (via a video-camera link, see above), who was blind to the experimental hypotheses (and not informed about the manipulations), and who was instructed to imagine being the actual addressee interacting with the participant and to press a button whenever it felt appropriate and natural to provide addressee feedback. Which of the confederate's button presses triggered a nod and which a brow furrow (within the nod/brow furrow condition) was varied automatically by the computer program. To avoid unnatural repetitions of eyebrow furrows, we made sure that following each eyebrow furrow, the next one or two feedback responses (randomly varied) would be a nod before a next eyebrow furrow could be produced.

The crucial experimental manipulation in the present study was the feedback responses the avatar produced when she was in the addressee role (see Fig. 1 for example stills). Critically, the form of these feedback responses were modelled on feedback behavior that occurs in natural conversation and they consisted of head nods (duration of 500 milliseconds from nod onset to nod offset) and in one condition eyebrow furrows (duration of 500 milliseconds from eyebrow furrow onset to eyebrow furrow offset). In the control condition, the avatar produced 'non-responses', periods in which the avatar did not and could not produce any feedback response. That is, during a 'non-response', the avatar was just still (default behavior). Note that the duration of 'non-responses' matched the durations of the other feedback responses precisely (i.e., 500 milliseconds). The timing of these feedback responses were based on the confederate's behaviour, as described above.

Questionnaires. The questionnaires used on this study consisted of the Dutch version of the Empathy Quotient questionnaire (test-retest reliability: r = 0.97, as reported by 69) and the Dutch version of the brief 'Fear of Negative Evaluation Scale' (test-retest reliability: r = 0.75, as reported by 72). To control for the possibility that any differences in answer length might be driven by differences in perceived naturalness, perceived ease of understanding by the avatar of the participant, and perceived likability of the avatars depending on the different feedback behaviors they produced, we asked participants to fill in three additional questionnaires tapping these three aspects (one for each avatar each participant interacted with, that is, one per addressee-feedback condition). The avatar evaluation questionnaires consisted of statements designed to assess the participants' perception of the avatar's (1) humanness (*Ik vond deze avatar menselijk overkomen*; ‘This avatar appeared human’), (2) ease of understanding by the avatar of the participant (*Ik denk dat deze avatar mij makkelijk te begrijpen vond*; ‘I think this avatar found me easy to understand’), and (3) likability (*Ik vond deze avatar sympathiek overkomen*; ‘This avatar appeared nice’; *Ik zou vrienden kunnen zijn met deze avatar*; ‘I could be friends with this avatar’; *Ik vond deze avatar egocentrisch overkomen*; ‘This avatar appeared selfish’) as their conversational partner (adapted from 77), and the Dutch translations used in the Relationship Questionnaire of (78). Participants indicated on a 6-point Likert scale their degree of agreement for each statement (1 = *I do not agree at all*, 6 = *I absolutely agree*). Statistical tests confirmed that the perceived humanness, ease of understanding by the avatar of the participant, and likability (rated through scores for niceness, friendship, selfishness, see above) of the avatars did not differ across addressee-feedback conditions (see Appendix).

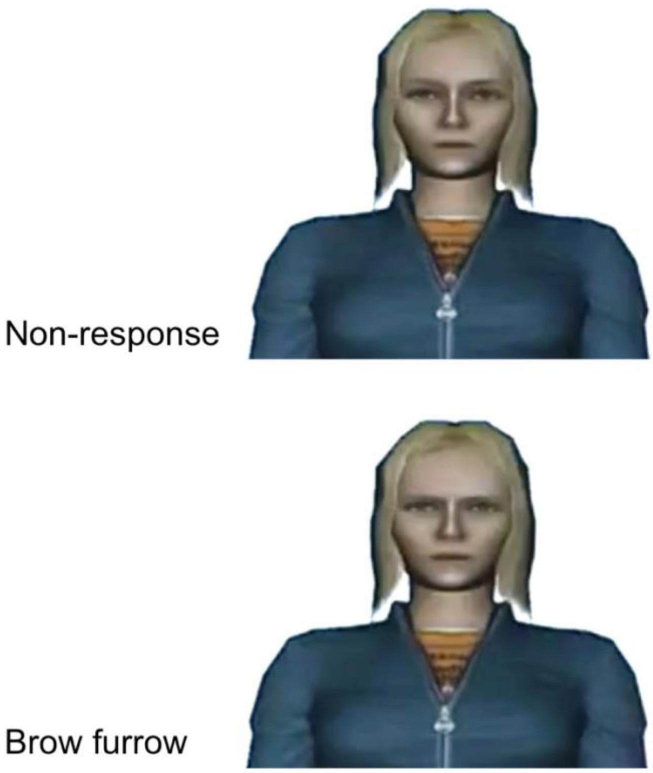

Figure 9. Example stills of a virtual addressee producing different types of addressee feedback responses (non-response, brow furrow) that were varied across condition.

Procedure

Participants were seated in front of the computer screen and were asked to meet and have a conversation with three different avatars (see Fig. 9) by responding to their questions. After a short personal introduction, the avatar asked questions and produced different types of visual feedback responses while participants answered (see Avatar characteristics and behavior). Upon each answer completion by the participant, the avatar produced a response to the participant’s answer (e.g., ‘Oh, how interesting!’). After having finished the conversation with the third avatar, the experiment was over and participants were asked to complete questionnaires before they were debriefed on the purpose of the experiment. The study was approved by the Social Sciences Faculty Ethics Committee, Radboud University Nijmegen and informed consent was obtained before and after the experiment.

Behaviour analysis

Answer length. Answer length was measured in seconds (in ELAN 4.9.3, 47), from the first to the last vocalization produced by the speaker in response to each question.

Hesitations. To differentiate changes in answer length due to content from changes in answer length due to hesitations, we measured different types of hesitations, namely the frequency and average duration of filled pauses (uh's and uhm's, 79) and unfilled pauses (79,80).

Statistical Analysis

We used R (R Core Team, 2012) and *lme4* (57) to test in a linear mixed-effects model whether answer length differed depending on addressee feedback. The initial model was an intercept-only model estimating the mean answer length including intercepts for items (question stimuli) and participants as random effects (more complex models including random slopes for participants did not converge). Using a likelihood ratio test (using the 'anova' function), the intercept model was compared to a model which differed only in that addressee feedback (nod, nod/eyebrow furrow, nod/non-response) was included as a fixed effect. To test whether any effect of addressee feedback on answer length was modulated by the speakers' empathy, we first entered addressee feedback (nod, nod/eyebrow furrow, nod/non-response) and speaker empathy (EQ score as a scaled and centered continuous variable) as fixed effects (without interaction term), and intercepts for items (question stimuli) and participants as random effects into the model. This model was then compared to a model that only differed in that addressee feedback and speaker empathy were entered as fixed effects *with* interaction term, again using a likelihood ratio test (with the 'anova' function). To test whether any effect of addressee feedback on answer length was modulated by the speakers' fear of negative evaluation, we first entered addressee feedback (nod, nod/eyebrow furrow, nod/non-response) and fear of negative evaluation (FNE score as a scaled and centered continuous variable) as fixed effects (without interaction term), and intercepts for items (question stimuli) and participants as random effects into the model. This model was then compared to a model that only differed in that addressee feedback and fear of negative evaluation were entered as fixed effects *with* interaction term, again using a likelihood ratio test (with the 'anova' function). To test whether any differences in answer length could be explained by differences in hesitations, we subtracted all filled and unfilled pauses—that is, the sum of durations of all filled and unfilled pauses produced within each answer—from the total length of each answer. Then, we ran the same model comparisons again, as described above, with the only difference that the dependent variable now was 'answer length minus hesitations'.

Results – Study 2

Speakers' answer length

Did speakers' answer length differ depending on addressee feedback? As one can see in Figure 10 showing the overall mean answer length by addressee feedback condition, speakers indeed produced longer answers in the nod/brow furrow condition than in the nod condition, and answers in the control nod/non-response condition were not longer than in the nod condition[6].

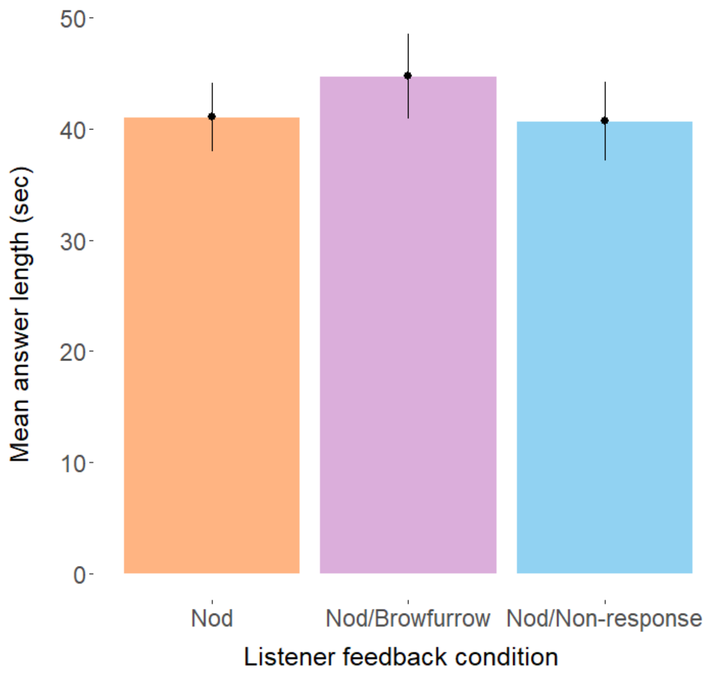

Figure 10. Mean answer length (sec) by addressee feedback. Standard errors are represented in the figure by the error bars attached to each column.

We used a mixed-effects model to test whether answer length (M = 42.16 sec; SD = 25.05) differed depending on addressee feedback. Including 'addressee feedback' as fixed effect provided a model with a significantly better fit ($\chi^2(2)$ = 6.03, p = .048), revealing that—relative to avatars that only nodded (41.4 seconds ± 3.62 [standard error])—the presence of addressees' eyebrow furrows increased speakers' answer length by about 3.77 seconds ± 1.67 standard error (t = 2.25, p = .025), that is, by approximately eight to eleven words (based on an average of two to three words produced per second in conversation, 81). Also relative to speakers' answer length in the nod/non-response control condition (41.91 seconds ± 3.61 [standard error]) speakers' answer length was significantly longer in the nod/eyebrow furrow condition (β = 3.34, SE = 1.68, t = 1.98, p = .047).

Speakers' answer length in the nod condition and the nod/non-response control condition was statistically indistinguishable (β = 0.43, SE = 1.68, t = 0.25, p = .798), suggesting that it was not the relatively reduced number of nods in the nod/eyebrow furrow condition that increased the answer length but, as predicted, the presence of eyebrow furrows. Overall,

[6] Predicting confederate's feedback button press frequency (number of button presses per answer divided by the length of the same answer in minutes; M = 10.74; SD = 3.52) by feedback condition (nod, nod/brow furrow, nod/non-response), including random intercepts for participants and items, confirmed that button press frequency was consistent across conditions (nod vs. nod/brow furrow: β = -0.05, SE = 0.25, t = -0.255, p = 0.822; nod vs. nod/non-response: β = 0.231, SE = 0.253, t = 0.913, p = 0.362; nod/brow furrow vs. nod/non-response: β = 0.288, SE = 0.253, t = 1.138, p = 0.256).

these results support the hypothesis that addressee brow furrows can signal "I've *not* received sufficient information for current purposes", such that speakers provide more information, overall resulting in longer answers.

However, rather than providing additional semantic information, speakers may have produced more hesitations (unfilled, silent pauses and filled pauses like *uh* and *uhm*) when facing an avatar who occasionally furrowed her brows, which may explain the overall longer answers in the brow furrow condition, compared to the nod condition. To address this issue, we subtracted all filled and unfilled pauses—that is, the sum of durations of all filled and unfilled pauses produced within each answer—from the total length of each answer and then tested again in a linear-mixed effects model whether answer length, now disregarding all filled and unfilled pauses, differed depending on addressee feedback. Again, including 'addressee feedback' as fixed effect provided a model with a significantly better fit ($\chi^2(2) = 9.38$, $p = .009$), revealing that, relative to avatars that only nodded (27.32 seconds ± 2.5 [standard error]), the presence of addressees' eyebrow furrows increased speakers' answer length by about 3.26 seconds ± 1.17 standard error ($t = 2.77$, $p = .005$). Also relative to speakers' answer lengths in the nod/non-response control condition (27.57 seconds ± 2.50 [standard error]) speakers' answer length was significantly longer in the nod/eyebrow furrow condition ($\beta = 3.00$, $SE = 1.18$, $t = 2.54$, $p = .011$). Again, speakers' answer lengths in the nod condition and the nod/non-response control condition were statistically indistinguishable ($\beta = 0.25$, $SE = 1.18$, $t = 0.21$). These results indicate that the observed differences in answer length cannot be explained by differences in hesitations, suggesting that, rather than hesitating more, speakers indeed provided more semantic information when facing an avatar who occasionally furrowed her brows compared to an avatar who nodded throughout, further supporting the hypothesis that addressee brow furrows can signal "I've *not* received sufficient information for current purposes", such that speakers provide more information.

Speakers' answer length and individual differences in empathy and fear of negative evaluation

We have seen that speakers provided longer answers when talking to a brow-furrowing addressee than when talking to a addressee who nodded throughout, and we have also seen that the reason for this was not because they hesitated more when talking to a brow-furrowing addressee. Here we investigate whether the effect we found is modulated by individual differences in social sensitivity, focusing on the Empathy Quotient and the Fear of Negative Evaluation scale (see Method).

Did the relationship between addressee feedback and speakers' answer length depend on speakers' Empathy Quotient? As one can see in Figure 11, high-empathy speakers and low-empathy speakers show similar patterns of results.

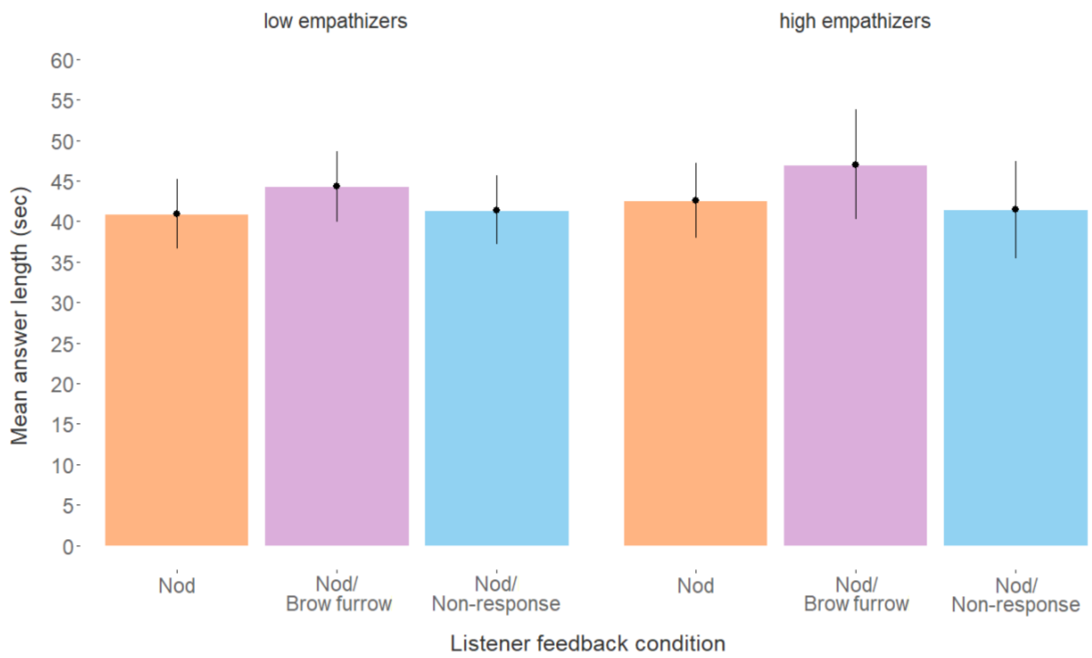

Figure 11. Mean answer length (sec) by addressee feedback in low-empathy and high-empathy speakers (median split). Error bars represent Standard Error.

We used a linear mixed-effects model to statistically test whether answer length by addressee feedback condition differed depending on the speakers' degree of empathy. We entered addressee feedback (nod, nod/brow furrow, nod/non-response) and speaker empathy (EQ score as a scaled and centered continuous variable; M = 43.21; SD = 10.67) as fixed effects (without interaction term), and intercepts for items and participants as random effects into the model. This model was compared to a model that only differed in that addressee feedback and speaker empathy was entered as fixed effects *with* interaction term. Including addressee feedback and speaker empathy *with* interaction term did not provide a model with a significantly better fit ($\chi^2(2) = 1.77$, $p = .40$), revealing that the effect of addressee feedback on speakers' answer length was unaffected by speakers' degree of empathy, also when disregarding filled and unfilled pauses, that is, when predicting 'answer length minus hesitations' ($\chi^2(2) = 1.77$, $p = .32$).

Did the relationship between addressee feedback and speakers' answer length depend on speakers' degree of fear of negative evaluation (see Fig. 12)?

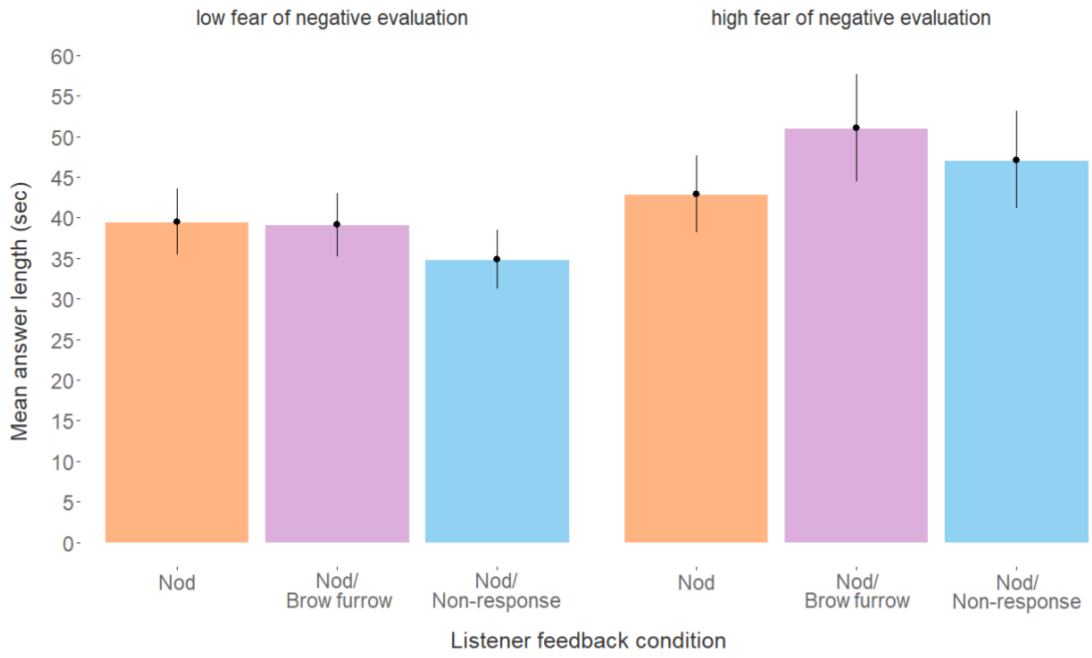

Figure 12. Mean answer length (sec) by addressee feedback in speakers with high fear of negative evaluation versus speakers with low fear of negative evaluation (median split). Error bars represent Standard Error.

To test this, we entered addressee feedback (nod, nod/eyebrow furrow, nod/non-response) and speaker fear of negative evaluation (FoNE score as a scaled and centered continuous variable) as fixed effects (without interaction term), and intercepts for items and participants as random effects into the model. This model was compared to a model that only differed in that addressee feedback and speaker fear of negative evaluation was entered as fixed effects *with* interaction term. Including addressee feedback and speaker fear of negative evaluation *with* interaction term improved the model fit marginally ($\chi^2(2) = 5.64$, $p = .058$, revealing that the effect of addressee feedback on speakers' answer length was not reliably modulated by the speakers' degree of fear of negative evaluation. However, there was a trend in the data revealing that, relative to the nod condition ($\beta = 41.6$, $SE = 3.55$), the higher the speakers' fear of negative evaluation, the longer the answer length in the nod/non-response condition ($\beta = 3.53$, $SE = 1.68$, $t = 2.1$), and the longer the answer length in the nod/brow-furrow condition ($\beta = 3.37$, $SE = 1.67$, $t = 2.01$). Crucially though, the main effect of addressee feedback on answer length was still significant ($\beta = 3.66$, $SE = 1.66$, $t = 2.19$) indicating that differences in answer length between the nod and the nod/brow furrow conditions cannot be fully explained by differences in speakers' degree of fear of negative evaluation. Interestingly, however, when predicting 'answer length minus hesitations' (i.e. excluding filled and unfilled pauses from the answer length measure), the marginally significant interaction effect with fear of negative evaluation disappeared. Including addressee feedback and fear of negative evaluation *with* interaction term did not provide a model with a significantly better fit ($\chi^2(2) = 3.4$, $p = .18$)[7]. Note that, to explore whether speakers adjusted

[7] This non-significant interaction of fear of negative evaluation raises the question whether the marginally significant interaction effect of fear of negative evaluation reported above, that is, using answer length including hesitations, actually reflects differences in the amount of semantic information provided, or rather differences in the amount of hesitations produced (see Appendix for additional analyses zooming into this possibility).

or marked their speech more locally in response to a addressee brow furrow, we also looked at a range of additional variables (speech rate, intensity, pitch change, hesitations), but that none of them explained a significant amount of the data variance (see Appendix for these additional analyses).

Discussion - Study 2

The central question of Study 2 was: Are speakers sensitive to addressee eyebrow furrowing as a communicative signal of insufficient understanding? The findings suggest that they are. In this study, speakers produced longer answers when talking to a brow-furrowing addressee than when talking to an addressee that only nodded, thus supporting the hypothesis that addressee eyebrow furrowing can indeed signal insufficient understanding. The observed differences in answer length could neither be alternatively explained by differences in hesitations, nor by differences in speakers' perception of how human or 'natural' the virtual addressees appeared as conversational partners in each addressee feedback condition, as assessed by the avatar evaluation questionnaires after the experiment.

We were also able to rule out one additional possible alternative explanation. Remember that in the nod/brow furrow condition, a nod was occasionally replaced with a brow furrow, meaning the two conditions did not only differ in the absence versus presence of brow furrows, but also in the overall number of nods. Since nods signal understanding, the relatively reduced overall number of nods in the nod/brow furrow condition rather than the presence of brow furrows could have caused speakers to design longer answers than in the baseline nod condition. If this was the case, one would also have expected longer answers in the control non-response condition than in the baseline nod condition, because the control non-response condition was identical to the nod/brow furrow condition except that the occasional brow furrows were replaced with no response at all (i.e., the control condition differed from the experimental condition only in that brow furrows were absent). However, answer length in the control nod/non-response condition did not differ from answer length in the baseline nod condition. This suggests that the difference in answer length between the nod/brow furrow condition and the nod condition cannot be explained by the reduced number of nods but indeed, as hypothesized, by the presence of eyebrow furrows.

General discussion

Do eyebrow movements play a functional role in signaling communicative problems in spoken face-to-face interaction? This article presents converging correlational and experimental evidence suggesting that they do indeed. In Study 1, we have shown that (1) in addition to the linguistic format of the verbal repair initiation, the type of co-occurring eyebrow movement independently predicted the type of repair solution, (2) the presence of an eyebrow movement as a visual preliminary to verbal repair initiations enhanced repair speed, and (3) eyebrow movements alone were sufficient to occasion clarification.

In Study 2, we followed-up on our corpus-based correlational findings and showed experimentally that speakers are indeed sensitive to addressee eyebrow furrowing as a communicative signal of insufficient understanding (as evidenced by speakers' longer answers when talking to a brow-furrowing addressee than when talking to an addressee that only nodded), thus supporting the hypothesis that addressee eyebrow furrowing is indeed interpreted as a signal of insufficient understanding.

Our findings have clear theoretical implications. We show that addressee's facial behavior can shape the speaker's speaking behavior, likely reflecting speaker adjustments at the 'message level' (82). As such, it provides further support for bilateral accounts of speaking, according to which the addressee is an active collaborator coordinating with the speaker moment by moment to maintain mutual understanding. It highlights that speakers in face-to-face communication not only rely on auditory self-monitoring (e.g., 83) but also on visual other-monitoring (see also 84). Although natural human language is multimodal and social-interactive in nature, traditional models of language processing have primarily focused on verbal language and on utterances produced outside of a social-interactive context. The studies presented in this article embrace the multimodal as well as the social-interactive, bilateral nature of language and it provides further motivation for a paradigm shift, an 'interactive turn' (85, p. 7) that is already taking place in psycholinguistics (43,86,87), but also in the cognitive sciences more generally (88–90).

While we are suggesting that eyebrow movements serve a communicative function, this does not necessarily entail that they are communicatively intended (91) and future experimental work is required to provide conclusive insights into the extent to which addressee brow furrowing is indeed a communicatively intended, conventional signal. Note, for example, that furrowing the brows, might merely be a symptom of the addressees' processing difficulty or high cognitive load, which is then interpreted and treated by the speaker as indicating a need for clarification. Darwin (7) already mentioned that eyebrow furrows (or 'frowns', as he called them) are not only associated with unpleasantness but also with a potentially related but distinct state of dealing with difficulty in thought:

"A man may be absorbed in the deepest thought, and his brow will remain smooth until he encounters some obstacle in his train of reasoning, or is interrupted by some disturbance, and then a frown passes like a shadow of his brow." (7, p. 221)

The observation that people—as individuals not engaged in conversation—also furrow their brows when dealing with cognitive difficulties suggests that such furrows may not only serve an other-oriented, communicative function in signaling a need for clarification in conversation, but that they may also serve a self-oriented, cognitive function (see Figure 7, for an illustration).

Photo on the left: retrieved from https://pxhere.com/en/photo/1127793, CCO.
Photo on the right: retrieved from https://www.flickr.com/photos/renaud-camus/8375622029, CCO, and cropped afterwards.

Figure 13. Rodin's sculpture *Le Penseur* ('The Thinker', 1880) and a facial close-up showing his furrowed eyebrows. Note that the philosopher Gilbert Ryle famously used *Le Penseur* in the mind-body debate, asking 'What is he doing?' (92), arguing against the privacy of cognitive states.

Social-communicative functions and potential cognitive, perceptual, and emotional functions of eyebrow movements are not mutually exclusive. It is possible that the cognitive, perceptual, and emotional functions underlie and precede the communicative signaling function, phylogenetically as well as ontogenetically (e.g., 93 reports eyebrow furrows during "concentration" already in one to three month old infants). The eyebrow furrow as a potential symptom of mental effort, for example, may have been co-opted for communicative purposes through processes of ritualization (7,22,94–96), which would point to a non-arbitrary, iconic relationship (97,98) between form and function in communicative eyebrow furrows. In the same way in which closing the eyes by blinking may signal "no need to see anymore" because sufficient understanding has been reached (45,67), furrowing the eyebrows—as if trying to see more clearly [8], [9]—appears to signal insufficient understanding, potentially shedding new light on the suggested "embodied" origin of the Understanding-Is-Seeing metaphor (100) and on visual origins of mental-state signaling (39).

Our results suggesting a communicative function of eyebrow movements in signaling informational needs in spoken Dutch are in line with examples from other spoken languages like English (37), Italian and Chapalaa (38), and Siwu (36), but also with studies on eyebrow movements in signed languages like Dutch Sign Language (26,101) and Argentine Sign Language (29,38). This suggests that eyebrow movements as signals of insufficient hearing or understanding may be independent from language modality—since they are used in spoken as well as signed language—as well as from language history—since they have been described in unrelated languages. If the use of eyebrow movements as a signal of insufficient hearing or understanding is stable across a variety of unrelated languages, this would be consistent with Darwin (7) who noted "the Australians, Malays, Hindoos, and Kafirs of South Africa frown, when they are puzzled" and who suggested that "men of all races frown when they are in any way perplexed in thought" (p. 221), but it may also suggest that eyebrow movements as signals of communicative problems have evolved from common pressures of a shared conversational infrastructure (102–106).

To conclude, the results suggest that—in addition to visual, emotional, and possible cognitive functions—eyebrow furrowing may serve as a cooperative signal of insufficient understanding. While closing the eyes by blinking may signal "no need to see anymore" because sufficient understanding has been achieved (45,67), furrowing the brows—as if trying to see more clearly—appears to signal insufficient understanding, potentially shedding new light on visual origins of mental-state signaling in face-to-face communication.

[8] According to Darwin (7), Prof. Donders already suggested that eyebrows are furrowed to see more clearly ("the corrugators are brought into action in causing the eyeball to advance in accommodation for proximity in vision", p. 221).
[9] See also (99) on the "thinking face", referring to the speaker marking a word search by turning away her gaze from the addressee with a distant look and with a facial gesture of someone thinking hard.

Acknowledgments:

We thank Jeroen Derks and Han Sloetjes for programming support, Kim Koopmans (KK) and Marina Koleva (MK) for coding support, Ruya Akdag who served as the confederate, and Herb Clark, Mark Dingemanse, Florian Hintz, Kobin Kendrick, Markus Ostarek, David Peeters, Emma Valtersson, the Language & Cognition Department and members of the Virtual Reality Focus Group at the MPI Nijmegen for valuable discussions. This research was funded by the European Research Council (Advanced Grant #269484 awarded to Prof. Levinson and Consolidator Grant #773079 awarded to Dr. Holler) and the Max Planck Gesellschaft.

References

1. Levinson SC, Holler J. The origin of human multi-modal communication. Phil Trans R Soc B. 2014 Sep 19;369(1651):20130302.

2. Goldin-Meadow S. Hearing gesture: How our hands help us think. Cambridge, MA, US: Belknap Press of Harvard University Press; 2003. xiv, 280 p. (Hearing gesture: How our hands help us think).

3. Enfield NJ. The anatomy of meaning: Speech, gesture, and composite utterances. Cambridge University Press; 2009. 267 p.

4. McNeill D. Hand and mind: What gestures reveal about thought. Chicago, IL, US: University of Chicago Press; 1992. (Hand and mind: What gestures reveal about thought).

5. Kendon A. Gesture: Visible action as utterance. Cambridge University Press; 2004.

6. Church RB, Alibali MW, Kelly SD. Why gesture?: How the hands function in speaking, thinking and communicating. John Benjamins Publishing Company; 2017. 443 p.

7. Darwin C. The expression of the emotions in man and animals. New York: Philosophical Library, 1955; 1872.

8. Ekman P. Facial expression and emotion. Am Psychol. 1993;48(4):384–92.

9. Bavelas J, Chovil N. Some pragmatic functions of conversational facial gestures1. Gesture. 2018;17(1):98–127.

10. Bolinger D. Intonation and Gesture. Am Speech. 1983;58(2):156–74.

11. Ekman P. About brows: Emotional and conversational signals. In: Cranach M, Foppa K, Lepenies W, Ploog D, editors. Human ethology. Cambridge: Cambridge University Press; 1979. p. 169–249.

12. Kraut RE, Johnston RE. Social and emotional messages of smiling: An ethological approach. J Pers Soc Psychol. 1979;37(9):1539–53.

13. Bavelas J, Gerwing J, Healing S. Hand and facial gestures in conversational interaction. In: The Oxford handbook of language and social psychology. New York, NY, US: Oxford University Press; 2014. p. 111–30. (Oxford library of psychology).

14. Clark HH. Depicting as a method of communication. Psychol Rev. 2016;123(3):324–47.

15. Sidnell J. Coordinating gesture, talk, and gaze in reenactments. Res Lang Soc Interact. 2006;39(4):377–409.

16. Bavelas J, Gerwing J, Healing S. Effect of dialogue on demonstrations: Direct quotations, facial portrayals, hand gestures, and figurative references. Discourse Process. 2014 Nov 17;51(8):619–55.

17. Stec K, Huiskes M, Redeker G. Multimodal quotation: Role shift practices in spoken narratives. J Pragmat. 2016 Oct 1;104:1–17.

18. Argyle M, Cook M. Gaze and mutual gaze. Oxford, England: Cambridge U Press; 1976. xi, 210 p. (Gaze and mutual gaze).

19. Senju A, Johnson MH. The eye contact effect: mechanisms and development. Trends Cogn Sci. 2009;13(3):127–34.

20. Ekman P, Friesen WV. Facial action coding system: manual. Palo Alto, Calif.: Consulting Psychologists Press; 1978.

21. Ekman P, Friesen WF, Hager JC. Facial Action Coding System. Manual and investigator's guide. Salt Lake City, UT: Research Nexus; 2002.

22. Eibl-Eibesfeldt I. Similarities and differences between cultures in expressive movements. In: Non-verbal communication. Oxford, England: Cambridge U. Press; 1972. p. xiii, 443–xiii, 443.

23. Ekman P, Friesen WV. Unmasking the face: A guide to recognizing emotions from facial clues. Oxford, England: Prentice-Hall; 1975. xii, 212 p. (Unmasking the face: A guide to recognizing emotions from facial clues).

24. Wierzbicka A. Cross-cultural pragmatics: the semantics of human interaction. Berlin; New York: Mouton de Gruyter; 1991.

25. Baker-Shenk C. A microanalysis of the nonmanual components of questions in American Sign Language (Doctoral dissertation) [Internet]. University of California, Berkeley.; 1983 [cited 2022 Mar 6]. Available from: https://escholarship.org/uc/item/7b03x0tz

26. Coerts J. Nonmanual grammatical markers: an analysis of interrogatives, negations and topicalisations in Sign Language of the Netherlands (Doctoral Dissertation). Amsterdam: Universiteit van Amsterdam.; 1992.

27. Zeshan U. Interrogative constructions in signed languages: Crosslinguistic perspectives. Language. 2004;80(1):7–39.

28. Dachkovsky S, Sandler W. Visual Intonation in the prosody of a sign language. Lang Speech. 2009;52(2–3):287–314.

29. Manrique E. Other-initiated repair in Argentine Sign Language. Open Linguist [Internet]. 2016 [cited 2022 Mar 7];2(1). Available from: https://www.degruyter.com/document/doi/10.1515/opli-2016-0001/html

30. Clark HH, Brennan SE. Grounding in communication. In: Perspectives on socially shared cognition. Washington, DC, US: American Psychological Association; 1991. p. 127–49.

31. Clark HH, Schaefer EF. Contributing to discourse. Cogn Sci. 1989 Apr 1;13(2):259–94.

32. Clark HH, Wilkes-Gibbs D. Referring as a collaborative process. Cognition. 1986;22(1):1–39.

33. Schegloff EA, Jefferson G, Sacks H. The preference for self-correction in the organization of repair in conversation. Language. 1977;53(2):361–82.

34. Dingemanse M, Kendrick KH, Enfield NJ. A coding scheme for other-initiated repair across languages. Open Linguist [Internet]. 2016 [cited 2022 Mar 7];2(1). Available from: https://www.degruyter.com/document/doi/10.1515/opli-2016-0002/html

35. Enfield NJ, Dingemanse M, Baranova J, Blythe J, Brown P, Dirksmeyer T, et al. Huh? What? – a first survey in twenty-one languages. In: Hayashi M, Raymond G, Sidnell J, editors. Conversational repair and human understanding [Internet]. Cambridge: Cambridge University Press; 2013. p. 343–80. Available from: /core/books/conversational-repair-and-human-understanding/huh-what-a-first-survey-in-twentyone-languages/75CEA07D85DB0826F94EB9B5AD9F8BD0/core-reader

36. Dingemanse M. Other-initiated repair in Siwu. Open Linguist [Internet]. 2015 Feb 26 [cited 2022 Mar 7];1(1). Available from: https://www.degruyter.com/document/doi/10.1515/opli-2015-0001/html

37. Kendrick KH. Other-initiated repair in English. Open Linguist [Internet]. 2015 Feb 24 [cited 2021 Oct 2];1(1). Available from: https://www.degruyter.com/document/doi/10.2478/opli-2014-0009/html

38. Floyd S, Manrique E, Rossi G, Torreira F. Timing of visual bodily behavior in repair sequences: Evidence from three languages. Discourse Process. 2016 Apr 2;53(3):175–204.

39. Lee DH, Mirza R, Flanagan JG, Anderson AK. Optical origins of opposing facial expression actions. Psychol Sci. 2014 Mar;25(3):745–52.

40. Kaukomaa T, Peräkylä A, Ruusuvuori J. Foreshadowing a problem: Turn-opening frowns in conversation. J Pragmat. 2014;71:132–47.

41. Clark HH. Using language. Cambridge University Press; 1996. 452 p.

42. Bavelas JB, Chovil N. Visible acts of meaning: An integrated message model of language in face-to-face dialogue. J Lang Soc Psychol. 2000;19(2):163–94.

43. Holler J, Levinson SC. Multimodal language processing in human communication. Trends Cogn Sci [Internet]. 2019 Aug [cited 2021 Mar 14];23(8). Available from: https://pubmed.ncbi.nlm.nih.gov/31235320/

44. van Son R, Wesseling W, Sanders E, van den Heuvel H. The IFADV corpus: a free dialog video corpus. In: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08) [Internet]. Marrakech, Morocco: European Language Resources Association (ELRA); 2008 [cited 2022 Mar 7]. Available from: http://www.lrec-conf.org/proceedings/lrec2008/pdf/132_paper.pdf

45. Hömke P, Holler J, Levinson SC. Eye blinking as addressee feedback in face-to-face conversation. Res Lang Soc Interact. 2017;50:54–70.

46. Mandelbaum J. Storytelling in conversation. In: Sidnell J, Stivers T, editors. The handbook of conversation analysis [Internet]. John Wiley & Sons, Ltd; 2012 [cited 2022 Mar 7]. p. 492–507. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118325001.ch24

47. Wittenburg P, Brugman H, Russel A, Klassmann A, Sloetjes H. ELAN : a professional framework for multimodality research. In 2006 [cited 2022 Mar 7]. p. 1556–9. Available from: https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_60436

48. Dingemanse M, Enfield NJ. Other-initiated repair across languages: towards a typology of conversational structures. Open Linguist [Internet]. 2015 Feb 5 [cited 2022 Mar 7];1(1). Available from: https://www.degruyter.com/document/doi/10.2478/opli-2014-0007/html

49. Curl TS. Practices in other-initiated repair resolution: The phonetic differentiation of 'repetitions'. Discourse Process. 2005;39(1):1–43.

50. Schegloff EA. On dispensability. Res Lang Soc Interact. 2004 Apr 1;37(2):95–149.

51. Mazeland HJ, Zaman-Zadeh M. The logic of clarification: Some Observations about Word-Clarification Repairs in Finnish-as-a-Lingua-Franca Interactions. In: Gardner R, Wagner J, editors. Second Language Conversations. New York/London: Continuum Publishers; 2004. p. 132–56.

52. Dingemanse M, Roberts SG, Baranova J, Blythe J, Drew P, Floyd S, et al. Universal principles in the repair of communication problems. PLoS One [Internet]. 2015;10(9). Available from: https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2179606

53. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.

54. Holle H, Rein R. EasyDIAg: A tool for easy determination of interrater agreement. Behav Res Methods. 2015 Sep;47(3):837–47.

55. Mondada L. The local constitution of multimodal resources for social interaction. 2014;

56. Mcgurk H, Macdonald J. Hearing lips and seeing voices. Nature. 1976 Dec;264(5588):746–8.

57. Bates D, Maechler M, Bolker B. lme4: Linear mixed-effects models using S4 classes. R Package URL HttpCRANR-Proj. 2012;

58. Vesper C, van der Wel RPRD, Knoblich G, Sebanz N. Making oneself predictable: reduced temporal variability facilitates joint action coordination. Exp Brain Res. 2011;211(3):517–30.

59. Holler J, Kendrick KH, Levinson SC. Processing language in face-to-face conversation: Questions with gestures get faster responses. Psychon Bull Rev. 2018 Oct 1;25(5):1900–8.

60. Sacks H, Schegloff EA, Jefferson G. A simplest systematics for the organization of turn-taking for conversation. Language. 1974;50(4):696–735.

61. Levinson SC. Action formation and ascription. In: Stivers T, Sidnell J, editors. The handbook of conversation analysis [Internet]. Malden, MA: Wiley-Blackwell; 2013 [cited 2019 Jan 4]. p. 101–30. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118325001.ch6

62. Brown P, Levinson SC. Politeness: Some universals in language usage. New York, NY, US: Cambridge University Press; 1987. xiv, 345 p. (Politeness: Some universals in language usage).

63. Manrique E, Enfield N. Suspending the next turn as a form of repair initiation: evidence from Argentine Sign Language. Front Psychol. 2015;6:1326.

64. Krahmer E, Swerts M. More about brows. In: Ruttkay Z, Pelachaud C, editors. From brows to trust: Evaluating embodied conversational agents [Internet]. Dordrecht: Springer Netherlands; 2004 [cited 2022 Mar 8]. p. 191–216. (Human-computer interaction series). Available from: https://doi.org/10.1007/1-4020-2730-3_7

65. Flecha-García ML. Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. Speech Commun. 2010 Jun 1;52(6):542–54.

66. Ekman P. Telling lies: clues to deceit in the marketplace [Internet]. New York, NY: Berkeley Books; 1985 [cited 2022 Jun 15]. Available from: https://www.semanticscholar.org/paper/Telling-lies%3A-clues-to-deceit-in-the-marketplace-Ekman/0a05decd630c7ad168dcb754f2d2c05d51c33d18

67. Hömke P, Holler J, Levinson SC. Eye blinks are perceived as communicative signals in human face-to-face interaction. PLOS ONE. 2018 Dec 12;13(12):e0208030.

68. Heerey EA. Decoding the dyad: Challenges in the study of individual differences in social behavior. Curr Dir Psychol Sci. 2015 Aug 1;24(4):285–91.

69. Baron-Cohen S, Wheelwright S. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. J Autism Dev Disord. 2004 Apr;34(2):163–75.

70. Watson D, Friend R. Measurement of social-evaluative anxiety. J Consult Clin Psychol. 1969;33(4):448–57.

71. Schlenker BR. Impression management : the self-concept, social identity, and interpersonal relations. Monterey, Calif. : Brooks/Cole; 1980.

72. Leary MR. A brief version of the Fear of Negative Evaluation Scale. Pers Soc Psychol Bull. 1983 Sep 1;9(3):371–5.

73. Rossignol M, Campanella S, Bissot C, Philippot P. Fear of negative evaluation and attentional bias for facial expressions: an event-related study. Brain Cogn. 2013 Aug;82(3):344–52.

74. Winton EC, Clark DM, Edelmann RJ. Social anxiety, fear of negative evaluation and the detection of negative emotion in others. Behav Res Ther. 1995 Feb;33(2):193–6.

75. Tanaka H, Ikegami T. Fear of negative evaluation moderates effects of social exclusion on selective attention to social signs. Cogn Emot. 2015;29(7):1306–13.

76. Fay N, Page AC, Serfaty C, Tai V, Winkler C. Speaker overestimation of communication effectiveness and fear of negative evaluation: being realistic is unrealistic. Psychon Bull Rev. 2008 Dec;15(6):1160–5.

77. Weatherholtz K, Campbell-Kibler K, Jaeger TF. Socially-mediated syntactic alignment. Lang Var Change. 2014 Oct;26(3):387–420.

78. Heyselaar E, Hagoort P, Segaert K. In dialogue with an avatar, language behavior is identical to dialogue with a human partner. Behav Res Methods. 2017 Feb 1;49(1):46–60.

79. Lickley RJ. Fluency and disfluency. In: Redford MA, editor. The handbook of speech production [Internet]. John Wiley & Sons, Ltd; 2015 [cited 2022 Mar 8]. p. 445–74. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118584156.ch20

80. Eklund R. Disfluency in Swedish human–human and human–machine travel booking dialogues. [Lund, Sweden]: Lund University; 2004.

81. Levelt WJM. Models of Word Production. Trends Cogn Sci. 1999;3(6):223–32.

82. Levelt WJM. Speaking: From intention to articulation. Cambridge, MA, US: The MIT Press; 1989. xiv, 566 p. (Speaking: From intention to articulation).

83. Levelt WJM. Monitoring and self-repair in speech. Cognition. 1983 Jul 1;14(1):41–104.

84. Clark HH, Krych MA. Speaking while monitoring addressees for understanding. J Mem Lang. 2004 Jan 1;50(1):62–81.

85. Kendrick KH. Using conversation analysis in the lab. Res Lang Soc Interact. 2017 Jan 2;50(1):1–11.

86. Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. Behav Brain Sci. 2004 Apr;27(2):169–90.

87. Levinson SC. Turn-taking in human communication – origins and implications for language processing. Trends Cogn Sci. 2016 Jan 1;20(1):6–14.

88. Sebanz N, Bekkering H, Knoblich G. Joint action: bodies and minds moving together. Trends Cogn Sci. 2006 Feb 1;10(2):70–6.

89. Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T, et al. Toward a second-person neuroscience 1. Behav Brain Sci. 2013 Aug;36(4):393–414.

90. De Jaegher H, Di Paolo E, Gallagher S. Can social interaction constitute social cognition? Trends Cogn Sci. 2010 Oct;14(10):441–7.

91. Brennan SE, Galati A, Kuhlen AK. Two minds, one dialog: Coordinating speaking and understanding. In: The psychology of learning and motivation: Advances in research and theory, Vol 53. San Diego, CA, US: Elsevier Academic Press; 2010. p. 301–44. (The psychology of learning and motivation).

92. Ryle G. The thinking of thoughts: What is ‘le Penseur’ doing? In: Collected Essays 1929 - 1968. Routledge; 2009.

93. Oster H. Facial Expression and Affect Development. In: Lewis M, Rosenblum LA, editors. The Development of Affect [Internet]. Boston, MA: Springer US; 1978 [cited 2022 Jun 15]. p. 43–75. (Genesis of Behavior). Available from: https://doi.org/10.1007/978-1-4684-2616-8_3

94. Tinbergen N. Derived activities; their causation, biological significance, origin, and emancipation during evolution. Q Rev Biol. 1952 Mar;27(1):1–32.

95. Bruner J. Learning how to do things with words. In: Aaronson D, Rieber R, editors. Psycholinguistic research (PLE: psycholinguistics): Implications and applications. Psychology Press; 1979.

96. The gestural communication of apes and monkeys. New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates; 2007. viii, 256 p. (Call J, Tomasello M, editors. The gestural communication of apes and monkeys).

97. Grice HP. Meaning. Philos Rev. 1957;66(3):377–88.

98. Perniss P, Vigliocco G. The bridge of iconicity: From a world of experience to the experience of language. Philos Trans R Soc B Biol Sci. 2014;369(1651):20130300.

99. Goodwin M, Goodwin C. Gesture and coparticipation in the activity of searching for a word. Semiotica. 1986;62(1–2):51–76.

100. Lakoff G, Johnson M. Philosophy in the flesh the embodied mind and its challenge to Western thought. Basic Books; 1999.

101. de Vos C, van der Kooij E, Crasborn O. Mixed signals: Combining linguistic and affective functions of eyebrows in questions in sign language of the Netherlands. Lang Speech. 2009;52(2–3):315–39.

102. Levinson SC. On the human ‘interaction engine’. In: Enfield NJ, Levinson SC, editors. Roots of human sociality: Culture, cognition and interaction. Oxford: Berg; 2006. p. 39–69.

103. Schegloff EA. Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In: Enfield NJ, Levinson SC, editors. Roots of human sociality. Routledge; 2006. p. 70–96.

104. Levinson SC. Interactional foundations of language: The interaction engine hypothesis. In: Hagoort P, editor. Human language: From genes and brain to behavior. Cambridge, MA: MIT Press; 2019. p. 189–200.

105. Dingemanse M, Torreira F, Enfield NJ. Is “Huh?” a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items. PLOS ONE. 2013 Nov 8;8(11):e78273.

106. Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, Heinemann T, et al. Universals and cultural variation in turn-taking in conversation. Proc Natl Acad Sci. 2009;106(26):10587–92.

Study 2: Appendix

Additional statistical analyses

Speaker’s answer length, fear of negative evaluation, and hesitations.

The non-significant interaction of fear of negative evaluation raises the question whether the marginally significant interaction effect of fear of negative evaluation reported above, that is, using answer length including hesitations, actually reflects differences in the amount of verbal information provided, or rather differences in the amount of hesitations produced. Zooming in on this possibility, we tested in a mixed-effects model whether the proportion of filled or unfilled pauses within each answer (that is, the sum of durations of filled or unfilled pauses within each answer divided by the answer’s total length) differed depending on addressee feedback, and especially whether there was an interaction with fear of negative evaluation. We entered addressee feedback (nod, nod/brow furrow, nod/non-response) as a fixed effect, fear of negative evaluation as interaction term, and intercepts for items and participants as random effects into the model. This model was compared to a reduced model without the interaction term of ‘fear of negative evaluation’ using a Likelihood Ratio Test.

When predicting the proportion of filled pauses per answer, including 'fear of negative evaluation' as interaction term did not improve the model fit significantly ($\chi^2(2) = 2.07$, $p = .35$). Relative to the nod condition ($\beta = 8.06$, $SE = 0.6$), there was neither a significant main effect of addressee feedback on the proportion of filled pauses (nod/brow-furrow condition: $\beta = -0.17$, $SE = 0.28$, $t = -0.629$; nod/non-response condition: $\beta = 0.24$, $SE = 0.28$, $t = 0.85$), nor a significant interaction effect of fear of negative evaluation (nod/brow-furrow condition * fear of negative evaluation: $\beta = -0.31$, $SE = 0.28$, $t = -1.098$; nod/non-response condition * fear of negative evaluation: $\beta = -0.38$, $SE = 0.28$, $t = -1.356$).

However, when predicting the proportion of unfilled pauses per answer, including 'fear of negative evaluation' as interaction term did improve the model fit significantly ($\chi^2(2) = 6.68$, $p = .034$). Relative to the nod condition ($\beta = 25.92$, $SE = 1.41$), there was no significant main effect of addressee feedback on the proportion of unfilled pauses (nod/brow-furrow condition: $\beta = -0.52$, $SE = 0.74$, $t = -0.701$; nod/non-response condition: $\beta = -0.59$, $SE = 0.74$, $t = -0.796$), but there was a significant interaction effect of fear of negative evaluation, revealing that the higher the speakers' fear of negative evaluation, the higher the proportion of unfilled pauses she produced in the in the nod/non-response condition ($\beta = 1.92$, $SE = 0.75$, $t = 2.553$), relative to the nod condition, but critically, not in the nod/brow furrow condition ($\beta = 0.66$, $SE = 0.75$, $t = 0.882$), relative to the nod condition.

What underlies the FNE-dependent higher overall proportion of unfilled pauses in the nod/non-response condition relative to the nod condition? Does it reflect a longer average duration of unfilled pauses and/or a higher frequency of unfilled pauses? We tested in a mixed-effects model whether the effect of addressee feedback on average duration (in milliseconds) or frequency of unfilled pauses (number of unfilled pauses divided by answer length) was modulated by the speakers' degree of fear of negative evaluation. We entered addressee feedback (nod, nod/brow furrow, nod/non-response) as a fixed effect, fear of negative evaluation as interaction term, and intercepts for items and participants as random effects into the model. This model was compared to a reduced model without the interaction term of 'fear of negative evaluation' using a Likelihood Ratio Test. When predicting the average duration of unfilled pauses, including 'fear of negative evaluation' as interaction term did not improve the model fit significantly ($\chi^2(2) = 0.3$, $p = .858$). However, when predicting the frequency of unfilled pauses, including 'fear of negative evaluation' as interaction term did improve the model fit significantly ($\chi^2(2) = 15.89$, $p = .000$), revealing that the higher the speakers' fear of negative evaluation, the higher the frequency of unfilled pauses in the nod/non-response condition ($\beta = 0.26$, $SE = 0.06$, $t = 3.75$) but not in the nod/brow furrow condition ($\beta = 0.04$, $SE = 0.06$, $t = 0.63$), relative to the nod condition ($\beta = 3.79$, $SE = 0.11$).

These results suggest that, regarding the difference in answer length between the nod condition and the nod/non-response control condition, the marginally significant interaction effect of fear of negative evaluation reported above (using answer length including hesitations) appears to reflect differences in the amount of hesitations, specifically the frequency of unfilled pauses produced. This suggests that, the higher the speakers' fear of negative evaluation, the more unfilled pauses in the nod/non-response condition relative to the nod condition. However, regarding the difference in answer length between the nod condition and the nod/brow furrow condition—the main contrast of interest—the marginally significant interaction effect of fear of negative evaluation reported above (using answer length including hesitations) indeed appears to reflect differences in the amount of semantic information rather than differences in hesitations. Overall, these results from answer length and hesitations further support the hypothesis that addressee brow furrows can signal "I've *not* received sufficient information for current purposes", such that speakers in general provide more verbal information when facing a brow-furrowing avatar addressee than when facing an avatar addressee that nodded throughout.

Speakers' local adjustments in response to addressee brow furrows

We have seen that speakers indeed provided overall longer answers when talking to a addressee who occasionally furrowed her brows than when talking to a addressee who nodded throughout, suggesting that addressee eyebrow furrowing can indeed signal "I've *not* received sufficient information for current purposes". But did speakers adjust or mark their speech more locally in response to a addressee brow furrow? Did speakers slow down or speed up, increase or decrease the loudness or the pitch of their speech, did they provide more or less information, or hesitate more or less?

To address this issue, we zoomed in on the nod/brow furrow condition and used several measures comparing speech produced between the onset of a addressee nod and the onset of the subsequent addressee response (nod segments; n = 865) versus speech produced between the onset of a addressee brow furrow and the onset of a subsequent addressee response (brow furrow segments; n = 632), resulting in a total of 1497 speech segments. We then tested in linear-mixed effects models whether speech rate (syllables per second), intensity (average, minimum, maximum intensity), pitch change (speaker-specific fundamental frequency minus the median pitch measured over the first 700, 1000, 1500, or 2000 ms of each speech segment), duration without hesitations (subtracting the sum of durations of all filled and unfilled pauses from the total duration of each segment), or the proportion of hesitations (duration of filled or unfilled pauses divided by segment duration) of the speech differed depending on addressee feedback (nod segment, brow furrow segment) within the nod/brow furrow condition. We entered addressee feedback (nod, brow furrow) as a fixed effect and intercepts for items and participants as random effects into the model. This model was compared to a reduced model without the fixed effect of 'addressee feedback' using a Likelihood Ratio Test.

There were no significant main effects of addressee feedback. Including 'addressee feedback' as fixed effect did not provide a model with a significantly better fit, neither when predicting speech rate ($\chi^2(1) = 9e\text{-}04$, $p = .976$), nor intensity (average: $\chi^2(1) = 1.76$, $p = .184$, minimum: $\chi^2(1) = 2.38$, $p = .122$, maximum intensity: $\chi^2(1) = 0.21$, $p = .639$), nor pitch change (first 700ms: $\chi^2(1) = 1.12$, $p = .288$; first 1000 ms: $\chi^2(1) = 0.072$, $p = .787$; first 1500 ms: $\chi^2(1) = 0.1$, $p = .748$; first 2000 ms: $\chi^2(1) = 0$, $p = .995$), speech segment duration ($\chi^2(1) = 3.3$, $p = .069$), nor the proportion of hesitations (proportion of filled pauses ($\chi^2(1) = 2.04$, $p = .153$; proportion of unfilled pauses ($\chi^2(1) = 1.79$, $p = .18$).

There were also no significant interaction effects with speakers' fear of negative evaluation. Including 'fear of negative evaluation' as interaction term did not improve the model fit significantly when predicting speech rate ($\chi^2(1) = 0.83$, $p = .364$), intensity (average: $\chi^2(1) = 0.68$, $p = .407$, minimum: $\chi^2(1) = 0.52$, $p = .469$, maximum intensity: $\chi^2(1) = 2.44$, $p = .117$), pitch change (first 700ms: $\chi^2(1) = 2.77$, $p = .09$; first 1000 ms: $\chi^2(1) = 2.57$, $p = .108$; first 1500 ms: $\chi^2(1) = 0.07$, $p = .788$; first 2000ms: $\chi^2(1) = 0.851$, $p = .356$), speech segment duration without hesitations ($\chi^2(1) = 3.3$, $p = .069$), or the proportion of hesitations (proportion of filled pauses ($\chi^2(1) = 3.66$, $p = .055$; proportion of unfilled pauses ($\chi^2(1) = 0.176$, $p = .674$). Thus, within the nod/brow furrow condition, speakers did not change their speech rate, intensity, pitch, the amount of verbal information or hesitations based on whether they received a nod or a brow furrow as addressee feedback.

Avatar evaluations: Perceived humanness, ease of understanding by the avatar of the participant, and likability

We have seen that speakers provided longer answers when talking to a brow-furrowing addressee than when talking to a addressee who nodded throughout, and we have also seen that the reason for this was not because they hesitated more when talking to a brow-furrowing addressee. However, perhaps the differences in answer length might be driven by the perceived humanness, ease of understanding by the avatar of the participant, and perceived likability of the avatars as conversational partners in the different addressee feedback conditions. To address this issue, we asked participants to fill in three questionnaires tapping these three aspects (see Method).

We tested in linear-mixed effects models whether the scores on each item of the questionnaire (humanness, ease of understanding by the avatar of the participant, and likability [rated through scores for niceness, friendship, selfishness, see Method]) differed depending on addressee feedback condition (nod, eyebrow furrow, non-response) and whether this depended on speakers' empathy or fear of negative evaluation. For all models, we entered intercepts for items and participants as random effects. When testing for main effects of 'addressee feedback condition', we compared a full model including 'addressee feedback condition' with a reduced model without 'addressee feedback condition' using a Likelihood Ratio Test. When testing for interaction effects of 'addressee feedback condition' with speakers' empathy or fear of negative evaluation, we compared a full model including 'addressee feedback condition' with empathy or fear of negative evaluation as interaction term with a reduced model without empathy or fear of negative evaluation as interaction term using a Likelihood Ratio Test.

Humanness. Adding addressee feedback condition did not improve the model fit of ratings of 'humanness' ($\chi^2(2) = 0.35$, $p = .835$. Also, there were no significant interaction effects, neither for empathy ($\chi^2(2) = 0.059$, $p = .97$) nor for fear of negative evaluation ($\chi^2(2) = 4.76$, $p = .092$). Overall, this suggests that all speakers perceived all three avatars—whether producing nods, brow furrows, or non-responses—as equally human.

Ease of understanding by the avatar of the participant. Adding addressee feedback condition did not improve the model fit of ratings of 'ease of understanding by the avatar of the participant' ($\chi^2(2) = 3.64$, $p = .161$). Also, there was no significant interaction effect for empathy ($\chi^2(2) = 3$, $p = .222$). However, there was a significant interaction effect of addressee feedback and fear of negative evaluation on ease of understanding by the avatar of the participant ($\chi^2(2) = 8.4$, $p = .014$), revealing that the higher the speakers' fear of negative evaluation, the more they rated the brow-furrowing avatar and the non-response avatar as having difficulty understanding them, relative to the nodding avatar.

Likability. Adding addressee feedback condition did not improve the model fit of ratings of 'likability', that is neither of ratings of niceness ($\chi^2(2) = 3.9$, $p = .141$), friendship ($\chi^2(2) = 4.75$, $p = .092$), nor selfishness ($\chi^2(2) = 5.1$, $p = .077$). Also, there were no interaction effects of addressee feedback condition and empathy (niceness: $\chi^2(2) = 0.23$, $p = .889$; friendship: $\chi^2(2) = 1.9$, $p = .385$; selfishness: $\chi^2(2) = 0.11$, $p = .994$).

However, there were significant interaction effects of addressee feedback condition and fear of negative evaluation (niceness: $\chi^2(2) = 10.26$, $p = .005$; friendship: $\chi^2(2) = 11.54$, $p = .003$; selfishness: $\chi^2(2) = 6.89$, $p = .031$). Note that these significant interaction effects only regard differences between the non-response (control) condition and the nod and the brow furrow condition, respectively. They reveal that the higher the speakers' fear of negative evaluation, the lower the speakers' ratings of the non-response avatar's niceness, friendship potential, and the higher the ratings of the non-response avatar's selfishness, compared to the nodding avatar (niceness: $\beta = -0.78$, $SE = 0.29$, $t = -2.7$; friendship: $\beta = -0.92$, $SE = 0.26$, $t = -3.458$; selfishness $\beta = 0.62$, $SE = 0.23$, $t = 2.686$), as well as compared to the brow-furrowing avatar (niceness: $\beta = -0.87$, $SE = 0.29$, $t = -3.012$; friendship: $\beta = -0.64$, $SE = 0.27$, $t = -2.396$; but note the non-significant effect for selfishness: $\beta = 0.27$, $SE = 0.23$, $t = 1.173$).