**Compensis, Paul**

# Using AI-Generated Avatar Interactions To Study Socio-Affective Functions of Feedback Signals in Language : an Exploratory Experimental Study on Head Nods

Date of secondary publication: 02.06.2026

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-115359x

**Primary publication**

Compensis, Paul (2026): Using AI-Generated Avatar Interactions To Study Socio-Affective Functions of Feedback Signals in Language : an Exploratory Experimental Study on Head Nods, in: Journal of nonverbal behavior, New York, NY: Springer, Vol. 50, No. 1, pp. 117–136, doi: 10.1007/s10919-026-00500-y.

Journal of Nonverbal Behavior (2026) 50:117–136
https://doi.org/10.1007/s10919-026-00500-y

ORIGINAL PAPER

# Using AI-Generated Avatar Interactions To Study Socio-Affective Functions of Feedback Signals in Language: an Exploratory Experimental Study on Head Nods

**Paul Compensis**[1,2]

Received: 1 June 2025 / Accepted: 2 January 2026 / Published online: 12 February 2026


**Abstract**
Feedback signals are a central aspect of human interaction that often fulfil not only an interaction-structuring function (e.g., prompting a speaker to continue his story) but also serve socio-affective purposes. While both head nods and verbal continuers are used to signal affirmation, approval and continuation in a given interaction, head nods are also claimed to index social affiliation and stance-taking – in contrast to verbal continuers where this function is less clear. To investigate this potential functional difference further, I present the results of a pilot study that used AI-generated videos of dyadic interactions in which one avatar narrates a positive or negative life event while another avatar (the recipient) simply listens and occasionally provides feedback in the form of head nods, verbal continuers, a combination of both, or no feedback. While watching the videos, participants continuously rated how affiliative they perceived the behaviour of recipients in response to narrators and afterwards rated how socially adequate and empathic the recipient was. While recipients displaying head nods only or both feedback signals were rated as behaving more affiliative, more socially adequate and more empathic, avatars using verbal continuers only received lower ratings, with these ratings varying also depending on the story's valence. The results of this study support the idea that head nods and verbal continuers differ in their socio-affective function and illustrate that using AI-generated avatars is a useful addition to interactional research, particularly to shed light on specific functions underlying different feedback signals.

**Keywords** Feedback in interaction · Backchannels · Head nods · Verbal continuers · Affiliation · Empathy · Virtual avatars

Portions of this research were presented at the LingCologne2025 conference "Feedback in Interaction" on May 22.

✉ Paul Compensis
paul.compensis@uni-bamberg.de

1 Institute of Psychology, University of Bamberg, Bamberg, Germany

2 Department of Psychology, University of Erlangen-Nuremberg, Erlangen, Germany

## Introduction

When we are interacting with another person in a conversation, we need to not only closely keep track of the unfolding discourse and the intentional goals of the interlocutor but also to let the speaker know that we are still listening and that we understand and possibly agree with what they are saying. For this purpose, humans rely on several interactional feedback signals, either in the form of full responses (such as *I see*, *Yes*) and assessments (e.g., *That's great*!) or as brief verbal continuers/backchannels (such as *Mhmm*, *Uh*, *Aha*) and non-verbal signals, in particular hand movements and gestures. As long as we can follow our interlocutor and generally agree with what they are saying, we usually indicate that we are still listening and invite the other to continue by nodding our head or verbally affirming them to continue. In contrast, it would make a strange impression if a communication partner provided no feedback at all in a conversation. Initially, we would probably question whether we said something wrong or if the interlocutor is struggling to follow the narrative. But if that was not the case and the person still provided any feedback in response to our narration, we would arguably at some point start feeling a little awkward in the presence of such a person. This illustrates that feedback signals do not only structure an interaction but also extend into the domain of social affiliation, bonding and mutual emotional responding.

In general, feedback signals are a central aspect of human interaction (Groß et al., 2024; Ruusuvuori & Peräkylä, 2009) and especially head movements are highly relevant for signalling feedback or backchannelling in interaction (Bauer et al., in press; Cerrato, 2007; Duncan, 1972; Maynard, 1987; Yngve, 1970). With respect to structuring an ongoing interaction, head nods, similar to verbal continuers, are primarily used to signal affirmation, approval and continuation (De Stefani, 2021; McClave, 2000; Schegloff, 1982). In addition to these interaction-structuring functions, there is some indication that head nods also fulfil socio-affective functions, particularly by indexing social affiliation with an interlocutor and expressing stance towards the narrator's content (Cerrato, 2005, 2007; Stivers, 2008).

For verbal continuers, in contrast, the picture is less clear. There is some indication that verbal continuers are considered a weaker response in comparison to head nods and that they might constitute a non-affiliative response (Selting, 2017). Also, it is not clear how head nods and verbal continuers relate to each other with respect to this socio-affective function, in particular, whether they accumulate when being used together or whether they contribute differently or independently to index social affiliation.

Therefore, in the present study I investigated this function through an experimental design employing AI-generated videos of dyadic interactions in which one avatar narrates a positive or negative life event while another avatar (the recipient) simply listens and occasionally provides feedback in form of head nods, verbal continuers, a combination of both, or no feedback. Participants rated the recipient's behaviour with respect to social affiliation, social adequacy and empathy.

The remainder of this introduction is structured as follows. First, I briefly discuss affiliative and empathic responding in social interaction. Second, I discuss the acclaimed socio-affective functions of head nods and verbal continuers before I outline the present exploratory experimental study.

### Social Affiliation and Empathic Responding in Social Interaction

Human interaction is not characterized only by an exchange of ideas, concepts and needs but also by affiliative and emotional dimensions expressed as part of an ongoing conversation. In other words, when speaking with each other, humans usually do not only talk about semantic content but also express mutual affection and bonding in order to build or elaborate on social relationships. An important question in this regard is how affiliation and related socio-affective dimensions are expressed and manifest in other-oriented behaviour and which verbal or non-verbal signals are used for this in a given conversation.

Broadly speaking, affiliation is defined as a "social relationship in which a person joins or seeks out one or more other individuals, usually on the basis of liking or a personal attachment rather than perceived material benefit" (American Psychological Association, n.d). Affiliation is typically described as a basic human desire or need (cf. the concept of *need for affiliation*, Murray, 1938) and constitutes a form of active behaviour directed towards others. In a conversation, usually both interaction partners contribute to and elaborate on social affiliation by shaping the dialogue in an other-centred and understandable way, as for example reflected in common ground management (Enfield, 2008). Depending on the respective role (narrator or recipient), an interlocutor can use different verbal and non-verbal cues to guide the other person through the discourse or to provide adequate feedback to reassure the interlocutor that everything is going (and understood) as planned.

In a narrower sense, affiliation in the context of feedback in interaction was described as the degree to which a recipient endorses a narrator's perspective and supports their stance (Stivers, 2008), especially by drawing on a number of verbal and non-verbal cues. This function is linked to a socio-affective dimension and contributes to mutual emotional and evaluative experience in interaction. Therefore, the degree to which a person seeks to behave affiliatively in social contexts should be reflected in how they provide feedback signals to an interlocutor.

Another important dimension of emotional responding in social interaction is empathy, often defined as the ability or process to "put oneself in the shoes of another person" or "imagining how one would think and feel in the other's place" (Batson, 2011). Despite the ubiquity of this concept, it is important to mention here that there is an overlap with related concepts such as *perspective-taking* and *theory of mind* (for summaries of conceptual debates and a more detailed discussion of different concepts associated with empathy, see Cuff et al., 2016, Hall & Schwartz, 2018, 2022; and Stietz et al., 2019). For the purpose of the present investigation, I understand empathy and related forms of interpersonal behaviour as situationally determined and motivated processes emerging in social interaction with the degree and strength of this emotional response being determined by an interplay of personal, situational and contextual features (de Vignemont & Singer, 2006; Hein & Singer, 2008; Shamay-Tsoory & Hertz, 2022; Zaki et al., 2008). In the context of conversation, the emotional valence of the dialogue or narrative alters the empathic response, with negative events typically eliciting stronger empathy in a recipient than positive events (see, for instance, McCrackin & Itier, 2021).

To express affiliation and to respond in a socially adequate and empathic way, a recipient should be expected to respond with feedback signals indicating affiliation but also stance-taking while listening to a story by another person. Furthermore, depending on valence, negative events with a particular feedback signal might be different from responses to positive events. These two questions are addressed in the present contribution with a focus on the two feedback signals head nods and verbal continuers.

### Interactional and Socio-Affective Functions of Head Nods and Verbal Continuers

It is important for our understanding of communication to take into account the plethora of multimodal cues that play a meaningful role in interaction, especially also non-verbal signals such as head movements (Bauer et al., in press). As was briefly mentioned above, head nods and verbal continuers (such as *Mhmm*, *Uh*, *Aha*), are used to signal affirmation and approval with what is being said and usually invite a speaker to continue their story (Aoki, 2011; Cerrato, 2005; McClave, 2000; Schegloff, 1982) and both are generic responses in the sense of Bavelas et al. (2000). Interestingly, head nods can also be used by a recipient to initiate and negotiate the completion of a story (Li, 2019). Similarly, head nods are described as relevant for feedback and turn management and it was shown that their occurrence usually aligns closely with the interlocutor's prosodic and stress patterns (Hadar et al., 1985; McClave, 2000). In addition, particular variations of head nods (e.g., a series of head nods) can be used to signal that the message is new or unfamiliar to the recipient (Bauer et al., in press). For a broad classification of different types of head nods and different functional aspects associated with these nods, see Poggi et al. (2010). Interesting in this regard is also the observation that head nods usually start a little earlier than the respective speech sequence (Dittmann & Llewellyn, 1968; Paggio, 2016), illustrating that nods are not only a responsive feedback signal but actively contribute to the structuring of an interaction.

With respect to socio-affective functions, the story is more complex. Heads nods were described as conveying affiliation and stance-taking with a reported event (Stivers, 2008). Interestingly, they can express affiliation irrespective of agreeing as was shown in a study of psychotherapy sessions (Muntigl et al., 2012). Also, head nods indicate interest and communicative attention (Nguyen et al., 2012), a function that also entails affiliative components. A body of evidence adds to the idea that (adequate) head nodding in interaction clearly extends into the socio-affective domain. For instance. Freiermuth and Hamzah (2023) have shown that empathy-associated head nodding becomes more frequent among communication partners when mutual trust increases. Also, human subjects produced more head nods in response to a robot when the robot nodded back or when they were told that the robot recognizes the meaning of head nods (Sidner et al., 2006). Similarly, humans agree more with virtual avatars that mimic their nodding (Zhang, 2020) and avatars producing nods in a behaviourally adequate way are liked and trusted more by human study subjects (Aburumman et al., 2022). In addition, avatars that showed head movements that closely aligned with affective content of a story they told were perceived as more natural (Busso et al., 2007; Lee & Marsella, 2010).

In contrast, it is less clear whether and to what extent verbal continuers fulfil a comparable or distinguishable function in the socio-affective domain. Stivers (2008) argues that verbal continuers are functionally distinct from head nods in that they do not express stance-taking (but only express alignment with the story itself). Similarly, Selting (2017) describes verbal continuers as weaker and non-affiliative type of response (Selting, 2017). Nonetheless, this view is challenged by some researchers (e.g., Guardiola & Bertrand, 2013), especially because both feedback signals often co-occur at the same time point (Bavelas, 2002; Dittmann & Llewellyn, 1968). It is therefore highly interesting to investigate further whether head nods and verbal continuers differ in their socio-affective function and whether they contribute differently or only partly to affiliation in an interaction. To this end, it is the goal of the present study to address this research question with an exploratory experimental design.

## The Present Investigation

In the present contribution, I present the results of an exploratory study that addressed the proclaimed functional difference between verbal continuers and head nods from an experimental perspective. The main goal of this study was to investigate how affiliative, empathic and socially adequate the behaviour of a recipient is perceived depending on the type of feedback signals they provide in response to listening to emotional life events told by a narrator. In other words, to explore socio-affective functions of selected feedback signals I draw on a design that relied on the rating of a passively perceived interactional process – isolating specific functions using an oversimplified experimental approach.

The second goal of this study was to explore how automatically AI-generated avatars can be adapted to research on feedback in language. AI-generated virtual avatars can be used as a dynamic, yet more controllable form of stimuli and are therefore ideal for research on social cognition and social interaction (Peres et al., 2023). This makes it particularly practical for the study of subtle social signals, such as socio-affective cues and feedback signals, as well as the interactional alignment of these signals, for which static experimental stimuli (e.g., images with text) might lack the necessary dynamicity and naturalness to study such phenomena. With recent text-to-image and text-to-video generation tools, generating a large number of audio or video stimuli is relatively cost-efficient and fast, especially compared to producing videos with human actors. Using such videos, I explored in a previous study currently under review (Compensis, 2025) how event agency and the emotional valence of a story affect how empathic we perceive a virtual avatar narrator. Given the ability to control some parameters while still generating dynamic and relatively natural videos, videos displaying AI-generated avatars are also suitable for studying the integration and processing of congruent and incongruent socio-affective and linguistic cues (Compensis, in press). Producing material that imitates interactions of two (or more) people is more challenging since most tools focus on the generation of videos showing one avatar. In this contribution, I therefore explore possibilities to adapt (ready-made) text-to-video avatar generation tools to imitate dyadic interaction to study interactional functions of feedback signals. Of course, these tools cannot fully capture real-life interaction completely, but they provide a useful approach for experimentally examining functional differences between different feedback signals and modalities.

Based on previous research, I hypothesize that a recipient is evaluated differently depending on the type of feedback they show when listening to a positive or negative life event. More specifically, I assume that head nods are perceived as indexing affiliation (Muntigl et al., 2012; Stivers, 2008) and therefore predict that avatars providing nods as feedback are expected to be rated as affiliative, especially in comparison to trials with no feedback signals or only verbal continuers. Also, I assume that more affiliative recipients are perceived as more empathic. Following the idea that continuers are a weaker, non-affiliative response (Selting, 2017), I assume that recipients showing only this feedback signal alone are perceived as less affiliative and less empathic. For trials with both feedback signals, I expect a similar pattern as for head nods only, that is, perceiving the recipient as affiliative and empathic. Regarding the social adequacy rating, I expect no difference between the two types of feedback since both are socially accepted forms of feedback. In contrast, for the absence of any feedback, I expect low ratings for affiliation, empathy and adequacy, reflecting negative effects of unaligned or absent feedback. Regarding emotional valence of the stories, I hypothesise that (adequate) feedback in response to negative events (that is, head nods or both feedback signals) is perceived as slightly more affiliative because negative events typically also elicit a more empathic reaction in listeners (e.g., McCrackin & Itier, 2021).

## Method

### Material

For this study, I generated 24 videos depicting a brief interaction of two virtual avatars (for an illustration, see Fig. 1; all videos are available on https://osf.io/ajwn7). In each video, a narrator (visualized in a small, round circle) either told a positive or negative life event about themselves that is typically expected to engender at least some emotional response in a listener. While the narrator talks, the other visual avatar (the recipient) simply listens to the narrative and occasionally provides feedback. This feedback consists either of head nods or verbal continuers only, a combination of both feedback signals, or no feedback signal. Each video lasted roughly about 45 s. Avatar gender and apparent ethnocultural backgrounds were balanced across the stimuli. The set included avatars reflecting a broad and diverse range of commonly represented demographic groups, without claiming comprehensive global coverage. Interactions involved either same-gender or different-gender avatar pairings. All stimuli were generated entirely using AI-based tools.

For each emotional valence, I wrote one example story and then prompted ChatGPT-4o (OpenAI, 2024) to generate the remaining 22 stories based on the example. https://osf.io/ajwn7/overview?view_only=dc10bfd1a4104a9e8ed28bca422879e6 Positive stories typically consisted of a brief report about trying something new (e.g., trying out a new hobby and feeling good or motivated as a result). Negative events, in contrast, consisted of failed plans

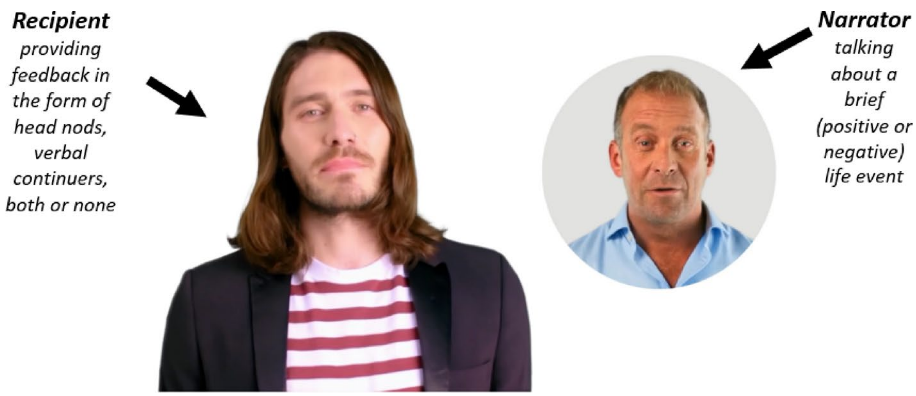


**Fig. 1** Example Stimuli of Interacting Avatars

or missed appointments with moderately negative outcomes and a certain dissatisfaction with the event. All stories are expected to engender a moderate emotional response but should not cause an excessive or disturbing emotional response. A research assistant and myself assessed each story independently for plausibility and made selective changes where necessary (for instance, when the wording was unusual).

Then, avatar interactions were generated by using D-ID's Video Studio (www.d-id.com; based on GPT-3.5-turbo; OpenAI, 2023) and Synthesia (www.synthesia.io; Express-1 model; Synthesia, 2024). Four videos were (successfully) generated completely with Synthesia by using their graphical user interface. First, two avatars (from a random but balanced selection of avatars with different gender and ethnicity) were positioned next to each other. For the purpose of this study, facial emotion expressions were kept neutral to avoid interference with this important social signal. Then, one of the story texts (described above) was entered and attributed to the narrator and a voice was selected for the narrator from the voice database included in Synthesia. The story was then split into several parts and the behaviour of the recipient was defined between the single textual paragraphs, either by adding verbal continuers (such as *Mhm*, *Oh*, *Aha*, *Ok*) during these brief pauses between segments of the story or by defining the nodding behaviour in the same positions.

Synthesia allows for the specification of different nods for a subset of their avatars by adding a nodding symbol to the script. In each story, the respective feedback signal (or signals, in the case of co-occurring nods and continuers) was added eight times during the story at different positions (for instance, either after a full sentence or between two coordinate sentences). In each video, the recipient showed the feedback signal for the first time after the first sentence and each trial concluded with the respective feedback signal. Length of the nodding or the continuer (e.g., *Mhm* vs. *Mhmmm*) was altered to make the feedback response appear more natural. Head nods also continued briefly after a turn, that is, when the narrator continues the story. While this procedure does not allow for a full replication of natural feedback behaviour, it offers a straightforward way to generate approximations of feedback behaviour that allow for the experimental assessment of functional differences between signals.

Due to technical issues and strict video quality control mechanisms, only four videos could be generated using this straight-forward approach with one tool alone. Therefore, to generate an additional 20 videos I had to try another approach and first generated only the feedback behaviour of the listener with Synthesia (that is, I basically only generated videos of the recipient alone). In a second step, I then generated video sequences of the narrators telling a life event using D-ID's video studio. During video generation with this tool, it is possible to merge the newly generated avatar sequence with previously generated videos, basically in form of an overlay of both videos. Despite the extra workload, an advantage of this procedure is that the feedback signals overlap more with the speech of the narrator, thereby creating an even more natural appearance of the interaction, especially in the case of verbal continuers. Also, with this procedure it was possible to control the volume of the verbal continuers more. Besides that, I did not further control for turn-taking or the behaviour of the narrator, and I did not collect ratings on how natural the interaction was perceived in this exploratory study.

### Sample and Procedure

For this exploratory web-based study, 30 monolingually raised speakers of German (living in Germany) were recruited online via Prolific (www.prolific.co), using this tool's pre-screening function to only select participants matching the language and nationality criteria. The sample had a mean age of 34.03 (*SD* = 6.87) with the majority self-identifying as (cisgender) men (*n* = 20). In this (admittedly small) sample, no participants identified as transgender, nonbinary, or intersex according to a brief demographic questionnaire before the experiment.

After registering for the study on Prolific, participants were automatically directed to this web-based study, that was developed with Gorilla's task builder (version 2; www.gorilla.sc) and also hosted on their servers. Before starting the actual experiment, participants had to provide informed consent and then filled in basic demographic questions (age, gender, and whether they were monolingually raised in German). After that, they saw the instructions. Participants were told to watch videos with interactions of two people and were instructed to imagine that these two people are interacting with each other, irrespective of the fact that both avatars were facing the viewer. In each trial, participants had to start the video by pushing a play button (I opted for this manual start option because Gorilla's documentation advises against using autoplay because it is blocked by some browsers).

Underneath each video, there was a slider element, ranging from 0 (*"gar nicht zugewandt"* / 'not attentive/affiliative') to 100 (*"zugewandt"* / 'attentive/affiliative'), with the starting position at 50. While watching the avatars interact, participants were asked to continuously rate how affiliative they perceived the recipient's behaviour to be in response to the narrator ("*Sie sollen zunächst bewerten*, *wie sehr die zuhörende Person in einem sozialen Sinne zugewandt zur erzählenden Person ist*.", 'Please assess to what extent the listening person is socially attentive toward the person telling the story'). Participants were instructed to immediately start moving the slider after they started the video and then continuously make use of it until the video ended. Gorilla's slider element can be adjusted in a way that it stores the slider rating every 500 ms (or larger time ranges) – resulting in a continuous, time-series rating.

After each video, participants were also asked to provide ratings on 7-point Likert scales, ranging from zero ("überhaupt nicht", 'not at all to') to six ("sehr", 'very much') on how socially adequate they found the recipient's behaviour ("Bewerten Sie, wie sozial angebracht Sie das Verhalten der zuhörenden Person empanden.", 'Please rate how socially appropriate you found the behavior of the listening person.') and how empathic they evaluated the recipient to be ("Wie empathisch nahmen Sie die zuhörende Person wahr?", 'How empathetic did you perceive the listening person to be?'). In each question, the term "zuhörende Person" (lit. 'listening person' instead of 'listener') was used as a gender-neutral form to refer to the recipient of the story in German.

This study has a 2 × 4 within-subject design with the two variables, emotional valence of the event story (*Positive Event* or *Negative Event*) and feedback type (*Head Nods Only*, *Verbal Continuer Only*, *Both Feedback Signals* or *No Feedback Signal*). Each subject saw all 24 videos, with three videos in each of the six conditions. After watching and rating each of the videos, participants were thanked for their participation and redirected to Prolific where they were reimbursed for their participation.

### Data Preprocessing and Analysis

All data pre-processing and analyses were conducted using *R* (R Core Team, 2023). For social adequacy and empathy, means and standard deviations were calculated based on the raw values. Mean affiliation was calculated by averaging all single values from the continuous ratings over time and then calculating means and standard deviations based on the average values of each subject. Also, I calculated Pearson's correlations for these three variables. To compare the continuous affiliation ratings across subjects, the raw data had to be normalized to account for different onsets of the ratings (for instance, when subjects were slightly faster or slower in using the slider or when they simply waited for the interaction to unfold a little longer) and for differences in making use of the slider continuously (since some subjects only move the slider a couple of times – resulting in only a few measures at specific time points – while others constantly moved the slider – resulting in relatively fine-grained and constant ratings). To account for these differences, I divided each timepoint a response was recorded by the final (i.e., maximum) timepoint of the last rating, thereby re-scaling the time variable into a relative or normalized time. Here, each rating is then bound to its relative time point in the rating process and can be expressed (and visualized) in relation to its relative timing of rating the video – providing an approximation to make ratings comparable.

In a next step, I calculated linear mixed effects models using the *lmer()* function of the *lme4* package (Bates et al., 2015). For mean affiliation, social adequacy, and empathy, I calculated models with the fixed factors valence and feedback type and random effects for item and subject. For pairwise comparisons, I calculated estimated marginal means for feedback (since this factor showed a significant main effect for each dependent variable) using the *emmeans* package (Lenth, 2021). In addition, I calculated pairwise comparisons for feedback split by valence. Note that these follow-up analyses were conducted even when the interaction term was not significant, as the exploratory nature of the present research made it informative to examine potential patterns across valence. For the same reason, no adjustment for multiple comparisons was applied. For (continuous) affiliation, I calculated a linear mixed effects model with the same fixed and random effects as in the previous models but additionally added (normalized) time as a fixed factor. To compare estimated slopes as pairwise contrasts, I used the *emtrends()* function of the *emmeans* package.

## Results

### Means and Standard Deviations for Each Study Variable

Means and standard deviations for mean affiliation, social adequacy, and empathy are provided in Table 1. As can be seen there, trials with head nods only received the highest ratings on each of these dimensions, followed by trials with both feedback signals. Trials with verbal continuers only received comparatively lower ratings, but overall higher than avatars that displayed no feedback at all. Regarding valence, the pattern is less uniform. For head nods (that overall received the highest ratings) and trials without feedback signals (that received mainly lower ratings), the differences across valence are relatively negligible. In

**Table 1** Means and standard deviations of each study variable by condition

| Variable | Mean Affiliation[a] | | Social Adequacy[b] | | Empathy[b] | |
|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD |
| No Feedback Signal | 37.85 | 18.17 | 2.64 | 1.61 | 1.76 | 1.43 |
| Positive Event | 38.11 | 16.87 | 2.83 | 1.57 | 1.86 | 1.39 |
| Negative Event | 37.59 | 19.46 | 2.44 | 1.65 | 1.66 | 1.47 |
| Head Nod Only | 57.06 | 20.10 | 3.88 | 1.45 | 3.31 | 1.56 |
| Positive Event | 55.75 | 19.97 | 3.96 | 1.39 | 3.27 | 1.56 |
| Negative Event | 58.38 | 20.26 | 3.81 | 1.52 | 3.34 | 1.57 |
| Verbal Continuer Only | 45.00 | 21.79 | 2.45 | 1.75 | 2.37 | 1.70 |
| Positive Event | 38.62 | 20.76 | 1.80 | 1.58 | 1.89 | 1.60 |
| Negative Event | 51.38 | 21.01 | 3.10 | 1.67 | 2.86 | 1.66 |
| Both Feedback Signals | 52.72 | 22.73 | 3.05 | 1.83 | 2.96 | 1.88 |
| Positive Event | 49.16 | 23.02 | 2.60 | 1.85 | 2.72 | 1.92 |
| Negative Event | 56.29 | 21.98 | 3.50 | 1.70 | 3.19 | 1.82 |

[a]Group average calculated from the overall average over time of the continuous rating (on a scale from 0 to 100) of each participant. [b]Rated on a 7-point Likert scale ranging from 0 to 6

contrast, in the case of verbal continuers and both feedback signals, ratings on each dimension were higher for negative events compared to positive events.

Regarding the distinctiveness of each measure, it should be noted here that the three socio-affective rating variables are not statistically independent of each other. The average of the continuous affiliation rating (averaged over time by subject and over subject) is significantly correlated with social adequacy ($r$=.75***) and empathy ($r$=.80***). Also, the two last-mentioned variables were positively correlated ($r$=.79***) with each other.

The mean continuous affiliation ratings in relation to the normalized time-series are plotted in Fig. 3. On average, the affiliation ratings increased while watching the videos in the case of head nods and both feedback signals. This is true for both positive and negative events, but the effect is more pronounced in the latter case. The decrease toward the end might reflect wrap-up effects when using the slider. For trials with no feedback signals or verbal continuers in the case of positive events, the ratings decreased continuously and followed comparable trajectories over the course of watching the videos. In other words, the differences between head nods and both feedback signals, and between these two and verbal continuers as well as no feedback signal, remain relatively stable, with the latter two not differing from each other.

Special attention should be given to the case of verbal continuers in response to negative events. After a slower start, with a slight lag, videos in this condition were gradually rated as more affiliative compared to no feedback signals or verbal continuers in response to positive events, and the difference from head nods and both feedback signals became much smaller at later time points (ultimately resulting in a comparable wrap-up as for these other two conditions). This more specific result may indicate some valence-based divergence in the case of verbal continuers, suggesting a more nuanced socio-affiliative function.

## Main and Interaction Effects of Feedback Type and Valence on Mean Affiliation, Social Adequacy, and Empathy

The results of the significance tests for the linear mixed effects models for mean affiliation, social adequacy and empathy are provided in Table 2. Type of feedback had a significant

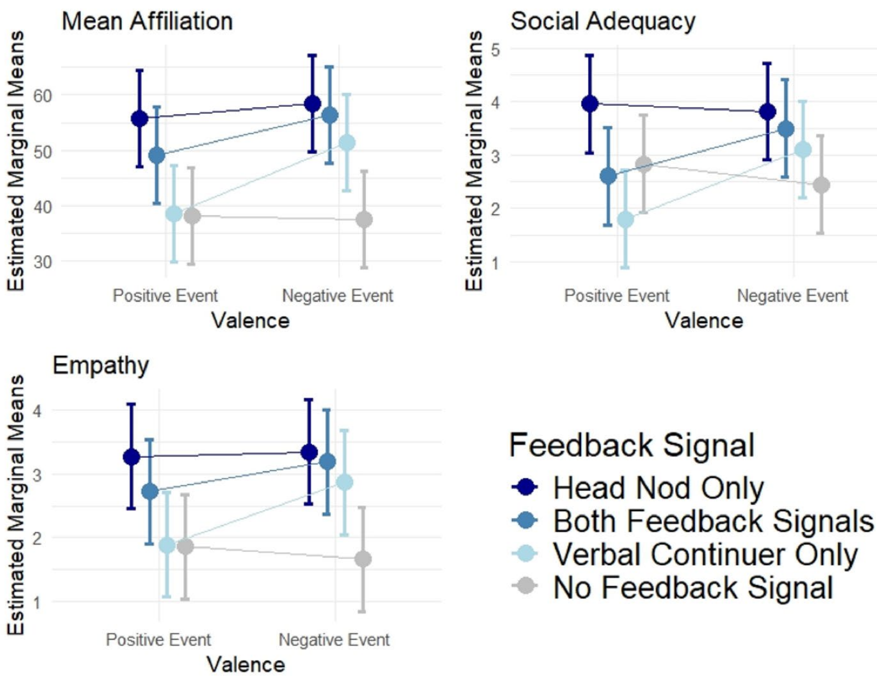


**Fig. 2** Visualization of Estimated Marginal Means for Each Linear Mixed Effects Model. Mean Affiliation = Group average calculated from the overall average over time of the continuous rating (on a scale from 0 to 100) of each participant. Error bars represent the 95% confidence intervals of the estimated marginal means

**Table 2** Analysis of variance for the linear mixed effects models for each study variable by condition

| Predictor | Mean Affiliation[a] | | | Social Adequacy | | | Empathy | | |
|---|---|---|---|---|---|---|---|---|---|
| | df | $\chi^2$ | *p* | df | $\chi^2$ | *p* | df | $\chi^2$ | *p* |
| Main Effects | | | | | | | | | |
| Feedback | **3** | **34.50** | **<0.001*** ** | **3** | **14.59** | **0.002** ** | **3** | **21.66** | **<0.001*** ** |
| Valence | **1** | **4.83** | **0.028*** | 1 | 2.08 | 0.149 | 1 | 1.72 | 0.190 |
| Interaction | | | | | | | | | |
| Feedback*Valence | 3 | 3.98 | 0.264 | 3 | 5.90 | 0.117 | 3 | 3.02 | 0.389 |

Significant effects marked in bold. [a]Group average calculated from the overall average over time of the continuous rating (on a scale from 0 to 100) of each participant

main effect on each of the ratings while the emotional valence of the stories only had a significant main effect on mean affiliation. The interaction of both variables was not significant. Estimated marginal means are plotted for visualization in Fig. 2.

For mean affiliation, pairwise comparison of estimated marginal means for feedback split by valence (given the significant interaction of both variables in this case) revealed that in the case of positive events, the contrasts between head nods and verbal continuers (mean difference = 17.13, SE = 5.01, $t(16) = 3.420$, $p = .017$*) and head nods and no feedback signal (mean difference = 17.64, SE = 5.01, $t(16) = 3.522$, $p = .014$*) was significant. In contrast, for negative events, also the contrast between head nods and no feedback signal (mean difference = 20.79, SE = 5.01, $t(16) = 4,149$, $p = .004^{**}$) as well as the contrast between both feedback signals and no feedback signal (mean difference = 18.70, SE = 5.01, $t(16) = 3.732$, $p = .009^{*}$) were significant.

For social adequacy, pairwise comparison of estimated marginal means for feedback revealed that the contrasts between head nods and verbal continuers (mean difference = 1.435, SE = 0.408, $t(16) = 3.514$, $p = .014^{*}$) and head nods and no feedback signal (mean difference = 1.244, SE = 0.409, $t(16) = 3.044$, $p = .035^{*}$) were responsible for the main effect of feedback. When calculating pairwise contrasts for feedback split by valence, these contrasts were significant only in the case of positive events with a mean difference of 2.156 (SE = 0.578, t(16) = 3.732, $p = .009^{**}$) for head nods contrasted with verbal continuers.

For empathy, the contrast between head nods and no feedback signal (mean difference = 1.547, SE = 0.357, $t(16) = 4.331$, $p = .003^{**}$) and between both feedback signals and no feedback signal (mean difference = 1.197, SE = 0.357, $t(16) = 3.351$, $p = .019^{*}$) was significant. When comparing contrasts split be emotional valence of the story, both significant contrasts were only present for negative events, with a mean difference of 1.682 for the contrast between head nods and no feedback signal (SE = 0.505, $t(16.1) = 3.330$, $p = .020^{*}$) and a mean difference of 1.527 for the contrast between both and no feedback signals (SE = 0.505, $t(16.1) = 3.022$, $p = .036^{*}$).

## Interaction Effects with (Normalized) Time

The results of the linear mixed effects model with continuous affiliation ratings as the dependent variable and feedback type, valence and (normalized) time as the fixed factors are presented in Table 3. As with the static ratings (and also mean affiliation), there was a significant main effect of feedback type but no main effect of valence. In addition, there was a main effect of normalized time on affiliation rating (for a visualization of the average time course of each rating, see Fig. 3). Also, there was a significant two-way interaction of feedback and time as well as a significant three-way interaction of the three fixed factors.

To examine further how the rating changes over time within each combination of feedback and valence, I assessed the estimated slopes of normalized time for each feedback type disaggregated by valence. Also, I examined contrasts pairwise between feedback types, again disaggregated by valence (given the significance of the three-way interaction). In the case of positive events, ratings increased significantly over time in the case of head nods (b = 5.49, SE = 1.67, 95% CI [2.22, 8.76]), but decreased in trials with *only verbal continuers*

**Table 3** Analysis of variance for the linear mixed effects models for the continuous affiliation rating

| Predictor | Affiliation[a] | | |
|---|---|---|---|
| | df | $\chi^2$ | *p* |
| Main Effects | | | |
| Feedback | **3** | **25.38** | **< 0.001*****|
| Valence | 1 | 2.87 | 0.090 |
| Time[b] | **1** | **24.57** | **< 0.001***** |
| Two-way Interactions | | | |
| Feedback*Valence | 3 | 1.66 | 0.646 |
| Feedback*Time[b] | **3** | **93.29** | **< 0.001***** |
| Valence*Time[b] | 1 | 3.10 | 0.078 |
| Three-way Interaction | | | |
| Feedback*Valence*Time[b] | **3** | **67.83** | **< 0.001***** |

Significant effects marked in bold. [a]Continuous ratings of affiliation (on a scale from 0 to 100) of each participant. [b]Time was normalized within subject for each trial by dividing each time point by the maximum time recorded for the respective trial, resulting in a proportional time metric ranging from 0 to 1

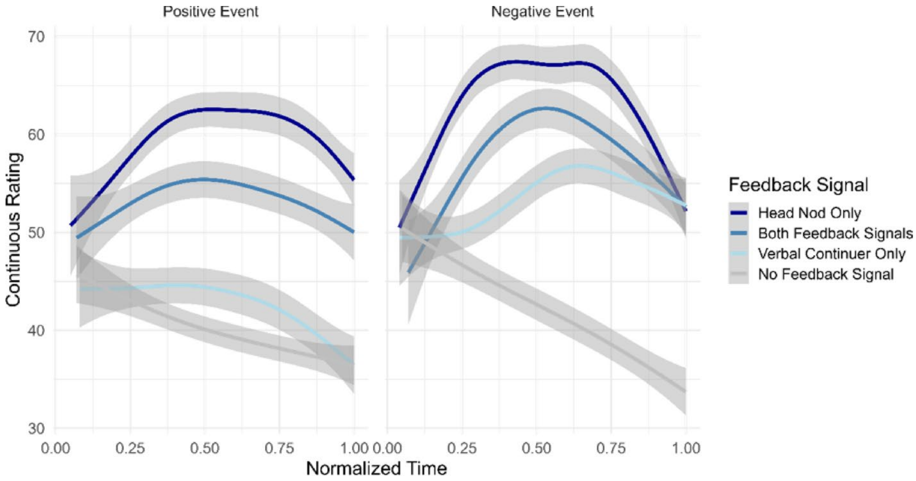


**Fig. 3** Continuous Rating of Perceived Affiliation between Interlocutors (by Valence and Feedback). Solid lines represent smoothed regression lines calculated from the continuous ratings. Time was normalized within subject for each trial by dividing each time point by the maximum time recorded for the respective trial, resulting in a proportional time metric ranging from 0 to 1. Shaded areas represent the 95% confidence intervals based on the standard errors

(b = -10.87, SE = 1.69, 95% CI [-14.17, -7.56]) or *no feedback* (b = -9.00, SE = 1.75, 95% CI [-12.43, -5.57]). For negative events, ratings increased over time for verbal continuers (b = 4.83, SE = 1.48, 95% CI [1.92, 7.73]) but decreased for trials with no feedback (b = -16.57, SE = 1.69, 95% CI [-19.88, -13.27]).

In the case of positive events, pairwise comparisons of each contrast between feedback types, disaggregated by valence and normalized time revealed significant differences between all feedback types (all $ps < 0.005$**), except for the contrast between both feedback signals and no feedback ($p = .057$) and verbal continuers and no feedback ($p = .867$). For negative events, the contrast between head nods and verbal continuers was significant ($p = .050$*), the contrasts between no feedback and all other feedback types were highly significant (all $ps < 0.001$***) and the remaining contrasts, between head nods and both feedback types and both feedback types and verbal continuers was not significant.

## Discussion

In this exploratory experimental study, I investigated whether a recipient is evaluated differently in terms of affiliation, social adequacy and empathy depending on the feedback signals they use when listening to a slightly emotional narration of a life event by another person. In particular, I examined whether head nods, contrasted with the use of verbal continuers or no feedback signal and a combination of head nods and verbal continuers together alter the perception of the socio-affective behaviour of the recipient. To address this research question, I used an experimental design with videos of interacting avatars, generated with AI-based image-to-video tools, that focusses on the evaluation of a perceived (and simplified) interaction.

Despite limitations in capturing interaction fully naturally, this experimental work supports the idea that head nods and verbal continuers differ in their socio-affective function. On average, avatar recipients displaying only head nods or both feedback signals in response to an avatar narrating a life event were rated as behaving more affiliative, more socially adequate and more empathic. Head nods received the highest ratings on each dimension and ratings for this feedback signal were highly comparable for positive and negative events. Also, continuous ratings of how affiliative the feedback-providing avatar was perceived increased while watching each video (and only decreased toward the end of each rating, arguably due to wrap-up or conclusion effects). That is, head nods are apparently perceived as a socially adequate feedback response when listening to a narrator telling an emotional life event and lead to the recipient being perceived as more affiliative and more empathic.

Trials with both feedback signals followed a similar pattern but the continuous rating was smaller in comparison to head nods. Importantly, there was also a within-condition difference. Continuous, mean and static ratings were higher for both feedback signals in response to negative events compared to positive events. A similar difference was found in the case of verbal continuers only. While avatars in this condition were rated as less affiliative (on average), less empathic and less socially adequate, ratings were higher in the case of negative events (and only slightly smaller than for trials with both feedback signals). This difference is even more striking in the case of the continuous affiliation ratings. For positive events, ratings gradually decreased in the course of the narration while for negative events, ratings slowly but continuously increased (but also decreased toward the end due to the aforementioned wrap-up effects). That is, while providing only verbal feedback without any specific head movement appeared to be relatively inadequate and socially non-affiliative, this type of feedback seems to be more acceptable or expectable when responding to a negative life event. This difference might arguably occur because it could be tolerable not to express affiliation and stance-taking with a negative event when it is not necessarily possible to truly “share” or grasp it fully. Still, this behaviour appears to be less expected, given the flatter slope in comparison to trials where verbal continuers were accompanied by head nods.

Responding without any feedback was clearly dispreferred, irrespective of the event’s valence. Avatars that did not provide any explicit feedback were rated as being drastically less empathic, behaving less socially adequate and also being less affiliative. Even more, continuous affiliation ratings decreased across emotional valence.

The findings of this pilot study are in line with previous claims in the literature that head nods and verbal continuers are different with respect to their socio-affective function, especially to the degree to which they signal affiliation. In terms of interaction-structuring functions, both feedback signals indicate affirmation and approval with what an interlocutor says and usually invite the narrator to continue their story (Aoki, 2011; Cerrato, 2005; McClave, 2000; Schegloff, 1982). In terms of socio-affective functions, it was argued that head nods license stance-taking and convey affiliation with the reported event (Stivers, 2008) and can express affiliation even when the listener does not agree with the narrator (Muntigl et al., 2012). In contrast, verbal continuers were described as weaker and arguably non-affiliative responses (Selting, 2017; Stivers, 2008).

The findings of the present study add to this perspective and suggest, at least tentatively, that verbal continuers do not exhibit a socio-affective function to the same extent as head nods. Interestingly, even when verbal continuers were combined with head nods, ratings did not exceed those for head nods only – suggesting that head nods and verbal continuers do not accumulate with respect to socio-affective meaning. It is interesting to note however, that the socio-affective function of verbal continuers was also dependent on the emotional valence of the story of the narrator (while head nods were largely unaffected by valence). Noteworthily, verbal continuers are perceived of as more adequate when used as feedback in response to negative events and in this particular case, verbal continuers apparently also express affiliation to some degree, a finding that requires further investigation.

## Limitations and Future Avenues

In the present contribution, I presented the results from an oversimplified, exploratory study to investigate a proclaimed functional difference of head nods and verbal continuers. Besides some more general experimental limitations (in particular, a very limited number of participants and items due to the exploratory nature of the present research), there are specific limitations with respect to the stimuli material.

This study fully relied on ready-made AI-based text-to-video and image-to-video tools that are available online. While these tools are relatively simple to use and allow for the fast generation of a high number of stimuli without requiring any knowledge of coding, they only allow for limited control over some relevant parameters. The number of avatars is limited, reducing the number of items with different individuals that can be generated (but with some tools it is usually possible to also generate avatars based on text-to-image prompting, allowing for the generation of more individuals). In contrast, most tools allow for the selection of a relatively large number of voices or allow for the integration of audio files. Therefore, it is also possible to first generate audio files with other (text-to-speech) tools and then streamline audios and images into a video generation tool – allowing for a more creative generation of stimuli.

In addition, all tools usually apply some pre-defined movement drivers that determine the basics of how the avatars move (in terms of body and facial movement as well as micro-gestures). These drivers are usually trained on recordings or videos showing human beings and should express proto-typical, average movements. Nonetheless, regarding feedback in interaction research, these micro-movements might affect how the avatars are perceived. Similarly, facial emotional expressions were kept neutral, fully relying on the built-in specification parameters. Nevertheless, it would clearly be highly relevant to consider this social signal and its interaction with the feedback signals discussed here.

Another limitation concerns the length and design of the videos with respect to the usual frequency of occurrence of the feedback signals under discussion. Practical considerations for experimental testing were prioritised in the present study and the videos were not specifically designed in terms of the usual length of an interaction or the frequency of occurrence of the feedback signals being studied. Consequently, the results of this study cannot be interpreted as directly reflecting the realities of natural interactions but rather they provide insight into functional mechanisms that underlie feedback behaviours found in real-life social exchanges. Even more, the experimental approach used here—particularly the controlled, avatar-based design—may serve as a useful starting point for future studies with higher ecological validity.

Besides these stimuli-inherent aspects, it is clearly a downside that many tools apparently rushed to market without fully developing a broadly applicable toolbox that fits all needs of social-interactional research. Most tools available online are relatively expensive but only offer limited support (in one case, it was basically impossible to get in touch with the actual developers because e-mails were automatically answered by an AI bot). Also, there is often no version control available – meaning that (potentially unexpected) upgrades might affect the way stimuli are produced or change the data basis generations are based on (meaning that images of avatars might appear slightly but noticeably different after an update, rendering it useless for application in sensitive experimental designs, for instance, studies with neurophysiological measures). Occasionally, generation of images or videos is relatively unstable without a determinable pattern based on prompts used.

Given some of these technical limitations (and as outlined in the methods section), I could not create all stimuli of this study using exactly the same workflow. In some cases, I had to rely on manually integrating into one video generations of two avatars that were produced independently of each other. In regard of feedback signals – that often overlap with the verbal behaviour of a narrator – this might be acceptable enough to study an overall research question as in the present study. Yet, these limitations clearly limit the comparability across items.

I did not control for (or evaluate in a pre-study) alignment or turn-taking between the avatars, and I only roughly matched feedback to each story based on subjective intuition (only independently checked by another person) and within the technical limits described above. In general, research of this kind should control for, or at least assess, the (degree of) naturalness of the interaction through pre-testing, in order to ensure comparability across items. This applies not only to the full stimuli but also to single components, such as the narratives or voices.

In particular, with regard to the behaviour of verbal continuers, unnaturalness could in theory provide an alternative explanation for the lower ratings with this feedback signal. While unnaturalness is certainly a relevant factor that needs to be controlled for in future research, I still believe that the difference across condition holds in this case since the combined condition received ratings only slightly lower than those for head nods alone, rather than almost as low as the verbal-continuer–only condition. Also, using a within-subjects design with different conditions presented in the same way still allows for capturing functional differences between feedback signals experimentally, even though this approach necessarily simplifies the complexity of real interactions.

While (ready-made) AI-based text-to-video technology that has recently become available offers innovative and affordable solutions to generate large numbers of stimuli items for social-interactional research and can be straightforwardly be used to examine broader functions of feedback signals, it is probably even more beneficial for social-interactional research to use and adapt transformer models directly, that is, without using third-party platforms. Using models locally usually also allows for a more fine-grained control over relevant parameters. Especially regarding feedback signals, movement control parameters should be used more strictly and alignment between the avatar interlocutors be matched more closely to mimic human interaction in an even more naturalistic sense. Most likely, more powerful text-to-video generation tools that are currently becoming available and that are not restricted to the generation of avatars alone will probably allow for even more complex stimulus material. Nonetheless, this study demonstrates that major functions of feedback and other socially relevant feedback signals can successfully be investigated also with simpler and more easily accessible tools.

Regarding statistical analysis, two limitations should be noted. First, interactions between feedback and valence were significant only for the continuous ratings, but not for the static ratings. Therefore, the exploratory follow-up analyses on these factors should be interpreted with caution, and the tentative valence differences discussed here require further investigation in the future. Second, the rating procedure used may not have been the most appropriate solution, as one anonymous reviewer rightly pointed out. As suggested, it might be more beneficial in future work to segment the timeline and calculate ratings as categorical variables, or to apply models that capture the apparent U-shaped distribution over time more adequately. While this points to an interesting avenue for future research, these methodological issues do not undermine the main effect highlighted here—namely, the functional differences in how feedback signals differentially contribute to affiliation in interaction.

## Conclusion

In this exploratory experimental study, I investigated how affiliative, empathic and socially adequate the behaviour of a recipient in an interaction is perceived depending on the type of feedback signals they provide in response to listening to emotional life events told by a narrator. More specifically, I focused on the two feedback signals head nods and verbal continuers and experimentally investigated their proclaimed functional difference with respect to expressing social affiliation. Unlike previous research focusing on human – human interaction, I addressed this research question by asking study subjects to rate the behaviour of avatars in videos that were generated with AI-based image-to-video tools – approximating natural interaction in an experimental design.

This study supports previous claims that head nods and verbal continuers differ in their socio-affective function. Avatars that responded with head nods (or both feedback signals) while listening to the narrator were rated as behaving more affiliative, more socially adequate and more empathic in contrast to avatars that only used verbal continuers or no feedback signal. Yet in the case of negative events, verbal continuers were rated as comparably more adequate and the behaviour of the avatar as more affiliative and empathic in comparison to using this feedback signal in response to positive events. Nonetheless, ratings were below the level of head nods on any measure.

Despite methodological limitations, this experimental study provides evidence for a difference in the socio-affective function expressed by head nods and verbal continuers. Also, it illustrates that AI-generated avatars and videos displaying these avatars are a useful addition to research into social interaction and can also be successfully used to shed light on the use and function of feedback signals and other socially relevant signals.

**Acknowledgements** I would like to thank Laura Müller for her assistance in generating the stimulus material of the present study, and the participants of the LingCologne2025 conference for their valuable feedback and insightful comments.

**Author Contributions** P.C. conceptualised, designed, and conducted the study, analysed the data, and wrote the manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL. This study received financial support by the Emerging Talent's Initiative of the University of Erlangen-Nuremberg as part of the seed funding to Paul Compensis for the project "Individual differences in empathy and social orientation: Evidence from language behaviour and neurophysiology".

**Data Availability** The data are available from the author upon reasonable request.

## Declarations

**Competing interests** The authors declare no competing interests.



## References

Aburumman, N., Gillies, M., Ward, J. A., & Hamilton, A. F. C. (2022). Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions. *International Journal of Human-Computer Studies*, *164*, 102819. https://doi.org/10.1016/j.ijhcs.2022.102819

American Psychological Association. (n.d.). Affiliation. In *APA dictionary of psychology*. Retrieved May 31 (2025). from https://dictionary.apa.org/affiliation

Compensis, P. (in press). Assessing the integration of social information using AI-generated videos: A brief research report. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2025.1505843

Compensis, P. (2025). *Low agency in reported negative events increases empathic response but decreases memory of AI-generated avatars*. A pilot study. Manuscript submitted for publication.

Aoki, H. (2011). Some functions of speaker head nods. In J. Streeck, C. Goodwin, & C. LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 93–105). Cambridge University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Batson, C. D. (2011). These things called empathy: Eight related but distinct phenomena. In J. Decety, & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 3–16). MIT Press.

Bauer, A., Gipper, S., Herrmann, T. A., & Hosemann, J. (in press). Rethinking linguistic feedback: A modality-agnostic and holistic approach to multimodal addressee signals in spoken and signed dyadic interaction. *Glossa*.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, *52*(3), 566–580. https://doi.org/10.1111/j.1460-2466.2002.tb02562.x

Busso, C., Deng, Z., Grimm, M., Neumann, U., & Narayanan, S. (2007). Rigid head motion in expressive speech animation: Analysis and synthesis. *IEEE Transactions on Audio Speech and Language Processing*, *15*(3), 1075–1086.

Cerrato, L. (2005). Linguistic functions of head nods. In J. Allwood, B. Dorriots & S. Nicholson (Eds.), *Proceedings from the Second Nordic Conference on Multimodal Communication* (pp. 137–152). Göteburg University. https://gup.ub.gu.se/publication/36429

Cerrato, L. (2007). *Investigating communicative feedback phenomena across languages and modalities* [Doctoral dissertation, KTH Royal Institute of Technology]. KTH DiVA. https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A11962&dswid=2500

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144–153. https://doi.org/10.1177/1754073914558466

De Stefani, E. (2021). Embodied responses to questions-in-progress: Silent nods as affirmative answers. *Discourse Processes*, *58*(4), 353–371. https://doi.org/10.1080/0163853x.2020.1836916

de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences*, *10*(10), 435–441. https://doi.org/10.1016/j.tics.2006.08.008
Dittmann, A. T., & Llewellyn, L. G. (1968). Relationship between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology*, *9*(1), 79–84. https://doi.org/10.1037/h0025722
Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283–292. https://doi.org/10.1037/h0033031
Enfield, N. J. (2008). Common ground as a resource for social affiliation. In I. Kecskes, & J. L. Mey (Eds.), *Intention, common ground and the egocentric Speaker-Hearer* (pp. 223–254). Mouton de Gruyter.
Freiermuth, M. R., & Hamzah, N. H. (2023). I agree! Empathetic head-nodding and its role in cultural competences development. *Lingua*, *296*, 103629. https://doi.org/10.1016/j.lingua.2023.103629
Groß, A., Dix, C., Ruusuvuori, J., & Peräkylä, A. (2024). Facial gestures in social interaction: Introduction to the special issue. *Social Interaction: Video-Based Studies of Human Sociality*, *6*(3). https://doi.org/10.7146/si.v6i3.142894
Guardiola, M., & Bertrand, R. (2013). Interactional convergence in conversational storytelling: When reported speech is a cue of alignment and/or affiliation. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2013.00705. *4*.
Hadar, U., Steiner, T. J., & Rose, F. C. (1985). Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, *9*(4), 214–228. https://doi.org/10.1007/BF00986881
Hall, J. A., & Schwartz, R. (2018). Empathy present and future. *The Journal of Social Psychology*, *159*(3), 225–243. https://doi.org/10.1080/00224545.2018.1477442
Hall, J. A., & Schwartz, R. (2022). Empathy, an important but problematic concept. *The Journal of Social Psychology*, *162*(1), 1–6. https://doi.org/10.1080/00224545.2021.2004670
Hein, G., & Singer, T. (2008). I feel how you feel but not always: The empathic brain and its modulation. *Current Opinion in Neurobiology*, *18*(2), 153–158. https://doi.org/10.1016/j.conb.2008.07.012
Lee, J., & Marsella, S. C. (2010). Predicting speaker head nods and the effects of affective information. *IEEE Transactions on Multimedia*, *12*(6), 552–562. https://doi.org/10.1109/tmm.2010.2051874
Lenth, R. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.11.1. https://CRAN.R-project.org/package=emmeans
Li, X. (2019). Negotiating activity closings with reciprocal head nods in Mandarin conversation. In E. Reber, & C. Gerhardt (Eds.), *Embodied activities in face-to-face and mediated settings: Social encounters in time and space* (pp. 369–396). Palgrave Macmillan.
Maynard, S. K. (1987). Interactional functions of a nonverbal sign: Head movement in Japanese dyadic casual conversation. *Journal of Pragmatics*, *11*(5), 589–606. https://doi.org/10.1016/0378-2166(87)90181-0
McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, *32*, 855–878. https://doi.org/10.1016/S0378-2166(99)00079-X
McCrackin, S. D., & Itier, R. J. (2021). Feeling through another's eyes: Perceived gaze direction impacts ERP and behavioural measures of positive and negative affective empathy. *Neuroimage*, *226*, 117605. https://doi.org/10.1016/j.neuroimage.2020.117605
Muntigl, P., Knight, N., & Watkins, A. (2012). Working to keep aligned in psychotherapy. *Language and Dialogue*, *2*(1), 9–27. https://doi.org/10.1075/ld.2.1.01mun
Murray, H. A. (1938). *Explorations in personality*. Oxford University Press.
Nguyen, L., Odobez, J. M., & Gatica-Perez, D. (2012). Using self-context for multimodal detection of head nods in face-to-face interactions. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)* (pp. 289–292). Association for Computing Machinery. https://doi.org/10.1145/2388676.2388734
OpenAI (2024). *GPT-4o* [Large language model]. https://platform.openai.com/docs/models/gpt-4o
OpenAI (2023). *GPT-3.5-turbo* [Large language model]. https://platform.openai.com/docs/models/gpt-3.5-turbo
Paggio, P. (2016). Coordination of head movements and speech in first encounter dialogues. In E. Gilmartin, L. Cerrato, & N. Campbell (Eds.), *Proceedings from the 3rd European Symposium on Multimodal Communication, Dublin, September 17–18, 2015* (Vol. 105, pp. 69–74). Linköping Electronic Conference Proceedings. http://www.ep.liu.se/ecp/105/012/ecp16105012.pdf
Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGPT and beyond: How generative artificial intelligence May affect research, teaching, and practice. *International Journal of Research in Marketing*, *40*(2), 269–275. https://doi.org/10.1016/j.ijresmar.2023.03.001
Poggi, I., D'Errico, F., & Vincze, L. (2010). Types of nods: The polysemy of a social signal. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (pp.2570–2576). European Language Resources Association. https://aclanthology.org/L10-1407/.

R Core Team. (2023). *R: A Language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Ruusuvuori, J., & Peräkylä, A. (2009). Facial and verbal expressions in assessing stories and topics. *Research on Language and Social Interaction*, *42*(4), 377–394. https://doi.org/10.1080/08351810903296499

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71–93). Georgetown University Press.

Selting, M. (2017). The display and management of affectivity in climaxes of amusing stories. *Journal of Pragmatics*, *111*, 1–32. https://doi.org/10.1016/j.pragma.2017.01.008

Shamay-Tsoory, S. G., & Hertz, U. (2022). Adaptive empathy: A model for learning empathic responses in response to feedback. *Perspectives on Psychological Science*, *17*(4), 1008–1023. https://doi.org/10.1177/17456916211031926

Sidner, C. L., Lee, C., Morency, L. P., & Forlines, C. (2006). The effect of head-nod recognition in human-robot conversation. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (pp. 290–296). Association for Computing Machinery. https://doi.org/10.1145/1121241.1121291

Stietz, J., Jauk, E., Krach, S., & Kanske, P. (2019). Dissociating empathy from perspective-taking: Evidence from intra- and inter-individual differences research. *Frontiers in Psychiatry*, *10*, 126. https://doi.org/10.3389/fpsyt.2019.00126

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, *41*(1), 31–57. https://doi.org/10.1080/08351810701691123

Synthesia (2024). *EXPRESS-1* [Large language model]. https://www.synthesia.io/post/expressive-avatars-powered-by-synthesias-new-express1-model-are-here

Yngve, V. H. (1970). On getting a word in Edgewise. In M. A. Campbell (Ed.), *Papers from the sixth regional Meeting, Chicago linguistics society* (pp. 567–578). University of Chicago.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, *19*(4), 399–404. https://doi.org/10.1111/j.1467-9280.2008.02099.x

Zhang, L. (2020). Coordination of Nods in Dialogue [Doctoral dissertation, Queen Mary University of London]. Queen Mary Research Online. https://qmro.qmul.ac.uk/xmlui/handle/123456789/71130