

Nods of Agreement: Webcam-Driven Avatars Improve Meeting Outcomes and Avatar Satisfaction Over Audio-Driven or Static Avatars in All-Avatar Work Videoconferencing

FANG MA, Goldsmiths, University of London, United Kingdom
JU ZHANG, Goldsmiths, University of London, United Kingdom
LEV TANKELEVITCH, Microsoft Research, United Kingdom
PAYOD PANDA, Microsoft Research, United Kingdom
TORANG ASADI, Microsoft, United States
CHARLIE HEWITT, Microsoft, United Kingdom
LOHIT PETIKAM, Microsoft, United Kingdom
JAMES CLEMOES, Microsoft Research, United Kingdom
MARCO GILLIES, Goldsmiths, University of London, United Kingdom
XUENI PAN, Goldsmiths, University of London, United Kingdom
SEAN RINTEL, Microsoft Research, United Kingdom
MARTA WILCZKOWIAK, Microsoft, United Kingdom

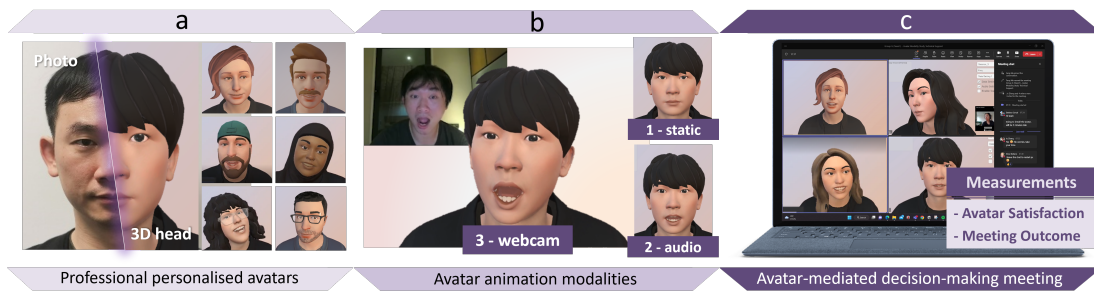


Fig. 1. An overview of the experimental study. a) A personalized 3D stylized avatar was generated for each participant, using their photos with facial landmarks, and preserving diverse personal features like beards, glasses, and hairstyles. b) Participants used their avatars in 3 modalities: static picture, audio-animated and webcam-animated, affecting their head movements and expressions (see supplementary video). c) Participants used the avatars in a videoconferencing meeting, during which they engaged in three group decision-making tasks, one for each avatar modality. Meeting outcomes and avatar satisfaction were measured after each task and compared.

Authors' Contact Information: **Fang Ma**, f.ma@gold.ac.uk, Goldsmiths, University of London, London, United Kingdom; **Ju Zhang**, j.zhang@gold.ac.uk, Goldsmiths, University of London, London, United Kingdom; **Lev Tankelevitch**, lev.tankelevitch@microsoft.com, Microsoft Research, Cambridge, United Kingdom; **Payod Panda**, payod.panda@microsoft.com, Microsoft Research, Cambridge, United Kingdom; **Torang Asadi**, torang.asadi@microsoft.com, Microsoft, San Francisco, United States; **Charlie Hewitt**, chewitt@microsoft.com, Microsoft, Cambridge, United Kingdom; **Lohit Petikam**, lohithpetikam@microsoft.com, Microsoft, Cambridge, United Kingdom; **James Clemoes**, jaclemoe@microsoft.com, Microsoft Research, Cambridge, United Kingdom; **Marco Gillies**, m.gillies@gold.ac.uk, Goldsmiths, University of London, London, United Kingdom; **Xueni Pan**, x.pan@gold.ac.uk, Goldsmiths, University of London, London, United Kingdom; **Sean Rintel**, serintel@microsoft.com, Microsoft Research, Cambridge, United Kingdom; **Marta Wilczkowiak**, mawilczk@microsoft.com, Microsoft, Cambridge, United Kingdom.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the

Avatars are edging into mainstream videoconferencing, but evaluation of how avatar animation modalities contribute to work meeting outcomes has been limited. We report a within-group videoconferencing experiment in which 68 employees of a global technology company, in 16 groups, used the same stylized avatars in three modalities (static picture, audio-animation, and webcam-animation) to complete collaborative decision-making tasks. Quantitatively, for meeting outcomes, webcam-animated avatars improved meeting effectiveness over the picture modality and were also reported to be more comfortable and inclusive than both other modalities. In terms of avatar satisfaction, there was a similar preference for webcam animation as compared to both other modalities. Our qualitative analysis shows participants expressing a preference for the holistic motion of webcam animation, and that meaningful movement outweighs realism for meeting outcomes, as evidenced through a systematic overview of ten thematic factors. We discuss implications for research and commercial deployment and conclude that webcam-animated avatars are a plausible alternative to video in work meetings.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**.

Additional Key Words and Phrases: videoconferencing, work, avatar animation modalities, satisfaction, outcomes, effectiveness, alignment, comfort, inclusivity, expression, perception, preference

ACM Reference Format:

Fang Ma, Ju Zhang, Lev Tankelevitch, Payod Panda, Torang Asadi, Charlie Hewitt, Lohit Petikam, James Clemoes, Marco Gillies, Xueni Pan, Sean Rintel, and Marta Wilczkowiak. 2025. Nods of Agreement: Webcam-Driven Avatars Improve Meeting Outcomes and Avatar Satisfaction Over Audio-Driven or Static Avatars in All-Avatar Work Videoconferencing. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW142 (April 2025), 28 pages. <https://doi.org/10.1145/3711040>

1 Introduction

In 2023, both Microsoft Teams [44] and Zoom [81] introduced stylised avatars for all users as an alternative to webcam video in their videoconferencing services. The ability to use avatars is a major change in representational affordances for users, especially in a work context, offering a crucial new choice for people who prefer not to use video or are unable to use video. From a commercial perspective, completely static avatars are obviously the cheapest to develop and use, albeit likely to have low user satisfaction. The currently available avatars in Zoom and Microsoft Teams are limited to viseme animation (audio-driven mouth positions) and some pre-designed animations, which limit the naturalness and fidelity of their expression compared with actual video. User satisfaction and meeting outcomes are likely to improve with more signals for animating avatars, but increased expressiveness is also associated with increased effort and time to develop at production quality. On the other hand, minimal animation also offers benefits, such as protecting user privacy and reducing bandwidth consumption in virtual meetings [2, 49]. Navigating the trade-off between increased expressiveness and the required development effort remains a challenge.

Previous studies have explored various aspects of avatar-mediated communication. While older studies supported the notion that stylized avatars are inappropriate for professional settings [32], recent studies of avatars in workplace and classroom settings show that this trend might be changing [51, 78]. One key result from recent avatar studies is that the visual fidelity of avatars is less relevant than their motion fidelity, which holds for both self-identification [27] and trusting other avatar participants in a meeting [49]. However, the link between avatar motion fidelity and work meeting outcomes remains unclear. More broadly, avatar research on meeting outcomes is

full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2573-0142/2025/4-ARTCSCW142

<https://doi.org/10.1145/3711040>

very limited [23]. Critically, both the scientific and commercially-driven questions are the same: **what are the differences between avatar animation modalities in terms of their impact on meeting outcomes and user preferences?**

To answer these research questions, we conducted a within-subjects, mixed-methods experiment in which 68 employees of a global technology company (in 16 groups), represented as stylized avatars (with a professional appearance), engaged in collaborative decision-making tasks akin to those in work meetings. The avatars all used the same illustration style, but had three modalities of animation, randomised between tasks in a counterbalanced order across groups: no animation (picture), audio-animation (visemes), and webcam-animation (visemes, facial expressions, and head movements).

Our quantitative results demonstrate that webcam-animated improves meeting outcomes in terms of meeting effectiveness, comfort, and inclusiveness. This is supported by qualitative feedback that mostly favours the head and face motion of webcam-animated avatars over voice-animated lip-syncing only or a static picture. Qualitative results are presented as a systematic thematic framework that allows for comparison of the effects of avatar animation modality preferences and meeting outcomes. We discuss the implications of these results for research and commercial deployment, arguing that webcam-animated avatars are a plausible alternative to video in work meetings, albeit with some nuances, and that visual fidelity is less important than motion fidelity for meeting outcomes. In summary, this paper's key contributions are:

- An empirical comparison of avatar animation modalities in terms of their impact on meeting outcomes and user preferences in a controlled study with ecological validity (participants were employees of a global technology company who had worked with one another, and tasks were comparable to work scenarios). This study, designed to replicate the intricacies of professional interactions, presented a unique challenge in recruitment due to the involvement of individuals from diverse corporate backgrounds.
- Quantitative and qualitative evidence for webcam-animated avatars providing the highest support for decision agreement in work meetings, suggesting that they are a plausible alternative to video in work meetings.
- A systematic thematic qualitative framework of ten factors of user responses about avatars in three categories, which allows for comparison of factors around user preferences versus those around meeting outcomes.

2 Background and Related Work

While video in online meetings is valued for its apparent richness of expression, its use is not without problems. Situational issues with using video include desiring privacy, confidentiality or anonymity, coping with low bandwidth, or an inability to use video due to a need for mobility [13, 16, 33]. Other issues are more personal, such as appearance anxiety [16, 35, 51] or physical, cognitive, emotional burden due to disability or neurodiversity [20, 66]. Many of these issues also contribute to videoconferencing fatigue [5, 55]. Video-on meetings have become a source of tension between those who wish to use it to establish inclusion [18], which involves rapport, presence, or attention, and those who cannot or will not use video [10, 36, 73]. One solution to expand beyond the binary choice of turning video on or off in online meetings is to use avatars for visual representation of participants [49]. Ratan et al. [53] suggest that avatars may reduce videoconferencing fatigue by occluding the user's video and thus reducing self-focused attention. However, the question remains as to how this will affect meetings themselves. In the following sub-sections, we review prior work which has investigated the appearance, realism, and animation of avatars, as well as their use in professional and collaborative settings (Section 2.1). We also

review relevant research focused on measuring aspects of avatar satisfaction and meeting outcomes (Section 2.2). Based on this, we propose our research questions and hypotheses (Section 2.3).

2.1 Avatars: Appearance, Animation, Professionalism, and Use in Collaboration

Appearance and realism. Research on avatars for meetings stretches at least from the late 1990s [21], picks up in the 2010s [1, 30, 32, 61], and retains a strong focus on evaluating the relationship between realism and presence [21, 38, 42, 72]. Research has found that increased avatar realism does not always correspond to heightened communicative efficacy [47], and that, in fact, there is a threshold in avatar realism beyond which the enhancement of social presence plateaus [48]. On the other hand, personalized avatars can foster a deep sense of self-identification and self-presence [68, 79].

Since realism is very hard to accomplish, customised stylized avatars, which blend abstraction with representation, are easier to create and deploy, but this depends on whether such avatars are regarded as appropriate for work scenarios. In the 2010s, workers were not ready for stylized avatars for work meetings [30, 32] and strongly preferred video or at least more realistic avatars. More recently, the trend appears to be shifting, although not without complications. Anecdotally, during the COVID-19 pandemic, individual experiments with add-ons such as virtual webcams and filters [63] that produce the effect of avatars for a specific user in a video conference had infamous unprofessional complications [8, 69]. However, in 2022, Dobre et al. [23] found that avatars do not necessarily require advanced realism for work meetings (in Mixed Reality). Users can become accustomed to cartoon avatars over a few days, and may even prefer them. Phadnis et al. [52] report a 2023 survey of 2,509 knowledge workers responding that while less realistic avatars may still have acceptance issues, modern realistic avatars also suffer from uncanny valley effects. They argue that the optimal avatar for knowledge work would be stylized with a professional appearance.

Animation. Beyond stylization and customization, other considerations include the ways in which an avatar moves and the signals which it uses. Although the cutting edge research into producing visual fidelity is exciting [41], outcome-oriented research on avatars has consistently shown that motion fidelity is more important than visual fidelity [70]. This holds for both self-identification [27] and trust in others in meetings [49]. Similarly, and in line with the current study's focus on meeting outcomes, decades of videoconferencing research has shown that while people like video, good audio is the most important *functional* communicative requirement [31, 64]. Video can be heavily perturbed, frozen, or non-existent and a video call will still be possible [56, 71]. As such, it should be possible for a conversation to occur between meeting participants with just a picture of a customized avatar as long as the audio is operating. This is very similar to the way that most commercial videoconferencing systems display a picture of audio-only participants in videoconferencing. Audio signals can quite easily drive animation in the form of mouth movement. This may range from simple lip-flapping [7], through to viseme shaping of the mouth [65], and even other facial features [40, 59]. Dobre et al. [23] used simple lip-flapping along with canned eye-blinking to achieve minimally naturalistic motion in their study of cartoon versus realistic avatars in workplace meetings.

Non-verbal movement is key to rich communication and core to our identity. Beyond the obvious component of facial expressions for displaying emotion [14], head motion turns out to be the next most important movement component, crucial for persuasion [74], understanding of talk content [45], and both personal and cultural identity [28]. While previous research has explored low-compute methods like IMUs on head-worn devices (like headphones [50]) to drive avatar head movement, using webcam-driven animation allows the use of facial anchors and blendshapes to drive mouth movement and facial expression [75]. Additionally, body pose for head motion is a

well-solved problem [80]. While more complex methods can animate near photorealistic avatars [41], these are still nascent and computationally costly, so simple animation is most scalable for commercial systems.

Professional settings. Previous studies on the use of stylized avatars in professional situations are limited, as are studies conducted with actual organizational employees (as opposed to students). Historically, stylized avatars have been considered inappropriate for work meetings. Users have reported being worried about avatars looking “unprofessional” (e.g. [32]). However, trends may be changing. For instance, a 2022 study by Yasuoka et al. [78] used a virtual meeting platform that offers avatars representing professional roles like secretaries and officers, exploring one-on-one managerial meetings with avatars. Abdullah et al. [2] explored the potential for personalized avatars in enhancing user engagement and facilitating more meaningful interactions in professional contexts. In that study, factors such as user preferences, psychological impact, and the establishment of a virtual identity that aligns with an individual’s professional persona were key to the effectiveness of avatars. As reported above, Dobre et al. [23] found that both cartoon and realistic avatars could be used to hold real work meetings in Mixed Reality, although they did not use a specific measure of success. Phadnis et al. [52] have surveyed 2,509 knowledge workers about avatar acceptance. They report that while less realistic avatars may still have acceptance issues, hyper-realistic avatars also suffer from uncanny valley effects, and thus argue that the optimal avatar for work contexts is an illustration with professional face, hair, and clothing styles.

Collaboration. In terms of collaboration, avatars with non-verbal cues such as facial expressions and gestures have been found to enhance brainstorming and negotiation tasks among younger users, suggesting that the interactivity and expressiveness of avatars can be beneficial in certain collaborative contexts [4]. Similarly, comparing communication patterns in videoconferences and embodied VR environments, research indicated that the medium of interaction influences the dynamics of communication. In VR settings, avatars can offer a more embodied and engaging experience, potentially leading to different outcomes in teamwork and collaboration [2].

2.2 Avatar Satisfaction and Meeting Outcome Factors

The final point above brings us to what we know about measuring the impact of avatar modalities in videoconferencing. While there has been significant prior research into meeting effectiveness, there is a lack of consensus on its precise measurement [29]. Further, many identified factors, such as meeting type, meeting hygiene, or specific meeting features, are not directly applicable to avatar animation modalities [3, 64]. Collaboration outcomes using avatars have been reported in previous research (e.g., [4]), but, again, there are no standardised metrics. As such, for the purposes of this study, we propose a set of factors drawn from meeting science, videoconferencing research, and avatar research, categorised into two sets: *meeting outcomes* and *avatar satisfaction*.

Meeting outcome factors. The two core considerations of meeting outcomes in all meeting types and communication modalities are decision agreement and alignment [46, 54, 67]. For simplicity of expression in this study, we term these *effectiveness* and *alignment*. Effectiveness is determined by questions about whether a group makes a final decision and whether any modalities make it easier or harder to reach agreement. Alignment is determined by questions about whether participants align with the group decision.

With respect to avatars and meeting outcomes, it has been found in research on both avatar self-presentation (e.g. [25]) and videoconferencing fatigue (e.g. [24, 55]) that a comfortable experience is key to engaging with others. For our study, we consider *comfort* to encompass ease of *expressing* oneself using a particular animation modality, and how *tiring* a given animation modality is.

The final factor pertaining to meeting outcomes is *inclusivity*. Inclusion in the meeting context [11, 18, 29] is the feeling that oneself and others can contribute to discussion, and it has been found to be a key issue in users' choice to turn video on or off when videoconferencing. Given that avatars may replace video, we propose inclusivity as having two aspects, *taking part* and *considering contributions* (each paired for self and others).

Avatar satisfaction factors. There are three key factors integral to avatar satisfaction, which encompass both satisfaction with one's own avatar and the avatars of others. *Self-expressive perception* [77] emphasizes the importance of how users express themselves through avatars. *Other-expression perception* [4], focuses on the perceived expressiveness of others in avatar-mediated settings. Finally, overall avatar *preference* [76] pertains to the users' preferences for certain types of avatars, underlining the impact of avatar choice and design on user satisfaction.

In sum, an empirical examination of the impact of avatar-based communication in meetings needs to focus on *meeting outcome* factors (effectiveness, alignment, comfort, and inclusivity) and *avatar satisfaction* factors (self-expressive perception, other-expression perception, and avatar preference). The specific items that we used to explore these two categories of factors are detailed in Table 1.

2.3 Research Questions and Hypotheses

As reviewed above, prior work has shown that representational motion fidelity is important in social contexts, such as for generating trust in others in meetings [49], displaying emotion [14], persuasion [74], understanding of talk content [45], and both personal and cultural identity [28]. However, how avatar motion fidelity impacts work meeting outcomes remains unclear. Therefore, our first research question and hypotheses focus on *meeting outcomes*:

RQ1. How does avatar animation modality impact meeting outcomes?

H1a: Meetings conducted with the *webcam-animated* (*W*) modality would outperform those with the other two modalities, *static picture* (*S*) and *audio-animated* (*A*), on meeting outcome factors.

H1b: Meetings conducted with the *audio-animated* (*A*) modality would outperform those with the *static picture* (*S*) modality on meeting outcome factors.

Alongside impact on meeting outcomes, user satisfaction with avatars remains an important and evolving factor for adoption. Prior work has shown that body and head tracking along with facial features increases self-identification with avatars (i.e., the enfacement illusion) [27], and in another recent study, avatars with head motion were preferred over those without [49]. More broadly, as society's perception of avatars evolves (e.g., see the appropriateness of stylized avatars in professional settings: 2012 [32] vs. 2023 [51]), it is important to maintain an understanding of user preferences as newer capabilities are introduced, including webcam-driven facial features. We thus propose our second research question and hypotheses, focused on *avatar satisfaction*:

RQ2. Which avatar animation modality do people prefer?

H2a: The *webcam-animated* (*W*) modality would be the preferred modality, compared to the other two modalities, *static picture* (*S*) and *audio-animated* (*A*).

H2b: The *audio-animated* (*A*) modality would be the preferred modality over the *static picture* (*S*) modality.

3 Methods

3.1 Participants and Allocation for Meeting Groups

We recruited 68 participants from a global technology company to take part in one-hour experimental sessions¹. Participants were recruited via bulk emails to employees who were located in different offices across various countries (e.g., United States, Spain, Italy, United Kingdom etc.). Participants who expressed interest were asked to recruit an additional 3-5 employees with whom they had professional familiarity. We chose to leverage existing relationships among participants, a choice which intersected with that to create personalized avatars instead of random, generic, or self-created avatars. Since familiarity among participants has the potential to enhance group dynamics [15, 26], we wanted people to be unique and recognizable among group members. This would reduce uncertainty about who was talking, and reduce the confounding influences of avatar animation with the novelty of both appearance and acquaintance [57] (e.g. if someone chose an avatar or avatar attributes that were novel to other group members, the group might fixate on those differences as an issue rather than the animation level or the task). We subsequently formed 16 groups, 4 groups with 4 members each and the rest with 5 members. Participant ages ranged from 18 to 64, comprising 44 males and 24 females (see details in supplementary material).

3.2 Personalized Avatars

For avatar development, participants submitted one self video and two self photographs. Avatar generation was based on several advanced techniques such as 3D face reconstruction [75]. The procedure is outlined in Figure 2 (a). The generated personalized avatar had head bones to enable head movements and utilizes blendshapes, a commonly used method for facial animation and tracking. The researchers chose accessories for participants based on their photographs, to provide cues to similarity beyond the facial landmarks. Participants were sent three versions of their avatar in advance, allowing them to experiment with movement and select the avatar that they found most comfortable. This approach aimed to mitigate potential dissatisfaction of avatars, which might arise due to limitations in the generation process (such as different angles of the provided photographs) and to reduce the novelty effects of seeing their avatars. Photographs and videos were used exclusively for creating avatars and deleted immediately after the participant approved an avatar.

3.3 Animation Implementation for Face and Head Movement

Participants were sent a Unity3D application which enabled them to use their avatar in three animation conditions, and also to change their background to a neutral and clutter-free gradient to avoid attentional confounds. Through Unity Capture², the visuals from the application were used as a virtual camera in Microsoft Teams meetings (Figure 2 (b)). In the picture condition, the application just showed a single frame of the avatar in neutral mode for the entire time. In the audio-animated modality, the application used voice for viseme-driven lip animation (blendshape animations). In the webcam-animated modality, the application used video to drive facial expressions (blendshape animations) and head motion (head pan-tilt-yaw).

¹This study was approved by the IRB (Ethical ERP ID 10489, RCT ID 5711) of the global technology company and all participants completed an informed consent procedure prior to participation.

²<https://github.com/schellingb/UnityCapture>

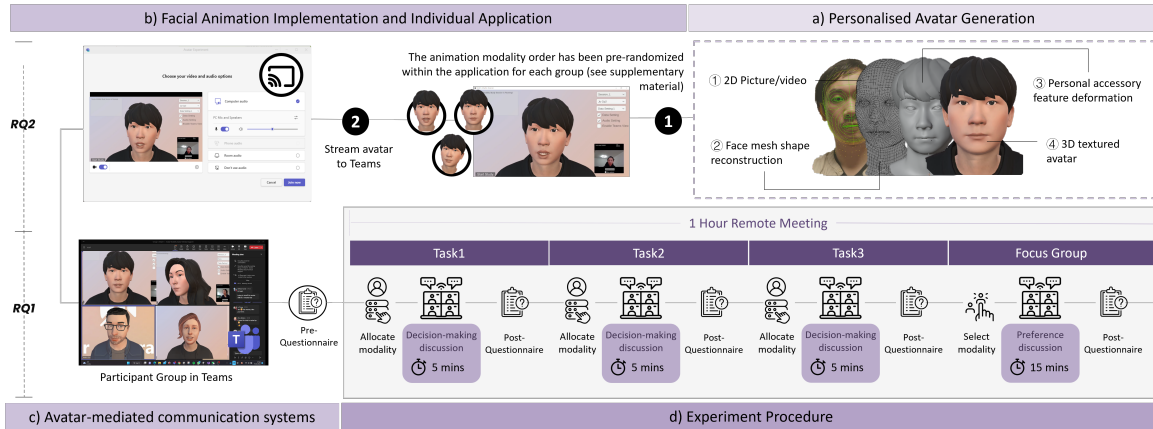


Fig. 2. Overview of the experimental protocol based on the research questions. To adequately address RQ2 (avatar satisfaction), (a) personalized avatars were generated based on participants' self videos and photographs. (b) The animation modality order for the avatars was randomized in advance for each group. Avatars were streamed to a Microsoft Teams videoconferencing meeting for the experiment. To address RQ1 (meeting outcomes), (c) participants in a group joined a Microsoft Teams meeting using their avatars to complete the experimental task sessions. Before starting the tasks, participants individually completed an onboarding questionnaire. (d) They then performed three, five-minute group decision-making tasks, each using a different avatar animation modality: static picture mode (S), audio-animated mode (A), and webcam-animated mode (W). Each task was followed by a post-task questionnaire. Lastly, participants completed a 15-minute focus group, for which participants could use any of the three avatar modalities they individually preferred.

3.4 Procedures

The experiment was conducted using the following procedures, outlined in Figure 2 (d). Participants were asked to complete an onboarding questionnaire at the start of their experiment. The experiment began with three decision-making tasks, each using a distinct avatar animation modality (static picture, audio-animated and webcam-animated). We used the Latin square [12] design to counterbalance the order of avatar modalities used in the tasks systematically across groups (i.e., groups differed in the order of avatar modalities used in the tasks; see order layout in Supplementary materials), minimizing the influence of presentation order on their responses for each task. The order of the tasks was fixed, as tasks were carefully structured to progress from relatively simple to more selective. This design was intended to eliminate fatigue effects, as in this setup, if the fatigue effect was significant, participants might be more inclined to consider audio during the final focus group.

All task sessions were presented in a gallery format for consistency and a male experimenter acted as a moderator throughout, using his own personalized avatar. All three sessions started with a warm-up question. The warm-up questions were thematically related to the subsequent group decision-making tasks. The warm-up reduced some of the immediate novelty of each animation condition, and also helped familiarise participants with the topic. As highlighted by Schilling et al. [60], group alignment effects can also prime participants for specific types of thinking, leading to more effective and focused group discussions and outcomes in the main tasks.

Warm-up questions were followed by a 5-minute group decision-making task. The group decision-making tasks had three different emphases from McGrath's circumplex [2], which models different forms of group interaction. The emphases were Creativity, Planning, and Decision Rationale. The difference in emphases provided a variety of engagement with the decision-making component to

reduce the potential for boredom or repetition for the participant groups. The decision-making tasks were as follows:

Q_{Task1} : If you were going to put one thing in your organization's welcome pack for new starters, what would it be? Assume a budget of \$50 per person. (Creativity emphasis)

Q_{Task2} : Imagine you're planning one social gathering for your team, what type of activity would that be? Assume a budget of \$100 per person. Ignore the geography constraint. Please consider accessibility needs of people in the organisation. (Planning emphasis)

Q_{Task3} : Imagine you are setting up a new <organization> office, where would that be and what is the most important factor in making this decision? (Decision Rationale emphasis)

After each task session, participants completed a 3-minute survey about *meeting outcome factors* (Q_{MO}) and *avatar satisfaction factors* (Q_{AS}), as outlined in Table 1. The decision-making task sessions were followed by a 15-minute focus group session. Instead of being allocated an avatar animation modality in the focus group, users were empowered to choose their preferred modality from any of the three modalities they had just experienced, which provided an additional preference metric beyond those captured in the task sessions. The focus group was tailored to extract qualitative insights into participants' experiences, preferences, and feedback. Finally, as the focus group ended, participants were offered the opportunity to complete an *optional* post-session questionnaire, which served as supplementary data for participants to voice any thoughts or concerns they may have been reluctant to share during the group discussion. The entire experiment lasted approximately one hour.

3.5 Quantitative and Qualitative Methods

3.5.1 Quantitative Methods. As noted in the review of prior work on avatars and meeting outcomes (Section 2.2), our measurement of *meeting outcome* factors included the assessment of *effectiveness*, *alignment*, *comfort*, and *inclusiveness*. The majority of these questions were adapted from prior work. Questions on comfort and inclusiveness were adapted from Cutler et al. [19] and Bluedorn et al. [11]. Comfort ($Q_{MO-comf}$) was measured with a 5-point response scale ranging from 'Very uncomfortable' to 'Very comfortable'. The question on meeting fatigue was adapted from Fauville et al. [24]. Inclusiveness ($Q_{MO-incl}$) consisted of four questions measured with a 5-point response scale (ranging from 'Not at all' to 'Very much'). We created questions for effectiveness and alignment that suited the decision-making tasks (partly inspired by [54, 67]). Effectiveness (Q_{MO-eff}) was presented as a binary Yes/No question. Alignment ($Q_{MO-align}$) was measured using a 5-point Likert scale. Avatar satisfaction (Q_{AS}) was measured by asking participants about the perceived effectiveness of the avatars in terms of their self-expression ($Q_{AS-self}$) and the expression of other team members ($Q_{AS-other}$), each measured on a 5-point response scale ranging from 'Not at all' to 'Very much'. Finally, participants' *preference* ($Q_{AS-pref}$) was measured by their actual choice of avatar animation modality for the focus group. These questions were integrated into task questionnaires. Specific questions can be found in Table 1.

3.5.2 Qualitative Methods. Incorporating qualitative findings alongside our quantitative findings enabled a nuanced understanding of participant reasoning for their choices and unpacking distinctions between personal preferences and meeting outcomes. Qualitative analysis was conducted on the free text questions from the questionnaires (as shown in Table 1). There were three free text questions relating to *preferences* ($Q_{AS-pref-text}$), *effectiveness* ($Q_{MO-eff-text}$) and *comfort* ($Q_{MO-comfort-text}$), all factors are crucial for Q_{MO} . The questions were presented after participants

Meeting Outcome Factor (Q_{MO})			
Q_{MO}	Method	Measure	Question
effectiveness	Quantitative	Yes/No	Q_{MO-eff} : Did the group make a final decision?
	Qualitative	Free text	$Q_{MO-eff-text}$: Did any modality make it easier or harder to reach agreement? Why?
alignment	Quantitative	5-point response scale	$Q_{MO-align}$: After the discussion, how much do you agree or disagree with the group's decision?
comfort	Quantitative	5-point response scale	$Q_{MO-comf}$: How comfortable did you feel expressing yourself via this visual modality?
	Qualitative	Free text	$Q_{MO-comf-text}$: Were any modalities more or less tiring than others?
inclusivity	Quantitative	5-point response scale	$Q_{MO-incl}$: 1. To what extent do you feel you had a chance to take part in the discussion? 2. To what extent do you feel others had a chance to take part in the discussion? 3. To what extent do you feel you considered contributions from others? 4. To what extent do you feel others considered your contributions?
Avatar Satisfaction Factor (Q_{AS})			
Q_{AS}	Method	Measure	Question
self-expressive perception	Quantitative	5-point response scale	$Q_{AS-self}$: To what extent do you feel your representation was effective in expressing yourself?
other-expression perception	Quantitative	5-point response scale	$Q_{AS-other}$: To what extent do you feel the representation of others was effective in expressing themselves?
preference	Quantitative	Choice from 3 modalities	$Q_{AS-pref}$: The actual avatar animation modality choice participants made to represent them in the focus group.
	Qualitative	Free text	$Q_{AS-pref-text}$: Why did you choose to appear in your current modality?

Table 1. Meeting outcome and avatar satisfaction factors measurement and questions. All quantitative questions were asked in a survey after each task conducted in the different modality. All qualitative questions were asked in the final optional post-questionnaire after the Focus Group session. For *comfort*, the response scale ranged from ‘Very uncomfortable’ to ‘Very comfortable’. For *inclusivity*, *self-expressive perception*, and *other-expression perception*, it ranged from ‘Not at all’ to ‘Very much’.

had chosen their avatar modality for the focus group, providing detailed reasoning for their preference ($Q_{AS-pref-text}$) and how the different modalities influenced meeting outcomes ($Q_{MO-eff-text}$ and $Q_{MO-comf-text}$).

Initially, we used a bottom-up inductive thematic analysis method [17] starting with the *preferences* data ($Q_{AS-pref-text}$). Themes were derived from the data, starting with low-level codes, which were then combined into higher-level themes, allowing the data to guide the identification of bottom-up patterns. However, when this approach was applied to the Q_{MO} questions ($Q_{MO-eff-text}$ and $Q_{MO-comf-text}$) it was clear that we had reached saturation, in the sense that codes from the new data were a good fit with existing themes rather than introducing new themes. We nonetheless continued with the analysis to understand which of the existing themes were important for *meeting outcomes*. We therefore then took a top-down, deductive approach in which we used the existing themes as pre-defined codes which were applied to the data, rather than extracting new codes from the data. This enabled us to analyse where participants discussed these themes in relation to *effectiveness* and *comfort*.

We used NVivo 12 to assign quotes and generate initial codes from post-questionnaire text (see coding process in supplementary material). We then identified themes from these codes, which helped us distill meaningful insights from the participants' responses. NVivo 12 also facilitated effective code counting, which helped us understand the relative frequency of themes and compare individuals' perceptions of Avatar and Meeting Outcome. Thus, code counts generated during the coding process were used to assign meaning to specific portions of data. Some quotes are counted multiple times because they encompass multiple meanings assigned to different codes. For instance, when a participant reported that "*the [webcam-animated modality] was the most expressive, and made me feel most 'like myself' in a meeting,*" it encompassed both the expression of avatars and the sense of resembling one's self in the representation. In this situation, this quote will be counted twice in two different themes.

4 Results

We first present our quantitative analysis for meeting outcome and avatar satisfaction factors. We then present a qualitative thematic analysis for the responses provided by textual questions after the focus group sessions.

4.1 Quantitative Results

Our quantitative results (Table 2) cover four *meeting outcome* factors and three *avatar satisfaction* factors, based on surveys conducted immediately after each of the three task sessions. Other than *effectiveness*, which is a Yes/No question, all the other questions were measured by a 5-point scale. For *alignment*, the response scale was a Likert scale. For *comfort*, the response scale ranged from 'Very uncomfortable' to 'Very comfortable'. For *inclusiveness*, *self-expressive perception*, and *other-expression perception*, it ranged from 'Not at all' to 'Very much'. All tests were conducted in SPSS version 27.

4.1.1 Avatar Animation Modality and Meeting Outcome Factors (Q_{MO}). For *effectiveness*, after each modality, each participant independently indicated whether they thought the group had reached an agreement (Yes/No answer). Sixty-five participants indicated that they had reached agreement in the webcam-animated modality (96%), 59 for the audio-animated modality (87%), and 54 for the picture modality (79%). A Test of Proportions (using our custom Matlab function, see supplementary material) revealed a significant difference between webcam-animated & picture modality ($z = 2.85, p = 0.004$), but not between webcam & audio-animated modality ($z = 1.81, p = 0.070$) nor audio & picture modality ($z = 1.14, p = 0.25$). This suggested that in the webcam-animated modality, more participants thought they reached group agreement compared to the picture modality. Each participant also indicated their level of *alignment*, *comfort*, and *inclusivity* after each modality on a 5-point scale. $Q_{MO-align}$ & $Q_{MO-comf}$ were both single-item questions. For $Q_{MO-incl}$, we took the

average of the four questions. All four measurements failed the Shapiro-Wilk test for normality ($p < 0.05$), so we used the Wilcoxon Signed-Rank Test (non-parametric equivalent to paired t-test).

As shown in Figure 3, for *alignment*, there was no difference between the three conditions. We also ran tests excluding participants who said that their group did not reach an agreement in *effectiveness*, and received a similar result. For *comfort* & *inclusivity*, the webcam-animated modality was perceived to be the best. However, although the audio-animated modality was higher in both measurements than picture modality, the differences were not significant.

Overall, our *meeting outcome* factors results supported **H1a**, in terms of *comfort*, *alignment*, and sense of *inclusivity*, suggesting that the *webcam-animated* (W) modality outperformed the other two modalities in general. However, in terms of *effectiveness*, the *webcam-animated* (W) modality was statistically better than the static picture (S), but only marginally better than audio-animated (A). Also, in terms of *alignment*, no statistical differences were found between any of the modalities. **H1b** was not supported, as we found no difference between audio-animated and picture modality, in any of the *meeting outcome* factors measurements.

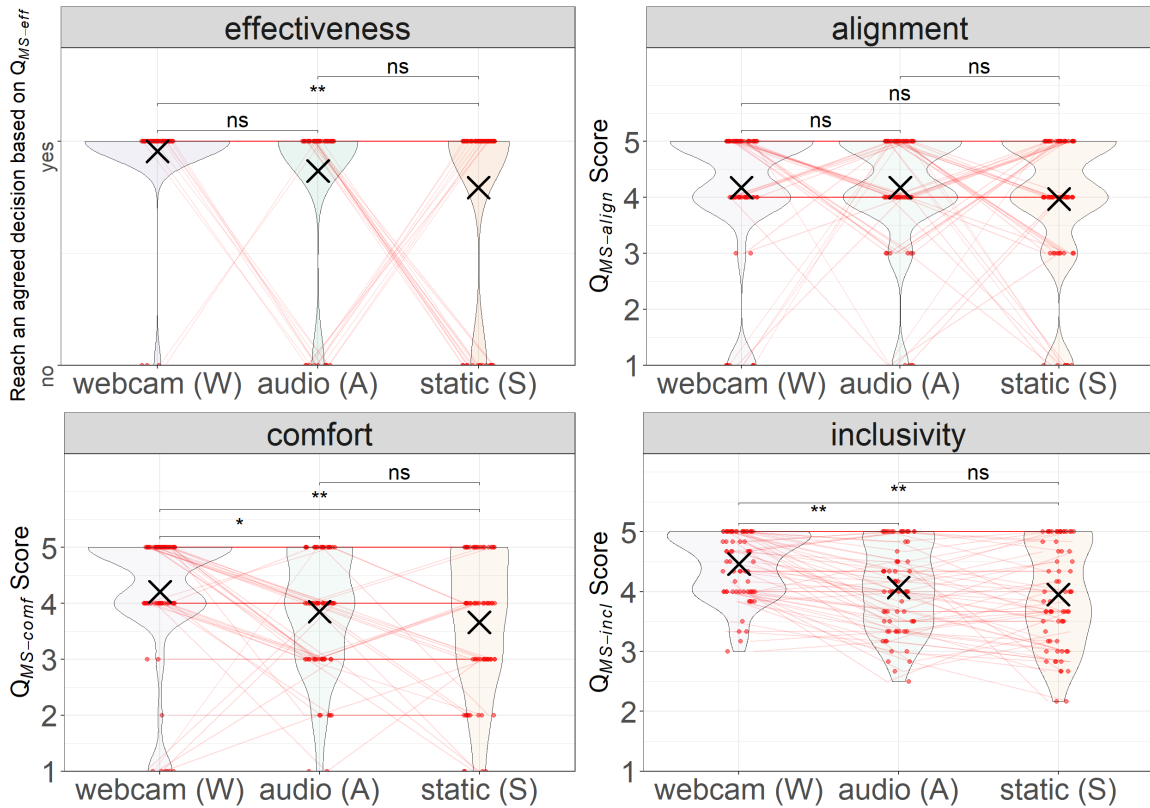


Fig. 3. Quantitative results comparing the three avatar modalities on the outcomes of meeting effectiveness, alignment, comfort, and inclusivity. Violin plots depict distributions of numeric data using density curves. Red dots and lines depict individual responses, illustrating comparisons across the three modalities. Black crosses indicate mean values. For effectiveness, a yes/no response scale was used to measure whether the meeting reached an agreed decision. For alignment, comfort, and inclusivity, a five-point response scale was used, where 1 indicates 'low' and 5 indicates 'high' on the response scale, respectively. * indicates $p < 0.05$, ** indicates $p < 0.01$, 'ns' indicates not statistically significant.

	Avatar Animation Modality			Comparison		
	webcam	audio	static	webcam vs audio	webcam vs static	audio vs static
	Proportion of effective meetings			Test of Proportions		
effectiveness	96%	87%	79%	$z = 1.81, p = .070$	$z = 2.85, p = .004^{**}$	$z = 1.14, p = .25$
	Median			Wilcoxon Test		
alignment	4.18	4.18	3.97	$T = 249, z = .35, p = .729$	$T = 372, z = .96, p = .339$	$T = 436, z = .96, p = .335$
comfort	4.21	3.85	3.66	$T = 146, z = 2.46, p = .014^*$	$T = 474, z = 3.05, p = .002^{**}$	$T = 179, z = 1.77, p = .076$
inclusivity	4.61	4.41	4.46	$T = 581, z = 2.68, p = .007^{**}$	$T = 541, z = 2.92, p = .003^{**}$	$T = 290, z = .68, p = .496$
self-expressive perception	4.15	3.34	2.93	$T = 703, z = .494, p = .000^{***}$	$T = 896, z = .516, p = .000^{***}$	$T = 418, z = 3.00, p = .003^{**}$
other-expression perception	4.16	3.41	2.96	$T = 841, z = 5.10, p = .000^{***}$	$T = 885, z = 5.03, p = .000^{***}$	$T = 601, z = 3.07, p = .002^{**}$

Table 2. Quantitative survey analysis results comparing impact of three avatar animation modalities on meeting outcome and avatar satisfaction factors.

4.1.2 Avatar Animation Modality and Avatar Satisfaction Factors (Q_{AS}). Similar to *meeting outcome* factors, after each modality, participants responded to questions about their perceived level of expressiveness for their self-representation (*self-expressive perception*) and representation of others (*other-expression perception*). Both measurements failed the Shapiro-Wilk test for normality ($p < 0.05$) so we used the non-parametric Wilcoxon Signed-Rank Test. As shown in Table 2, for both *self-expressive perception* and *other-expression perception*, the webcam-animated modality was perceived to be the best. Audio-animation was also perceived to be significantly better than a static picture.

After experiencing all three modalities, participants were free to choose one for the focus group, which we took as an expression of their avatar *preference*. Webcam-animation was most popular, selected by 65 participants (95.6%), indicating a very strong preference for representation that combines visual and auditory elements. Just two participants opted for audio-animation (2.9%), and only one participant chose the picture modality (1.5%). Overall, the *webcam-animated* (W) modality was the most preferred. In questionnaires, *webcam-animation* (W) is preferred over the other two

modalities, and *audio-animation* (A) is preferred over the *static picture* (S) modality, supporting **H2a** and **H2b**.

To summarise our quantitative results: we found that webcam-animation was significantly preferred to the picture modality on three factors related to Q_{MO} factors, except for alignment, and also significantly more comfortable and inclusive than the other two modalities. There was, however, no statistically significant difference between the audio-animated and picture modalities. For Q_{AS} , there was a similar preference for webcam-animation on all measurements compared to the other two modalities. Audio-animation was also significantly preferred over the picture modality for both self- and other-expression, which was not the case for any of the meeting outcome factors. In the following section we use qualitative findings to explain and contextualize these results.

4.2 Qualitative Thematic Analysis

To understand the key motivations behind the trends uncovered in the quantitative analysis, we performed topic analysis of the free text comments from the text input in the post-focus group questionnaire. For the purpose of this analysis, the meeting outcome factors questions on *effectiveness* and *comfort* were explored through a deductive approach, while the avatar satisfaction factor of *preference* was explored inductively, as explained in [Section 3.5.2](#). The decision to narrow our focus to these specific factors, omitting *alignment* and *inclusivity*, was deliberate to maintain a more manageable scope for our study. Despite not encompassing all potential factors, the results obtained from the selected aspects provide sufficient explanatory detail. Participation in the final questionnaire was optional, and had a 73.5% response rate (50 of 68 participants).

Theme	Initial codes
1 [Avatar] Naturalness	<i>real, natural, life-like</i>
2 [Avatar] Movement control	<i>control by motion, movement</i>
3 [Avatar] Aesthetic	<i>looks good, face less fat, head leaning forward</i>
4 [Social Interaction] Expressiveness	<i>expressive, head nod/shake, smile</i>
5 [Social Interaction] Non-verbal Communication	<i>dynamic, interactive, talking timing</i>
6 [Social Interaction] Representation, Personalisation and Identity	<i>keep identity, personal representation, mimic actions</i>
7 [Psychological Effects] Engagement and Interest	<i>engaging, fun, interesting</i>
8 [Psychological Effects] Emotional Awareness and Reaction	<i>convey emotions, see reactions, ability to understand others</i>
9 [Psychological Effects] Cognitive Load and Attention	<i>multitasking, focus on task, lower cognitive</i>
10 [Psychological Effects] Psychological Safety and Trust	<i>trust, privacy with avatar face, relatable experience, self-aware, self-conscious, aware of how I was appearing</i>

Table 3. Initial codes derived from the question ‘Why did you choose to appear in your current modality?’ and identified themes which reorganised into three categories after inductive coding

4.2.1 Inductive Thematic Development (Q_{AS}). The inductive coding process drew upon 95 coding references. In [Table 3](#), we show how the initial codes from $Q_{AS-pref-text}$ were then organized into ten key themes, which were further into three categories (Avatar, Social Interaction and Psychological

Effects) to aid interpretation. As well as considering the perspective of *preference*, categorizing these thematic results in this way can also provide a clearer perspective on the technical concerns in designing avatars.

- (1) **Avatar Themes:** These themes related to the visual and movement aspects of the avatar. The design of the avatar held significant import in shaping user interactions and experiences (32 coding references).
 - **1 - Naturalness:** Participants explicitly stated that their choice for the modality was influenced by its lifelike and natural characteristics with 17 coding references, as captured in the statements: *"I chose [webcam-animated modality] because it was the most life-like"* and *"It's [webcam-animated modality] the most natural one because I can see the feedback from my movements and it just seems more natural."* Additionally, the concept of 'realness' emerged as a dimension. It was not just the real appearance of the avatars that mattered but also an overall experience: *"[webcam-animated modality] closest to real video stream while having movement."*
 - **2 - Movement Control:** Another key reason cited by participants for selecting the webcam-animated modality pertained to the avatar's movement. Specifically, participants emphasised the importance of responsiveness in avatar movement. Illustrative comments include: *"I was constantly using [...] webcam-animated modality] to nod or shake my head to make things seem more natural," "Session 2 [webcam-animated modality] - having my avatar move with my movements is critical vs having a static version or just mouth move - that felt distracting,"* and *"I also found my Session 1 avatar [webcam-animated modality] to be really accurate to my face! I couldn't stop staring at it!"*.
 - **3 - Aesthetics:** The appearance of the avatars was consistent across three modality conditions, but specific observations were made nevertheless. One participant mentioned that the webcam-animated modality made her feel that her *"Face was less fat,"* leading her to opt for this particular modality. More negatively, another participant pointed out that *"My head was leaning forward [in the webcam-animated modality],"* which led him to choose the audio-animated modality as it offered a more natural head angle. This leaning forward is likely a result of a technical issue with capturing head pose, as opposed to the designed experience, which shows how sensitive users are to errors.
- (2) **Social Interaction Themes:** This factor encompassed themes related to communication and social interactions within the meetings. Themes involve non-verbal communication, social behaviours, and the role of Avatar Satisfaction factors in facilitating or hindering social interaction. This category underscores the critical role that social interactions play in shaping user experiences (37 coding references).
 - **4 - Expressiveness:** The word 'expressive' came up a lot when people talked about why they chose a particular modality. Participants exhibited a preference for webcam-animated avatars, as indicated by statements such as, *"I like the expressivity," "It is more interactive than the others,"* and *"I feel like the [webcam-animated modality] better represents me and enables me to better express myself with some facial expressions and movements"*. Also, quotes like *"I thought I could express myself better."* are indicative of a connection to the perception of self-expression ($Q_{AS-self}$). This theme had the most coding references (17).
 - **5 - Non-verbal Communication:** This theme centred on the significance of non-verbal communication for selecting a modality. While the audio-animated modality offered some visual feedback (a moving mouth while speaking), participants drew attention to the comprehensive advantages offered by the webcam-animated modality. Responses such as, *"We know when people are talking in [webcam-animated modality]"*, *"The [picture modality]"*

was somewhat exhausting, as it presented difficulties in discerning the appropriate moments to contribute to the discussion," and "In [picture] and [audio-animated modalities] cannot tell whether people are ready to start or finish talking," accentuate the value of dynamic non-verbal cues. Non-verbal communication elements like nodding and smiling were said to provide additional value. One participant explicitly emphasized the utility of the webcam-animated modality by stating, *"I appreciated that I was able to convey emotions and reactions using non-verbal communication (head nodding, smiling, etc) without having to rely on a video call."*

- **6 - Representation, Personalization, and Identity:** Participants valued modalities that allowed for a better representation of their personality and identity. One noted that *"The [webcam-animated modality] felt like it allowed my personality to come across more in the discussion."* Similar comments included *"The [webcam-animated modality] was the most expressive, and made me feel most 'like myself' in a meeting"* and *"it's important to try new experience not losing the identity so I chose the [webcam-animated modality]"*.
- (3) **Psychological Effects Themes:** The psychological effects of avatars were expressed in 26 coding references, comprising themes such as engagement, emotional connection, cognitive load and the sense of privacy. Although this had the fewest codes among the three categories, its importance should not be overlooked. Indeed, it holds considerable potential for influencing users' attitudes and behaviours [6].
 - **7 - Engagement and Interest:** This theme highlighted the importance of a new kind of meeting experience through avatars. Users found it exciting, as one participant noted, *"I like the [webcam-animated modality] animations. They remind me that I'm speaking with real people in real time. I also found my avatar to be really accurate to my face! I couldn't stop staring at it!"* This sense of enjoyment contributed to engaging experiences, affecting choice of modality. This was further corroborated by participant remarks like, *"The [webcam-animated modality] is the most engaging and lifelike."* *"This [webcam-animated modality] was fun!"* and *"The [webcam-animated modality] seemed like the most interesting one because it is dynamic."*
 - **8 - Emotional Awareness and Reaction:** The theme was underscored as an essential reason to choose a modality, especially for improving perception and expression of emotions. Participants appreciated the ability to convey emotions, with statements like, *"I appreciated that I was able to convey emotions and reactions using non-verbal communication by [webcam-animated modality];"* and *"That [webcam-animated modality] makes me feel more comfortable and confident while I can see my peers' emotions."*
 - **9 - Cognitive Load and Attention:** Participants preferred modalities that minimised cognitive load and helped them focus on tasks. Participants mentioned, *"The [webcam-animated modality] encouraged me to stay focused on the task and discussion."* and *"I like the expressively and lower cognitive burden of trying to deduce other people's states / make my own clear by voice only."*
 - **10 - Psychological Safety and Trust:** This theme stood out even though it was only mentioned four times. Participants said things like, *"I chose the [webcam-animated modality] because it feels more natural and trustworthy,"* and *"I appeared as the [webcam-animated modality] because it felt expressive but didn't invade my privacy like a full-on camera would."*

4.2.2 Deductive Thematic Analysis (across Q_{AS} and Q_{MO}). Since the analysis of effectiveness and comfort did not introduce new themes by inductive process, we took a deductive approach, mapping the 10 pre-existing themes derived from preference $Q_{AS-pref-text}$ to the data from effectiveness $Q_{MO-eff-text}$ and comfort $Q_{MO-comf-text}$ (excepts from quoted content in Section 4.2.1; more coding

references from Q_{MO} can be found in the supplementary material). This allowed us to determine which of the themes were seen as relevant for *effectiveness* and *comfort* while also seeing if there were any themes that were seen as important for *modality preference*, but not because of their influence on *effectiveness* or *comfort*.

Figure 4 shows the results of this analysis. We reordered these themes based on their influence on user preferences and satisfaction which revealed a clear pattern. We identified four **primary themes** that impacted both Meeting Outcome and Avatar Satisfaction. Primary themes were those that consistently referenced across *all three* focus group questions: $Q_{AS-pref-text}$, $Q_{MO-eff-text}$, and $Q_{MO-comf-text}$. The four primary themes were *Expressiveness*, *Non-verbal communication*, *Emotional Awareness and reaction*, and *Cognitive Load and attention*. We further identified three **secondary themes**. Secondary themes referenced *any two of the three* focus group questions: $Q_{AS-pref-text}$ plus $Q_{MO-eff-text}$ or $Q_{MO-comf-text}$. The two secondary themes were *Naturalness*, *Psychological safety and trust*, and *Engagement and interest*. Three further **tertiary themes** were referenced only in $Q_{AS-pref-text}$, calling into question their relevance to Meeting Outcome, as we will discuss later. These themes were *Movement control*, *Representation*, *personalization*, and *identity*, and *Aesthetics*.

These findings emphasize the crucial role of communication dynamics on the Social Interaction and Psychological Effects all-avatar meetings. Expressiveness was mentioned 17 times, Non-verbal Communication 13 times, and Emotional Awareness and Reaction 7 times. Although the impact of these themes on the factors of Effectiveness and Comfort was comparatively less pronounced (less than 5 times), people still mentioned it.

Further, the theme of Cognitive Load and Attention, which spans all three satisfaction factors, exhibited an uneven distribution in its mentions. Its prominence in the domain of comfort was most pronounced, with 11 mentions, but much fewer mentions for preference and effectiveness. This disparity suggests that while Cognitive Load and Attention are crucial for meeting *comfort*, their impact on *effectiveness* (1 mention) is relatively limited. Thus, we identify Cognitive Load and Attention and Engagement and Interest emerge as themes influencing both preference and comfort factors, but with less influence on effectiveness.

Additionally, themes such as Naturalness, Psychological safety, and trust, though mentioned in the context of *effectiveness*, were not as frequently cited. This observation led us to categorize these themes as more specific to avatar preference than to overall Meeting Outcome (similar to the themes of Movement control, Representation, Personalization and Identity, and Aesthetics). These findings indicate that these factors are important for choice of modality but not because of their influence on either *effectiveness* or *comfort*.

4.2.3 Connections between the Quantitative results and Qualitative analysis.

Impact on effectiveness. We found that the effectiveness of decision-making varied depending on the type of questions asked. After each task, users were asked Q_{MO-eff} ("Did your group make a final decision?") and $Q_{MO-eff-text}$ ("Did modality make it easier or harder to reach agreement?"). In our quantitative analysis, a statistically significant number of participants reported that the group agreement was reached with the *webcam-animated* (W) modality compared to the audio-animated (A) and picture (S) modalities. However, in the qualitative analysis, 66% of participants (33 out of 50) believed that the avatar modality had no influence on their group decision making. To gain a deeper understanding of the dynamics, we analysed the actual quotes.

Participants emphasised support for the impact of avatar animation modality when discussing the role of Naturalness, Expressiveness, Non-verbal communication, Emotional awareness and reaction and Psychological safety and trust in the decision-making meeting contexts. For example, quotes mentioned that "*It [webcam-animated modality] was helpful to see smiles and head nods as*

Category	Themes	$Q_{AS-pref-text}$	$Q_{MS-eff-text}$	$Q_{MS-comf-text}$
Social Interaction	Expressiveness	17	3	4
Social Interaction	Non-verbal communication	13	5	2
Psychological effects	Emotional awareness and reaction	7	5	3
Psychological effects	Cognitive load and attention	4	1	11
Avatar	Naturalness	17	1	-
Psychological effects	Engagement and interest	11	-	4
Psychological effects	Psychological safety and trust	4	1	-
Avatar	Movement control	12	-	-
Social Interaction	Representation, personalization and identity	7	-	-
Avatar	Aesthetics	3	-	-

Fig. 4. Thematic interrelations between the meeting outcome: *effectiveness* ($Q_{MO-eff-text}$) and *comfort* ($Q_{MO-comf-text}$) and avatar satisfaction factors: *preference* ($Q_{AS-pref-text}$) ranked in descending order in impact themes (primary, secondary and tertiary sections). Additional qualitative Sankey Flow interpretation see supplementary material.

feedback to my statements", and "first modality [webcam-animated modality] is for sure the more complete to create a trust with other people joining the call".

On the other hand, participants de-emphasised the impact of avatar animation modality due to two factors. First, the complexity of the decision-making tasks themselves were said to have a greater influence on the likelihood of reaching an agreement than the modality used. One participant stated, *"I felt the ease of decision-making was more aligned with the question at hand than the modality."* Second, familiarity with one another was said to be important, such that even if they did not have visually expressive cues, they already knew how to collaborate. One participant pointed out, *"Not for this group because we all know each other very well so the non-animated avatars were just like being in a phone call."*

Comfort and fatigue. We encountered a range of opinions on the matter of whether a particular modality leads to increased or reduced fatigue. These differing perspectives give rise to distinct expressions in Expressiveness and Cognitive Load and Attention, which can be categorized into pros and cons associated with the tracking that enabled the *webcam-animated* (W) modality. On the pro side (more tracking leads to less fatigue), participants said, *"The [picture modality] session was more tiring because it was only picture therefore requiring a little more work on my end to effectively express myself"* and *"I found the [picture modality] session where folks were frozen to be more tiring. I had to force myself to pay attention to the conversation"* On the con side (less tracking leads to less fatigue), participants said, *"The [webcam-animated] one was the most tiring because it encouraged me to move more."* and *"Video was more natural for communication and less cognitive load, but I became more self-aware and self-conscious about what my avatar was doing at any given moment (was my mouth open, weird face angle, etc.)"* For the remaining themes (Non-verbal communication, Engagement and interest, the consensus was pro webcam-animation (more tracking lead to less

fatigue), with statements such as, "*the pictures were more tiring as we could not see the reactions of the people*".

5 Discussion

In this study, we addressed gaps in the existing literature in understanding the influence of avatar modalities on both meeting outcomes and avatar satisfaction for work videoconferencing meetings, particularly in the context of decision-making.

5.1 Meeting Outcome Factors

We hypothesized and found that webcam-animated (W) avatars positively influence meeting outcomes. They have statistically significantly higher scores than the static picture (S) modality for *effectiveness*, and higher than both static (S) and audio-animated (A) modalities for *comfort* and *inclusivity*. Thematic analysis of participants' comments also demonstrated that satisfying one's preferences for avatars is related to expressive qualities best provided with accurate facial and head motion, and these also contribute to meeting effectiveness and meeting comfort. At a base level, then, these findings highlight the potential of webcam-animated (W) animation in enhancing decision-making.

However, the findings about meeting outcomes are more complex. We introduced our interest in meeting outcomes as being whether any modalities make it easier or harder to reach a final decision (*effectiveness*) and easier or harder for participants to align with the group decision (*alignment*). For *effectiveness*, participants' independent indications of whether they thought the group had reached agreement showed significant difference between the webcam-animated (W) and static picture (S) modalities, but not between webcam and audio-animated modalities, nor audio and picture modalities. For alignment, there was no difference between the three conditions. Further, qualitative responses emphasised support for the impact of avatar animation modality for naturalness, expressiveness, non-verbal communication, emotional awareness, and trust-building. However, in the qualitative data, participants also said that the complexity of the decision-making tasks had a greater influence on the likelihood of reaching an agreement than the modality used, and they also said that even when they did not have visually expressive avatars, their pre-existing collegial familiarity meant that they already knew how to collaborate. Thus, whereas our quantitative findings show an impact of avatar modality on some meeting outcomes (i.e., *effectiveness*), participants' perceptions emphasize the relative importance of task complexity and team familiarity for meeting outcomes.

Importantly, our counterbalancing of avatar modalities across tasks eliminates task complexity as a confound in our study. Similarly, our within-subjects experimental design means that team familiarity is fixed across conditions. Thus, our study design accounted for these latter factors, isolating the influence of avatar animation modality on the meeting outcomes of *effectiveness* and *alignment*. Nevertheless, despite our study's focus on avatar modality, we do not suggest that this factor is *more* important than factors such as task complexity and team familiarity in affecting meeting outcomes. Indeed, participants' comments rightfully acknowledge the complexities of workplace collaboration and social dynamics. Avatar modality may play a relatively smaller yet still influential role in this context. Motivated by participants' comments, future work should explore how team familiarity and task complexity moderate the influence of avatar modality on meeting outcomes.

Why did participants perceive avatar modality to play a relatively smaller role in meeting outcomes? One possible explanation is that avatar animation modalities affect the social presence and interpersonal trust of the participants, *facilitating* reaching agreement but are not perceived to affect the decision itself. Some older support for this comes from a study by Bente et al. [9], which found that avatars increased the social presence and interpersonal trust of collaborators, especially

when they were animated with both video and audio. They suggest that avatars provide nonverbal cues that enhance the impression of being together and the feeling of mutual understanding and reliability, which contribute to decision-making, but are not themselves primary parts of the decision-making process. Our qualitative findings around the value of webcam-animated avatars for social interaction align with this idea. The current study precluded systematically exploring the link between social presence, interpersonal trust, and meeting outcomes due to its limited measurement of meeting *effectiveness* and *alignment*, a result of the study session's time constraints. We leave this important question for future work on avatars.

5.2 Avatar Satisfaction Factors

In the context of maintaining consistent personalized avatar appearance, our hypothesis and observations indicate the substantial impact of level of expression on participants' avatar choices for meetings. Notably, there exists a pronounced *preference* for webcam-animated (W) avatars in actual focus group meeting selections (see also [76]), with a significant majority (95.6%). Moreover, our analysis reveals statistically significant differences in both *self-expressive perception* and *other-expression perception* between audio-animated (A) and static picture (S) modality. Our results highlight the importance of head movement [45, 74, 80] and facial expressions [4, 14] in the webcam-animated modality, along with mouth motion in both the webcam- and the audio-animated modalities [23]. The upshot is that partial or subtle expressive cues can enhance interaction and expressiveness when contrasted with entirely static representations. Additionally, our findings on social interaction align with the literature's emphasis on the importance of non-verbal movement for meetings that benefit from rich communication modalities for optimal outcomes, e.g. those to do with building or maintaining relationships, or resolving conflicts [64].

Also in line with prior research, our participants connected their perceived level of self-expression to better self-avatar representation [77] in responses such as *"I feel like it better represents me and enables me to better express myself with some facial expressions and movements."* Within the context of decision-making, perceptions of effectively conveying information to others align with the concept of the expressivity of others [4, 45], e.g., in responses such as *"Movements add the ability to guess more body language/nonverbal communication, which is important to understand the feelings of others"* and *"[static picture modality] was a little tiring in the sense that I had a harder time knowing when to hop into the conversation since I couldn't see facial expressions for feedback"*. These observations of avatar satisfaction further imply the pivotal role of modality expressivity in enhancing the overall avatar-mediated communication experience [4].

One tension bubbling somewhat under the surface of the avatar satisfaction results was how participants felt about being seen versus seeing others. Specifically, some people seem to want lower fidelity self-representations, potentially making them 'harder to read' (*"The webcam driven one was a bit tiring since I felt like I had to be more 'on.'"*), while also wanting higher fidelity representations of others so that they are 'easier to read' (*"Movements adds the ability to guess more body language / nonverbal communication, which is important to understand feelings of the others"*). This issue has been reported previously in avatar research: people can feel "exposed" with high-fidelity self-representations, and uncomfortable or uncanny when the movements don't match their own, while preferring high-fidelity representations of others [49]. It has also been anecdotally reported in videoconferencing research using standard video, especially during the COVID-19 pandemic, when people reported fatigue about seeing their own video, and thus wanting to turn it off, while also reporting fatigue stemming from not seeing the video of others who have turned their video off [33]. However, self-representation aversion in videoconferencing is a complex phenomenon. Kuhn [35] finds that it has correlations with the dispositional trait of the user's degree of public self-consciousness, while Shockley et al. [62] notes that gender and organizational tenure also

correlate with camera usage and fatigue. Future research should conduct a direct test of preferences for self and other representation, and compare avatar modalities to traditional modalities, and check any correlations of modality with both psychological traits and demographic categories.

5.3 Implications

Our results have implications for designing avatar-mediated communication systems and remote work meetings.

5.3.1 Meaningful movement outperforms realism for meeting outcomes. The **primary themes** that emerged from the qualitative analysis were all related to motion. Primary themes, as a reminder, are the themes with comments about all three issues of avatar preferences, meeting effectiveness, and meeting comfort. These themes included Expressiveness, Non-verbal communication, Emotional awareness and reaction, as well as Cognitive load and attention. Collectively, they demonstrate that webcam-animated animation enriches communication by perceived expressiveness and reduces meeting fatigue [3, 31]. Realism and fidelity are also frequently mentioned in **secondary themes**, which is in line with previous findings of importance of natural movement [27, 70] for the acceptance of self-expression and expressions of others. That is, the motion of smiles, head nods, and head shakes are considered more critical for achieving effective communication than realism. Our findings suggest that the incorporation of dynamic movements and expressive gestures in webcam-animated modalities can significantly enhance meeting effectiveness by improving comfort and facilitating non-verbal communication. A further study extending our research to delve deeper into the realm of hand gestures would be recommended. Participants themselves expressed this in ways such as, "Would love to see the incorporation of hand gesture in the future roadmap" and "it would be great if the avatar also had hands and could manage reactions with gestures."

5.3.2 Webcam-animated avatars are a plausible alternative in remote work meetings. As noted in our review of prior work (Section 2), historically, stylized avatars have been considered inappropriate for work meetings. The avatars used in our study were stylized, but in our thematic analysis, "Naturalness" stood out prominently, *but only for the webcam-animated avatars*. Indeed, we found a significant difference in user preferences, with webcam-animated (W) avatars being significantly higher rated than audio-animated (A) or picture (S) avatars. The high acceptance of the stylized avatars, providing that they are animated, might be due to the fact that with a rise in the number of platforms that offer avatars, the acceptance of the avatars that are not photo-realistic has also gone up. And, as we note, the most recent survey of avatar acceptance by knowledge workers reported that hyper-realistic avatars suffer from uncanny valley effects. The authors conclude that the optimal avatar for knowledge work should be a stylized illustration [52]. Additionally, in our quantitative analysis, the use of webcam-animated avatars resulted in significantly higher *effectiveness*, *comfort* and *inclusiveness* in decision-making. This suggests that in collaborative remote work environments, especially in situations where the utilisation of video is impractical, webcam-animated avatars could be more advantageous compared to merely deactivating the camera.

5.3.3 Different technical aspects of the experience implementation impact different experience objectives. Finally, we have found that the top level theme categories that emerged from our thematic analysis – Avatar characteristics, Social interaction, Psychological effects – relate to personal user preferences and the success of the meeting in different ways. By understanding how these themes interconnect within these categories, researchers and practitioners can make more informed decisions about crafting avatars that meet the diverse needs and preferences of users. For example, meeting outcomes (*meeting effectiveness* and *meeting comfort*) are mostly satisfied by avatars that

support expressiveness, non-verbal communication, emotional awareness and reaction, and cognitive load and attention. At the same time, users' avatar preferences are dominated by themes related to avatar and social interactions (*expressiveness, naturalness, non-verbal communication, movement control*). However, issues of visual realism in terms of representation, personalization, and identity, and aesthetics do not intersect with issues around meeting outcomes. This means that when developing avatars, focusing on visual fidelity might increase the users' immediate satisfaction, but not necessarily lead to a more effective meeting. As we noted in our review of prior work, the functional communicative aspects of meetings rely more on the fidelity of how people move than the fidelity of how they look [27, 49, 70]. While developers will need to balance these to ensure that users at least accept their avatars, time with avatars also plays a role. Users can learn to accept less realistic avatars, and even prefer them, as long as they have a motivation to interact using avatars (such as avoiding appearance anxiety, e.g. [33, 35]) and learn that they can accomplish their tasks [23].

5.4 Limitations and Future Research

A key limitation of the study resulted from the trade-off of ecological validity [34] against experimental consistency. We recruited real distributed global teams, but in doing so we had to rely on participants' own technology (including computers, webcams, and internet access) both for conducting the meetings and recording the experimental application. As such, not all groups experienced optimal computer performance, video, and audio quality, and remote debugging of problems was complicated. In order to ensure smooth meeting progress, some data was discarded and some sessions were split into two separate 30-minute slots rather than the ideal single one-hour block.

Given that the experimental setup required coordinating teams, balancing time became a limitation. In order to make the experiment feasible for all participants to attend during their busy work schedule, we had to limit our study (including all three tasks and post-task questionnaires) to a total of one hour, leaving the discussion time for each condition to be capped at five minutes. There were also occasions when one participant had to leave 10 minutes early, or the meeting was inadvertently prolonged. This led to inconsistencies in the duration of some focus group discussions. However, we tried to compensate for these issues in our post-task questionnaire with more textual qualitative data.

Arguably, many of the issue above are upper-bound limitations. That is, had we been able to control all technology and provide the highest quality conditions, the results may have shown more differentiation between conditions and more significant positive results for animation conditions. However, real-life performance is always likely to be more degraded and subject to both technical and logistical idiosyncrasies. As such, the trade-off of experimental consistency for ecological validity provides results that may enable reasonably robust comparisons in future practitioner and academic work.

A different form of limitation stems from our choice to study teams in which members were already familiar with one another. Some participants noted that this made it easier for them to understand each other during the meeting. Exploring the impact of avatars when participants are unfamiliar with each other (e.g. in situations of onboarding new members to teams or 'swift trust' teams) and how this might influence meeting outcomes by using animated avatars would be an intriguing direction to investigate. Combining this with work on existing teams would provide a more comprehensive understanding of the implications of our findings across diverse organizational contexts.

We recognize the potential bias of using data from only one company in our experiment, which may limit the generalizability of our findings. Despite both diversity and cultural dynamics present

within this global entity, the restricted sample, unique culture, processes, and organizational structures of this company are nevertheless a constraint. To address this limitation, future research should prioritize expanding the sampled cohort to include data from multiple companies.

We further acknowledge that cultural background and gender could potentially impact the outcome of meetings, and we were unable to recruit a balanced group of participants (24 female, 44 male) from global offices. Out of the 16 groups, three were male-only, and the other 13 were mixed-gender groups (see supplementary materials). However, the impact of gender and culture were not the focus of our current study, and we do not feel that we have enough data to reach any conclusions in these areas. Future work should explore the influence of these areas, for example, given that meeting practices and expectations vary across cultures [37, 39].

Our study focuses on the impact of conditions with all avatars from the same modality. However, in a real-life situation, should participants been given the freedom to choose, it is possible to have mixed avatar modalities in one setting. As we controlled the avatar modality for each discussion conditions, the only mixed-avatar type session appeared in our focus group session, when one of the participants chose anything other than the webcam-animated modality. That is a total of 3 groups (or 3 participants) out of the 16 groups (or 65 participants). Therefore, we do not have enough data to comment on mixed avatar modalities. Future work should be conducted to understand the differences between mixed and single modalities.

Relatedly, our study examined the specific modalities of video-animated, audio-animated, and picture modalities, to provide a baseline understanding of clear differences between animation conditions. However, better overall avatar animation might be achieved by blending animation technologies. There is potential in exploring the mid-level of animation between the video-animated and audio-animated modality. For example, a more natural, expressive audio-animated algorithm could replace simple lip movement replication and add natural head motion. Further, additional non-verbal capture, such as tracking hand gestures tracked, would likely enhance the overall functional communicative value of avatars and have a positive effect on meeting outcomes.

Finally, we introduced the relevance of understanding the impact of avatars in work meetings by noting that avatar usage is nascent in mainstream communication platforms (Zoom and Microsoft Teams have only recently introduced them, while Slack Huddle and Google Meet have yet to introduce them at the time of writing). Avatars driven by similar technologies are further along the adoption journey in entertainment and consumer contexts, such as use by vTubers (video Youtubers) for personal content creation (e.g. exploring performative gender identity [43]) and marketing (e.g. avatar marketing persuasiveness [22, 58]). These uses are quite different to the professional contexts and outcomes of work meetings, but the technical overlap and salience of identity satisfaction could prompt further exploration into adoption issues and also pave the way for future advancements in virtual collaboration experiences, benefiting both professional and creative domains.

6 Conclusion

The findings of this study show support for webcam-animated avatars as an appropriate replacement for webcam video for situations where users do not want to share their live webcam feed. This finding was supported both quantitatively in reports on alignment in meetings using the webcam-animated modality, and in qualitative responses about how specific aspects of webcam animation contributed to aspects of meeting outcomes. Beyond that core finding, our thematic analysis also provided categorizations to consider when tailoring avatar design choices and features to address specific aspects related to avatar appearance, social interaction dynamics, and the psychological impact on users. By understanding how these themes interconnect within these categories, we can make more informed decisions in crafting avatars that meet the diverse needs and preferences of

users. This categorization serves as a valuable resource not only for researchers seeking a deeper understanding of user preferences but also for industrial designers involved in avatar development.

Acknowledgments

We would like to thank Microsoft for the support of this project, as well as employees from global offices who participated in the experiments. In particular, John Tang and Tadas Baltrusaitis provided valuable experimental suggestions and technical support, while Tom Cashman and Antonio Criminisi offered their support of the overall project. Additionally, Fang Ma, Ju Zhang, Xueni Pan, and Marco Gillies received funding from the Arts and Humanities Research Council (AH/T011416/1).

References

- [1] Safurah Binti Abdul Jalil, Brabbyn Osburn, Jingwen Huang, Mike Barley, Marin Markovich, and Robert Amor. 2012. Avatars at a Meeting. In *Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction* (Dunedin, New Zealand) (CHINZ '12). Association for Computing Machinery, New York, NY, USA, 84–87. doi:10.1145/2379256.2379270
- [2] Ahsan Abdullah, Jan Kolkmeier, Vivian Lo, and Michael Neff. 2021. Videoconference and Embodied VR: Communication Patterns Across Task and Medium. 5 (2021), 1–29. doi:10.1145/3479597
- [3] Joseph A. Allen and Nale Lehmann-Willenbrock. 2022. The key features of workplace meetings: Conceptualizing the why, how, and what of meetings at work. *Organizational Psychology Review* (2022), 20413866221129231. doi:10.1177/20413866221129231 arXiv:https://doi.org/10.1177/20413866221129231
- [4] Chee Siang Ang, Ania Bobrowicz, Panote Siriaraya, Joshua Trickey, and Kate Winspear. 2013. Effects of gesture-based avatar-mediated communication on brainstorming and negotiation tasks among younger users. 29, 3 (2013), 1204–1211. doi:10.1016/j.chb.2012.10.013
- [5] Jeremy N. Bailenson. 2021. Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior* 2, 1 (feb 23 2021). https://tmb.apaopen.org/pub/nonverbal-overload.
- [6] Jeremy N Bailenson. 2021. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, mind, and behavior* 2, 1 (Feb 2021). doi:10.1037/tmb0000030
- [7] Jeremy N. Bailenson, Andrew C. Beall, and Jim Blascovich. 2002. Gaze and task performance in shared virtual environments. *The Journal of Visualization and Computer Animation* 13, 5 (2002), 313–320. doi:10.1002/vis.297 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/vis.297.
- [8] Sabrina Barr. 2020. Boss accidentally transforms into a potato during hilarious video team meeting. *The Independent* (April 2020). https://www.independent.co.uk/life-style/coronavirus-video-chat-potato-snapchat-microsoft-meeting-boss-twitter-a9440076.html Section: Lifestyle.
- [9] Gary Bente, Sabine Rüggenberg, Nicole C. Krämer, and Felix Eschenburg. 2008. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Human Communication Research* 34, 2 (04 2008), 287–318. doi:10.1111/j.1468-2958.2008.00322.x arXiv:https://academic.oup.com/hcr/article-pdf/34/2/287/22325251/jhumcom0287.pdf
- [10] Rachel Bergmann, Sean Rintel, Nancy Baym, Advait Sarkar, Damian Borowiec, Priscilla Wong, and Abigail Sellen. 2023. Meeting (the) Pandemic: Videoconferencing Fatigue and Evolving Tensions of Sociality in Enterprise Video Meetings During COVID-19. *Computer Supported Cooperative Work (CSCW)* 32, 2 (June 2023), 347–383. doi:10.1007/s10606-022-09451-6
- [11] Allen Bluedorn, Daniel Turban, and Mary Love. 1999. The effects of stand-up and sit-down meeting formats on meeting outcomes. *Journal of Applied Psychology* 84 (04 1999), 277–285. doi:10.1037/0021-9010.84.2.277
- [12] James V. Bradley. 1958. Complete Counterbalancing of Immediate Sequential Effects in a Latin Square Design. *J. Amer. Statist. Assoc.* 53 (1958), 525–528. https://api.semanticscholar.org/CorpusID:121947016
- [13] Jed R. Brubaker, Gina Venolia, and John C. Tang. 2012. Focusing on Shared Experiences: Moving beyond the Camera in Video Communication. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) (DIS '12). Association for Computing Machinery, New York, NY, USA, 96–105. doi:10.1145/2317956.2317973
- [14] Judee K Burgoon, Valerie Manusov, and Laura K Guerrero. 2021. *Nonverbal communication*. Routledge.
- [15] Peter W. Cardon, Haibing Ma, and Carolin Fleischmann. 2023. Recorded Business Meetings and AI Algorithmic Tools: Negotiating Privacy Concerns, Psychological Safety, and Control. 60, 4 (2023), 1095–1122. doi:10.1177/23294884211037009
- [16] Frank R. Castelli and Mark A. Sarvary. 2021. Why students do not turn on their video cameras during online classes and an equitable and inclusive plan to encourage them to do so. *Ecology and Evolution* 11, 8 (2021), 3565–3576.

- [doi:10.1002/ece3.7123](https://doi.org/10.1002/ece3.7123) arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.7123>
- [17] Victoria Clarke and Virginia Braun. 2013. *Successful Qualitative Research: A Practical Guide for Beginners*. Sage.
 - [18] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, Yuan Tu, and Johannes Gehrke. 2021. Meeting Effectiveness and Inclusiveness in Remote Collaboration. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 173 (apr 2021), 29 pages. [doi:10.1145/3449247](https://doi.org/10.1145/3449247)
 - [19] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, Yuan Tu, and Johannes Gehrke. 2021. Meeting Effectiveness and Inclusiveness in Remote Collaboration. *Proceedings of the ACM on Human-Computer Interaction* 5 (04 2021), 1–29. [doi:10.1145/3449247](https://doi.org/10.1145/3449247)
 - [20] Maitraye Das, John Tang, Kathryne Ringland, and Anne Marie Piper. 2021. Towards Accessible Remote Work: Understanding the Practices of Neurodivergent Professionals in Working from Home. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. event-place: Yokohama, Japan.
 - [21] Alanah Davis, John Murphy, Dawn Owens, Deepak Khazanchi, and Ilze Zigurs. 2009. Avatars, People, and Virtual Worlds: Foundations for Research in Metaverses. *Journal of the Association for Information Systems* 10, 2 (Feb. 2009). [doi:10.17705/1jais.00183](https://doi.org/10.17705/1jais.00183)
 - [22] Marianny Jessica de Brito Silva, Lorena de Oliveira Ramos Delfino, Kaetana Alves Cerqueira, and Patrícia de Oliveira Campos. 2022. Avatar marketing: a study on the engagement and authenticity of virtual influencers on Instagram. *Social Network Analysis and Mining* 12, 1 (Sept. 2022), 130. [doi:10.1007/s13278-022-00966-w](https://doi.org/10.1007/s13278-022-00966-w)
 - [23] Georgiana Cristina Dobre, Marta Wilczkowiak, Marco Gillies, Xueni Pan, and Sean Rintel. 2022. Nice is Different than Good: Longitudinal Communicative Effects of Realistic and Cartoon Avatars in Real Mixed Reality Work Meetings. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 437, 7 pages. [doi:10.1145/3491101.3519628](https://doi.org/10.1145/3491101.3519628)
 - [24] Géraldine Fauville, Mufan Luo, Anna C. Queiroz, Jeremy Bailenson, and Jeff Hancock. 2021. Zoom Exhaustion & Fatigue Scale. *SSRN Electronic Journal* (01 2021). [doi:10.2139/ssrn.3786329](https://doi.org/10.2139/ssrn.3786329)
 - [25] Guo Freeman and Divine Maloney. 2021. Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 239 (jan 2021), 27 pages. [doi:10.1145/3432938](https://doi.org/10.1145/3432938)
 - [26] David C. Funder and C. Randall Colvin. 1988. Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology* 55, 1 (1988), 149–158. [doi:10.1037/0022-3514.55.1.149](https://doi.org/10.1037/0022-3514.55.1.149) Place: US Publisher: American Psychological Association.
 - [27] Mar Gonzalez-Franco, Anthony Steed, Steve Hoogendyk, and Eyal Ofek. 2020. Using Facial Animation to Increase the Enfacement Illusion and Avatar Self-Identification. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2023–2029. [doi:10.1109/TVCG.2020.2973075](https://doi.org/10.1109/TVCG.2020.2973075)
 - [28] Simon Harrison. 2014. Head shakes: Variation in form, function, and cultural distribution of a head movement related to “no”. In *Volume 2*, Cornelia Müller, Alan Cienki, Ellen Fricke, Silva Ladewig, David McNeill, and Jana Bressem (Eds.). De Gruyter Mouton, Berlin, München, Boston, 1496–1501. [doi:10.1515/9783110302028.1496](https://doi.org/10.1515/9783110302028.1496)
 - [29] Yasaman Hosseinkashi, Lev Tankelevitch, Jamie Pool, Ross Cutler, and Chinmaya Madan. 2024. Meeting Effectiveness and Inclusiveness: Large-scale Measurement, Identification of Key Features, and Prediction in Real-world Remote Meetings. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 93 (apr 2024), 39 pages. [doi:10.1145/3637370](https://doi.org/10.1145/3637370)
 - [30] Kori M. Inkpen and Mara Sedlins. 2011. Me and My Avatar: Exploring Users’ Comfort with Avatars for Workplace Communication. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (Hangzhou, China) (CSCW '11)*. Association for Computing Machinery, New York, NY, USA, 383–386. [doi:10.1145/1958824.1958883](https://doi.org/10.1145/1958824.1958883)
 - [31] Ellen A. Isaacs and John C. Tang. 1994. What video can and cannot do for collaboration: A case study. *Multimedia Systems* 2, 2 (Aug. 1994), 63–73. [doi:10.1007/BF01274181](https://doi.org/10.1007/BF01274181)
 - [32] Sasa Junuzovic, Kori Inkpen, John Tang, Mara Sedlins, and Kristie Fisher. 2012. To See or Not to See: A Study Comparing Four-Way Avatar, Video, and Audio Conferencing for Work. In *Proceedings of the 2012 ACM International Conference on Supporting Group Work (Sanibel Island, Florida, USA) (GROUP '12)*. Association for Computing Machinery, New York, NY, USA, 31–34. [doi:10.1145/2389176.2389181](https://doi.org/10.1145/2389176.2389181)
 - [33] Katherine A. Karl, Joy V. Peluchette, and Navid Aghakhani. 2022. Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly. *Small Group Research* 53, 3 (2022), 343–365. [doi:10.1177/10464964211015286](https://doi.org/10.1177/10464964211015286) arXiv:<https://doi.org/10.1177/10464964211015286>
 - [34] Suzanne Kieffer. 2017. ECOVAL: Ecological Validity of Cues and Representative Design in User Experience Evaluations. *AIS Transactions on Human-Computer Interaction* 9, 2 (June 2017), 149–172. <https://aisel.aisnet.org/thci/vol9/iss2/4>
 - [35] Kristine M. Kuhn. 2022. The constant mirror: Self-view and attitudes to virtual meetings. *Computers in Human Behavior* 128 (2022), 107110. [doi:10.1016/j.chb.2021.107110](https://doi.org/10.1016/j.chb.2021.107110)
 - [36] Anastasia Kuzminykh and Sean Rintel. 2020. Low Engagement As a Deliberate Practice of Remote Participants in Video Meetings. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. [doi:10.1145/3334480.3383080](https://doi.org/10.1145/3334480.3383080)

- [37] Tine Köhler and Markus Götz. 2015. Meetings across cultures: Cultural differences in meeting expectations and processes. In *The Cambridge handbook of meeting science*. Cambridge University Press, New York, NY, US, 119–149. doi:10.1017/CBO9781107589735.007
- [38] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The Effect of Avatar Realism in Immersive Social Virtual Realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) (VRST '17). Association for Computing Machinery, New York, NY, USA, Article 39, 10 pages. doi:10.1145/3139131.3139156
- [39] Nale Lehmann-Willenbrock and Annika L. Meinecke. 2017. Team-Meeting Behaviors in Germany and the United States. In *The International Encyclopedia of Intercultural Communication*. John Wiley & Sons, Ltd, 1–6. doi:10.1002/9781118783665.ieicc0262 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118783665.ieicc0262>
- [40] Yuxuan Li, Han Zhang, Shaozhong Cao, Dan Jiang, Meng Wang, and Weiqi Wang. 2023. Literature Review of Audio-Driven 2D Avatar Video Generation Algorithms. In *IEIS 2022*, Menggang Li, Guowei Hua, Xiaowen Fu, Anqiang Huang, and Dan Chang (Eds.). Springer Nature Singapore, Singapore, 85–96.
- [41] Zhe Li, Zerong Zheng, Yuxiao Liu, Boyao Zhou, and Yebin Liu. 2023. PoseVocab: Learning Joint-Structured Pose Embeddings for Human Avatar Modeling. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) (SIGGRAPH '23). Association for Computing Machinery, New York, NY, USA, Article 8, 11 pages. doi:10.1145/3588432.3591490
- [42] Matthew Lombard and Theresa Ditton. 1997. At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication* 3, 2 (Sept. 1997), JCMC321. doi:10.1111/j.1083-6101.1997.tb00072.x
- [43] Zhicong Lu, Chenxinran Shen, Jiannan Li, Hong Shen, and Daniel Wigdor. 2021. More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Yokohama</city>, <country>Japan</country>, </conf-loc>) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 137, 14 pages. doi:10.1145/3411764.3445660
- [44] Microsoft. 2023. Join a meeting as an avatar in Microsoft Teams - Microsoft Support. <https://support.microsoft.com/en-us/office/join-a-meeting-as-an-avatar-in-microsoft-teams-5384e7b7-30c7-4bcb-8065-0c9e830cc8ad>
- [45] K. G. Munhall, Jeffery A. Jones, Daniel E. Callan, Takaaki Kuratate, and Eric Vatikiotis-Bateson. 2004. Visual Prosody and Speech Intelligibility: Head Movement Improves Auditory Speech Perception. *Psychological Science* 15, 2 (2004), 133–137. doi:10.1111/j.0963-7214.2004.01502010.x _eprint: <https://doi.org/10.1111/j.0963-7214.2004.01502010.x>
- [46] Carol T. Nixon and Glenn E. Littlepage. 1992. Impact of meeting procedures on meeting effectiveness. *Journal of Business and Psychology* 6, 3 (1992), 361–369. doi:10.1007/BF01126771
- [47] Kristine L. Nowak and Jesse Fox. 2018. Avatars and Computer-Mediated Communication: A Review of the Uses and Effects of Virtual Representations. *Review of Communication Research* 6 (2018), 30–53. <https://rcommunicationr.org/index.php/rcr/article/view/39>
- [48] Kristine L. Nowak and Christian Rauh. 2005. The Influence of the Avatar on Online Perceptions of Anthropomorphism, Androgyny, Credibility, Homophily, and Attraction. *Journal of Computer-Mediated Communication* 11, 1 (Nov. 2005), 153–178. doi:10.1111/j.1083-6101.2006.tb00308.x
- [49] Payod Panda, Molly Jane Nicholas, Mar Gonzalez-Franco, Kori Inkpen, Eyal Ofek, Ross Cutler, Ken Hinckley, and Jaron Lanier. 2022. AllTogether: Effect of Avatars in Mixed-Modality Conferencing Environments. In *2022 Symposium on Human-Computer Interaction for Work* (Durham, NH, USA) (CHIWORK 2022). Association for Computing Machinery, New York, NY, USA, Article 8, 10 pages. doi:10.1145/3533406.3539658
- [50] Payod Panda, Molly Jane Nicholas, David Nguyen, Eyal Ofek, Michel Pahud, Sean Rintel, Mar Gonzalez-Franco, Ken Hinckley, and Jaron Lanier. 2023. Beyond Audio: Towards a Design Space of Headphones as a Site for Interaction and Sensing. In *DIS 2023 - Proceedings of the 2023 ACM Conference on Designing Interactive Systems*. Association for Computing Machinery (ACM), 904–916. doi:10.1145/3563657.3596022
- [51] Hanqing Pang, Shiyu Tang, Jia Yi Han, and Fun Man Fung. 2023. Exploring the Use of an Avatar-Based Online Platform to Facilitate Social Interaction in Laboratory Sessions. *Journal of Chemical Education* 100, 10 (2023), 3832–3840. doi:10.1021/acs.jchemed.3c00093 arXiv:<https://doi.org/10.1021/acs.jchemed.3c00093>
- [52] Vrushank Phadnis, Kristin Moore, and Mar Gonzalez Franco. 2023. Avatars in Work Meetings: Correlation Between Photorealism and Appeal. arXiv:2304.01405 [cs.HC]
- [53] Rabindra Ratan, Dave B. Miller, and Jeremy N. Bailenson. 2022. Facial Appearance Dissatisfaction Explains Differences in Zoom Fatigue. *Cyberpsychology, Behavior, and Social Networking* 25, 2 (Feb. 2022), 124–129. doi:10.1089/cyber.2021.0112
- [54] Bruce A Reinig. 2003. Toward an understanding of satisfaction with the process and outcomes of teamwork. *Journal of Management Information Systems* 19, 4 (2003), 65–83.
- [55] René Riedl. 2022. On the stress potential of videoconferencing: definition and root causes of Zoom fatigue. *Electronic Markets* 32, 1 (March 2022), 153–177. doi:10.1007/s12525-021-00501-3

- [56] E. Sean Rintel. 2010. Conversational management of network trouble perturbations in personal videoconferencing. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI '10)*. Association for Computing Machinery, New York, NY, USA, 304–311. doi:10.1145/1952222.1952288
- [57] Sean Rintel. 2007. Maximizing environmental validity: remote recording of desktop videoconferencing. Springer-Verlag, 1772593, 911–920.
- [58] Hiroshi Sakuma, Ao Hori, Minami Murashita, Chisa Kondo, and Yoshinori Hijikata. 2023. YouTubers vs. VTubers: Persuasiveness of human and virtual presenters in promotional videos. *Frontiers in Computer Science* 5 (2023). doi:10.3389/fcomp.2023.1043342
- [59] Jack Saunders and Vinay Namboodiri. 2023. READ Avatars: Realistic Emotion-controllable Audio Driven Avatars. arXiv:2303.00744 [cs.CV]
- [60] Martin Schilling, Nadine Oeser, and Cornelius Schaub. 2007. How Effective Are Decision Analyses? Assessing Decision Process and Group Alignment Effects. *Decision Analysis* 4 (12 2007), 227–242. doi:10.1287/deca.1070.0101
- [61] N. Sadat Shami, Li-Te Cheng, Steven Rohall, Andrew Sempere, and John Patterson. 2010. Enhancing Distributed Corporate Meetings with Lightweight Avatars. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (Atlanta, Georgia, USA) (CHI EA '10)*. Association for Computing Machinery, New York, NY, USA, 3829–3834. doi:10.1145/1753846.1754064
- [62] Kristen M. Shockley, Allison S. Gabriel, Daron Robertson, Christopher C. Rosen, Nitya Chawla, Mahira L. Ganster, and Maira E. Ezerins. 2021. The fatiguing effects of camera use in virtual meetings: A within-person field experiment. *Journal of Applied Psychology* 106, 8 (2021), 1137–1155. doi:10.1037/apl0000948 Place: US Publisher: American Psychological Association.
- [63] Snapchat. 2023. How to use Snap Camera to transform yourself in Zoom, Skype, and Teams calls. <https://www.pcworld.com/article/399015/how-to-use-snap-camera-in-zoom-skype-and-teams-calls.html>
- [64] Willem Standaert, Steve Muylle, and Amit Basu. 2021. How shall we meet? Understanding the importance of meeting mode capabilities for different meeting objectives. *Information & Management* 58, 1 (Jan. 2021), 103393. doi:10.1016/j.im.2020.103393
- [65] Hao Tang, Yuxiao Hu, Yun Fu, Mark Hasegawa-Johnson, and Thomas S. Huang. 2008. Real-time conversion from a single 2D face image to a 3D text-driven emotive audio-visual avatar. In *2008 IEEE International Conference on Multimedia and Expo*. 1205–1208. doi:10.1109/ICME.2008.4607657
- [66] John Tang. 2021. Understanding the Telework Experience of People with Disabilities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 30:1–30:27. doi:10.1145/3449104
- [67] Karine Toupin-April, Jennifer Barton, Liana Fraenkel, Linda C. Li, Peter Brooks, Maarten De Wit, Dawn Stacey, France Légaré, Alexa Meara, Beverley Shea, Anne Lyddiatt, Cathie Hofstetter, Laure Gossec, Robin Christensen, Marieke Scholte-Voshaar, Maria E. Suarez-Almazor, Annelies Boonen, Tanya Meade, Lyn March, Christoph Pohl, Janet Elizabeth Jull, Sigogini Sivarajah, Willemina Campbell, Rieke Alten, Suvi Karuranga, Esi Morgan, Jessica Kaufman, Sophie Hill, Lara J. Maxwell, Vivian Welch, Dorcas Beaton, Yasser El-Miedany, and Peter S. Tugwell. 2017. Toward the Development of a Core Set of Outcome Domains to Assess Shared Decision-making Interventions in Rheumatology: Results from an OMERACT Delphi Survey and Consensus Meeting. *The Journal of Rheumatology* 44, 10 (2017), 1544–1550. doi:10.3899/jrheum.161241 arXiv:https://www.jrheum.org/content/44/10/1544.full.pdf
- [68] Asimina Vasalou, Adam Joinson, Tanja Bänziger, Peter Goldie, and Jeremy Pitt. 2008. Avatars in social media: Balancing accuracy, playfulness and embodied messages. *International Journal of Human-Computer Studies* 66, 11 (2008), 801–811. doi:10.1016/j.ijhcs.2008.08.002
- [69] Jane Wakefield. 2021. Viral 'I'm not a cat' filter is decades-old software. *BBC News* (Feb. 2021). <https://www.bbc.com/news/technology-56010156>
- [70] Shensheng Wang, Scott O. Lilienfeld, and Philippe Rochat. 2015. The Uncanny Valley: Existence and Explanations. *Review of General Psychology* 19, 4 (2015), 393–407. doi:10.1037/gpr0000056 arXiv:https://doi.org/10.1037/gpr0000056
- [71] Anna Watson and M. Angela Sasse. 2000. The good, the bad, and the muffled: the impact of different degradations on Internet speech. In *Proceedings of the eighth ACM international conference on Multimedia (MULTIMEDIA '00)*. Association for Computing Machinery, New York, NY, USA, 269–276. doi:10.1145/354384.354503
- [72] Florian Weidner, Gerd Boettcher, Stephanie Arevalo Arboleda, Chenyao Diao, Luljeta Sinani, Christian Kunert, Christoph Gerhardt, Wolfgang Broll, and Alexander Raake. 2023. A Systematic Review on the Visualization of Avatars and Agents in AR & VR displayed using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2596–2606. doi:10.1109/TVCG.2023.3247072
- [73] Susie Weller. 2017. Using internet video calls in qualitative (longitudinal) interviews: some implications for rapport. *International Journal of Social Research Methodology* 20, 6 (2017), 613–625. doi:10.1080/13645579.2016.1269505 arXiv:https://doi.org/10.1080/13645579.2016.1269505

- [74] Gary L. Wells and Richard E. Petty. 1980. The Effects of Over Head Movements on Persuasion: Compatibility and Incompatibility of Responses. *Basic and Applied Social Psychology* 1, 3 (1980), 219–230. doi:10.1207/s15324834basp0103_2 Publisher: Routledge _eprint: https://doi.org/10.1207/s15324834basp0103_2.
- [75] Erroll Wood, Tadas Baltrušaitis, Charlie Hewitt, Matthew Johnson, Jingjing Shen, Nikola Milosavljević, Daniel Wilde, Stephan Garbin, Toby Sharp, Ivan Stojiljković, Tom Cashman, and Julien Valentin. 2022. 3D Face Reconstruction with Dense Landmarks. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Lecture Notes in Computer Science, Vol. 13673. Springer, Cham. https://doi.org/10.1007/978-3-031-19778-9_10
- [76] Chi-Lan Yang, Shigeo Yoshida, Hideaki Kuzuoka, Takuji Narumi, and Naomi Yamashita. 2023. Affective Profile Pictures: Exploring the Effects of Changing Facial Expressions in Profile Pictures on Text-Based Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany). ACM, 1–17. doi:10.1145/3544548.3581061
- [77] Martin Yarmand, Jaemarie Solyst, Scott Klemmer, and Nadir Weibel. 2021. “It Feels Like I am Talking into a Void”: Understanding Interaction Gaps in Synchronous Online Classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama Japan). ACM, 1–9. doi:10.1145/3411764.3445240
- [78] Mika Yasuoka, Marko Zivko, Hiroshi Ishiguro, Yuichiro Yoshikawa, and Kazuki Sakai. 2022. Effects of Digital Avatar on Perceived Social Presence and Co-presence in Business Meetings Between the Managers and Their Co-workers. In *Collaboration Technologies and Social Computing: 28th International Conference, CollabTech 2022, Santiago, Chile, November 8–11, 2022, Proceedings* (Santiago, Chile). Springer-Verlag, Berlin, Heidelberg, 83–97. doi:10.1007/978-3-031-20218-6_6
- [79] Nick Yee, Jeremy N. Bailenson, and Nicolas Ducheneaut. 2009. The Proteus Effect: Implications of Transformed Digital Self-Representation on Online and Offline Behavior. *Communication Research* 36, 2 (2009), 285–312. doi:10.1177/0093650208330254 arXiv:<https://doi.org/10.1177/0093650208330254>
- [80] Ce Zheng, Wenhan Wu, Chen Chen, Taojiannan Yang, Sijie Zhu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah. 2023. Deep Learning-Based Human Pose Estimation: A Survey. *ACM Comput. Surv.* 56, 1, Article 11 (aug 2023), 37 pages. doi:10.1145/3603618
- [81] Zoom. 2023. Using Avatars in meetings and webinars. <https://support.zoom.us/hc/en-us/articles/4642184011917-Using-Avatars-in-meetings-and-webinars>

Received January 2024; revised July 2024; accepted October 2024

Nods of Agreement: Webcam-Driven Avatars Improve Meeting Outcomes and Avatar Satisfaction Over Audio-Driven or Static Avatars in All-Avatar Work Videoconferencing

FANG MA, Goldsmiths, University of London, United Kingdom
JU ZHANG, Goldsmiths, University of London, United Kingdom
LEV TANKELEVITCH, Microsoft Research, United Kingdom
PAYOD PANDA, Microsoft Research, United Kingdom
TORANG ASADI, Microsoft, United States
CHARLIE HEWITT, Microsoft, United Kingdom
LOHIT PETIKAM, Microsoft, United Kingdom
JAMES CLEMOES, Microsoft Research, United Kingdom
MARCO GILLIES, Goldsmiths, University of London, United Kingdom
XUENI PAN, Goldsmiths, University of London, United Kingdom
SEAN RINTEL, Microsoft Research, United Kingdom
MARTA WILCZKOWIAK, Microsoft, United Kingdom

Supplementary materials.

ACM Reference Format:

Fang Ma, Ju Zhang, Lev Tankelevitch, Payod Panda, Torang Asadi, Charlie Hewitt, Lohit Petikam, James Clemoes, Marco Gillies, Xueni Pan, Sean Rintel, and Marta Wilczkowiak. 2025. Nods of Agreement: Webcam-Driven Avatars Improve Meeting Outcomes and Avatar Satisfaction Over Audio-Driven or Static Avatars in All-Avatar Work Videoconferencing. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW142 (April 2025), 8 pages. <https://doi.org/10.1145/3711040>

A Experiment Design

A.1 Latin Square counterbalancing method

Authors' Contact Information: **Fang Ma**, f.ma@gold.ac.uk, Goldsmiths, University of London, London, United Kingdom; **Ju Zhang**, j.zhang@gold.ac.uk, Goldsmiths, University of London, London, United Kingdom; **Lev Tankelevitch**, lev.tankelevitch@microsoft.com, Microsoft Research, Cambridge, United Kingdom; **Payod Panda**, payod.panda@microsoft.com, Microsoft Research, Cambridge, United Kingdom; **Torang Asadi**, torang.asadi@microsoft.com, Microsoft, San Francisco, United States; **Charlie Hewitt**, chewitt@microsoft.com, Microsoft, Cambridge, United Kingdom; **Lohit Petikam**, lohitpetikam@microsoft.com, Microsoft, Cambridge, United Kingdom; **James Clemoes**, jaclemoe@microsoft.com, Microsoft Research, Cambridge, United Kingdom; **Marco Gillies**, m.gillies@gold.ac.uk, Goldsmiths, University of London, London, United Kingdom; **Xueni Pan**, x.pan@gold.ac.uk, Goldsmiths, University of London, London, United Kingdom; **Sean Rintel**, serintel@microsoft.com, Microsoft Research, Cambridge, United Kingdom; **Marta Wilczkowiak**, mawilczk@microsoft.com, Microsoft, Cambridge, United Kingdom.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2573-0142/2025/4-ARTCSCW142

<https://doi.org/10.1145/3711040>

Order#	1st	2nd	3rd	Group#
1	3 - webcam-animated	2 - audio-animated	1- static picture	1, 7, 13
2	3 - webcam-animated	1- static picture	2 - audio-animated	2, 8, 14
3	2 - audio-animated	3 - webcam-animated	1- static picture	3, 9, 15
4	2 - audio-animated	1- static picture	3 - webcam-animated	4, 10, 16
5	1- static picture	2 - audio-animated	3 - webcam-animated	5, 11
6	1- static picture	3 - webcam-animated	2 - audio-animated	6, 12

Table 1. Latin Square layout and distribution of six orders for 16 meeting groups

Table 1 represents all six orders of the three animation modality conditions (static picture, audio-animated, webcam-animated), ordered by the balanced Latin Square, ensuring that the conditions are systematically balanced across the meeting groups in the experiment.

A.2 An interface example of a group meeting discussion



Fig. 1. A Microsoft Teams meeting window snapshot during task session.

Figure 1 shows an example during the actual experimental meeting, featuring a timer that counts down for decision-making tasks. The purple highlighted frames on the windows of the left two participants indicate the current speaker during ongoing discussions. This is a feature in Microsoft Teams.

B Quantitative Study

B.1 Demographics

The following Table 2 presents the demographic data collected from participants via the pre-session questionnaire, which provides a detailed demographic breakdown of the participants involved in the study, organized by gender and age group across the 16 formed groups.

Group#	Female	Male	18-24	25-34	35-44	45-54	55-64	Total
1	3	1	-	3	-	1	-	4
2	3	1	-	4	-	-	-	4
3	1	3	-	1	1	3	-	4
4	1	3	-		1	2	1	4
5	3	1	-	2	2	-	-	4
6	1	3	-	1	1	2	-	4
7	1	3	-	1	2	1	-	4
8	1	4	-	-	4	1	-	5
9	2	3	-	1	2	2	-	5
10	2	2	-	2	-	2	-	4
11	2	2	1	2	1	-	-	4
12	0	4	-	-	2	2	-	4
13	0	4	-	-	2	2	-	4
14	2	2	-		2	2	-	4
15	0	4	-	1	2	1	-	4
16	2	3	1	-	2	1	-	5
Total	24	43	2	18	24	22	1	67

Table 2. Demographic breakdown of participants by gender and age group across 16 formed groups.

B.2 Equation and Matlab function

Test of Two Proportions equation is:

$$z = \frac{p1 - p2}{\sqrt{p(1-p)(\frac{1}{n1} + \frac{1}{n2})}} \quad (1)$$

The Matlab function is:

```

1 function [p1,p2,z] = proportionsTest(r1,n1,r2,n2)
2 p1 = r1/n1;
3 p2 = r2/n2;
4 q1 = 1 - p1;
5 q2 = 1 - p2;
6 z = (p1-p2)/sqrt((p1*q1/n1) + (p2*q2/n2));
7 p1 = normcdf(-abs(z),0,1); % One tail P Value
8 p2 = 2*p1; %two tail P Value

```

C Qualitative Study

C.1 Quotes in Qualitative Analysis

Table 3 provides sample sentences illustrating Meeting Satisfaction factors (*effectiveness* and *comfort*) coding references, as well as Avatar Satisfaction factors (*preference*), derived from open-ended textual questions in the post-focus group questionnaire, used in thematic analysis.

C.2 Additional Qualitative Sankey Flow Interpretation

This qualitative analysis flow is an extra interpretation through the use of a Sankey graph for **Figure 2** in the paper. This visualized graph facilitates both inductive (*preference*) and deductive (*effectiveness* and *comfort*) thematic analysis flow for understanding the impact of themes of Avatar Satisfaction on Meeting Satisfaction. This approach allows the exploration and discovery of patterns, themes, and insights within qualitative themes.

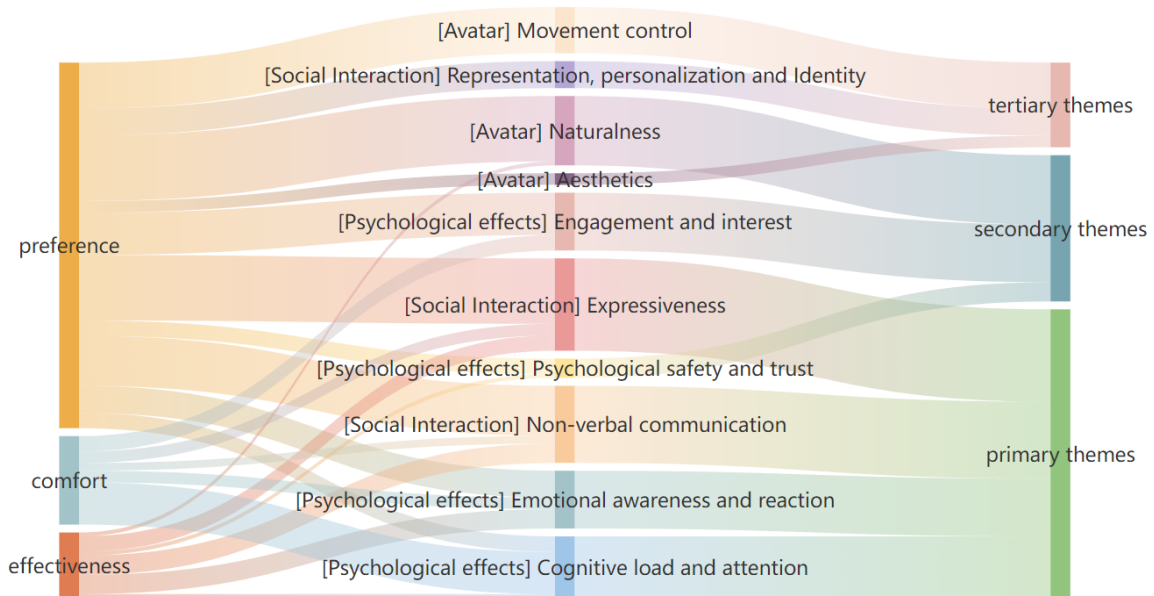


Fig. 2. Sankey diagram illustrating the thematic flow recategorized into primary themes, secondary themes, and tertiary themes

Theme name	Factor	Sample sentences
1 - Naturalness	preference	"most life-like", "most natural one because I can see the feedback from my movements", "closest to real video stream while having movement"
	effectiveness	"I like to see people's expressions, understand what they think as well as what they say... and the [video driven] seems more natural to me"
2 - Movement control	preference	"I was constantly using [webcam-animated modality] to nod or shake my head to make things seem more natural", "having my avatar move with my movements is critical vs having a static version or just mouth move - that felt distracting", "I found [webcam-animated modality] to be really accurate to my face! I couldn't stop staring at it!"
3 - Aesthetic	preference	"Face was less fat", "My head was leaning forward in the other modalities"
4 - Expressiveness	preference	"I like the expressivity", "It is more interactive than the others", "I feel like it better represents me and enables me to better express myself with some facial expressions and movements"
	effectiveness	"I felt a slight preference for webcam driven because I am someone who relies heavily on facial expressions to track the conversation"
	comfort	"The [webcam-animated] was the most tiring because it encouraged me to move more", "The [picture modality] was more tiring because it was only picture therefore requiring a little more work on my end to effectively express myself"
5 - Non-verbal Communication	preference	"We know when people are talking [in webcam-animated modality modality]", "The [picture modality] was somewhat exhausting, as it presented difficulties in discerning the appropriate moments to contribute to the discussion", "[In picture and audio-animated modalities] cannot tell whether people are ready to start or finish talking", "I was able to convey emotions and reactions using non-verbal communication (head nodding, smiling, etc) without having to rely on a video call"
	effectiveness	"It was helpful to see smiles and head nods as feedback to my statements"
	Comfort	[picture and audio-animated modality] was less tiring for my body, but more tiring for my brain because I had to think about what people might be feeling or taking away from the meeting without any visual confirmation. Also, it was harder to know when to jump into the conversation (when no one else was talking)
6 - Representation, Personalization, and Identity	preference	"the [webcam-animated modality] felt like it allowed my personality to come across more in the discussion", "the [webcam-animated modality] was the most expressive, and made me feel most 'like myself' in a meeting", "because it's important to try new experience not losing the identity so I chose the [webcam-animated modality]"
7 - Engagement and Interest	preference	"I like the [webcam-animated modality] animations. They remind me that I'm speaking with real people in real time. ", "[The webcam-animated modality is] the most engaging and lifelike." "This was fun!", "the [webcam-animated modality] seemed like the most interesting one because it is dynamic."
	Comfort	I don't that any were more tiring, but [audio-animated and picture modality] were both boring. [picture modality], in particular. If it's just a picture, I'd rather see their profile photo"
8 - Emotional Awareness and Reaction	preference	"I appreciated that I was able to convey emotions and reactions using non-verbal communication by [webcam-animated modality]", "That [webcam-animated modality] makes me feel more comfortable and confident while I can see my peers' emotions"
	effectiveness	"the topics with picture, and with broader options were more difficult, as on scenario 3, as we could not see the people reactions"
	Comfort	"the pictures were more tiring as we could not see the reactions of the people"
9 - Cognitive Load and Attention	preference	"The [webcam-animated modality] encouraged me to stay focused on the task and discussion", "I like the expressively and lower cognitive burden of trying to deduce other people's states / make my own clear by voice only"
	comfort	"I found the [picture modality] where folks were frozen to be more tiring. I had to force myself to pay attention to the conversation".
	effectiveness	"Video was more natural for communication and had less cognitive load, but I became more self-aware and self-conscious about what my avatar was doing at any given moment (was my mouth open, weird face angle, etc)", "Session 2 is more tiring as it makes you more conscious of what you are doing"
10 - Psychological Safety and Trust	preference	"I chose the [webcam-animated modality] because it feels more natural and trustworthy", "I appeared as the [webcam-animated modality] because it felt expressive but didn't invade my privacy like a full-on camera would."
	effectiveness	"[webcam-animated modality] is for sure the more complete to create a trust with other people joining the call"

Table 3. Quotes associated with the 10 identified themes across the three focus group questions ($Q_{AS-pref-text}$, $Q_{MS-eff-text}$ and $Q_{MS-comf-text}$)

C.3 Thematic Analysis Coding Process: Branching Diagram

These coding process diagrams that are generated by Nvivo12, specifically pertain to Avatar and Meeting Satisfaction. The branching directions are provided to demonstrate a deeper understanding of initial codes and 10 themes in this paper which are analysed from both deductive coding and inductive coding procedures.

Inductive Coding Process Bottom-up process in Figure 3, Figure 4, Figure 5, shows three categories (Avatar, Social Interaction and Psychological Effects) of Avatar Satisfaction factor *preferences* ($Q_{AS-pref-text}$) with the reference count of each theme:

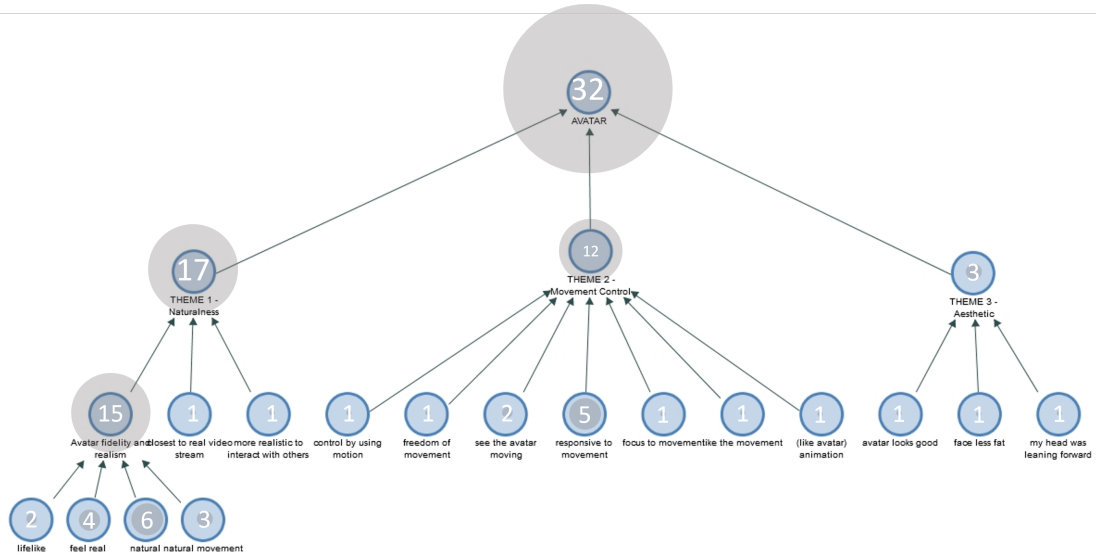


Fig. 3. Inductive coding branch for Avatar Satisfaction Factor *preference* ($Q_{AS-pref-text}$): Naturalness, Movement control and Aesthetic in category Avatar

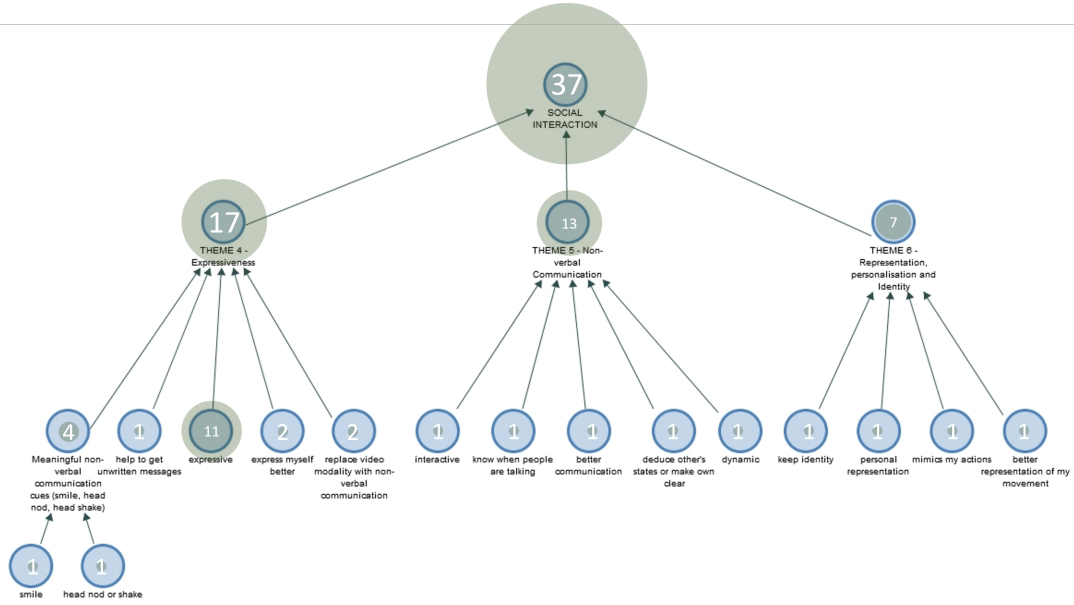


Fig. 4. Inductive coding branch for Avatar Satisfaction Factor *preference* ($Q_{AS-pref-text}$): Expressiveness, Non-verbal communication and Representation, personalisation and Identity in category Social Interaction

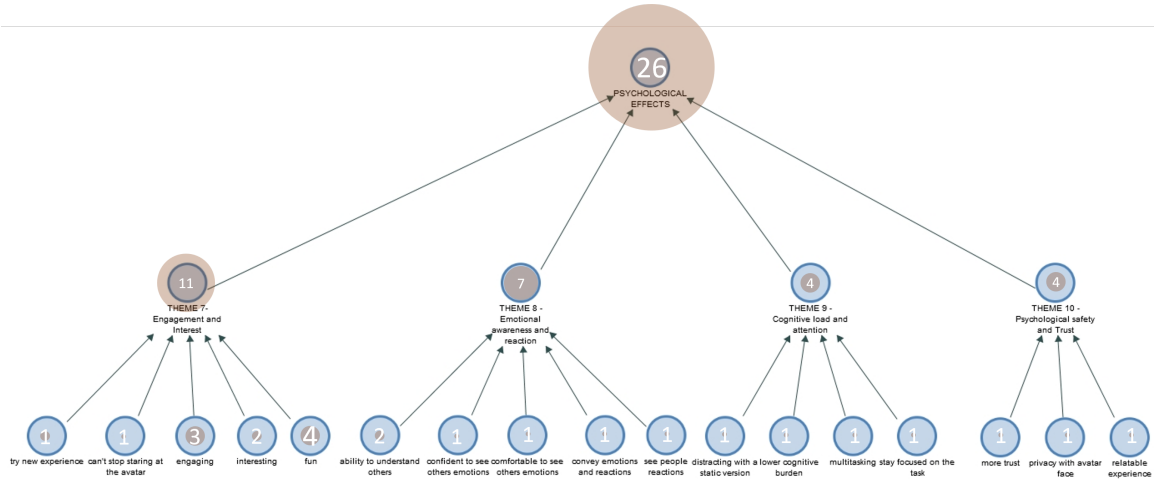


Fig. 5. Inductive coding branch for Avatar Satisfaction Factor *preference* ($Q_{AS-pref-text}$): Engagement and interest, Emotional awareness and reaction, Cognitive load and attention and Psychological safety and trust in category Psychological Effects

Deductive Coding Process Top-down process in Figure 6 and Figure 7 for two Meeting Satisfaction factors *effectiveness* ($Q_{MS-eff-text}$) and *comfort* ($Q_{MS-comf-text}$) with the reference count of each theme:

Received January 2024; revised July 2024; accepted October 2024

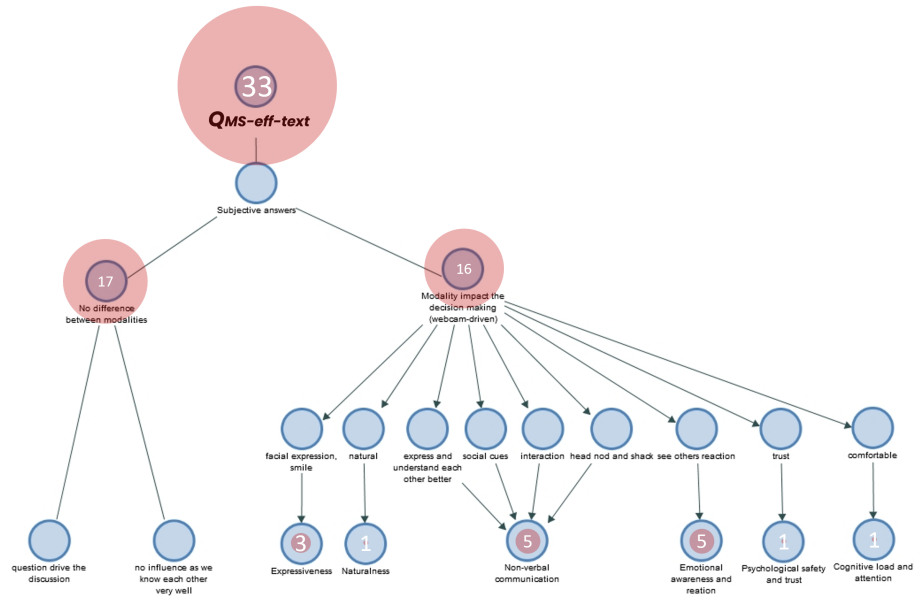


Fig. 6. Deductive coding branch for Meeting Satisfaction Factor *effectiveness* ($Q_{MS-eff-text}$) includes 6 themes: Expressiveness, Non-verbal communication and Representation, Emotional awareness and reaction, Cognitive load and attention, Naturalness, Psychological safety and trust

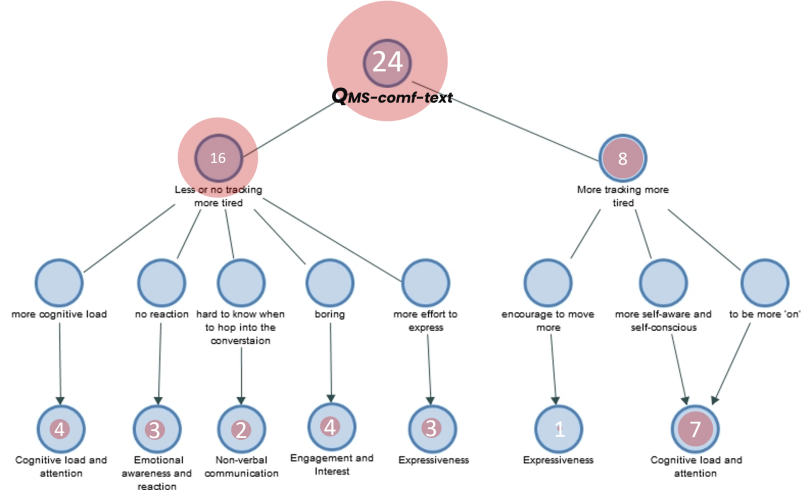


Fig. 7. Deductive coding branch for Meeting Satisfaction Factor *comfort* ($Q_{MS-comf-text}$) includes 5 themes: Expressiveness, Non-verbal communication and Representation, Emotional awareness and reaction, Cognitive load and attention, Engagement and interest