

FaceValue: Exploring Real-Time Self-View Overlays to Prompt Meaning-Oriented Self-Awareness in Remote Meetings

GUN WOO (WARREN) PARK, Department of Computer Science, University of Toronto, Canada

ANTHONY TANG, School of Computing and Information Systems, Singapore Management University, Singapore

FANNY CHEVALIER, Department of Computer Science, University of Toronto, Canada

In remote video meetings, visual non-verbal cues, such as facial expressions or head movements, are seen continuously but often only partially. This increases ambiguity compared to in-person settings and can cause misinterpretation or misalignment between intended and perceived meaning. Motivated by communication theories, we designed *FaceValue*, a technology probe that augments the self-view with private, real-time overlays. These overlays are subtle, suggestive prompts intended to help attendees reflect on how their cues might be interpreted by others. To invite personal interpretation, *FaceValue* avoids behavioral labeling and instead aims to support meaning-oriented self-awareness: recognizing when visible cues may unintentionally (mis)communicate intent. We deployed *FaceValue* in the wild with thirteen knowledge workers over multiple weeks, capturing perceived changes in self-awareness and behavior, and impressions on the design concepts, as self-reported by participants through diary entries and exit interviews. Participants felt *FaceValue* increased their awareness of potentially misaligned cues and motivated in-meeting adjustments, which they believe resulted in improved communication with other attendees. We contribute a conceptual framing that positions visual non-verbal cues as a manipulable communication resource, a technology probe that aims to foster meaning-oriented self-awareness, and empirically-grounded design insights for future meeting systems.

CCS Concepts: • **Human-centered computing** → **Interactive systems and tools**; *Empirical studies in interaction design*; *Collaborative and social computing systems and tools*.

Additional Key Words and Phrases: remote meetings, self-view, non-verbal cues, communication

ACM Reference Format:

Gun Woo (Warren) Park, Anthony Tang, and Fanny Chevalier. 2026. *FaceValue: Exploring Real-Time Self-View Overlays to Prompt Meaning-Oriented Self-Awareness in Remote Meetings*. *Proc. ACM Hum.-Comput. Interact.* 10, 6, Article CSCW086 (October 2026), 27 pages. <https://doi.org/10.1145/3816934>

1 Introduction

When attending remote meetings, we continuously convey meaning not just through what we say, but also through how we look. Our facial expressions, head movements, and gestures are on display, and other attendees use these visual non-verbal cues, like a nod, a frown, or a smile, to interpret how we feel and think [7], decide how to respond, and shape the conversation [70, 71, 103].

Remote meetings make it more difficult for attendees to interpret one another's visual non-verbal cues. Small video tiles and cropped views of the body can make some cues easy to miss or misread [20]. When attendees encounter incomplete or ambiguous cues, they fill in the gaps,

Authors' Contact Information: Gun Woo (Warren) Park, warren@dgp.toronto.edu, Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; Anthony Tang, tonyt@smu.edu.sg, School of Computing and Information Systems, Singapore Management University, Singapore, Singapore; Fanny Chevalier, fanny@cs.toronto.edu, Department of Computer Science, University of Toronto, Toronto, Canada.



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

© 2026 Copyright held by the owner/author(s).

ACM 2573-0142/2026/10-ARTCSCW086

<https://doi.org/10.1145/3816934>

sometimes incorrectly [9, 39, 52]. This can hinder collaboration: a frown might cause someone to prematurely dismiss a good idea, or a lack of visible reaction might falsely convey agreement. These ambiguities and misinterpretations can erode trust and communication [79]. Given the prevalence of remote meetings [109], we ask: *How can we assist meeting attendees in enhancing their self-awareness of their own non-verbal behaviors in remote meetings?*

The self-view, a built-in feature in most remote video meeting tools, gives attendees the ability to see their own video feeds alongside those of others. Seeing oneself can support self-monitoring [23]; however, there are two aspects to consider. First, individuals often hold the mistaken belief that others can readily perceive their internal emotional states—a phenomenon known as the Illusion of Transparency [31]. Second, when attendees use the self-view, they tend to focus more on managing their appearance than on what their expressions are communicating [17, 58]. Our work aims to help attendees notice and reflect on what their cues might be communicating to others, which we characterize as *meaning-oriented self-awareness*. With better awareness, attendees can choose to adjust their cues or clarify verbally when needed.

Prior work has shown that systems that highlight certain visual non-verbal cues can help individuals become more aware of their own non-verbal behavior. Chow et al. [18] introduced a sidebar interface that notified users of variations to (and instances of) low-level non-verbal behaviors like smiling, nodding, gaze, and posture. This real-time information helped users become more aware of their self-presentation in remote meetings. While Chow et al. focused on monitoring low-level behaviors from the user's own perspective, we take a complementary approach by asking a self-reflective question: how might others perceive these behaviors, particularly in terms of communicative intent? Our work extends this trajectory by not only detecting the presence of cues but also surfacing their likely interpretations. This real-time feedback helps users notice when they may be sending unintended signals and adjust their behavior in the moment. Our goal is to support people in maintaining this *meaning-oriented self-awareness*.

To explore the concept of meaning-oriented self-awareness, we propose to augment the self-view with private, real-time overlay elements that highlight visual non-verbal cues when those cues may communicate meanings to others. To realize this approach, we derive design goals (Section 3) grounded in communication and design theories (Section 2): detecting non-verbal cues that might convey specific meanings, and through real-time overlays, encouraging users to reflect on these behaviors.

To examine how meaning-oriented self-monitoring influences attendees' awareness and adjustment of non-verbal cues across varied meeting contexts, we implement these design goals through a technology probe, *FaceValue*, that analyzes facial expressions and head movements, and augments the self-view with private comic-style overlays that can evoke attendees' own meaning-oriented self-awareness (Section 4). We deployed *FaceValue* in an in-the-wild study with 13 knowledge workers (Section 5). Participants used *FaceValue* as part of their regular remote meetings, professional or casual. Our focus is on the perceived impact of *FaceValue* on one's awareness and behavior as self-reported by participants, collected through self-administered diary entries and post-study semi-structured interviews. Participants used *FaceValue* for five meetings, over the course of one to three weeks. Our study focused on the following questions:

RQ1: How does real-time feedback on the potential interpretations of one's visual non-verbal cues help remote meeting attendees *notice* when their behavior may align or misalign with their communicative intent?

RQ2: How do attendees describe *acting* on meaning-oriented feedback in order to realign their visible cues with their communicative intent?

RQ3: How do meeting characteristics (e.g., stakes, formality) shape attendees' perceptions of the feedback's usefulness and potential for distraction?

Participants felt that using *FaceValue* helped them observe their self-view more intentionally. They reported feeling more able to notice moments when their on-screen behavior might misalign with their intended messages. They also emphasized that *FaceValue* was most useful during high-stakes or goal-oriented meetings, where the perceived importance of communicative clarity was higher. These findings suggest that private, metaphorical overlays can give attendees a provisional reference point for monitoring and adjusting their visual non-verbal behavior, supporting meaning-oriented self-awareness without imposing prescriptive feedback (Section 6).

Our contributions are:

- (1) Conceptual framing. We frame visual nonverbal cues in video as a way to reinforce intended meaning, while recognizing that they may also unintentionally convey unintended meaning. Through this conceptualization, we motivate the need for meaning-oriented self-feedback.
- (2) *FaceValue*, a lightweight technology probe that aims to foster meaning-oriented self-awareness and reflection through private, comic-style overlays.
- (3) Empirical insights from an in-the-wild deployment with 13 participants, indicating participants feel that meaning-oriented feedback can help them align their expressions with their intentions. We discuss scenarios where *FaceValue* is most effective, and based on a reflection about potential unintended consequences, we outline implications for designing future context-sensitive meeting tools.

2 Background

We first outline theories of communication that ground our work (Table 1). Then, we discuss relevant tools and current practices for self-monitoring in remote meeting, with a focus on the self-view. We also review prior systems supporting interpretation of visual non-verbal cues.

2.1 Theoretical Grounding: Visual Non-verbal Cues in Remote Meetings

Visual non-verbal cues such as facial expressions, gestures, and head movements are integral part of communication [7] (Table 1–T1). These cues indicate understanding, enrich verbal explanations, and convey attitudes toward the conversation content [1, 47]. To support effective communication, visual non-verbal cues should ideally be both clear and congruent with one's intended message. However, such cues are inherently ambiguous (Table 1–T5,T6): different observers can derive different meanings from the same cue [1, 4], and these perceived meanings may not always align with one's communicative intent [32, 105]. Because people tend to over-estimate how well their cues are understood by others (Table 1–T4), they may not consciously refine their cues. This can be a problem, as research suggests that individuals often rely on visual non-verbal cues to interpret communicative meaning, sometimes placing greater trust in them than in speech, especially when the two conflict [70, 71].

This ambiguity is even more pronounced in remote meetings due to reduced media richness (Table 1–T2). Video tiles are often small, crowded, or cropped, causing many subtle visual cues to be lost or only partially conveyed [79]. Because people are fundamentally motivated to reduce uncertainties about others' emotions or intentions, they actively engage in drawing meaning from these limited or unclear visual non-verbal cues (Table 1–T6), intentionally or not, as long as video is enabled [103]. The problem is that the gap-filling process associated with interpreting missing, partial, or leaked cues is often biased [39, 52], resulting in inaccurate conclusions and misjudgment. Although misinterpretations are often overlooked and inconsequential, empirical studies demonstrate tangible harm. For example, a study showed that remote attendees exhibit

Table 1. Overview of the theories that inform our perspective on meaning-oriented self-awareness. Visual non-verbal cues have significant roles in communication, yet remote meeting environments often degrade these cues. This uncertainty affects the *perceiving* attendee, who may misinterpret ambiguous cues, and the *cueing* attendee, who may be unaware of how their cues are being perceived. Our work identifies gaps supporting **cueing attendees'** self-awareness and draws on theories and design lenses to guide our approach.

ID	Theory	Perspective	Role in this Work
T1	Transactional Model of Communication [7]	General Communication	Explains why visual non-verbal cues remain relevant even when verbal cues are available.
T2	Media Richness Theory [20]	Meeting Medium	Highlights that remote meeting media can hinder the effective conveyance of visual non-verbal cues, thereby reducing communication richness and potentially increasing ambiguity in the interpretation of those cues.
T3	Objective Self-Awareness [23, 108]	Cueing Attendee	Describes how individuals focus on themselves during communication, affecting their awareness and adjustment of expressive behavior.
T4	Illusion of Transparency [31]	Cueing Attendee	Explains how individuals overestimate how accurately others perceive their internal states and intentions.
T5	Social Information Processing [103]	Perceiving Attendee	Describes how people adaptively extract and interpret cues from available information sources in mediated interactions.
T6	Uncertainty Reduction Theory [9]	Perceiving Attendee	Explains how individuals seek to reduce uncertainty by interpreting available cues, sometimes leading to over-interpretation or bias.
T7	Affect-as-Interaction [10]	Design Lens	Encourages designing for emotional expression as socially situated and interpretable, rather than as a static, classifiable signal.
T8	Ambiguity-as-Design-Resource [30]	Design Lens	Positions ambiguity not as a flaw, but as a productive feature that invites reflection and deeper interpretation in user experience.

more negative affect and a tendency to perceive hostile intents from others [79]. Misattributions have also been shown to undermine team success [107].

Efforts to reduce cue ambiguity in remote meetings include refined symbolic cues, such as emoji reactions, button-press feedback [2], or explicit turn-taking intent displays [44]. Another strategy is to stream only a single visual cue, such as gaze [38], or machine-interpreted emotional states [37]. While both approaches reduce communicative noise, they also reduce richness. This tradeoff can be acceptable for simple communication tasks, but in more complex tasks, it can increase uncertainties [9, 20]. Professionals' preference for videos [27]—hence visual non-verbal cues—underscores the importance of supporting this form of communication in remote meetings.

Our work takes a different approach to reducing cue ambiguity. Motivated by Chow et al.'s suggestion that “the evaluation of one's own non-verbal behavior from the perspective of the other attendees [could] inform actionability” [18], we provide private feedback that shows how a participant's cues *might* be interpreted by others. We emphasize that the feedback reflects *possible* meanings inferred from automated analysis of the attendee's own cues, and does not incorporate actual data from other attendees' interpretations during the meeting, as these remain internal. This feedback is designed to help individuals consider potential (mis)alignments with communicative intent, leaving them free to decide whether any adjustment is needed. In other words, our objective is to encourage self-assessment and deliberate action toward reducing ambiguity at the source by evoking meaning-oriented self-awareness in the cueing attendee (Table 1–T3).

2.2 Self-Monitoring in Remote Meetings

When interacting with others, we naturally engage in *self-monitoring*, i.e. aligning our expressive behavior with social expectations and communicative goals [93]. In this paper, we use the “self-monitoring” construct in a situated, behavioral sense, i.e., moment-to-moment attention to and adjustment of one's expressive cues during interaction [59, 72]. This is related to, but distinct from the original psychological construct of self-monitoring as a stable *trait* [93], which describes

individual differences in how people monitor their social context, and adapt their self-presentation, but does not refer to any specific behavior (e.g., literally looking at one's own face more often). We return to this distinction in [subsection 7.6](#) when considering how trait self-monitoring may shape people's experiences with our system. In this behavioral framing, self-monitoring goes hand in hand with *self-awareness* of expressive cues. Consistent with this view, some studies suggest that awareness of one's expressions can help calibrate perceived demeanor [23] and user control [87], and thus enhance clarity, rapport [96, 102], and relational communication [73].

In remote meetings, the self-view panel, focused on the face of attendees, can facilitate this process, by enhancing self-awareness and allowing real-time adjustment of visual non-verbal cues. However, prolonged exposure to one's own image through the self-view can lead to fatigue, anxiety [5, 91, 92], discomfort [22], or cognitive load [87]. These effects vary across individuals [56] and may not arise at all for some [49, 60]. For others, however, aspects of the self-view can undermine its benefits or introduce additional challenges.

Psychology often frames self-awareness as a component of mindfulness, typically associated with well-being [35, 95, 101], which contrasts with reports that self-view may induce stress [83]. There are reports that self-views often result in monitoring for more cosmetic cues e.g., hair, background [17, 58], which serves little function in communication. Building on these insights, we hypothesize that focusing on communicative cues (e.g., expressions of agreement) may support more purposeful interaction with the self-view. Our work explores this approach by augmenting the self-view with feedback on potentially communicative visual non-verbal cues (e.g., subtle displays of agreement or confusion), as a mechanism to supporting attendees in recognizing, interpreting, and adjusting their own cues in remote meetings.

While not our primary focus, we note that leveraging the self-view also allows to preserve other benefits this view affords, such as coordinating objects relevant to conversations within the camera's frame [36, 45, 62, 63, 77]. This aspect of communication is important and complementary to our focus on visual non-verbal cues, albeit beyond our scope.

2.3 Visualizing Visual Non-verbal Cues

Many remote collaboration tools aim to visualize participants' emotional or engagement states to help interpret group dynamics. These include emoji reactions, active-speaker highlighting [76], and experimental dashboards [2, 8, 21, 26, 57, 67, 110]. While helpful for group awareness, such interfaces offer limited support for *personal* self-monitoring. Feedback is often public, encouraging shared interpretation rather than private reflection. Our work introduces a personal reflective tool to support individual self-monitoring.

When dealing with visual non-verbal cues, there are two design approaches that we can consider. Under the affect-as-information paradigm [10], visual non-verbal cues are treated as classifiable data [2, 8, 18, 21, 26, 57, 84, 110]. Visualizations here are often literal, using labels or percentage scores presenting discrete, context-free observations, such as "three head nods in the past two minutes" or "a 55% probability of smiling." Such systems often still avoid prescriptive *guidance* (e.g., "nod twice now"), and instead adopt non-normative stances for feedback, which leaves the action plan to the users. For example, Chow et al. [18] presents counts of detected behaviors like smiles or nods, but leaves interpretation entirely to the users, to preserve authenticity in expressions. This paradigm supports analytical perception of facial expressions [11, 16, 55, 88], but risks over-confidence in Facial Expression Recognition (FER) accuracy [84, 111].

By contrast, the affect-as-interaction paradigm (Table 1–T7) treats visual cues as meaningful through interaction. For instance, AffectiveSpotlight [76] enlarges video feeds rather than labeling them, and other systems [43, 81] evoke reflection through ambient feedback. This paradigm is not inherently superior to affect-as-information; each supports different user needs. If the goal is

quick, quantifiable awareness, the information paradigm may work better. However, if the aim is to support users in reflecting on how their cues might be perceived holistically and contextually, the interaction paradigm offers a more interpretive lens [10]. We adopt this latter approach to evoke user's holistic awareness on not just what cues occurred, but how those cues may align, or misalign, with intended meaning. Facial expressions are often perceived holistically [14, 15, 97], so prompting reflection on multiple integrated cues can be important. Designing for ambiguity, via color gradients, abstract visuals, or symbolic shapes [69, 89], can support this paradigm by encouraging interpretation. We similarly use ambiguity as a resource for design (Table 1–T8) to stimulate attendees' reflection.

Closest to our work is the *Novecs* system that automatically detects and displays users' visual non-verbal cues, to support workers' self-presentation in remote meetings [18]. Our work shares similar design choices: providing real-time personal feedback in a non-normative way, which Chow et al.'s experiment showed can enhance perceived self-awareness of the participants [18]. Motivated by these findings, we extend this line of research by exploring two novel aspects, drawing on theories of communication pertaining to ambiguity in cues (Section 2.1). First, *Novecs* displays feedback in glanceable sidebar of facial and postural cues—a form of separate dashboards [57, 110]. Here, we explore in-situ overlays, as they may help attendees connect feedback with real-time cues [80]. Second, *Novecs* focused on smiling, nodding, gaze, and posture in isolation. Our work looks into ways to enable more holistic interpretation of one's own visual non-verbal cues, with respect to how others may (mis)interpret the meaning behind these cues—a future direction suggested by Chow et al. [18]. Finally, Chow et al.'s longitudinal study focused on work meetings. We follow a similar methodology, but invited participants to experiment with our probe in meetings of their choosing, allowing us to expand insight on the type of meetings where personal meaning-oriented feedback might be valuable.

2.4 Recognizing Visual Non-verbal Cues

To provide feedback, we need to detect visual non-verbal cues and infer their potential meanings. Prior work in remote communication has largely focused on verbal cues [41, 104], but visual non-verbal cues, especially facial expressions, are also key [7]. Advances in affective computing have yielded Facial Expression Recognition (FER) systems that can match or surpass untrained human judges on constrained benchmarks [61]. Seminal work comparing posed versus spontaneous expressions [75] warns, however, that benchmark superiority may not transfer to natural settings. FER has been applied in education [94, 98, 110], mental health [29, 48], and robotics [53, 64].

FER commonly uses categorical (e.g., happiness, anger [24]) or dimensional (valence–arousal [85]) models. Categorical models suit users seeking direct feedback [89], while dimensional models capture nuance. However, FER accuracy depends on training data, which may lack real-world complexity [51, 65, 74, 78], and systems still lack contextual understanding. Privacy, demographic-bias, and other ethical concerns persist [3, 82], prompting some companies to discontinue FER services (e.g., Microsoft's 2022 withdrawal of emotion recognition from Azure Face [19]).

To balance ambiguous, reflective feedback [10, 30] with these challenges, we display FER output only to the attendee, not to others. This private feedback helps attendees refine expression with reduced risk of misinterpretation. We incorporate categorical and dimensional outputs to support varied preferences. We focus on facial expressions and head movements due to their salience in communication [1, 40, 86]. Future work may expand to hand movements or posture.

3 Design Goals

Drawing on theory, we identify two insights that motivate our approach and inform the design goals that follow. (1) People tend to overestimate how clearly their internal states are understood (Table 1-T4). We hypothesize that presenting feedback supporting this understanding could help, since theory suggests that (2) making oneself an object of attention supports regulative self-monitoring (Table 1-T3). To achieve this, we reconceptualize the self-view as an active, reflective communication aid that supports meaning-oriented self-awareness. We articulate concrete design goals:

- **DG1: Detect and Surface Visual Non-verbal Cues that Convey Meaning.** Actively detect and highlight visual non-verbal cues most salient to conversational meaning, to help users gauge how others may interpret their non-verbal cues. The purpose of doing this is to re-center users' attention on communicative aspects of their appearance, rather than cosmetic self-checks [17, 58]. Our implementation focuses on facial expressions and head movements—such as nods, frowns, and smiles—as these are cues interlocutors use to infer emotional state and stance [1, 40, 86].
- **DG2: Provide Subtle, Suggestive Feedback.** Invite user's interpretation and reflection without overstepping into definitive judgments about their internal state. Doing so guards against prescribing normative behavior [18]. Given that FERs are imperfect, recognition results can vary. Thus, we treat its output as one plausible but fallible reading of visual non-verbal cues, rather than as ground truth. Informed by the affect-as-interaction paradigm [10], we choose against rigid categorizations of emotional states and stance (e.g., “you look angry”) that can reinforce overly-analytic self-monitoring. Instead, we realize this design following principles of ambiguous design [30], through suggestive display of the detected visual non-verbal cues that invites a wider range of interpretations. For instance, we could overlay gently suggestive visual hints for valence shifts within the self-view.
- **DG3: Sustain Real-Time, In-Context, Continuous Feedback.** Deliver immediate, ongoing, in-situ feedback about how visual non-verbal cues could be interpreted by others. Doing so enables in-the-moment evaluation, thereby making it possible for the cueing attendee to iteratively adjust their visual non-verbal cues or offer verbal clarification in the flow of conversation, aligning more closely with their communicative intentions. We use in-situ self-view overlays to facilitate connections between feedback and one's real-time cues [80].

Applying these design goals, we turn the self-view into an *active* mirror that draws attention to visual non-verbal cues when attention is needed (DG1). We provide suggestive overlays rather than prescriptive judgments on what these cues could mean to others (DG2), in-situ and in real-time (DG3). This meaning-oriented feedback approach aims to invite reflection, which could encourage attendees to iteratively adjust their visual non-verbal cues to match their communicative intentions.

4 Design and implementation of the FaceValue probe

We implemented *FaceValue* as a proof-of-concept technology probe [46] to explore how people experience and incorporate meaning-oriented feedback. *FaceValue* operationalizes our design goals (DG1, DG2, DG3) to explore meaning-oriented self-awareness. The system interprets users' communicative visual non-verbal cues (e.g. agreement, neutrality, confusion, and so forth) by detecting facial expressions and head movements, and augments the attendee's self-view so they are aware of how their expressions may be interpreted by others. For example, if a user smiles, a thick green band encircles their head in the self-view (Figure 1), suggesting (but not prescribing) agreement or positivity. For certain expressions or head movements, the user's self-view will show overlaid icons or lines. For example, if a user is nodding, they will see speed lines at the top and bottom of the face (Figure 3). The system also monitors for sudden shifts in expression valence and



Fig. 1. *FaceValue* augments the user’s self-view based on its interpretation of their communicative expressions. For example, when the user appears to be expressing agreement—expressed with a smile—their face is outlined in green. Conversely, when they appear to express disagreement—expressed with a frown—their face is outlined in yellow. Some specific expressions or head movements are accompanied by an icon. For instance, a smile is paired with a sparkle icon, while an angry face is paired with a popping vein icon. The possible interpretations, such as “Happy” or “Angry” are not visible to users.

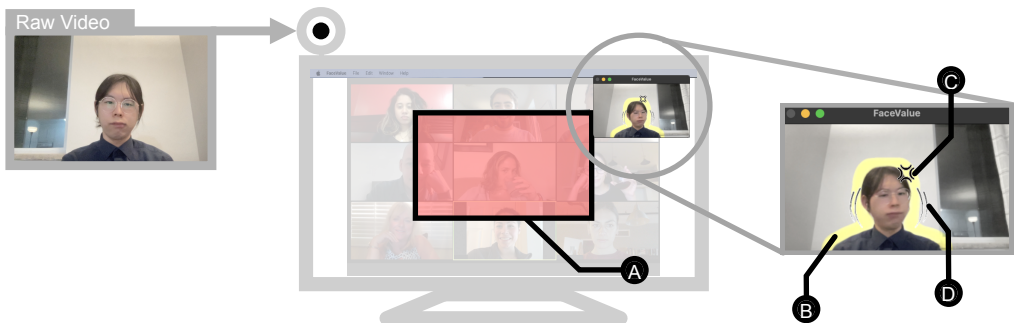


Fig. 2. Conceptual illustration of the augmented self-view in *FaceValue*. (A) A brief red overlay signals a significant change in valence; (B) Yellow lines around the face; (C) Subtle icons hinting at possible interpretive cues (e.g., anger, surprise); (D) line overlays indicating minor movements that might otherwise be overlooked.

alerts the user when a major change occurs—for example, when a smile or nod quickly turns into a frown or head shake. Since these augmentations are only visible to the user in their self-view, our aim is to promote self-awareness of the potential meaning of their communicative expressions.

4.1 An Example Scenario

To illustrate how the *FaceValue* probe works, consider a graduate student, Taylor, joining a team meeting with Ash. Ash is presenting a new UI design idea for their project.

Moment 1: Unintended Negative Expression. When Taylor becomes briefly confused, they inadvertently frown for a moment. *FaceValue* detects a shift to negative valence and briefly displays a red, semi-transparent overlay (Figure 2 (A)). Looking at the augmented self-view, Taylor notices *yellow lines* around their face and a “popping vein” icon (Figure 2 (B, C)), suggesting that their expression may be interpreted by others as frustration or negativity. Realizing this might send the wrong signal, Taylor consciously relaxes their face, preventing a potential misunderstanding.

Table 2. Details of the *FaceValue* mapping between recognition and the visual overlays

Measure	Input	Mapping
Valence (positive vs. negative)	$V > 0, A > 0$	Outline colour: Green
Arousal (intensity)	$V > 0, A < 0$	Outline colour: Blue
	$V < 0, A < 0$	Outline colour: Red
	$V < 0, A > 0$	Outline colour: Yellow
	Euclidean distance	Outline opacity ($\text{opacity} \propto \text{distance}$)
Detected Facial Expression	Happy	Sparkle icon
	Sad	Droplet icon
	Surprised	Exclamation icon
	Angry	Popping vein icon
Head Movement	Nodding	Lines on the top and bottom of the face
	Shaking	Lines on the left and right of the face
	Tilting	Lines coming from the face

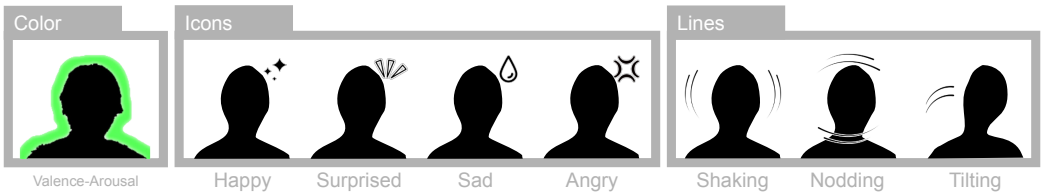


Fig. 3. The “Visualization Mapping” phase produces three main forms of visual cues: color-based overlays (left), icons (center), and lines near the face (right).

Moment 2: Subtle Agreement Expression. Later, Ash describes one of the ideas. Taylor nods slightly in agreement. *FaceValue* identifies a mild motion and displays faint lines around Taylor’s face (akin to Figure 2 (D) but for nods), hinting that the nod may be too subtle for others to notice. On realizing this, Taylor amplifies the nod, signaling clear support to Ash.

In this scenario, *FaceValue*’s private, nuanced cues in Taylor’s self-view help them to refine how they project agreement or uncertainty. Rather than prescribing a “correct” expression, the system merely alerts the user of shifts in facial expressions and head movements (DG2), giving them the freedom to interpret and adjust as deemed fit (DG1, DG3).

4.2 Sensing: Recognizing Facial Expressions and Head Movements

To direct users’ attention toward meaningful visual non-verbal cues (DG1), *FaceValue* first identifies facial expressions and head movements. We employ *EmoFAN* [99], a model that outputs categorical expressions (e.g., happy, angry) and dimensional expressions (valence/arousal). *EmoFAN* achieves high accuracy on ranges of common datasets, including $\approx 75\%$ on AffectNet [74] (categorical)¹ and an RMSE of 0.29/0.27 for valence/arousal regression, which we consider sufficient to indicate expressive fluctuations in real-world usage. These raw detection labels are not visible to users. To detect head movements, we extract the 3D transformation matrix from *MediaPipe* [66] (pitch, roll, yaw) which allows us to capture behaviors such as nodding, shaking, or tilting.

Our implementation runs at 10+ FPS on an Apple M1 chip with minimal latency, providing real-time responsiveness (DG3).

4.3 Visualization Mapping

In the sensing stage, *FaceValue* captures valence and arousal values, categorical emotion confidences, and head movement signals. We map these values to visualization elements that aim to be subtle and

¹This model has been tested for other major datasets as well, and shows excellent performance.

suggestive (**DG2**) to encourage user-driven interpretation, as described in [Table 2](#), and illustrated in [Figure 3](#). Inspired by the visual language of comics/manga [\[28, 90\]](#), our design lets users see hints of their expressions without receiving explicit emotion labels that can mislead.

Valence & Arousal (VA) $\Rightarrow$ Outline Color and Opacity. We adapt a hue scheme similar to prior work [\[50\]](#), where specific regions in VA space map to colors. For example, excited/happy tends toward green, sad/depressed toward red, angry/afraid toward yellow, and relaxed/calm toward blue. Intensity is encoded in the overlay’s opacity, derived from the Euclidean distance from the VA origin (i.e., stronger expressions lead to higher opacity). To avoid rapid color flicker from frame-to-frame noise, we apply a Kalman filter [\[106\]](#) for smoothing.

Detected Facial Expression $\Rightarrow$ Icons. When a detected emotion’s confidence exceeds 80%, *FaceValue* displays an icon: sparkle (happy), droplet (sad), exclamation (surprise), or popping vein (anger); see [Figure 3](#) (center). While explicit, these icons are intentionally ambiguous rather than literal, consistent with **DG2**’s philosophy of inviting contextual interpretation. For example, a popping vein icon may suggest anger or frustration, but we omit textual labels; interpretation is left to users, aligning with Chow et al.’s non-normative feedback design [\[18\]](#).

Head Movements $\Rightarrow$ Lines. Using pitch/roll/yaw data, we detect nods, shakes and tilts. For nods or shakes, we compute head position variance within a sliding window, accounting for their cyclic nature. If above a threshold, lines appear around the user’s face ([Figure 3](#), right). For tilts, we use the mean angle. Higher variance increases line opacity, signaling intensity. Thresholds can be tuned for different movement styles.

4.4 Composition: Placing Visual Elements on the Attendees’ Videos

FaceValue displays the icons, lines, and hue overlays directly over/near the user’s face [\[80\]](#) rather than off to the side [\[57, 110\]](#). This design supports **DG2** by allowing immediate comparison between the raw facial expression and the abstract cue. Because the system avoids explicit emotion labels, the user retains interpretive flexibility (e.g., a popping vein icon might prompt them to think, “Am I frustrated?”, “Am I concentrating?”, or “Did the model misinterpret my brow furrow?”).

A key design goal is to shift users from casual glances at their appearance to purposeful checks of communicative expressions. Conventional self-views do not explicitly support noticing significant changes in facial expression, so *FaceValue* detects sudden valence shifts and briefly displays a semi-transparent overlay ([Figure 2](#) (A)). These short-lived “flash” signals prompt users to reflect: “Am I aware my expression changed? Was it intentional?” Additionally, to help the face stand out, we reduce background saturation using MediaPipe [\[66\]](#), reinforcing focus on the face.

5 Deployment Study

The goal of the study was to better understand how an augmented self-view might support users’ self-awareness of their own communicative visual non-verbal cues during video conferencing. Specifically, we wanted to probe how people experience and interpret their own cues. To do so, we conducted an in-the-wild deployment study exploring how individuals integrated *FaceValue* into real-world remote meetings, and how this shaped their perceptions, experiences, and any perceived changes in behavior or communication practices they felt emerged through its use.

Following Chow et al. [\[18\]](#), we opted for an in-the-wild deployment to address our research questions. Short-term laboratory experiments often oversimplify meeting dynamics, are prone to the novelty effects [\[25\]](#), and are overly influenced by unique social dynamics [\[34\]](#), while lacking the time needed for users to integrate the tool into their routines [\[54\]](#). In contrast, in-the-wild studies excel at capturing real-life usage conditions, increasing ecological validity, and revealing genuine patterns of tool adoption and use [\[13\]](#), which better align with our goals.

Since our research focuses on self-awareness of the cueing attendee, we relied on internal measures, i.e. participants' self-reports, to probe the impact on their own experiences. We employed a diary study approach [42], where participants, after having engaged in a meeting with *FaceValue*, reported on their experiences using *FaceValue* in a semi-structured way. We used this approach to capture participants' immediate reactions to the tool during deployment, and followed this up with a post-study interview to capture longer-term reflections. External measures, such as objective observations of behavior change, are challenging in this context because real-world meetings vary greatly in topics, attendees, and dynamics. Capturing meaningful patterns would require extensive controls and a consistent baseline of similar meeting contexts over time, which was beyond the scope of this study. We discuss the limitations of our approach in [subsection 7.6](#).

5.1 Participants

We recruited 15 participants via convenience sampling, snowballing, and advertisements on an undergraduate portal and a HCI Slack group. Two discontinued due to busy schedules, leaving 13 full participants. All but one used a Mac with Apple Silicon Chip², ensuring consistent performance for real-time video processing. All participants were fluent in English, 18 or older, and planned to attend at least five remote meetings over one to three weeks.

Participants included 5 women and 8 men, all Canada-based, with occupation spanning: business analyst (P4), assistant professor (P1), UX researcher (P2), postdoctoral researcher (P6), system administrator (P12), graduate (P3, P5, P8–11, P13) and undergraduate student (P7). Seven were aged³ 25–34 (P1–2, P5, P9–11, P13), four 18–24 (P3–4, P7–8), one 35–44 (P6), and one 45–54 (P12).

5.2 FaceValue Application

FaceValue was deployed as an executable macOS application running on participants' computers during their real meetings. All processing was done on-device with no logging. Participants used *FaceValue* alongside their preferred software (e.g., Zoom, Microsoft Teams, FaceTime). The *FaceValue* window was instructed to be placed over the standard self-view (moved and resized), allowing participants to receive real-time feedback on their facial expressions and head movements.

5.3 Procedure

We conducted a three-phase diary study:

- **Phase 1 (Onboarding)**, ~ 30min. In a Zoom session, participants provided informed consent, installed *FaceValue*, were instructed on what *FaceValue* does and how to use it, answered a pre-study questionnaire (demographics and general meeting experiences), and participated in a semi-structured interview about prior remote meeting experiences.
- **Phase 2 (In-the-Wild Usage)**, 1-3 weeks. Participants used *FaceValue* in five actual video conferencing meetings of their choice with respect to type (e.g. work related, casual meeting with friends, etc.), and size. After each meeting, they filled out a diary entry capturing key metadata (e.g., duration, topic, number of attendees) and describing any notable experiences they had while using *FaceValue*; see section 5.4. Not accounting for meeting time, we estimate participants spent 5min to fill out each diary.
- **Phase 3 (Post-Study Interview)**, ~ 30min. Participants completed a post-study questionnaire (including preference tests) and a semi-structured interview conducted on Zoom, about their overall experiences, focusing on perceived changes in communication and self-awareness of

²P12 used an Intel iMac, which we confirmed provides equivalent experience. The minimum hardware was a MacBook Air with M1 Chip and 8GB of RAM.

³Collected in ranges.

one’s communicative visual non-verbal cues. During the interview, participants were asked to compare experience *with* and *without FaceValue*.

Participants received a \$30 CAD gift card in compensation. The study was approved by our institutional ethics board and piloted to refine procedures before the main study. Study materials are provided in the supplemental.

5.4 Diary instrument

Participants were asked to complete a diary online after each meeting with *FaceValue* via a personalized link with pre-filled participant ID. These diaries aimed to capture their immediate reflections on using *FaceValue* in that meeting, and how their perceptions of *FaceValue* evolved over time. Table 3 shows the questions in this instrument.

Table 3. Questions in the diary, prompting short-form answers (with drop-down lists of possible values for Q1–6), Likert scores (strongly agree (1) to strongly disagree (5) for Q7–8), or long-form answers (Q9–13).

Question
Q1: Date and time of the meeting
Q2: Type of the meeting
Q3: Duration of the meeting
Q4: Usual frequency of the meeting
Q5: Number of attendees
Q6: Perception of speaking more, less, or about the same during the meeting, compared to others
Q7: Extent of agreement: I viewed myself more often
Q8: Extent of agreement: I viewed others more often
Q9: What were you trying to understand about yourself when you glanced at your <i>FaceValue</i> ?
Q10: Were there moments when <i>FaceValue</i> ’s augmented self-view felt particularly helpful or distracting?
Q11: In what ways could <i>FaceValue</i> have helped (you) or changed this meeting?
Q12: Did any attendee comment on your appearance, expressions, or demeanor in a way that might relate to your on-camera presentation?
Q13: Any other thoughts or observations?

5.5 Analysis

We performed a thematic analysis of the semi-structured interviews transcripts and written diaries. We inductively captured themes relevant to our research questions, following an interpretivist approach to reflexive thematic analysis [12]. Our analysis focused on participants’ explicit descriptions of their self-monitoring experiences and perceptions, considering presence, rather than prevalence, in line with this approach. First, the lead author reviewed the transcripts and diary entries in full, and performed open, line-by-line coding in MAXQDA⁴. This step generated codes that captured salient themes and concepts in the participants’ account. The lead author then iteratively reviewed and refined the initial codes, merging, splitting, and discarding codes as necessary with input from all authors in weekly discussions. The authors’ cultural backgrounds span East Asia, Europe, and North America, and each has lived and worked in at least two countries across these regions (see Supplementary Material). This may have shaped our interpretive lens (e.g., meeting norms centered towards “Western” meeting styles) during analysis.

6 Findings

We collected 65 diary entries, spanning diverse meeting contexts (see Figure 4). We report on key findings from our analysis of diaries and interviews, organized by research question.

6.1 Impact of Meaning-Oriented Feedback on Noticeability of Visual Non-Verbal Cues

Our RQ1 asked: How does real-time feedback on the potential meanings of visual non-verbal cues help remote meeting attendees *notice* when their behavior aligns or diverges from their

⁴<https://www.maxqda.com>

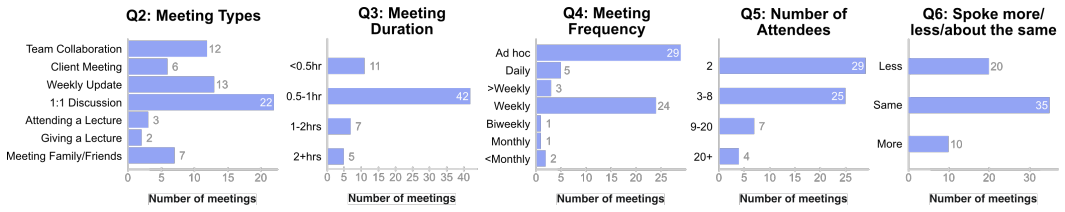


Fig. 4. Breakdown of meeting contexts reported in the diaries (65 entries - See also: Supplementary Material).

communicative intent? We found that not only did most participants perceive they engaged more frequently and for longer periods of time on their self-view, they recalled a perceived shift in focus from appearance self-checks to considering how their visual non-verbal cues influence communication. For some, this resulted in a perceived process of more deeply reevaluating how they routinely present themselves to others and the impacts on others' impressions and understanding.

Participants noted an increased perceived engagement with *FaceValue*, compared to traditional self-views with no feedback. They felt that using *FaceValue* transformed occasional self-checks into frequent (10/13), and longer (7/13) glances at their self-view. For example, P3 described how *FaceValue* prompted them to monitor themselves during the meeting rather than just during initial setup:

"I'm definitely looking at myself more because I think, when I join the meeting [without *FaceValue*], I make sure everything looks OK. And if I say something in the middle of the meeting and I'm not sure where to look, I would just look at myself. But with *FaceValue*, I feel it nudges me to look [at myself] more." – P3

Many participants (8/13) implied that it was *FaceValue*'s subtle nudges that prompted more frequent self-checks. Participants' diaries also indicate that they felt they viewed themselves more often (3.85 ± 1.01 , slight agreement) compared to their regular remote meeting experiences. Whether this increased engagement with the self-view is purposeful and useful is discussed next.

Participants conveyed the impression that the augmented self-view shifted their attention towards communicative meaning of their visual non-verbal cues. Participants reported that *FaceValue* led them to pay closer attention to the potential meanings conveyed through facial expressions and head movements (10/13 participants; 46/65 diary entries). They shared:

"I think what *FaceValue* added is that, before, when I looked at myself, it was more just for checking how I look. Unless I paid extra attention to my facial expression, I couldn't tell what I was communicating." – P6

"Compared to meetings without *FaceValue*, I now focus on how others will notice my facial expressions and how well I express my emotional feelings to the audience. Thanks to *FaceValue*, I keep my attention on visual non-verbal cues." – P10

Participants felt that meaning-oriented feedback enabled them to notice whether their visual non-verbal cues conveyed meaning in alignment with their communicative intent. For instance, P3 and P1 reported using *FaceValue* to help them maintain this self-awareness to ensure they communicated their engagement with the meeting:

"*FaceValue* allows me to notice when my attention drifts away from looking at other people, which can sometimes translate very poorly to the other side." – P3

"I need to make sure that I'm actively engaged as someone partly responsible for leading the project. So, there were times when I was trying to look at *FaceValue* and get that sense." – P1

Similarly, P9 commented on how *FaceValue* can help with conveyed emotions:

“I feel bad if I look grumpy during a meeting. Luckily, I don’t think the red box appeared too much. But when it did, I thought: ‘Okay, maybe I shouldn’t look annoyed.’ Sometimes it’s not that I’m actually angry, but it’s how others might perceive it, so I adjusted my expression.” – P9

Such an active assessment of how others may possibly misinterpret one’s communicative intent suggests that some participants may have begun to counter the Illusion of Transparency: the mistaken belief that one’s internal state is clearly understood by others (Table 1–T4). Critically, participants expressed that this self-awareness is focused on the communicative meaning implied by their non-verbal facial expressions—even when they are not speaking.

Some participants reported engaging in broader reflection on meaning-oriented self-presentation. In-the-moment realizations illustrated above could lead to a more profound reflection on how one’s cues might be interpreted more broadly, across meetings, roles, and audiences; and on one’s behavioral habits. For instance, P2 described how *FaceValue* led them to reconsider the role of visual non-verbal cues in their interactions with others:

“I learned a lot about myself. I assumed before *FaceValue* that I wouldn’t care about my behavior. I’d probably just sit there blank-faced. But with *FaceValue* I realized the importance of making more of an effort to look engaged.” – P2

This suggests that the self-awareness fostered by *FaceValue* helped this participant reflect on how their non-verbal expressions might communicate meaning in everyday interactions. Some participants reported beginning to reevaluate their expressive habits, not just in response to the tool, but in reflection on how they were routinely perceived by others.

Challenge: Algorithmic misapprehensions can hinder insight. Most participants (8/13) described the detection as predictable enough for real-time awareness, and found *FaceValue*’s feedback useful and actionable (8/13; 19/65 diaries). For instance:

“One thing I found really helpful is, I was able to see that I look really confused when I was thinking deeply. I think that helped me communicate ‘this part is a little confusing. Could you explain?’ ” – P4

However, some noted challenges with *FaceValue*’s interpretation of their cues. P8 reported:

“Whenever I’m thinking, it tends to flag me as negative emotion for whatever reason, even if that isn’t my intention.” – P8

For P12, the recognition appeared random:

“I couldn’t figure out what it was trying to detect in terms of positive or negative emotions. It seemed essentially random to me, so I couldn’t discern the pattern.” – P12

In these instances, participants were unable to reconcile *FaceValue*’s feedback with their cues, and attributed this to difficulties understanding what the algorithm did and why. P8’s quote suggests that feedback alone was not enough to trigger useful self-awareness, whereas P12’s points to how unpredictability can prevent meaningful insight altogether.

Challenge: Continuous feedback is prone to self-monitoring fatigue. While continuous overlays may enhance awareness for some, they can be tiring for others. Several participants (P1, P2, P3, P8) reported that *FaceValue* increased fatigue or mental effort due to the extra attention required to interpret its overlays. However, P3 noted that any added load could be positive:

“In terms of cognitive load, maybe there’s an increase. But again, I think that’s a good thing; it means I’m paying more attention to whatever is happening in the meeting.” – P3

P4 reported that they deliberately reduced how often they looked at the feedback to avoid becoming overwhelmed. These perspectives suggest variability in how users balance *FaceValue*'s benefits against the mental overhead of continuous self-monitoring.

Summary. Participants' reports suggest that *FaceValue* usage can support purposeful self-monitoring, encouraging reflection on how visual non-verbal cues might be interpreted by others. For many, this self-awareness was beneficial. However, others, found *FaceValue* unpredictable or hard to interpret, limiting the insight they gained from the tool. Some also reported increased mental fatigue from extra monitoring, raising questions about the tradeoff between providing continuous feedback to improve communication and the cognitive cost of processing such feedback.

6.2 From Self-Awareness to Action

RQ2 asked: How do attendees describe attempting to *act* on meaning-oriented feedback to realign their visible cues? Participants reported leveraging *FaceValue*'s feedback to reinforce or course-correct their visual non-verbal cues in real-time, with these adjustments perceived to have a positive impact on others. While *FaceValue*'s suggestive feedback was generally perceived as actionable, participants reported instances where the lack of guidance left them unsure of how to respond.

Meaning-oriented adjustment and realignment. Participants reported adjusting their visual non-verbal cues in response to *FaceValue*'s feedback, aiming to better align cues with their communicative intent. They described taking such actions immediately after noticing a color shift or overlay (6/13; 17/65 diary entries), which alerted them to sudden shifts in their cues.

"I was just listening [...] Sometimes I realized I wasn't paying attention, but when I saw a negative indication through *FaceValue*, I corrected my expression so I didn't appear negative." – P6

"If there's a sudden change, especially if there's a red thing that appears, I would immediately glance at myself, and then readjust, right? Correct course. [...] I would interpret the red as, 'Oh, there's been a sudden negative shift in my demeanor,' so [I] look at myself, and then immediately readjust." – P5

"I wanted to stay positive, but when *FaceValue* showed yellow, I thought 'I don't want to be neutral,' so I changed my expression to be more positive to convey my intent clearly." – P10

These examples illustrate *FaceValue*'s value in identifying unintentional drifts in expression (P6), misaligned cues (P5), or signals too subtle to be seen (P10), allowing immediate adjustments. Participants also reported intentionally amplifying some of their cues to trigger overlays which aligned with their communicative intent. P5 describes such an instance:

"When I nod I could see that I'm nodding with or without [by seeing the lines around my face]. Because of that feedback, I might be [intentionally] nodding more than usual." – P5

Perceived impacts of adjusting one's visual non-verbal cues on others. Beyond achieving clearer communication through realignment of their cues, some participants believed they saw positive changes in how others perceived them due to these adjustments. While we could not confirm this perception with other meeting attendees, three participants explicitly connected *FaceValue*-prompted adjustments with improved impressions or communication clarity. For instance:

"The interviewer said that I was lovely and brilliant and they were excited to work with me through this process. [...] Due to *FaceValue*, my real-time facial expressions reinforced the enthusiasm that I was vocalizing." – P2 (DIARY ENTRY)

"I was nodding and the *FaceValue* detected that I'm nodding. And I felt like, Yeah, I do want to nod. So I actually nodded more. And my participant, noticed this, as he was definitely smiling at my movement. So that is when I felt [the *FaceValue*] makes a difference." – P13

Challenge: Awareness may not translate to actionability. While *FaceValue* helped participants with a self-awareness of their expressions—both to help them to notice and correct cues (8/13)—some participants (3/13) explicitly reported uncertainty about how to respond to its prompts:

“I didn’t pay much attention to the icons. They weren’t distracting, but they weren’t particularly helpful either.” – P7

“When the interface reacts to head movement, is that good or bad? What should I do? Should I move more or less?” – P8

These comments are consistent with Chow et al. [18], who also found that some participants felt at a loss about what to do in response to the nudges and indications the tool provided.

Summary. *FaceValue* may help participants steer interactions intentionally through careful use of facial expressions and head movements, potentially reflecting Objective Self-Awareness [23] in action. However, some participants did not know whether and how to adjust their cues using feedback alone, raising the question of actionability of suggestive feedback.

6.3 Continuous Feedback Is Useful... Some of The Time

Our RQ3 asked: How do meeting characteristics (e.g., stakes, formality) shape how the feedback is perceived in terms of usefulness and distraction? Providing continuous feedback on one’s visual non-verbal cues may not always be relevant or appropriate [16], and a question that was posed by prior work [18]. We found that engaging in self-monitoring was particularly important in professional contexts, but less relevant for casual interactions with relatives and friends.

Self-Monitoring is deemed important in high-stakes and task-oriented meetings; less so in casual meetings with familiar people. Participants (7/13) found *FaceValue* especially helpful in high-stakes or task-oriented meetings (e.g., interviews, team projects) where one’s demeanor is important. For example, P4 appreciated it during sprints, and P6 in large-scale gatherings:

“It was a large-scale meeting with many people. [...] Even if I’m not speaking, if the other person is speaking and I agree, I want to communicate I’m in agreement. So I would monitor myself. I look at [*FaceValue*], and that was helpful.” – P6

Similarly, P5 valued the tool in stressful or sensitive discussions:

“During stressful meetings, it was especially important for me [to know] how I was coming across to the other person. I wouldn’t necessarily have the time to look at my *FaceValue* window as much because I’m so engrossed in the conversation. But then, if there was a sudden change in my expressions, there would be that pop up, the green or the red one. So that was helpful, because it served as a cue for me to readjust. For example, if I was stressed and then I see a red pop up. I was like, ‘Oh, no, I’m coming across stressed.’ So I would just readjust.” – P5

These are suggestive of how the self-awareness provided by *FaceValue* provided clarity to their impression management efforts. P9 found it also helped to support their intentional impression management efforts when encountering unfamiliar people:

“I think it can help if you’re less familiar with people to make sure that you’re appropriately communicating non-verbally.” – P9

Four participants found the feedback less relevant for casual or personal meetings; P10 describes:

“If you are in casual meetings, for example, if you’re chatting with your friends or family, of course it’s important to be genuine and not being rude towards the others but you can be true to yourself.” – P10

Overall, even if active self-monitoring is not called for, all but two participants indicated that they would prefer an augmented view over the regular self-view or no self-view (see Figure 5).

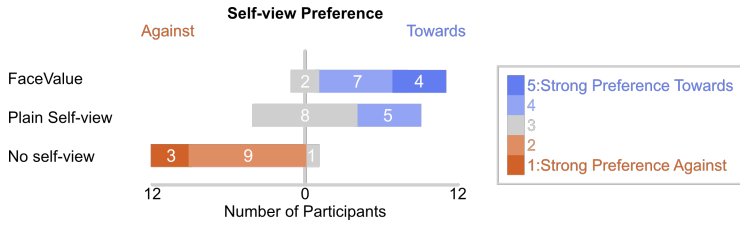


Fig. 5. Participants' self-view preferences collected in the post-study questionnaire.

Summary. *FaceValue* was perceived to be the most valuable in meeting contexts where nuanced communication through visual non-verbal cues had potential implications, such as one's making an impression during an interview or contributing to team's clear communication.

6.4 Design Reflection

Our deployment study allowed us to gain insight into the role and impacts of design choices behind *FaceValue*. In the following, we discuss participants' observations in relation to holistic feedback, suggestive (as opposed to prescriptive) prompts, and private overlay design.

Support for holistic self-monitoring. By overlaying attendees' faces with comic-style cues derived from multiple signals (Figure 3), our design aimed to encourage holistic, wide-ranged awareness of one's expressions, rather than an analytic one that would isolate instances of particular facial signals [10]. Ten participants referred to their cues using holistic language rather than analytical language. For instance, P1 said "I'm looking actively engaged," and P11 said "positive emotion." Occasionally, these participants used analytical terms to describe the overall state of their facial expression or head movements. In contrast, P8, P13, and P12 mostly described adjusting specific facial parts as their goal, suggesting a more analytical focus. For example, P13 said, "When I'm nodding and *FaceValue* detects I'm nodding, I feel like I do want to nod, so I actually nod more."

Support for predictable, and trustworthy self-monitoring. Acknowledging FER's limitations, our approach embraces imperfect sensor readings and models to create meaningful user interactions, treating ambiguity as a resource. This was only partially successful: 8/13 participants found detection predictable enough for real-time awareness, despite occasional false positives, while others found the feedback too unpredictable. P8 and P9 noted:

"I think whenever I'm confused, *FaceValue* is probably going to show me as angry. [...] Whenever it's not aligning with my actual behavior, I start to not look at it for a while." – P8

"I do think there were maybe a few false positives where, if I tilted my head a little bit, it would think I was shaking my head, but it would usually disappear pretty quickly. Because all these reactions are transient, I think the false positives don't really matter that much. And I think you can naturally build that in and just wait a little longer [for effects to disappear]." – P9

Although *FaceValue* was positively perceived with respect to trustworthiness (13/13) overall, there are still isolated instances where feedback was too unclear to be actionable. For some, it was easy to ignore, especially since this feedback was private. For two participants, that could have contributed to undermine the perceived value of such system: 2 participants (P8 and P12) rated the plain self-view higher (slight preference towards), than *FaceValue* (rated neutral) for remote meetings; with P12 rating no self-view also neutrally, whereas P8 reported a strong preference against no self-view. These two participants explicitly described *FaceValue* as distracting (P2, P12) primarily in relation to specific low-stakes or informal meetings where they felt it prompted more self-monitoring than they considered necessary. Otherwise, accuracy and actionability issues did

not appear to undermine the overall usefulness of *FaceValue*. These accounts are consistent with our intention to use ambiguity as a design resource: for many participants, the overlays' ambiguity seemed to make FER inaccuracies tolerable rather than fatal to the tool's perceived usefulness.

Preserving user privacy with private overlay We used private overlays for the visual effects, so only could see feedback on their visual non-verbal cues. As a result, privacy concerns around privacy did not emerge as a major barrier. Ten participants reported no issues under those conditions. P3 and P9 summarized:

“If the [*FaceValue* view] that I'm seeing is also going to be sent to the recipient, I might have second doubts about that. I feel like, ‘Oh, it's a gauge that I can choose to ignore, or I can choose to use that information to make sure that I'm still in control.’ ” – P3

“I was initially a bit concerned about privacy, but I think the local setup alleviated a lot of that.” – P9

Some (3/13) still acknowledged hypothetical risks if overlays were inadvertently shared or if organizers could access *FaceValue* data, underscoring the importance of user control over visibility.

7 Discussion

Our findings highlight *FaceValue*'s potential to evoke meaning-oriented self-awareness via real-time feedback on facial expressions and head movements. We discuss these results in relation to existing theories, outline implications for future systems, and review our study's scope and limitations.

7.1 Suggesting Nuances to Objective Self-Awareness: From Sensing to Adjusting

Our participants reported that the external cues (overlaid feedback) made monitoring feel more frequent, and shifted their attention to what was being monitored. They described reflecting more on how their expressions might be interpreted by others. Interpreting this through the Objective Self-Awareness (OSA) theory, we view *FaceValue* as an example of a computational tool that could support a shift in self-monitoring from cosmetic checking to meaning-oriented inspection (DG1), via continuous, real-time overlays (DG3). In contrast, prior OSA work has emphasized internal self-evaluation upon noticing oneself as a social object.

Further, real-time feedback was experienced as a bridge between internal awareness and expressive control, which participants described prompted them to “course-correct” their visual cues in the moment. From the perspective of OSA, this suggests that computation tools could serve as a behavioral mediator, helping users apply OSA in situ.

This is consistent with findings from Chow et al. [18], where a glanceable sidebar tallying detected smiles, nods, gaze, and postures increased self-awareness. According to our participants' reports, when feedback is more aligned with the rich, interpretive nature of visual non-verbal cues, attendees felt less need to fixate on a single trigger. These reports complement the insights of Chow et al., where participants typically responded with narrower actions like “don't smile” or “nod more often.” Chow et al. were motivated by reducing ‘non-verbal overload’ and Zoom fatigue, which resulted in favoring an analytic, count-based approach. In contrast, our participants used vocabulary describing their goals and intentions at a higher level. For instance, P1 reported wanting to appear more “engaged,” and P11 to express more “positive emotion.” The design approach we took in *FaceValue* invited participants to consider holistic interpretations of the communicative, meaning-oriented aspects of the non-verbal behavior. Participants reasoned in terms of “meaning” rather than easily-countable behaviors that sometimes serve as proxies for these meanings.

Further, privacy-preserving design choices, such as on-device processing and private overlays, were highly valued. Guaranteeing users full control over who sees feedback and when can reduce the emotional stakes, as users can choose to engage or ignore the feedback.

Design Implication 1: Subtle visual feedback can increase the perceived communicative usefulness of self-monitoring. Systems should support both detection (awareness) and enactment (adjustment) of meaningful cues through private feedback that preserves users agency and emotional safety.

7.2 Disrupting the Illusion of Transparency Through Reflective Awareness

Participants' comments indicated that *FaceValue* disrupted a tacit assumption that their communicative intent was obvious to others. This can be interpreted as supporting and extending prior research on the Illusion of Transparency, which argues that people overestimate how well their internal states are perceived externally. Rather than directly reporting overconfidence, our participants retroactively realized how their expressions might have been misinterpreted, implying that the illusion may stem not from confidence but from lack of attention. This addresses RQ1.

We thus posit that raising awareness of interpretation possibilities (DG1) may be an important precursor for breaking the illusion. *FaceValue* did this without judgment or exposure, suggesting a direction for feedback tools that improve communication clarity without increasing pressure.

Design Implication 2: Feedback tools can potentially help break the Illusion of Transparency, by prompting reflection on how visual non-verbal cues might be interpreted by others.

7.3 Interpreting Feedback: How Individual Users Made Sense of *FaceValue*

Participants varied in how they interpreted and responded to feedback from *FaceValue*. Most participants found *FaceValue* predictable enough to form their own interpretations (DG2), although others struggled with interpreting its feedback. These responses echo Gaver's notion of *ambiguity as a resource for design* [30], in which partial information can be valuable if it invites interpretation. This design ambiguity seemed to let participants engage more deeply with *FaceValue*'s overlays, enabling participants to form more holistic, wider-scoped meanings such as "looking engaged," "positive". As expected, while our results show that some participants saw ambiguity as a prompt for reflection, others perceived it as a failure of clarity. Therefore, designers would need to carefully consider user background and cognitive load, perhaps offering modes that adjust granularity or explanation level, concrete guidance, or pre-meeting training, to accommodate different user needs. Pre-meeting training could particularly focus on establishing a common understanding of what constitutes effective visual non-verbal cues in communication, thereby improving the actionability of the feedback provided by *FaceValue*.

Overall, participants expressed a strong preference for *FaceValue* over both a plain self-view and no self-view. This suggests that subtle, suggestive feedback, when private, can be more desirable than both absence and crude monitoring. However, our findings also suggest that for serving a broader population, systems should consider users' interpretive bandwidth. Future systems may benefit from adaptive modes that align with user needs, preferences, and cognitive load.

Design Implication 3: Effective meaning-oriented self-monitoring systems should dynamically adapt to users' momentary interpretive bandwidth.

7.4 Context-Dependent Utility

The utility of *FaceValue* varies by context. Participants emphasized that *FaceValue* was most helpful in formal or high-stakes meetings. This context-dependent perception of utility addresses RQ3 directly. This finding aligns with the Media Richness Theory [20], which suggests that ambiguous and complex tasks benefit from richer communication channels. Taken together, our results illustrate this theory by exemplifying how real-time refinements of visual non-verbal cue may help offset communication richness from the cueing attendee's perspective.

Participants also found *FaceValue* helpful in relationally ambiguous contexts (e.g., with unfamiliar participants), which aligns with the Social Information Processing (SIP) theory [103]. They reported little use in casual conversations, suggesting that relational stakes modulate perceived utility.

Taken together, these findings suggest that *FaceValue* is most valuable when (1) the stakes are high, (2) relational dynamics are uncertain, or (3) tasks require clear, nuanced communication. In casual or low-stakes settings, however, *FaceValue* risks feeling intrusive or distracting (“Clippy-like”).

The Task-Technology Fit (TTF) theory [33] helps explain participants’ impressions. In low-stakes situations, task demands are also low, so the perceivable benefit of launching and attending to *FaceValue* is likely small. In this case, the activation cost may seem high and users may choose to forgo using *FaceValue*. In contrast, in high-stakes or ambiguous situations, the benefits offered by *FaceValue* may be deemed to outweigh the activation cost, making its use appear worthwhile.

Guided by the TTF, we can also avoid “Clippy-like” intrusion. Our results suggest that participants encountered task characteristics in real-world meeting scenarios that match the technology’s characteristics, yielding task-technology fit. Under high-fit conditions, the theory predicts (1) greater utilization and (2) improved performance impacts (conditional on use), which are effects we desire to observe. Accordingly, *FaceValue* should be surfaced and easy to invoke when task characteristics match the technology’s capabilities, and otherwise should fade to avoid adding unnecessary complexity.

Designers should therefore ensure that feedback is context-sensitive and adaptive, appearing only when it supports the user’s goals and enhances, rather than disrupts, the flow of conversation.

Design Implication 4: Effective meaning-oriented self-monitoring systems should strive to encourage more authentic visual non-verbal expressions while dynamically adapting to users’ situational context.

7.5 Ethical Concerns

Several ethical considerations arise when providing feedback on facial expressions in remote meetings. In this section, we outline key concerns and discuss how the design choices in *FaceValue* aim to address, complicate, or limit their impact.

Cultural normativity of FER expression labels. Facial-expression recognition (FER) systems often rely on labels derived from specific cultural contexts, which can impose a normative interpretation of visual cues. This risks flattening culturally diverse forms of expressivity into a single reading. *FaceValue* mitigates this concern by using intentionally color-based overlays rather than fixed emotion labels, allowing users to interpret the feedback in ways that fit their own communicative norms. However, ambiguity does not eliminate underlying cultural assumptions in the training data. Future work should explore culturally-situated calibration or user-tailored interpretations.

Neurodivergent and atypical expressivity. Many neurodivergent individuals, including those on the autism spectrum, experience difficulties in producing or interpreting facial expressions [100]. Systems that imply a “typical” expressive baseline may inadvertently reinforce norms that do not fit all users. *FaceValue*’s reflective cues may help illuminate how their expressions could be perceived by others in remote settings. At the same time, such feedback may increase pressure to conform to neurotypical expressions, potentially contributing to emotional labor or masking. Further work is needed to understand when, and for whom, such feedback is supportive rather than burdensome.

Authenticity, performativity, dependency, and emotional labor. Highlighting moments of potential misalignment can support users who wish to communicate with greater clarity in high-stakes settings. Yet feedback on one’s appearance can also encourage performative self-monitoring, contributing to polished or “camera-ready” behavior that may feel less authentic. While *FaceValue* is designed as a reflective aid rather than a corrective tool, it may still prompt increased self-awareness or emotional labor [68], particularly during long or demanding meetings [5]. As such, users may

develop dependencies, in which they lose their innate self-monitoring skills and refuse to attend meetings without *FaceValue*. Systems that seek to enhance clarity must therefore balance this goal with the need to support authentic and sustainable self-presentation.

Together, these concerns position *FaceValue* as a voluntary, reflective aid rather than a source of ground truth about one's appearance. Any deployment should make the system's ambiguity, limitations, and optional nature explicit, ensuring users can decide whether such feedback is helpful, and can easily opt in or out depending on the interaction context. Similarly, deployments could present the system (or its feedback) less frequently, framing it as a training tool rather than an always-available mirror.

7.6 Limitations and Future Work

Our current study focused on $n = 13$ participants, most of whom were students and researchers comfortable with self-view. Future work should include participants with more diverse demographics, professional roles, and comfort levels with video, enabling the study of non-normative behaviours (e.g., camera-off attendees [6]) and testing whether meaning-oriented feedback supports or disrupts communication in different contexts.

While we obtained rich insights from participants addressing our research questions, we focused on internal instruments to collect subjective qualitative data obtained through one side of the conversation. As such, we acknowledge the limitations of not providing objective evidence of changes or group-level effects. Our participants may over- or understate effects, or misread others' reactions, and their accounts should be read as subjective, situated interpretations rather than ground truth about behavior. The research had to rely only on theory-based speculation and participants' direct quotes to gauge how many benefits *FaceValue* provided in terms of improving understanding of cues or group dynamics.

To address this limitation, future work could capture richer behavioral data and examine long-term use. Researchers could explore privacy-preserving hybrid deployments that collect richer interaction data under controlled conditions, allowing for detailed analysis of how participants actually engage with feedback in real time. Longitudinal studies across varied meeting types could also investigate how attitudes and behaviors change over time and whether sustained exposure leads to lasting improvements in self-awareness and cue adjustment. Similarly, future work could go beyond self-reports and examine group-level effects. Evaluations could triangulate self-reported findings with external measures (e.g., other meeting participants' direct feedback, conversation outcomes, or third-party ratings of communication effectiveness) to provide stronger evidence of impact. This would extend the perspective beyond the focal user and reveal how feedback systems influence group dynamics, shared understanding, and meeting outcomes.

Finally, future work could explicitly investigate individual differences in trait-level self-monitoring and how they interact with the behaviors that *FaceValue* supports. For example, people high in trait self-monitoring may experience *FaceValue* as a helpful resource for fine-tuning their self-presentation, whereas people low in trait self-monitoring may find the same feedback distracting or unnecessary. Although our study did not measure this trait directly, our findings also suggest that some participants cared more about ongoing self-monitoring than others. These differences in trait-level self-monitoring may help explain why some participants disliked using *FaceValue*. Measuring trait self-monitoring alongside behavioral and experiential outcomes would allow us to account for this confound in preference testing. Such measurement would also enable us to estimate the effect size of trait self-monitoring on behavior change, relative to baseline measures. Similarly, future work could explore how the system might be tuned or calibrated for users from different cultural backgrounds or for neurodivergent users.

8 Conclusion

This study introduced *FaceValue*, a lightweight self-view augmentation designed to support meaning-oriented self-monitoring in remote meetings. Building on theories of Objective Self-Awareness, Illusion of Transparency, and Social Information Processing, we explored how real-time, suggestive feedback on facial expressions might enhance attendees' perceived awareness and adjustment of visual non-verbal cues.

Our in-the-wild deployment with thirteen knowledge workers revealed that participants not only perceived themselves to self-monitor more frequently, but also perceived they had shifted their focus from cosmetic appearance to the communicative meanings of their cues. This heightened awareness, in many cases, may have translated into intentional real-time adjustments, improving how participants felt about how they might be perceived by other attendees. Importantly, interpretation of the feedback varied by user and context (which points to selective, context-dependent deployment, rather than as an always-on solution). While high-stakes and goal-oriented meetings amplified *FaceValue*'s utility, casual or familiar conversations often rendered it less relevant. Participants also emphasized the importance of design features, such as privacy preservation in developing remote meeting support tools. We present four design implications based on our study's results that can inform future designers of remote meeting systems.

Acknowledgments

This research was supported by Meta Reality Labs, the National Science and Engineering Research Council (RGPIN-2018-05072), and the Faculty of Information and the Department of Computer Science (University of Toronto). We would also like to thank the members of the Dynamic Graphics Project for their valuable advice and assistance, as well as the participants of the user study for their contributions. We also thank Prof. Jens Emil Sloth Grønbaek for helpful discussions at an early stage of this project.

We acknowledge the use of ChatGPT by OpenAI on proofreading the text, and refactoring/optimizing the code for developing *FaceValue*.

References

- [1] Ronald Brian Adler, George R Rodman, and Alexandre Sévigny. 2006. *Understanding human communication*. Vol. 10. Oxford University Press Oxford.
- [2] Bon Adriel Aseniero, Marios Constantinides, Sagar Joglekar, Ke Zhou, and Daniele Quercia. 2020. MeetCues: Supporting online meetings experience. In *2020 IEEE Visualization Conference (VIS)*. IEEE, 236–240.
- [3] Avi Asher-Schapiro. 2022. Zoom urged by rights groups to rule out 'creepy' AI emotion tech. <https://www.reuters.com/article/usa-tech-rights/zoom-urged-by-rights-groups-to-rule-out-creepy-ai-emotion-tech-idINL5N2X21UW/>.
- [4] Hillel Aviezer, Noga Ensenberg, and Ran R Hassin. 2017. The inherently contextualized nature of facial emotion perception. *Current opinion in psychology* 17 (2017), 47–54.
- [5] Jeremy N Bailenson. 2021. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. (2021).
- [6] Karolina Balogova and Duncan Brumby. 2022. How Do You Zoom?: A Survey Study of How Users Configure Video-Conference Tools for Online Meetings. In *2022 Symposium on Human-Computer Interaction for Work*. 1–7.
- [7] Dean C Barnlund. 1970. A Transactional Model of Communication in Sereno and Mortensen eds. *Foundations of Communication Theory*. Harper and Row 18 (1970), 50.
- [8] Ivo Benke, Maren Schneider, Xuanhui Liu, and Alexander Maedche. 2022. TeamSpiritous-A Retrospective Emotional Competence Development System for Video-Meetings. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.
- [9] Charles R Berger and Richard J Calabrese. 1974. Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human communication research* 1, 2 (1974), 99–112.
- [10] Kirsten Boehner, Rogério DePaula, Paul Dourish, and Phoebe Sengers. 2005. Affect: from information to interaction. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*. 59–68.

- [11] John L Bradshaw and Graeme Wallacei. 1971. Models for the processing and identification of faces. *Perception & Psychophysics* 9 (1971), 443–448.
- [12] Virginia Braun and Victoria Clarke. 2022. *Thematic Analysis: A Practical Guide*. SAGE Publications Ltd, London.
- [13] Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1657–1666.
- [14] Andrew J Calder and Jesse Jansen. 2005. Configural coding of facial expressions: The impact of inversion and photographic negative. *Visual cognition* 12, 3 (2005), 495–518.
- [15] Andrew J Calder, Andrew W Young, Jill Keane, and Michael Dean. 2000. Configural information in facial expression perception. *Journal of Experimental Psychology: Human perception and performance* 26, 2 (2000), 527.
- [16] Mei-Yen Chen and Chien-Chung Chen. 2010. The contribution of the upper and lower face in happy and sad facial expression classification. *Vision Research* 50, 18 (2010), 1814–1823.
- [17] Xinyue Chen, Si Chen, Xu Wang, and Yun Huang. 2021. "I was afraid, but now I enjoy being a streamer!" Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–32.
- [18] Kevin Chow, Roy Rutishauser, André N Meyer, Joanna McGrenere, and Thomas Fritz. 2025. Exploring a Real-time Feedback Display of Non-verbal Cues in Online Work Meetings to Support Self-Presentation. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–28.
- [19] Natasha Crampton. 2022. Microsoft's framework for building AI systems responsibly - Microsoft On the Issues — blogs.microsoft.com. <https://blogs.microsoft.com/on-the-issues/2022/06/21/microsofts-framework-for-building-ai-systems-responsibly/>. [Accessed 28-11-2024].
- [20] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.
- [21] Snigdha Das, Sandip Chakraborty, and Bivas Mitra. 2022. I Cannot See Students Focusing on My Presentation; Are They Following Me? Continuous Monitoring of Student Engagement through "Stungage". In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 243–253.
- [22] Jose Eurico de Vasconcelos Filho, Kori M Inkpen, and Mary Czerwinski. 2009. Image, appearance and vanity in the use of media spaces and video conference systems. In *Proceedings of the 2009 ACM International Conference on Supporting Group Work*. 253–262.
- [23] Shelley Duval. 1972. A theory of objective self-awareness.
- [24] Paul Ekman. 1992. Facial expressions of emotion: New findings, new questions.
- [25] Dirk M Elston. 2021. The novelty effect. *Journal of the American Academy of Dermatology* 85, 3 (2021), 565–566.
- [26] Mohamed Ez-Zaouia, Aurélien Tabard, and Elise Lavoué. 2020. EMODASH: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies* 139 (2020), 102411.
- [27] Korn Ferry. 2023. Cameras On or Off? Professionals Weigh in On Video Usage During Virtual Meetings in Korn Ferry Survey — kornferry.com. [Accessed 30-04-2025].
- [28] Charles Forceville, Elisabeth El Refaie, and Gert Meesters. 2017. Stylistics and comics. In *The Routledge handbook of stylistics*. Routledge, 485–499.
- [29] Inakshi Garg, Pradeep Thakur, Tarana Verma, and Amarnath Yadav. 2024. Affective Computing in Mental Health: The Role of Facial Expression Recognition. *IJFMR) International Journal for Multidisciplinary Research* (2024).
- [30] William W Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 233–240.
- [31] Thomas Gilovich, Kenneth Savitsky, and Victoria Husted Medvec. 1998. The illusion of transparency: biased assessments of others' ability to read one's emotional states. *Journal of personality and social psychology* 75, 2 (1998), 332.
- [32] Carlota Vazquez Gonzalez, Timothy Neate, and Rita Borgo. 2025. Trusting Tracking: Perceptions of Non-Verbal Communication Tracking in Videoconferencing. In *CHI Conference on Human Factors in Computing Systems*. ACM Conference on Human Factors in Computing Systems (CHI).
- [33] Dale L Goodhue and Ronald L Thompson. 1995. Task-technology fit and individual performance. *MIS quarterly* (1995), 213–236.
- [34] Jonathan Grudin. 1994. Groupware and social dynamics: Eight challenges for developers. *Commun. ACM* 37, 1 (1994), 92–105.
- [35] Jenny Gu, Clara Strauss, Rod Bond, and Kate Cavanagh. 2015. How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. *Clinical psychology review* 37 (2015), 1–12.
- [36] Brian D Hall, Lyn Bartram, and Matthew Brehmer. 2022. Augmented chironomia for presenting data to remote audiences. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

- [37] Ari Hautasaari, Minami Aramaki, Rintaro Chujo, and Takeshi Naemura. 2024. EmoScribe Camera: A Virtual Camera System to Enliven Online Conferencing with Automatically Generated Emotional Text Captions. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.
- [38] Zhenyi He, Keru Wang, Brandon Yushan Feng, Ruofei Du, and Ken Perlin. 2021. Gazechat: Enhancing virtual conferences with gaze-aware 3d photos. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 769–782.
- [39] Fritz Heider. 1958. *The Psychology of Interpersonal Relations* (1st ed.). Psychology Press.
- [40] Marie Helweg-Larsen, Stephanie J Cunningham, Amanda Carrico, and Alison M Pergram. 2004. To nod or not to nod: An observational study of nonverbal communication and status in female and male college students. *Psychology of Women Quarterly* 28, 4 (2004), 358–361.
- [41] Dustin Hillard, Mari Ostendorf, and Elizabeth Shriberg. 2003. Detection of agreement vs. disagreement in meetings: Training with unlabeled data. In *Companion Volume of the Proceedings of HLT-NAACL 2003-Short Papers*. 34–36.
- [42] Steven Houben, Connie Golsteijn, Sarah Gallacher, Rose Johnson, Saskia Bakker, Nicolai Marquardt, Licia Capra, and Yvonne Rogers. 2016. Physikit: Data engagement through physical ambient visualizations in the home. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 1608–1619.
- [43] Noura Howell, Laura Devendorf, Rundong Tian, Tomás Vega Galvez, Nan-Wei Gong, Ivan Poupyrev, Eric Paulos, and Kimiko Ryokai. 2016. Biosignals as social cues: Ambiguity and emotional interpretation in social displays of skin conductance. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 865–870.
- [44] Erzhen Hu, Jens Emil Sloth Grønbaek, Austin Houck, and Seongkook Heo. 2023. OpenMic: Utilizing Proxemic Metaphors for Conversational Floor Transitions in Multiparty Video Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [45] Erzhen Hu, Jens Emil Sloth Grønbaek, Wen Ying, Ruofei Du, and Seongkook Heo. 2023. ThingShare: Ad-Hoc digital copies of physical objects for sharing things in video meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–22.
- [46] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, et al. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 17–24.
- [47] Ellen A. Isaacs and John C. Tang. 1993. What video can and can't do for collaboration: a case study. In *Proceedings of the First ACM International Conference on Multimedia (Anaheim, California, USA) (MULTIMEDIA '93)*. Association for Computing Machinery, New York, NY, USA, 199–206. <https://doi.org/10.1145/166266.166289>
- [48] Zifan Jiang, Mark Luskus, Salman Seyedi, Emily L Griner, Ali Bahrami Rad, Gari D Clifford, Mina Boazak, and Robert O Cotes. 2022. Utilizing computer vision for facial behavior analysis in schizophrenia studies: A systematic review. *PloS one* 17, 4 (2022), e0266828.
- [49] Matthew K. Miller, Martin Johannes Dechant, and Regan L. Mandryk. 2021. Meeting You, Seeing Me: The Role of Social Anxiety, Visual Feedback, and Interface Layout in a Get-to-Know-You Task via Video Chat. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [50] Sepideh Kalateh, Sanaz Nikghadam Hojjati, and Jose Barata. 2024. Towards a framework for multimodal creativity states detection from emotion, arousal, and valence. In *International Conference on Computational Science*. Springer, 79–86.
- [51] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. 2000. Comprehensive database for facial expression analysis. In *Proceedings fourth IEEE international conference on automatic face and gesture recognition (cat. No. PR00580)*. IEEE, 46–53.
- [52] Harold H Kelley. 1967. Attribution theory in social psychology.. In *Nebraska symposium on motivation*. University of Nebraska Press.
- [53] Fahad Khan, Seemal Asif, and Phil Webb. 2024. Human Facial Emotion Recognition for Adaptive Human Robot Collaboration in Manufacturing. In *Annual Conference Towards Autonomous Robotic Systems*. Springer, 33–47.
- [54] Maria Kjærup, Mikael B Skov, Peter Axel Nielsen, Jesper Kjeldskov, Jens Gerken, and Harald Reiterer. 2021. *Longitudinal studies in HCI research: a review of CHI publications from 1982–2019*. Springer.
- [55] Yaroslav Konar, Patrick J Bennett, and Allison B Sekuler. 2010. Holistic processing is not correlated with face-identification accuracy. *Psychological science* 21, 1 (2010), 38–43.
- [56] Kristine M Kuhn. 2022. The constant mirror: Self-view and attitudes to virtual meetings. *Computers in Human Behavior* 128 (2022), 107110.
- [57] Ha Yeon Lee, Seora Park, Esther Hehsun Kim, Jiyeon Seo, Hajin Lim, and Joonhwan Lee. 2024. Investigating the Effects of Real-time Student Monitoring Interface on Instructors' Monitoring Practices in Online Teaching. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–11.
- [58] Joanne Leong, Pat Pataranutaporn, Yaoli Mao, Florian Perteneder, Ehsan Hoque, Janet M Baker, and Pattie Maes. 2021. Exploring the use of real-time camera filters on embodiment and creativity. In *extended abstracts of the 2021*

- CHI conference on Human Factors in Computing Systems*. 1–7.
- [59] Gilly Leshed, Dan Cosley, Jeffrey T Hancock, and Geri Gay. 2010. Visualizing language use in team conversations: designing through theory, experiments, and iterations. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 4567–4582.
 - [60] Benjamin J Li and Hui Min Lee. 2023. Filters uncovered: Investigating the impact of AR face filters and self-view on videoconference fatigue and affect. *Telematics and Informatics Reports* 11 (2023), 100088.
 - [61] Shan Li and Weihong Deng. 2020. Deep facial expression recognition: A survey. *IEEE transactions on affective computing* 13, 3 (2020), 1195–1215.
 - [62] Jian Liao, Adnan Karim, Shivesh Singh Jadon, Rubaiat Habib Kazi, and Ryo Suzuki. 2022. RealityTalk: Real-time speech-driven augmented presentation for AR live storytelling. In *Proceedings of the 35th annual ACM symposium on user interface software and technology*. 1–12.
 - [63] Xingyu'Bruce' Liu, Vladimir Kirilyuk, Xiuxiu Yuan, Peggy Chi, Alex Olwal, Xiang'Anthony' Chen, and Ruofei Du. 2023. Experiencing Visual Captions: Augmented Communication with Real-time Visuals using Large Language Models. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–4.
 - [64] Zhenhao Liu, Min Wu, Weihua Cao, Luefeng Chen, Jianping Xu, Ri Zhang, Mengtian Zhou, and Junwei Mao. 2017. A facial expression emotion recognition based human-robot interaction system. *IEEE CAA J. Autom. Sinica* 4, 4 (2017), 668–676.
 - [65] Steven R Livingstone and Frank A Russo. 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS one* 13, 5 (2018), e0196391.
 - [66] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuoling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).
 - [67] Kiyosu Maeda, Riku Arakawa, and Jun Rekimoto. 2022. Calmresponses: Displaying collective audience reactions in remote communication. In *Proceedings of the 2022 ACM International Conference on Interactive Media Experiences*. 193–208.
 - [68] Sandi Mann. 1997. Emotional labour in organizations. *Leadership & Organization Development Journal* 18, 1 (1997), 4–12.
 - [69] David Marino, Jiamin Dai, Pascal E. Fortin, Max Henry, and Jeremy Cooperstock. 2024. Co-Here: an expressive videoconferencing module for implicit affective interaction. In *Graphics Interface 2024*.
 - [70] Albert Mehrabian and Susan R Ferris. 1967. Inference of attitudes from nonverbal communication in two channels. *Journal of consulting psychology* 31, 3 (1967), 248.
 - [71] Albert Mehrabian and Morton Wiener. 1967. Decoding of inconsistent communications. *Journal of personality and social psychology* 6, 1 (1967), 109.
 - [72] Andre N Meyer, Gail C Murphy, Thomas Zimmermann, and Thomas Fritz. 2017. Design recommendations for self-monitoring in the workplace: Studies in software development. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–24.
 - [73] Matthew K Miller, Regan L Mandryk, Max V Birk, Ansgar E Depping, and Tushita Patel. 2017. Through the looking glass: The effects of feedback on self-awareness and conversational behaviour during video chat. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 5271–5283.
 - [74] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. 2017. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing* 10, 1 (2017), 18–31.
 - [75] Michael T Motley and Carl T Camden. 1988. Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Communication (includes Communication Reports)* 52, 1 (1988), 1–22.
 - [76] Prasanth Murali, Javier Hernandez, Daniel McDuff, Kael Rowan, Jina Suh, and Mary Czerwinski. 2021. Affectivespotlight: Facilitating the communication of affective responses from audience members during online presentations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
 - [77] Gun Woo (Warren) Park, Anthony Tang, and Fanny Chevalier. 2024. JollyGesture: Exploring Dual-Purpose Gestures and Gesture Guidance in VR Presentations. In *Proceedings of the 50th Graphics Interface Conference*. 1–14.
 - [78] Seho Park, Kunyoung Lee, Jae-A Lim, Hyunwoong Ko, Taehoon Kim, Jung-In Lee, Hakrim Kim, Seong-Jae Han, Jeong-Shim Kim, Soowon Park, et al. 2020. Differences in facial expressions between spontaneous and posed smiles: Automated method by action units and three-dimensional facial landmarks. *Sensors* 20, 4 (2020), 1199.
 - [79] So Yeon Park and Mark E Whiting. 2020. Beyond Zooming there: Understanding nonverbal interaction online. In *The New Future of Work Symposium*.
 - [80] Henning Pohl. 2024. Body-Based Augmented Reality Feedback During Conversations. *Proceedings of the ACM on Human-Computer Interaction* 8, PACMHCI (2024), 1–22.

- [81] Nina Rajcic and Jon McCormack. 2020. Mirror ritual: An affective interface for emotional self-reflection. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
- [82] Lauren Rhue. 2018. Racial influence on automated perceptions of emotions. *Available at SSRN 3281765* (2018).
- [83] René Riedl. 2022. On the stress potential of videoconferencing: definition and root causes of Zoom fatigue. *Electronic Markets* 32, 1 (2022), 153–177.
- [84] Camilo Rojas, Eugenio Zuccarelli, Alexandra Chin, Gaurav Patekar, David Esquivel, and Pattie Maes. 2022. Towards enhancing empathy through emotion augmented remote communication. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–9.
- [85] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.
- [86] EV Seidela, U Habela, M Kirschner, RC Gurd, and B Derntl. 2010. The impact of facial emotional expressions on behavioral tendencies in females and males. *J. Exp. Psychol. Hum. Percept. Perform* 36, 2 (2010), 500–507.
- [87] Julia Seitz, Ivo Benke, and Alexander Maedche. 2024. "Mirror, mirror in the call": Exploring the Ambivalent Nature of the Self-view in Video Meeting Systems with Self-Reported & Eye-Tracking Data. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–37.
- [88] Allison B Sekuler, Carl M Gaspar, Jason M Gold, and Patrick J Bennett. 2004. Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology* 14, 5 (2004), 391–396.
- [89] Sinem Şemsioglu, Pelin Karaturhan, and Asım Evren Yantaç. 2022. EMOTE: An interactive online tool for designing real-time emotional AR visualizations. In *13th Augmented Human International Conference*. 1–8.
- [90] Kazuko Shinohara and Yoshihiro Matsunaka. 2009. Pictorial metaphors of emotion in Japanese comics. *Multimodal metaphor* (2009), 265–293.
- [91] 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.
- [92] Channi Silence, Shauna M Rice, and Arianne Shadi Kourosh. 2022. Our Post-Zoom Reflection: Turning Off the Monitor. *Facial Plastic Surgery & Aesthetic Medicine* 24, 2 (2022), 73–74.
- [93] Mark Snyder. 1974. Self-monitoring of expressive behavior. *Journal of personality and social psychology* 30, 4 (1974), 526.
- [94] Ai Sun, Yingjian Li, Yueh-Min Huang, and Qiong Li. 2018. The exploration of facial expression recognition in distance education learning system. In *Innovative Technologies and Learning: First International Conference, ICITL 2018, Portoroz, Slovenia, August 27–30, 2018, Proceedings 1*. Springer, 111–121.
- [95] Anna Sutton. 2016. Measuring the effects of self-awareness: Construction of the self-awareness outcomes questionnaire. *Europe's journal of psychology* 12, 4 (2016), 645.
- [96] Beverly Davenport Sypher and Howard E Sypher. 1983. Perceptions of communication ability: Self-monitoring in an organizational setting. *Personality and Social Psychology Bulletin* 9, 2 (1983), 297–304.
- [97] James W Tanaka, Martha D Kaiser, Sean Butler, and Richard Le Grand. 2012. Mixed emotions: Holistic and analytic perception of facial expressions. *Cognition & emotion* 26, 6 (2012), 961–977.
- [98] Xiao-Yu Tang, Wang-Yue Peng, Si-Rui Liu, and Jian-Wen Xiong. 2020. Classroom Teaching Evaluation Based on Facial Expression Recognition. In *Proceedings of the 2020 9th International Conference on Educational and Information Technology* (Oxford, United Kingdom) (ICEIT 2020). Association for Computing Machinery, New York, NY, USA, 62–67. <https://doi.org/10.1145/3383923.3383949>
- [99] Antoine Toisoul, Jean Kossaifi, Adrian Bulat, Georgios Tzimiropoulos, and Maja Pantic. 2021. Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nature Machine Intelligence* 3, 1 (2021), 42–50.
- [100] Dominic A Trevisan, Maureen Hoskyn, and Elina Birmingham. 2018. Facial expression production in autism: A meta-analysis. *Autism Research* 11, 12 (2018), 1586–1601.
- [101] David R Vago and David A Silbersweig. 2012. Self-awareness, self-regulation, and self-transcendence (S-ART): a framework for understanding the neurobiological mechanisms of mindfulness. *Frontiers in human neuroscience* 6 (2012), 296.
- [102] Michelle Verheijden, Esther Giroldi, and Angelique Timmerman. 2023. Developing skilled communication: The power of self-monitoring. *Medical Education* 57, 9 (2023), 782–784.
- [103] Joseph B Walther. 1992. Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication research* 19, 1 (1992), 52–90.
- [104] Wen Wang, Kristin Precoda, Colleen Richey, Geoffrey Raymond, et al. 2011. Identifying Agreement/Disagreement in Conversational Speech: A Cross-Lingual Study.. In *INTERSPEECH*. 3093–3096.
- [105] Peter H Waxer. 1977. Nonverbal cues for anxiety: an examination of emotional leakage. *Journal of abnormal psychology* 86, 3 (1977), 306.
- [106] Greg Welch, Gary Bishop, et al. 1995. An introduction to the Kalman filter. (1995).

- [107] Mark E Whiting, Allie Blaising, Chloe Barreau, Laura Fiuza, Nik Marda, Melissa Valentine, and Michael S Bernstein. 2019. Did it have to end this way? Understanding the consistency of team fracture. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.
- [108] Robert A Wicklund. 1975. Objective self-awareness. In *Advances in experimental social psychology*. Vol. 8. Elsevier, 233–275.
- [109] Matthew Woodward. 2024. ZOOM USER STATISTICS: HOW MANY PEOPLE USE ZOOM IN 2024? <https://www.searchlogistics.com/learn/statistics/zoom-user-statistics/>.
- [110] Haipeng Zeng, Xinhuan Shu, Yanbang Wang, Yong Wang, Liguang Zhang, Ting-Chuen Pong, and Huamin Qu. 2020. Emotioncues: Emotion-oriented visual summarization of classroom videos. *IEEE transactions on visualization and computer graphics* 27, 7 (2020), 3168–3181.
- [111] Chunpeng Zhai, Santoso Wibowo, and Lily D Li. 2024. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments* 11, 1 (2024), 28.

Received May 13, 2025; revised January 13, 2026; accepted March 17, 2026