



The Look of (Un)confidence: Visual Markers for Inferring Speaker Confidence in Speech

Yondu Mori^{1,2*} and Marc D. Pell^{1,2}

¹ School of Communication Sciences and Disorders, McGill University, Montréal, QC, Canada, ² Centre for Research on Brain, Language and Music, Montréal, QC, Canada

Evidence suggests that observers can accurately perceive a speaker's static confidence level, related to their personality and social status, by only assessing their visual cues. However, less is known about the visual cues that speakers produce to signal their transient confidence level in the content of their speech. Moreover, it is unclear what visual cues observers use to accurately perceive a speaker's confidence level. Observers are hypothesized to use visual cues in their social evaluations based on the cue's level of perceptual salience and/or their beliefs about the cues that speakers with a given mental state produce. We elicited high and low levels of confidence in the speech content by having a group of speakers answer general knowledge questions ranging in difficulty while their face and upper body were video recorded. A group of observers watched muted videos of these recordings to rate the speaker's confidence and report the face/body area(s) they used to assess the speaker's confidence. Observers accurately perceived a speaker's confidence level relative to the speakers' subjective confidence, and broadly differentiated speakers as having low compared to high confidence by using speakers' eyes, facial expressions, and head movements. Our results argue that observers use a speaker's facial region to implicitly decode a speaker's transient confidence level in a situation of low-stakes social evaluation, although the use of these cues differs across speakers. The effect of situational factors on speakers' visual cue production and observers' utilization of these visual cues are discussed, with implications for improving how observers in real world contexts assess a speaker's confidence in their speech content.

Keywords: speech communication, memory, recall, nonverbal behavior, mental state attribution

INTRODUCTION

During conversation, speakers produce visual cues that (often inadvertently) demonstrate their confidence level in or commitment to the content of their speech. Here, confidence refers to a transient mental state indexing a speaker's subjective level of certainty in a concept and/or word as it is retrieved (or a retrospective metamemory judgment) (Nelson and Narens, 1990; Boduroglu et al., 2014). This demonstration of confidence is not to be confused with speakers (un)consciously communicating confidence related to their social status or personality traits (e.g., sitting up straight) (Tenney et al., 2008; Nelson and Russell, 2011; Locke and Anderson, 2015) and/or speakers confidently presenting information that was previously retrieved from memory

OPEN ACCESS

Edited by:

Marcela Pena,
Pontifical Catholic University of
Chile, Chile

Reviewed by:

Helene Kreysa,
Friedrich Schiller University
Jena, Germany
Ana Pinheiro,
University of Lisbon, Portugal

*Correspondence:

Yondu Mori
yondu.mori@mail.mcgill.ca

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 24 July 2019

Accepted: 23 October 2019

Published: 21 November 2019

Citation:

Mori Y and Pell MD (2019) The Look of (Un)confidence: Visual Markers for Inferring Speaker Confidence in Speech. *Front. Commun.* 4:63. doi: 10.3389/fcomm.2019.00063

and rehearsed. For example, politicians or TV news anchors generally display composure in their facial expressions and posture, and produce minimal facial movements to mark their high confidence and neutral emotional state (Coleman and Wu, 2006; Swerts and Krahmer, 2010). In these latter instances, a speaker's confidence is described as a static mental state without considering the variable speech content speakers spontaneously produce. Yet, speakers in these instances can still experience a transient mental state of confidence, reflecting the ongoing memory retrieval and dynamic emotional states speakers have during natural conversation.

Despite this differentiation in the factors underlying a speaker's conveyed confidence, interlocutors likely do not make this distinction when decoding a speaker's visual cues and drawing social inferences from them. Rather, interlocutors may simply try to detect a speaker's level of certainty in the information speakers present. Research suggests that visual cues broadly referring to a speaker's confidence level are automatically decoded by observers (Moons et al., 2013) and can impact enduring social assessments of a speaker, such as in job interviews (DeGroot and Motowidlo, 1999; DeGroot and Gooty, 2009) or courtrooms with expert witnesses (Cramer et al., 2009, 2014). In these contexts, speakers are asked questions they may not be expecting or prepared for, which can result in them producing visual and vocal cues that mark their confidence level in their speech content (Broisy et al., 2016). Interlocutors can then use these non-verbal cues to infer a speaker's credibility, trustworthiness or believability (Cramer et al., 2009; Birch et al., 2010; Jiang and Pell, 2015, 2017; Jiang et al., 2017), such as when determining if a speaker is lying (Depaulo et al., 2003) or persuading others to adopt their stance (Scherer et al., 1973). Thus, it is important to understand how the visual cues that speakers produce as a result of their transient confidence in their speech content, can impact observers' impressions of their confidence level. However, it is unclear *how* observers infer a speaker's confidence level strictly from the visual cues they produce while speaking, such as when answering questions that tap general (shared) knowledge (Swerts and Krahmer, 2005; Kuhlen et al., 2014).

When speakers spontaneously communicate their knowledge via answering questions, they may produce visual cues for two main reasons (Smith and Clark, 1993). One, speaker's visual cues can indicate their cognitive processes for semantic activation and lexical retrieval, whereby concepts are accessed from memory and communicated using language. The speed and success of this process is influenced by properties of the concepts retrieved. According to models of semantic memory, a concept is activated more strongly during lexical retrieval if it is encoded more frequently and stored longer in memory (Anderson, 1983a,b). For example, when responding to the trivia question, "In what sport is the Stanley cup awarded?" (Smith and Clark, 1993; Jiang and Pell, 2017), a speaker should take longer to respond when the target concept (i.e., hockey) is activated less strongly (e.g., if the speaker is not from a hockey-playing nation or is not interested in sports). The speaker may also produce visual cues to mark the ease (or difficulty) of this process, which can vary across speakers depending on their background knowledge, cultural background

and level of non-verbal expressiveness (Sullins, 1989; McCarthy et al., 2006, 2008; Zaki et al., 2009). However, this hypothesis for visual cue production does not necessarily involve speakers interacting with others.

Another reason why speakers may produce visual cues to indicate their confidence level in the content of their speech is for *pragmatic* purposes, as they consider how they appear to others during an interaction. According to the Gricean Maxim of Quality, during conversation speakers should not say information they believe is false or they have an insufficient amount of evidence for (Grice, 1975). To follow this maxim, speakers should indicate their level of certainty in their response to a question, which can be done through their (un)conscious production of visual cues. When speakers have low confidence in their speech content, the visual cues they produce may represent an unconscious mechanism that allows them to save face in the social context (Goffman, 1967, 1971; Visser et al., 2014). That is, speakers may furnish salient visual cues signaling their lack of commitment to the linguistic message so that their audience will be less critical of errors in the message. Conversely, speakers can produce visual cues to signal and pragmatically reinforce their certainty to others (Moons et al., 2013) or to feign certainty/false confidence. For example, speakers frequently gesture before speaking to retrieve words from memory (Rimé and Schiaratura, 1991; Krauss et al., 1996). A speaker's lexical retrieval may also be signaled by their facial expressions and changes in eye gaze (Goodwin and Goodwin, 1986; Bavelas and Gerwing, 2007). Goodwin and Goodwin (1986) examined gestures related to searching for a word in videotaped conversations produced in natural settings. They found that speakers produced a "thinking face" when word searching, which was often preceded by a change in gaze direction. This thinking face can involve the corners of the lips being turned downwards, the eyes widening, eyebrow movement, pursing of the lips or a stretching or slackening of the lips (Ekman and Friesen, 1978; Goodwin and Goodwin, 1986; Swerts and Krahmer, 2005). This facial expression demonstrates active involvement of various parts of the face during word retrieval. Based on these findings, one can postulate that speakers who have low confidence in the content of their speech are more likely to produce this face, possibly with an averted gaze, compared to speakers who have high confidence in their speech content. This pragmatic process can also interact with the cognitive process of lexical retrieval. Some researchers hypothesize that people avert their gaze from others when thinking to reduce their arousal level, so they can concentrate (Argyle and Cook, 1976; Patterson, 1976, 1982; Doherty-Sneddon and Phelps, 2005). This change in gaze can allow speakers to successfully retrieve and produce target words (Glenberg et al., 1998). These differences in gaze behavior could be a relevant sign of a speaker's confidence level.

When put together, different body and face movements can refer to a speaker's confidence level in the content of their speech in several ways: by demonstrating a speaker's retrieval of words from memory, their (un)conscious effort to communicate their (un)certainty to others for social purposes and/or their level of concentration in answering questions during face-to-face interactions. By focusing on how speakers respond to general

knowledge (trivia) questions, the current study is likely to capture those visual cues that are most relevant to processes for retrieving semantic information from memory and socially communicating a level of certainty to others.

Observers are hypothesized to use various visual cues when evaluating a speaker's social traits due to cognitive and social processes. Firstly, from a social ecological approach, observers are thought to attend to visual cues that are more perceptually salient (McArthur, 1981; Fiske and Taylor, 1984). Observers' attention to these cues can influence the social impressions they form of a speaker (McArthur and Solomon, 1978). Since speakers with low confidence may produce cues that involve more movement or are more marked (e.g., a thinking face) compared to when speakers have high confidence in the content of their speech, observers may be more attentive to using these cues when evaluating a speaker's confidence level. However, variability in speakers' level of non-verbal expressiveness may not allow observers to detect a large difference in marked features indicating high vs. low confidence. It is also unclear whether a speaker with high confidence is detected by the absence of cues indicating low confidence or if there are other types of cues that are uniquely produced and perceived.

Secondly, observers may differentiate speakers' confidence level based on their visual cues by considering the speaker's perspective. This process can allow observers to accurately detect a speaker's thoughts/feelings to try to understand why speakers are producing certain visual cues (i.e., mentalizing or empathizing). With this process, observers may also be affected by cues that are more perceptually salient (Kuhlen et al., 2014). For example, in Kuhlen et al. (2014) when observers watched muted videos of speakers responding to trivia questions and rated them to have low confidence, observers showed increased activity in the mentalizing network (e.g., medial prefrontal cortex and temporoparietal junction). The researchers suggest this activation occurred because speakers perceived to have low confidence produce more salient visual cues compared to speakers with high confidence (Kuhlen et al., 2014). However, it is unknown what visual cues observers attended to when mentalizing. Also, this brain activation was limited to speakers perceived to have low confidence, no significant effect was found when speakers were perceived to have high confidence.

Observers' evaluations of speakers with high confidence may be affected by another aspect of mentalizing: observers' expectations or beliefs about the visual cues a speaker with high confidence (stereotypically) produces (Schmid Mast et al., 2006). For example, in Murphy (2007), speakers were asked to appear intelligent (Acting condition) or were not given any instruction about their behavior (Control condition) as they engaged in an informal conversation with another speaker. Speakers in the Acting condition were more likely to display eye contact while speaking, a serious face and an upright posture compared to speakers in the Control condition. From these cues, only a speaker's eye contact while speaking was significantly correlated with their perceived intelligence (Murphy, 2007). This difference in the frequency of visual cues produced by the Acting and Control speakers was likely influenced by speakers' beliefs about displays of intelligence, which may be similar to the visual cues

indicating a speaker's high confidence in the content of their speech. Also, despite speakers' beliefs about producing a serious face and upright posture, these may not have been reliable cues of a speaker's perceived intelligence. This result demonstrates the reduced reliability of observers' beliefs about the cues a speaker with a given mental state produces. Not all visual cues produced by speakers may meaningfully contribute to evaluating a speaker's perceived confidence. Overall, not a lot is known about how speakers' confidence level in their speech content can impact the social impressions that observers form and the visual cues that observers use to evaluate a speaker's confidence level.

Research examining the visual cues that observers use to decode a speaker's mental state are often studied in literature on interpersonal sensitivity or empathic accuracy (Hall et al., 2001; Schmid Mast et al., 2006). It has been measured by participants providing a perceptual rating of an inferred state or trait in a speaker's and then indicating the types of cues that influenced their judgment. For example, in Mann et al. (2008), participants watched videos of speakers who were lying or telling the truth in one of three conditions, visual only (i.e., muted videos), audio only (i.e., audio recording, no video) or an audiovisual condition (i.e., full video). After seeing and/or hearing the speaker, participants indicated whether the speaker was lying or telling the truth, and then rated how often they paid attention to each speaker's speech content, vocal behavior (pauses, stutters, pitch etc.) or visual behavior (gaze, movements, posture etc.) in their judgments. The researchers only analyzed the audiovisual condition and found that participants reported to pay more attention to the speaker's visual cues compared to their speech or vocal cues (Mann et al., 2008). Although the researchers did not report the visual only condition, this methodology demonstrates the impact of the visual communication channel on observers' evaluations. With this study we aim to better understand the specific visual cues that affect observers' judgments of a speaker's transient state of confidence.

The purpose of this study was to identify major visual cues that observers use to evaluate a speaker's confidence level in the content of their speech, and to examine how these cues differ when speakers experience high vs. low confidence. Based on the literature, we focused on the production of facial expressions (e.g., a thinking face), facial movements (changes in gaze and eyebrow movements), and gross postural movements because of their reported association to speaker confidence. We also explored hand movements as indicating a speaker's confidence level, as they can cue lexical retrieval difficulties during speech production (Rimé and Schiaratura, 1991; Krauss et al., 1996). Following Swerts and Krahmer (2005), we posed general knowledge questions to spontaneously elicit high and low levels of confidence in a group of individuals, who were video recorded as they responded and then subjectively rated their own confidence level. A group of observers then watched muted versions of the videos and indicated the face or body areas they used when evaluating the speaker's confidence level.

We predicted there would be a strong correspondence between speaker's subjective confidence and how observers perceived their confidence level from visual cues (i.e., perceived confidence), demonstrating observers' accurate detection of a

speaker's transient mental state of confidence. We anticipated that observers would use more visual cues to decode speakers with low compared to high confidence because of the greater perceptual salience of these cues and observer's ability to mentalize with speakers conveying low confidence. These visual cues indicating low confidence may include a "thinking face," changes in gaze, eyebrow and head movements, and postural movements. In contrast, high confidence may be more marked by direct eye contact, a serious face and an upright posture. The relationship between visually-derived confidence impressions and more sustained social traits of the speaker (competence, attractiveness, trustworthiness, etc.) as well as variability in speakers' production of visual cues and observer's use of these visual cues were also explored.

PART I: PRODUCTION STUDY

Methods

Speakers

Ten native Canadian English speakers (McGill University students) volunteered for the study (five males and five females, Mean age = 21.5 years, $SD = 2.84$, different racial/ethnic identities). Speakers reported having normal or corrected-to-normal vision and normal hearing. Three speakers wore eyeglasses during testing which did not obscure visibility of the eye or eyebrow region in the video recordings.

Materials

Recordings were gathered while administering a trivia (or general knowledge) question task, following previous work (Smith and Clark, 1993; Brennan and Williams, 1995; Swerts and Krahmer, 2005; Visser et al., 2014). The stimuli were adapted from a corpus of 477 general knowledge statements (in English) constructed and presented in written format to 24 Canadian participants in a pilot study. In that study, items were first perceptually validated to specify whether each statement was a known general knowledge fact (based on the group accuracy rate) and to indicate how confident participants were in their knowledge for each statement (mean confidence rating on scale from 1 to 5). Statements covered a range of topics including science, history, culture, sports and literature. Based on the hit rate for the validation group, a set of 20 less-known general knowledge statements were selected with a hit rate range of 0.12–0.64 (out of 1) and a mean confidence rating range of 1.6 to 3.48 (out of 5). A set of 20 well-known general knowledge statements were selected with a hit rate range of 0.88 to 1 (out of 1) and a mean confidence rating range of 4.04–4.96 (out of 5). For the purposes of this study, the selected statements were then transformed into questions. For example, the statement: "*Carmine is a chemical pigment that is red in color*" became the question: "*What color is the chemical pigment, carmine?*" The well-known and less-known general knowledge questions were used to elicit a state of high and low confidence in the speaker regarding the speech content, respectively. The selected list of questions based on less-known or well-known general knowledge is provided in the **Supplementary Material**.

Elicitation Procedure

Each speaker was individually video recorded in a sound-attenuated recording booth in an experimental testing laboratory, to allow for high quality audiovisual recordings. Speakers sat at a table with a laptop computer and a Cedrus RxB60 response pad (with five buttons labeled from 01 to 05 consecutively) located directly in front of the computer. A video camera mounted on a tripod was positioned behind the laptop monitor along with a wall-mounted loudspeaker. The loudspeaker allowed speakers to communicate with the examiner, who was located outside the recording booth and could not see the speaker. The examiner faced a window to the recording booth that was covered by a white curtain, preventing their facial expressions from influencing the speaker's behavior (see Smith and Clark, 1993; Swerts and Krahmer, 2005, for similar methods). The experiment was controlled by SuperLab 5 presentation software (Cedrus Corporation, 2014).

To elicit naturalistic expressions associated with a speaker's confidence in the content of their speech, speakers engaged in a question-response paradigm with the hidden examiner who spoke in a neutral tone of voice. Participants answered a series of less-known and well-known general knowledge questions from the described corpus by formulating a complete sentence while looking at the video camera (e.g., Examiner: "*What color is a ruby?*," Speaker: "*A ruby is red*"). Speakers were instructed to guess if they did not know the answer to a question. No visual or verbal feedback on the quality of the speaker's performance was provided to the speaker. After answering each question, speakers used the response pad to rate how confident they were in their response on a 5-point scale (1 = not at all confident, 5 = very confident; the 5-point scale was simultaneously presented on the computer screen). Since speakers always had to provide an answer, a confidence rating of five can be interpreted as the speaker having high confidence in their response, not that they had high confidence in not knowing the answer. Speakers answered each question once. The 20 well-known and 20 less-known general knowledge questions were asked in a randomized order over four blocks with 10 questions/block, always by the same examiner (YM). For a small number of trials in which speakers requested clarification after hearing the general knowledge question, the examiner repeated the question but did not provide supplementary details. The examiner asked speakers to repeat their response when it was not formulated in a complete sentence, although this rarely happened. This procedure resulted in a total of 400 video recordings (10 speakers \times 2 types of general knowledge questions \times 20 questions) for analysis. Speakers were compensated \$10 CAD and the recording session took approximately 1 h.

Video Analysis

The video recordings were edited using Windows Movie Maker to isolate the response portion of each trial, defined as the offset of the examiner's question and the onset to offset of the speaker's verbal response. Videos of speaker's responses to the well-known and less-known trivia questions were an average duration of 5.05 s ($SD = 2.85$ s) and 9.88 s ($SD = 7.52$ s), respectively. The

audio channel of the individual files was then removed to create muted versions of each response. During editing, data from one female speaker had to be discarded due to a recording artifact (i.e., her videos showed a strong reflection of the computer screen against her eyeglasses). In addition, several trials using less-known general knowledge questions ($n = 30$) were excluded because the speaker initiated a dialogue with the examiner seeking clarification. We hypothesized that these videos would not reflect speaker's immediate and spontaneous expressions of low confidence via their visual cues.

For the remaining nine speakers, question accuracy and self-ratings of confidence (i.e., subjective confidence) for each question were then analyzed to ensure the questions consistently elicited intended high or low confidence states (See **Supplementary Material** for mean hit rate and subjective confidence ratings for each question). This step was crucial because our analyses sought to identify visual cues associated with unambiguous conditions that elicit high vs. low confidence. Question accuracy was determined by assessing the speech content of the response, for which a specific level of detail was required. For example, for the less-known general knowledge question, "Who invented the Theory of Relativity?," the response, "a scientist" was marked as incorrect. **Table 1** supplies data on the frequency and subjective confidence ratings of the nine speakers by trivia question type and accuracy. We retained data for 14 well-known general knowledge questions and 17 less-known general knowledge questions which consistently elicited the target confidence state without excessive variability across speakers (see section Results for more details). In total, this process resulted in 249 muted videos of single utterances from 9 different speakers to be used in the Perception Study (126 responses to well-known general knowledge questions (9 speakers \times 14 well-known general knowledge questions) + 123 responses to less-known general knowledge questions (9 speakers \times 17 less-known general knowledge questions - 30 excluded videos).

Coding of Visual Cues

A group of six coders specified the visual cues within each facial/body region characterizing speakers' confidence level. Coders were native speakers of Canadian or American English (4 female/2 male, mean age = 23.33, $SD = 2.58$) who were blind to the purpose of the experiment and did not

know any of the speakers, as determined at the onset of the study.

The coders were tested individually and were instructed to watch the muted videos and characterize specific visual cues they observed in the video without any associated context. Before the coding procedure, they learned a list of visual cues of interest: for changes in gaze direction, cues included, "sustained eye contact," "upward gaze," "downward gaze", and "sideways gaze"; for facial expressions, cues included, "thinking," "happy," "amused," "serious" and "embarrassed" expressions; and for shifts in posture (in the speaker's seat), cues included "still," "forward," "backward", or "sideways". Coders were provided a paper copy of descriptions for the different visual cue subcategories to use as a reference during the procedure. The subcategories of visual cues were determined based on previous studies (Goodwin and Goodwin, 1986; Krahmer and Swerts, 2005; Swerts and Krahmer, 2005; Cramer et al., 2009). See **Figure 1** for an illustration of some of these visual cues.

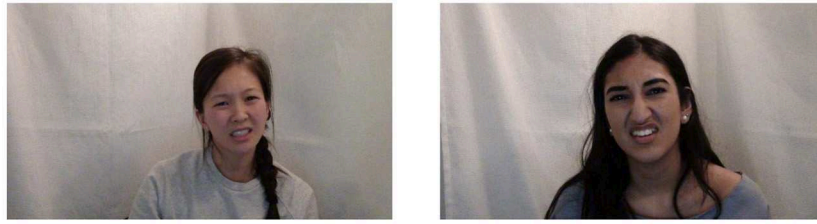
For the subcategories of changes in gaze direction, sustained eye contact was described as the speaker looking straight ahead at the camera. Descriptions of the different facial expressions were based on how they are typically described in the literature. A thinking expression (or thinking face) involved the corners of the lips turned downwards, widened eyes, pursed or stretched lips, wrinkled nose, and raised or furrowed eyebrows (Ekman and Friesen, 1978; Goodwin and Goodwin, 1986; Swerts and Krahmer, 2005). A happy expression involved a smile with the corners of the lips elongated and turned upwards, and raised cheeks (Ekman and Friesen, 1978; Sato and Yoshikawa, 2007). An amused expression involved similar movements as a happy expression as well as the head shifted backwards, an upward gaze, and tongue biting (Ekman and Friesen, 1976, 1978; Ruch, 1993; Keltner, 1996). A serious expression involved eye contact and minimal head movement without furrowed eyebrows or lip corner movement. An embarrassed expression involved a suppressed smile with minimal teeth showing, the head turned downwards or sideways, a downward gaze and face touching (Haidt and Keltner, 1999; Heerey et al., 2003; Tracy et al., 2009).

Each visual cue category (changes in gaze direction, facial expression, shifts in posture) was coded in separate blocks, and the order of these blocks was counterbalanced across coders. Each trial consisted of a fixation point in the middle of the screen (3,000 ms), a muted video (of variable duration), and the visual cue rating scale (presented until the coder responded). Coders could select all the subcategories of visual cues that they saw by clicking on these labels on the computer screen, unless they had chosen sustained eye contact or a still posture, in a block for changes in gaze direction or shifts in posture, respectively. In this case, they were instructed to not select any other gaze or posture cues, respectively. The coders' click responses indicated the presence of a cue, while not clicking indicated the absence of a cue. Stimuli were randomly assigned to ten blocks (≈ 20 videos/block), separated by a short break. No time limit was imposed, and each video could be repeatedly watched by the coder until they were satisfied.

TABLE 1 | Speaker's subjective confidence ratings (out of 5) as a function of the type of general knowledge question and their accuracy for the general knowledge questions and the frequency of occurrences (out of the 249 muted videos).

	General knowledge question type					
	Well-known			Less-known		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Accurate response						
Subjective confidence rating	119	4.39	1.12	7	2.00	1.42
Inaccurate response						
Subjective confidence rating	7	2.43	1.50	116	1.46	0.91

Thinking facial expression: Speaker furrows or raises their eyebrows, wrinkle their nose, and lip corners are turned downwards



Serious facial expression: Speaker has sustained eye contact and does not furrow their eyebrows or change their lip form



Change in eye gaze: Speaker averts their gaze upwards, sideways or downwards



FIGURE 1 | Still images of speakers (from Production Study) illustrating a few subcategories of visual cues and their descriptions.

Inter-rater Reliability

Gwet's AC1 (or Gwet's first-order agreement coefficient) (Gwet, 2008; Wongpakaran et al., 2013) was used to calculate inter-rater reliability of the coded visual cues because it can be used with categorical data involving more than two raters and more than two categories (McCray, 2013). This measure was used to calculate inter-rater reliability for the presence/absence of each visual cue subcategory. Gwet's AC1 is an agreement coefficient which ranges from 0 to 1 (McCray, 2013). Compared to other Cohen's Kappa measures for multiple raters, Gwet's AC1 provides a measure of statistical significance, is positively biased toward the agreement/disagreement between raters, and is able to handle missing data (i.e., it does not exclude items that were coded by less than six coders) (McCray, 2013). Gwet's AC1 was calculated using an R script called "agree.coeff3.dist.R" (Advanced Analytics LLC, 2010; Gwet, 2014). The magnitude of inter-rater agreement was determined by comparing the Gwet AC1 coefficient to the Altman benchmark scale (Altman, 1991).

For the subcategories of changes in gaze direction (eye contact, upward gaze, downward gaze, sideways gaze) there was fair agreement (Gwet AC1 coefficient = 0.26, SE = 0.02, 95% CI = 0.21, 0.30, $p < 0.001$). For the subcategories of facial expressions (thinking, serious, happy, amused, embarrassed expressions), there was also fair agreement (Gwet AC1 = 0.25, SE

= 0.02, 95% CI = 0.21, 0.29, $p < 0.001$). For the subcategories of shifts in posture (still posture, forward, backward, and sideways shifts), there was moderate agreement (Gwet AC1 = 0.58, SE = 0.03, 95% CI = 0.53, 0.63, $p < 0.001$).

Results

Analyses focused on characterizing speakers based on their confidence level by examining speaker's subjective confidence (i.e., self-ratings) in their responses for less-known and well-known general knowledge questions and the specific visual cues speakers produced during these states. Statistical analyses were performed and figures were created using R statistical computing and graphics software in RStudio (R Core Team, 2017).

Manipulation Check for Speaker's High and Low Confidence

The purpose of these analyses was to ensure that the two types of trivia questions (well-known and less-known general knowledge) posed to speakers elicited a state of high and low confidence, respectively. As expected, speakers were more accurate in responding to the well-known general knowledge questions ($M = 0.94$, $SD = 0.23$) than the less-known general knowledge questions ($M = 0.06$, $SD = 0.23$, $d = 3.83$). This effect was seen across all speakers based on each speaker's mean

accuracy by trivia question type, $t_{(8)} = 20.75$, $p < 0.001$, 95% CI [0.79, 0.99]. Seven (out of the nine) speakers answered all the well-known general knowledge questions correctly and six (out of the nine) speakers incorrectly answered all the less-known general knowledge questions. Speakers were also subjectively more confident following their response to well-known general knowledge questions ($M = 4.29$, $SD = 1.22$) than to less-known general knowledge questions ($M = 1.49$, $SD = 0.96$, $d = 2.92$). This effect was seen across all speakers based on each speaker's mean subjective confidence ratings by trivia question type, $t(8) = 16.71$, $p < 0.001$, 95% CI [2.39, 3.15]. Following their response to well-known general knowledge questions, speakers rated themselves as "very confident" (i.e., a rating of 5 out of 5) for 68.3% of trials and following their response to less-known general knowledge questions, speakers rated "not at all confident" (i.e., a rating of 1 out of 5) for 72.4% of trials.

When only correct responses to each trivia question type were considered, mean self-ratings of confidence were significantly higher ($M = 4.26$, $SD = 1.26$) than when they answered the question incorrectly ($M = 1.51$, $SD = 0.98$, $d = 2.18$). This effect was seen across all speakers based on each speaker's mean subjective confidence ratings by their trivia question accuracy, $t(8) = 14.87$, $p < 0.001$, 95% CI [2.30, 3.15]. When speakers answered correctly, they rated "very confident" for 68.3% of the trials and when speakers answered incorrectly, they rated "not at all confident" for 71.5% of the trials. Based on these data, it can be argued that the speakers' responses to well-known vs. less-known general knowledge questions elicited representative and highly differentiated states of high vs. low confidence in the speech content. For the remaining analyses of visual cues, we will therefore refer to responses to well- and less-known general knowledge questions as reflecting high and low confidence conditions, respectively.

Characterizing Speakers' Confidence Level From Coded Visual Cues

To further characterize speakers' subjective confidence, we analyzed the specific visual cues that speakers produced when they had high or low confidence in their speech content. Given that inter-rater reliability was fair to moderate, we adopted a relatively strict criterion: a visual cue was considered present if indicated by at least 5 out of the 6 coders (83% agreement), otherwise it was coded as absent. Coders most frequently observed speakers producing a still posture (57.3% of all items) and a change in gaze direction (upward, downward and sideways gaze) (46.1% of all items) irrespective of speaker confidence level, and a serious facial expression was most often observed in the high confidence condition (51.5% of items).

We performed chi-square tests to determine if the speaker's confidence level (high or low) varied independently of the types of visual cues speakers produced using their gaze, facial expression and posture, where these cues were either present or absent. For the speaker's gaze, the relationship between sustained eye contact and confidence level was significant, $\chi^2(1) = 5.24$, $p = 0.02$, Cramer's $V = 0.16$. When speakers produced sustained eye contact, it more often occurred when speakers had high confidence (20.8% of items) compared to low confidence (8.6%

of items). When speakers produced an upward gaze, it more often occurred when speakers had low confidence (17.1% of items) compared to high confidence (3.0% of items), $\chi^2(1) = 9.80$, $p = 0.002$, Cramer's $V = 0.22$. The other shifts in gaze (downward or sideways) were not significantly associated with a confidence condition, downward gaze: $\chi^2(1) = 2.10$, $p = 0.15$, Cramer's $V = 0.10$; sideways gaze: $\chi^2(1) = 3.77$, $p = 0.05$, Cramer's $V = 0.14$. A downward gaze was reported for 23.8% of low confidence items and 14.9% of high confidence items, and a sideways gaze occurred for 21.9% of low confidence items and 10.9% of high confidence items. These patterns suggest that upward eye movements are more prevalent when speakers experience low confidence, whereas sustained eye contact occurs more frequently in cases of high confidence.

When characterizing the speaker's facial expressions, it was first noted that coders rarely identified facial expressions being "amused," "embarrassed," or "happy" (always less than 2% of items in any condition). When speakers were characterized as having a "serious" face, this occurred significantly more often when speakers displayed high confidence, $\chi^2(1) = 41.11$, $p < 0.001$, Cramer's $V = 0.45$. A serious expression occurred for 51.5% of high confidence items, compared to only 9.5% of low confidence items. In contrast, when speakers had low confidence, there were significantly more instances of a "thinking face," $\chi^2(1) = 6.00$, $p = 0.01$, Cramer's $V = 0.17$ (low confidence = 13.3% of items, high confidence = 3.0% of items). These patterns suggest that the most prominent facial cues were a thinking expression for low confidence and a serious expression for high confidence.

For the speaker's posture, the relationship between speakers producing a still posture and the speaker's confidence level was significant, $\chi^2(1) = 5.93$, $p = 0.01$, Cramer's $V = 0.17$. When speakers had a still posture, it more often occurred when they had high confidence (66.3% of items) compared to low confidence (48.6% of items). Forward, backward, and sideways shifts in posture occurred more frequently when speakers had low confidence in their response; although these features were rarely identified in our stimuli for either confidence level (less than 3% of items in any condition). This pattern suggests that sitting still was the most salient postural feature in our data which was linked more often to a high confidence state.

We also analyzed whether speakers' self-ratings of confidence for the general knowledge questions related to the presence/absence of each visual cue and observed similar patterns (see **Supplementary Material**).

Discussion

This study allowed us to elicit a state of high and low confidence in speakers based on the speech content. Specifically, speakers were less accurate and had low subjective confidence ratings in their responses to the less-known general knowledge questions compared to the well-known general knowledge questions. This result replicates previous findings that used a similar task (e.g., Smith and Clark, 1993; Swerts and Kraemer, 2005).

We also replicated previous findings in terms of the visual cues that speakers with high and low confidence, respectively produce (Goodwin and Goodwin, 1986; Doherty-Sneddon and Phelps, 2005; Swerts and Kraemer, 2005; Murphy, 2007). For

example, speakers with high confidence in their speech content were more likely to produce sustained eye contact, a serious facial expression, and no shifts in posture. These cues are similar to those reported in Murphy (2007), where speakers who were asked to appear intelligent (Acting condition) during an informal conversation with another speaker, were also more likely to display eye contact while speaking, a serious face and an upright posture. Thus, the visual cues produced by speakers with high confidence in the content of their speech, may be related to the cues produced by speakers conveying high intelligence. Moreover, there was a medium to large effect size when speakers produced a serious facial expression. This may suggest that a serious facial expression is a reliable visual cue for differentiating speakers with high vs. low confidence. We also found that speakers with low confidence were more likely to produce an upward gaze and a thinking facial expression. This result supports previous findings that cite the presence of these visual cues when speakers are trying to retrieve words from memory (Smith and Clark, 1993) and/or to save face from others in a social context (Goffman, 1967, 1971; Visser et al., 2014). Overall, the effect sizes for most of the produced visual cues ranged from small to medium, potentially due to variability in speakers' production of these cues.

By creating a high and low confidence speaker condition based on speakers' level of certainty in their speech content, we then investigated how observers differentiated between these confidence levels using a speaker's visual cues in the Perception Study. The prevalence of the visual cues produced in this Production Study will impact the visual cues that observers use to discern speakers of high and low confidence.

PART II: PERCEPTION STUDY

Methods

Observers

To gather data on the perceived confidence level of speakers and to determine which visual cues observers used to decode a speaker's confidence level, a group of 27 observers rated the muted videos in three tasks. Observers were native speakers of Canadian or American English (19 female/8 male, mean age = 21.93 years, $SD = 3.38$). None of the observers were familiar with any of the speakers presented in the videos, as determined at the onset of the study.

Materials

The items included the 249 muted videos of single utterances from 9 different speakers, from the Production Study. Items were included irrespective of the speaker's accuracy in answering the trivia questions.

Tasks and Procedures

After providing written consent, observers were tested individually in a quiet testing laboratory containing several desktop computers. A testing session included one to four observers. Observers were seated in front of a desktop computer controlled by SuperLab 5 presentation software (Cedrus Corporation, 2014) and responses were recorded automatically

by the computer (by means of a mouse click). Each observer completed three tasks in a set order:

1. Social impressions rating task. While not the main focus of the study, we first collected information about broader social impressions of the speakers that could influence how observers rated their confidence level. Observers rated each of the nine speakers along six dimensions: competence, attractiveness, relaxedness, openness, trustworthiness, and friendliness (Cramer et al., 2009, 2014). Each social dimension was rated in a separate block using a 5-point scale ("not at all" to "very much," e.g., "not at all competent" to "very competent"). Blocks for each social dimension were randomized for presentation order across observers and trials of individual speakers were always randomized within blocks.

In each block, observers watched two consecutive muted videos of a given speaker separated by a short inter-stimulus interval: one in which the speaker was experiencing low confidence (i.e., low subjective confidence rating to a less-known general knowledge question) and one high confidence (i.e., high subjective confidence rating to a well-known general knowledge question). The same two videos for each speaker appeared when rating each social dimension and these videos were not used in the main tasks. Observers were not informed of a difference between the two videos or that the videos were demonstrating high or low confidence. By presenting each speaker in a high and low confidence state we aimed to partial out the effect of these social dimensions from a speaker's perceived confidence ratings, as similar videos were shown in the main task. After watching each speaker, observers clicked one of five circles along the target impression continuum (on the computer screen) to indicate their judgment. These data allowed us to analyze the mean rating of each speaker for each of the six social dimensions.

2. Speaker confidence rating task. Observers watched the 249 muted video stimuli and judged "how confident the speaker is" on a 5-point scale ("not at all confident" to "very confident") by clicking one of five circles on a computer screen. Observers were not informed which confidence condition they were watching. Each trial consisted of a fixation point in the middle of the screen (3,000 ms), a muted video (of variable duration), and the perceived confidence rating scale (presented until the observer responded). Stimuli were randomly assigned to eight blocks (≈ 31 – 32 videos/block), separated by a short break. The proportion of videos showing high vs. low confidence responses and different speakers was equally distributed within blocks and then randomized. The computer recorded the numerical perceived confidence rating for each trial.
3. Classification of visual cues task. After rating speaker confidence, observers watched the same 249 muted videos presented in a newly randomized order. Their goal was to categorize which visual regions they believed that they used to evaluate the speaker's confidence level in a seven forced-choice paradigm (See **Supplementary Material** for paradigm layout). After each video, observers encountered a screen showing six categories of visual cues (eyes, eyebrows, facial expression, head, hands, posture). They were instructed to

TABLE 2 | Speaker's general knowledge question accuracy, subjective confidence ratings (out of 5) (from Production Study), and perceived confidence ratings (out of 5) (from Perception Study) for high and low confidence conditions (formerly well- and less-known general knowledge questions, respectively).

	Speaker confidence condition			
	High confidence		Low confidence	
	M	SD	M	SD
Question accuracy	0.94	0.23	0.06	0.23
Subjective confidence ratings	4.29	1.22	1.49	0.96
Perceived confidence ratings	3.30	1.07	2.70	1.20

select all regions they believed helped them judge the speaker's confidence level, defined as either a movement or lack of movement in that body area. Alternatively, they could choose a seventh category, "Do not know". The structure and timing of trials was identical to the Speaker confidence rating task. The computer recorded all categories selected for each trial. Observers were compensated for their time and the experiment took approximately an hour and a half to complete.

Results

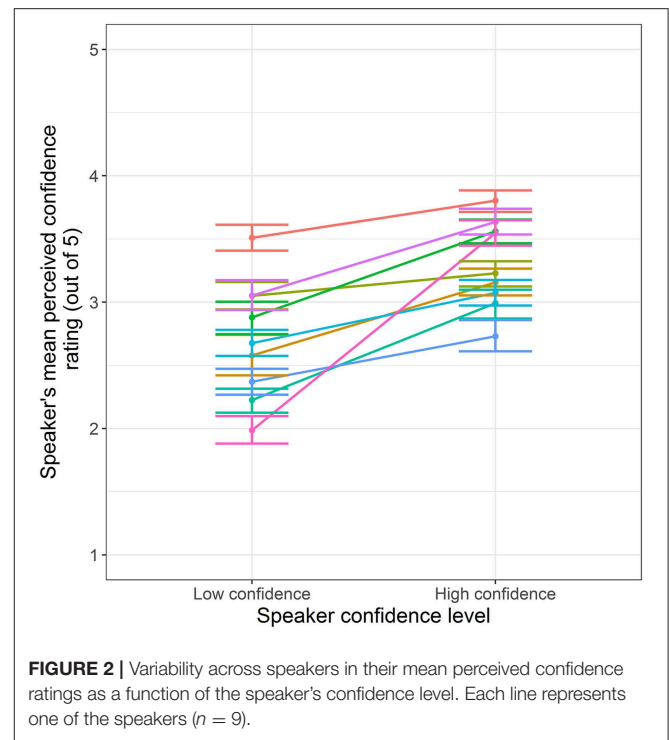
Analyses focused on determining the types of visual cues observers used to evaluate speakers' confidence level. Statistical analyses were performed, and figures were created using R statistical computing and graphics software in RStudio (R Core Team, 2017).

Speaker Perceived Confidence

Speakers were perceived to be visibly more confident when experiencing high confidence ($M = 3.30$, $SD = 1.07$) than low confidence ($M = 2.70$, $SD = 1.20$), $t_{(26)} = 7.06$, $p < 0.001$, 95% CI [0.43, 0.78], $d = 0.50$. This suggests that observers could accurately differentiate between the high and low confidence conditions from speakers' visual cues. **Table 2** compares the subjective and perceived confidence ratings of speakers in the high and low confidence conditions.

We then examined the relationship between the speakers' perceived confidence level and the speaker's self-rated confidence level using Spearman's rank correlation (ρ , ϱ). The speaker's perceived confidence level positively correlated with the speaker's self-rated confidence level, ($\varrho = 0.31$, $p < 0.001$), meaning that when speakers were subjectively more confident in their response, they visibly appeared more confident to observers. This weak-moderate relationship may be explained by **Figure 2**, showing variability in the difference between perceived confidence ratings for high and low confidence, for each speaker.

The relationship between the perceived confidence of a speaker and broader social impressions (attractiveness, relaxedness, friendliness, competence, openness, and trustworthiness) was briefly examined by calculating Spearman rank correlations (ρ , ϱ). The perceived confidence ratings were collapsed across the high and low confidence conditions. The confidence ratings assigned by observers correlated significantly with five of the six social dimensions: attractiveness ($\varrho = -0.07$, $p < 0.001$), relaxedness ($\varrho = 0.11$, $p < 0.001$), competence



($\varrho = 0.18$, $p < 0.001$), openness ($\varrho = 0.09$, $p < 0.001$) and trustworthiness ($\varrho = 0.07$, $p < 0.001$). Higher perceived confidence scores were weakly associated with speakers who were rated as more relaxed, competent, open, trustworthy and less attractive. There was no evidence of an association with speaker's perceived friendliness, $\varrho = -0.01$, $p = 0.31$. The intercorrelations between these social dimensions was low ($\varrho < 0.30$).

To assess whether speakers' perceived confidence ratings were predicted by their confidence condition and observers' broader social impressions, we conducted a linear mixed effects model using lme4 (Bates et al., 2015) in R (R Core Team 2017). We used the Satterthwaite (1946) approximation for computing p-values for t-statistics, as implemented in the lmerTest package version 2.0-6 (Kuznetsova et al., 2017). Speakers' perceived confidence ratings were the response variable.

The initial model included the following fixed effects: the speaker's confidence condition (baseline = high confidence), the ratings for all the social dimensions (attractiveness, relaxedness, friendliness, competence, openness, and trustworthiness) and each of their interactions with the speaker's confidence condition. Random effects included intercepts by Speaker and Observer. A speaker's perceived friendliness, openness, and trustworthiness were not significant predictors in this model and thus were excluded from the final model. The final model included the following fixed effects: the speaker's confidence condition, speaker's perceived attractiveness, relaxedness and competence ratings, and each of their interactions with a speaker's confidence condition. The same random effects as the initial model were also included. Compared to the minimal model (with speaker confidence condition as a fixed effect), this final model significantly contributed to the overall model likelihood, $\Delta D =$

TABLE 3 | Proportion of each visual cue selected by observers in Perception Study (Part II) in their confidence judgments of the speakers as a function of the confidence condition.

Visual cue	Speaker confidence condition		
	High confidence ^a	Low confidence ^b	Total trials
Eyes	0.63	0.68	0.65
Eyebrows	0.21	0.30	0.25
Facial expression	0.51	0.68	0.59
Head	0.25	0.43	0.34
Hands	0.08	0.06	0.07
Posture	0.29	0.26	0.28
Do not know	0.08	0.03	0.06

^aData is based on 3402 trials.

^bData is based on 3321 trials.

18886, $df = 6$, $p < 0.001$. The more competent speakers were perceived to be, the more confident they were also perceived to be ($B = 0.18$, $SE = 0.02$, $p < 0.001$), controlling for the effect of the other perceived social dimensions. Also, a speaker's confidence condition significantly interacted with a speaker's perceived attractiveness ($B = 0.15$, $SE = 0.02$, $p < 0.001$), and relaxedness ratings ($B = 0.05$, $SE = 0.02$, $p = 0.03$). These interactions indicate that when speakers had high confidence in their speech content, speakers who were initially perceived to be more attractive or relaxed, were more likely to be perceived as having high confidence. No other main effects or interactions were significant. These results demonstrate that observers' baseline impressions of speakers' perceived competence, attractiveness and relaxedness significantly affected their perceived confidence ratings.

Categorization of Visual Cues Used to Rate Speaker Confidence

Table 3 shows the total percentage of each type of visual cue (eyes, eyebrows, facial expressions, head, hands, posture, do not know) that the 27 observers reported using to rate the speaker's confidence level across trials in the high and low confidence conditions. Irrespective of the speaker's confidence level, observers most frequently reported using a speaker's eyes in their confidence evaluations (65.1% of all trials), followed by their facial expressions (59.4% of all trials). Observers reported using a speaker's hands with the lowest frequency (6.7% of all trials), although this result could partially reflect an artifact of our experimental setup in which speakers were seated alone in a booth and were required to press a button on the response pad following each trial. It is noteworthy that there were relatively few trials in which observers responded, "Do not know" (5.7% of all trials), which confirms that observers had strong impressions of the visual cues that guided their confidence ratings for most trials.

We analyzed the proportion of observers who selected each type of visual cue across trials in the high and low confidence conditions, for each of the nine speakers, using paired samples t -tests across speakers or the Mann-Whitney test where normality was violated. From Table 3, observers more often selected the speaker's eyes in their confidence judgments for low confidence

trials ($M = 0.68$, $SD = 0.13$) than high confidence trials ($M = 0.63$, $SD = 0.12$), $t_{(8)} = -5.27$, $p = 0.001$, 95% CI $[-0.07, -0.03]$, the speaker's facial expression in their confidence judgments for low confidence trials ($M = 0.68$, $SD = 0.21$) than high confidence trials ($M = 0.51$, $SD = 0.19$), $t_{(8)} = -4.31$, $p = 0.003$, 95% CI $[-0.26, -0.08]$, and the speaker's head in their confidence judgments for low confidence trials ($M = 0.43$, $SD = 0.22$) than high confidence trials ($M = 0.25$, $SD = 0.15$), $t_{(8)} = -5.13$, $p = 0.001$, 95% CI $[-0.27, -0.10]$. There was a marginally significant difference in observer's selection of the speaker's eyebrows in their confidence judgments for low confidence trials ($M = 0.30$, $SD = 0.18$) than high confidence trials ($M = 0.21$, $SD = 0.15$), $W = 63$, $p = 0.05$, and no significant difference in observers' selection of the speaker's hands in their confidence judgments (low confidence trials: $M = 0.06$, $SD = 0.19$, high confidence trials: $M = 0.07$, $SD = 0.20$), $W = 38$, $p = 0.86$, and the speaker's posture in their confidence judgments (low confidence trials: $M = 0.26$, $SD = 0.16$, high confidence trials: $M = 0.29$, $SD = 0.14$), $t_{(8)} = 1.26$, $p = 0.24$, 95% CI $[-0.02, 0.07]$. Lastly, observers more often did not know which visual cues they used in their confidence judgments for high confidence trials ($M = 0.08$, $SD = 0.07$) compared to low confidence trials ($M = 0.03$, $SD = 0.05$), $W = 12.5$, $p = 0.02$. These patterns suggest that observers more often use the speaker's eyes, facial expression and head when the speaker has low confidence compared to high confidence.

Discussion

We found that observers accurately perceived a speaker's confidence level relative to the speaker's subjective confidence level, as evidenced by a significant moderate correlation, supporting our hypotheses. Thus, by only accessing the visual communication channel observers' perception of speakers' visual cues can provide sufficient information for assessing their confidence level (Nelson and Russell, 2011) as previously reported (Krahmer and Swerts, 2005; Swerts and Krahmer, 2005; Visser et al., 2014). This moderate correlation is in line with previous studies where observers only had access to the visual communication channel when decoding a speaker's mental state (Zaki et al., 2009). However, it is important to note that the range of speakers' perceived confidence ratings overlapped when speakers had high vs. low confidence. This result indicates that when observers only have access to the visual communication channel it is difficult to distinguish between a speaker's confidence level in the content of their speech. Observers likely need cues from other communication channels to support/disprove their evaluations of a speaker's confidence. However, the extent of this differentiation of a speaker's confidence level in the speech content, also varied across speakers. As seen in Figure 2, a few speakers had a larger difference in their perceived confidence ratings when they had high vs. low confidence, likely because of a marked difference in the types and number of visual cues they produced in each confidence level. In contrast, for other speakers, they were consistently perceived as having relatively neutral or high confidence even when they had low confidence, or they were consistently perceived as having relatively neutral or low confidence, even when they had high confidence. This variability

in speakers' visual displays of their confidence level in the speech content supports the notion that speakers produce visual cues of confidence for pragmatic purposes, such as saving face (Goffman, 1967, 1971), or trying to persuade others of their stance (Moons et al., 2013), and relates to speakers varied expressivity of their mental state by producing visual cues.

For observers to perceive a speaker's confidence level in the content of their speech, they placed greater perceptual weight on cues from the speakers' facial expressions, head movements, and eyes, compared to the speakers' posture or hands. This result is somewhat consistent with our results from the Production Study where speakers more often produced an upward gaze, and a thinking facial expression when they had low confidence, and consistent eye contact, a serious facial expression and still posture when they had high confidence. The perceptual weight of these visual cues may be grounded in perceiving a speaker's eye movements, a common feature among many of these cues. Participants are more likely to believe a speaker's statement when produced with a direct gaze (or eye contact) before speaking compared to an averted gaze, while other facial features remain constant (Kreysa et al., 2016). This effect is explained by a motivation approach (Adams and Kleck, 2005; Bindemann et al., 2008) which postulates that direct gaze enhances the perception of approach-oriented emotions such as joy while an averted gaze enhances the perception of avoidance-oriented emotions such as sadness (Adams and Kleck, 2005). Thus, a speaker's direct gaze can signal an attempt to approach/persuade an observer, potentially because they are confident in the content of their speech, while an averted gaze signals their avoidance, potentially because they are less certain. Though this theory is based on speakers conveying emotions rather than a transient level of confidence, when observers are viewing muted videos of speakers producing utterances, they may not perceive the difference between a speaker conveying an emotion vs. a mental state.

Observers' use of the speakers' eye gaze, facial expressions and head movements may have allowed them to efficiently infer the speaker's low confidence due to mentalizing with the speaker (Kuhlen et al., 2014) as well as the increased perceptual salience of the visual cues produced in the eye and facial region. Though we did not predict the impact of the speaker's head movements on observer's evaluations of a speaker's confidence level, use of these movements along with speaker's facial expressions/movements may represent the speaker's lexical retrieval process and social communication of their mental state. For example, Goodwin and Goodwin (1986) found that when speakers changed their gaze direction before producing a thinking face, it also involved the speaker turning their head away from the interlocutor, possibly to avoid being distracted by the interlocutor and/or to concentrate on retrieving words from memory and planning their speech (Argyle and Cook, 1976). Also, lateral head shakes or small lateral tremors can be produced to mark a speaker's uncertainty in their speech where they may correct the words retrieved or produce a pause or hesitation in speech (McClave, 2000; Heylen, 2005).

Also, despite speakers with high confidence producing a still posture, postural cues were not used by observers to differentiate speaker's confidence level. This result is similar to Murphy (2007) where speakers consciously trying to display

intelligence produced an upright posture along with a serious face and eye contact while speaking, and upright posture was not significantly related to a speaker's perceived intelligence. Our result demonstrates that not all visual cues produced by speakers with high confidence are meaningfully used by observers to evaluate their confidence level. At the same time, observers' decreased use of the speaker's posture may have been influenced by speakers' limited range of movement while seated during the experiment, similar to Murphy (2007) where speakers were also seated. Our results may have differed if speakers were standing while answering the general knowledge questions, allowing for a greater range of movement that could be meaningfully used to decode their confidence level.

GENERAL DISCUSSION

During communication, speakers produce visual cues signaling their confidence level in the content of their speech, which can reflect their word retrieval (Goodwin and Goodwin, 1986; Rimé and Schiaratura, 1991; Krauss et al., 1996) and/or their (un)conscious desire to communicate their mental state to others (Goodwin and Goodwin, 1986; Visser et al., 2014). However, it is unknown if observers use the visual cues from these cognitive and social processes when perceiving a speaker's confidence level. There is a lack of research investigating whether there is a difference between the types of visual cues speakers produce because of their transient confidence level and the types of visual cues observers use to assess this mental state. This study tested this issue in relation to a speaker's confidence for general knowledge facts by observers indicating the visual cues they used after evaluating a speaker's confidence level.

Effect of Situational Factors on Visual Cue Production and Evaluation

Compared to previous studies that report the occurrence of visual cues in relation to a speaker's confidence level, our results shed light on the frequency of these visual cues that speakers produce, and consequently their effect on observers' confidence evaluations. Overall, we found that speakers do not often produce changes in eye gaze, facial expressions and postural shifts, but when salient changes in these cues are produced, they can impact observers' evaluations. We postulate that one reason for this low frequency production of visual cues is due to the stakes level of social evaluation created by our experiment. A situation involving high-stakes social evaluation occurs when an individual (or group) has significant power to impact another's future (Gifford et al., 1985; Cuddy et al., 2012) such as in a job interview. Here, a speaker's speech content and their confidence level in this content, through vocal and visual cues, can greatly impact the outcome of their social interaction. In contrast, a situation of low-stakes social evaluation does not pose a risk to a speaker's social status. For example, a speaker may mundanely report inaccurate information due to inaccurately recalling (highly or less accessible) information and they will not face major social consequences for this (Depaulo et al., 1996; Gozma et al., 2001; Azizli et al., 2016). Our trivia question task may have simulated

a situation of low-stakes social evaluation, in that if speakers answered a question incorrectly it did not have negative social consequences for them or their interaction with the examiner. The speakers did not know the examiner, the accuracy of their responses was not important for their social status, pausing due to word retrieval difficulties could occur because there was low social demand to respond quickly, and the questions were not used to elicit an emotional state in the speakers. This low-stakes context may explain why the speakers did not often produce postural shifts or happy, amused or embarrassed facial expressions based on their confidence level, as these cues are more often associated with situations involving social judgments and affect (Costa et al., 2001; Coleman and Wu, 2006; Hareli and Parkinson, 2008). Alternatively, the low frequency of embarrassed or amused facial expressions may have been related to the few instances where speakers answered questions about well-known general knowledge incorrectly (Table 1). Future research may explore the effect of shared knowledge that is known by the audience/observer on the visual cues that speakers produce and the visual cues that observers subsequently attend to. For example, if an observer knows the information that a speaker is trying to recall, or a speaker has difficulty recalling information that is well-known knowledge to an observer or audience; how do these situations affect a speaker's perceived confidence and the cues that observers attend to. This question may be addressed by examining the interaction between a speaker's visual and vocal cues on their perceived confidence.

This low-stakes context may also explain the small difference in speakers' perceived confidence ratings for high vs. low confidence. Since speakers were in a situation of low social stress, the quantity and types of visual cues they produced to indicate high vs. low confidence in their speech content, may have been less differentiated to observers. Future research may explore how the level of social evaluation in a given context can impact speaker's memory retrieval and visual cues as well as observers' evaluations. Nonetheless, we do not know if observers detected that speakers produced speech in a low-stakes context, and if detecting this context may have affected the types of visual cues used in their social evaluations.

Research on a speaker's transient confidence level based on their speech content has largely tested speakers' level of certainty for general knowledge facts (Smith and Clark, 1993; Brennan and Williams, 1995; Swerts and Kraemer, 2005). More research is needed to understand the impact of speakers' transient confidence level for other information, such as their opinions. This is particularly relevant when speakers are trying to persuade an audience of their stance, as the speech content may also elicit a greater emotional valence compared to responding to general knowledge questions. Here, speakers' responses, including their visual cues, are more likely to be socially evaluated by others such as an employer conducting a job interview (DeGroot and Motowidlo, 1999), a person on a date (Muehlenhard et al., 1986; Fichten et al., 1992; Ambady and Skowronski, 2008), a judge in court (Cramer et al., 2009), a customer interacting with a salesperson (Leigh and Summers, 2002), or a border security officer. Future research should explore this branch of a speaker's confidence in their speech content with greater

consideration for situational factors. In these contexts, speakers may (un)consciously produce visual cues to save face (Goffman, 1967, 1971; Visser et al., 2014) or reinforce their level of certainty (Moons et al., 2013), and observers' evaluations of a speaker's confidence level via their non-verbal cues can have real social implications.

Limitations and Conclusions

One limitation of our study was our inter-rater reliability for the subcategories of visual cues speakers produced. Though the inter-rater agreement between the coders was significantly greater than chance, the coefficients did not indicate strong agreement. Thus, the detection of a speaker's visual cues may require more training as individuals can vary in their conceptualization of and identification for visual cues. Also, a speaker's production of eye movements to signal their confidence level may be influenced by their cultural background, in accordance with cultural display rules in the eyes and facial expressions (McCarthy et al., 2006, 2008). Our results were based on native speakers of Canadian English with various cultural backgrounds; controlling for this factor could help explain some of the variability between the speakers in the quantity and types of visual cues produced. Moreover, examining the effect of observers' cultural attitudes based on a speaker's physical appearance would also allow us to examine other types of social factors that may have influenced observers' confidence evaluations of speakers and whether the types of visual cues that observers used in their evaluations differed based on a speaker's physical appearance. Also, having a larger sample of speakers would allow us to increase the generalizability of our findings. Lastly, our study operated under the assumption that when observers only have access to the visual communication channel, they can accurately recall what visual cues they used in their evaluation of a speaker's confidence. It may be argued that this manipulation created an unnatural context to evaluate a speaker's confidence level, since in everyday communication observers have access to many communication channels and can assess a speaker's mental state in real time. On the other hand, there are also real-life contexts where observers have heightened attention for/use of the visual communication channel, such as interacting with speakers of a foreign or second language (Sueyoshi and Hardison, 2005) or listening to speech in a noisy environment (Neely, 1956). Use of an eye-tracking paradigm could provide an objective, real-time measure of the visual cues that observers used in their confidence evaluations. Nonetheless, observers rarely reported they did not know which visual cues they used in their evaluations. This result may indicate that when observers only have access to the visual communication channel, they are to some extent, aware of the visual cues they perceive.

In conclusion, this study showed that observers can accurately perceive a speaker's confidence level in the content of their speech in a low-stakes context based on their visual cues, by primarily using the speaker's facial expressions and eye movements. Observers may be socially attuned to salient, yet infrequent cues in the facial regions so they can efficiently infer a speaker's confidence level in the speech content. This social attention may be important as the stakes of the situation increase for speakers.

Future research should explore using more video recordings of speakers' spontaneous speech and examining variability across speakers, to improve descriptions of the visual cues that speakers produce signaling their confidence level in the speech content. These methodological changes will better reflect the cues that aid in observers' assessments of a speaker's confidence level and their communicative interactions with others.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by McGill Faculty of Medicine Institutional Review Board in accordance with principles expressed in the Declaration of Helsinki. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

REFERENCES

- Adams, R. B. J., and Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion* 5, 3–11. doi: 10.1037/1528-3542.5.1.3
- Advanced Analytics LLC (2010). *R Functions for Calculating Agreement Coefficients*. Advanced Analytics LLC.
- Altman, D. G. (1991). *Practical Statistics for Medical Research*. Chapman and Hill.
- Ambady, N., and Skowronski, J. J. (Eds.). (2008). *First Impressions*. New York, NY: The Guilford Press.
- Anderson, J. R. (1983a). A spreading activation theory of memory. *J. Verbal Learning Verbal Behav.* 22, 261–295. doi: 10.1016/S0022-5371(83)90201-3
- Anderson, J. R. (1983b). *The Architecture of Cognition*. Cambridge: Harvard University Press.
- Argyle, M., and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.
- Azizli, N., Atkinson, B. E., Baughman, H. M., Chin, K., Vernon, P. A., Harris, E., et al. (2016). Lies and crimes: dark triad, misconduct, and high-stakes deception. *Pers. Individ. Dif.* 89, 34–39. doi: 10.1016/j.paid.2015.09.034
- Bates, D., Machler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bavelas, J., and Gerwing, J. (2007). "Conversational hand gestures and facial displays in face-to-face dialogue," in *Social Communication*, ed K. Fiedler (New York, NY: Psychology Press), 283–308.
- Bindemann, M., Burton, A. M., and Langton, S. R. H. (2008). How do eye gaze and facial expression interact? *Vis. Cogn.* 16, 708–733. doi: 10.1080/13506280701269318
- Birch, S. A. J., Akmal, N., and Frampton, K. L. (2010). Two-year-olds are vigilant of others' non-verbal cues to credibility. *Dev. Sci.* 13, 363–369. doi: 10.1111/j.1467-7687.2009.00906.x
- Boduroglu, A., Tekcan, A. I., and Kapucu, A. (2014). The relationship between executive functions, episodic feeling-of-knowing and confidence judgements. *J. Cogn. Psychol.* 26, 333–345. doi: 10.1080/20445911.2014.891596
- Brennan, S. E., and Williams, M. (1995). The feeling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *J. Mem. Lang.* 34, 383–398. doi: 10.1006/jmla.1995.1017

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: YM and MP. Performed the experiments, analyzed the data, and wrote the paper: YM. Contributed reagents, materials, and analysis tools and feedback on completed manuscript: MP.

FUNDING

This research was supported by an Insight Grant to MP from the Social Sciences and Humanities Research Council of Canada (G245592 SSHRC 435-2017-0885) and a Doctoral Fellowship to YM from the Social Sciences and Humanities Research Council of Canada (SSHRC 752-2017-1227).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2019.00063/full#supplementary-material>

Supplementary Videos 1 and 2 | Examples of the muted video stimuli presented to participants demonstrating the high confidence (**Video 1**) and low confidence conditions (**Video 2**).

- Brosy, J., Bangerter, A., and Mayor, E. (2016). Disfluent responses to job interview questions and what they entail. *Discourse Process.* 53, 371–391. doi: 10.1080/0163853X.2016.1150769
- Cedrus Corporation (2014). *SuperLab 5*. San Pedro, CA.
- Coleman, R., and Wu, H. D. (2006). More than words alone: incorporating broadcasters' nonverbal communication into the stages of crisis coverage theory - Evidence from September 11th. *J. Broadcast. Electron. Media* 50, 1–17. doi: 10.1207/s15506878jobem5001_1
- Costa, M., Dinsbach, W., Manstead, A. S. R., and Ricci Bitti, P. E. (2001). Social presence, embarrassment, and nonverbal behavior. *J. Nonverbal Behav.* 25, 225–240. doi: 10.1023/A:1012544204986
- Cramer, R. J., Brodsky, S. L., and Decoster, J. (2009). Expert witness confidence and juror personality: their impact on credibility and persuasion in the courtroom. *J. Am. Acad. Psychiatry Law* 37, 63–74.
- Cramer, R. J., Parrott, C. T., Gardner, B. O., Stroud, C. H., Boccaccini, M. T., and Griffin, M. P. (2014). An exploratory study of meta-factors of expert witness persuasion. *J. Individ. Differ.* 35, 1–11. doi: 10.1027/1614-0001/a000123
- Cuddy, A. J. C., Wilmoth, C. A., and Carney, D. R. (2012). The benefit of power posing before a high-stakes social evaluation. *Harvard Bus. Sch. Work. Pap.* 13, 1–18.
- DeGroot, T., and Gooty, J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *J. Bus. Psychol.* 24, 179–192. doi: 10.1007/s10869-009-9098-0
- DeGroot, T., and Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *J. Appl. Psychol.* 84, 986–993. doi: 10.1037/0021-9010.84.6.986
- Depaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979
- Depaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118. doi: 10.1037/0033-2909.129.1.74
- Doherty-Sneddon, G., and Phelps, F. G. (2005). Gaze aversion: a response to cognitive or social difficulty? *Mem. Cogn.* 33, 727–733. doi: 10.3758/BF03195338
- Ekman, P., and Friesen, W. V. (1976). Measuring facial movement. *Environ. Psychol. Nonverbal Behav.* 1, 56–75. doi: 10.1007/BF01115465

- Ekman, P., and Friesen, W. V. (1978). *Manual for the Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press. doi: 10.1037/t27734-000
- Fichten, C. S., Tagalakis, V., Judd, D., Wright, J., and Amsel, R. (1992). Verbal and nonverbal communication cues in daily conversations and dating. *J. Soc. Psychol.* 132, 751–769. doi: 10.1080/00224545.1992.9712105
- Fiske, S. T., and Taylor, S. E. (1984). *Social Cognition*. Reading, MA: Addison-Wesley.
- Gifford, R., Ng, C. F., and Wilkinson, M. (1985). Nonverbal Cues in the Employment Interview. Links between applicant qualities and interviewer judgments. *J. Appl. Psychol.* 70, 729–736. doi: 10.1037/0021-9010.70.4.729
- Glenberg, A. M., Schroeder, J. L., and Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Mem. Cogn.* 26, 651–658. doi: 10.3758/BF03211385
- Goffman, E. (1967). *Interaction ritual: Essays on Face to Face Behavior*. New York, NY: Double day.
- Goffman, E. (1971). *Relations in Public*. New York, NY: Harper and Row.
- Goodwin, M. H., and Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62, 51–76. doi: 10.1515/semi.1986.62.1-2.51
- Gozna, L. F., Vrij, A., and Bull, R. (2001). The impact of individual differences on perceptions of lying in everyday life and in a high stake situation. *Pers. Individ. Dif.* 31, 1203–1216. doi: 10.1016/S0191-8869(00)00219-1
- Grice, H. P. (1975). “Logic and conversation,” in *Syntax and Semantics 3: Speech Acts*, eds P. Cole and J. L. Morgan (New York, NY: Academic Press), 41–58.
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *Br. J. Math. Stat. Psychol.* 61, 29–48. doi: 10.1348/000711006X126600
- Gwet, K. L. (2014). *Some R Functions for Calculating Chance-Corrected Agreement Coefficients*. Available online at: <http://inter-rater-reliability.blogspot.ca/2014/03/some-r-functions-for-calculating-chance.html>
- Haidt, J., and Keltner, D. (1999). Culture and facial expression: open-ended methods find more expressions and a gradient of recognition. *Cogn. Emot.* 13, 225–266. doi: 10.1080/026999399379267
- Hall, J. A., Carter, J. D., and Horgan, T. G. (2001). Status roles and recall of nonverbal cues. *J. Nonverbal Behav.* 25, 79–100. doi: 10.1023/A:1010797627793
- Hareli, S., and Parkinson, B. (2008). What's social about social emotions? *J. Theory Soc. Behav.* 38, 131–156. doi: 10.1111/j.1468-5914.2008.00363.x
- Heerey, E. A., Keltner, D., and Capps, L. M. (2003). Making sense of self-conscious emotion: linking theory of mind and emotion in children with autism. *Emotion* 3, 394–400. doi: 10.1037/1528-3542.3.4.394
- Heylen, D. (2005). “Challenges ahead: head movements and other social acts in conversations,” in *AISB Convention: Virtual Social Agents: Social Presence Cues for Virtual Humanoids Empathic Interaction with Synthetic Characters Mind Minding Agents*, 45–52.
- Jiang, X., and Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex* 66, 9–34. doi: 10.1016/j.cortex.2015.02.002
- Jiang, X., and Pell, M. D. (2017). The sound of confidence and doubt. *Speech Commun.* 88, 106–126. doi: 10.1016/j.specom.2017.01.011
- Jiang, X., Sanford, R., and Pell, M. D. (2017). Neural systems for evaluating speaker (Un)believability. *Hum. Brain Mapp.* 38, 3732–3749. doi: 10.1002/hbm.23630
- Keltner, D. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: a study of recalled antecedents and facial expressions of emotion. *Cogn. Emot.* 10, 155–172. doi: 10.1080/026999396380312
- Krahmer, E., and Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Lang. Speech* 48, 29–53. doi: 10.1177/00238309050480010201
- Krauss, R. M., Chen, Y., and Chawla, P. (1996). “Nonverbal behavior and nonverbal communication: what do conversational hand gestures tell us?” in *Advances in Experimental Social Psychology*, ed M. Zanna (San Diego, CA: Academic Press), 389–450. doi: 10.1016/S0065-2601(08)60241-5
- Kreysa, H., Kessler, L., and Schweinberger, S. R. (2016). Direct speaker gaze promotes trust in truth-ambiguous statements. *PLoS ONE* 11:e0162291. doi: 10.1371/journal.pone.0162291
- Kuhlen, A. K., Bogler, C., Swerts, M., and Haynes, J. D. (2014). Neural coding of assessing another person's knowledge based on nonverbal cues. *Soc. Cogn. Affect. Neurosci.* 10, 729–734. doi: 10.1093/scan/nsu111
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13
- Leigh, T. W., and Summers, J. O. (2002). An initial evaluation of industrial buyers' impressions of salespersons' nonverbal cues. *J. Pers. Sell. Sales Manag.* 22, 41–53. doi: 10.1080/08853134.2002.10754292
- Locke, C. C., and Anderson, C. (2015). The downside of looking like a leader: power, nonverbal confidence, and participative decision-making. *J. Exp. Soc. Psychol.* 58, 42–47. doi: 10.1016/j.jesp.2014.12.004
- Mann, S. A., Vrij, A., Fisher, R. P., and Robinson, M. (2008). See no lies, hear no lies: differences in discrimination accuracy and response bias when watching or listening to police suspect interviews. *Appl. Cogn. Psychol.* 22, 1062–1071. doi: 10.1002/acp.1406
- McArthur, L. Z. (1981). “What grabs you: The role of attention in impression formation and causal attribution,” in *Social Cognition: The Ontario Symposium*, eds E. Higgins, C. Herman, and M. Zanna (Hillsdale, NJ: Erlbaum), 201–246.
- McArthur, L. Z., and Solomon, L. K. (1978). Perceptions of an aggressive encounter as a function of the Victim's salience and the Perceiver's arousal. *J. Pers. Soc. Psychol.* 36, 1278–1290. doi: 10.1037/0022-3514.36.11.1278
- McCarthy, A., Lee, K., Itakura, S., and Muir, D. W. (2006). Cultural display rules drive eye gaze during thinking. *J. Cross. Cult. Psychol.* 37, 717–722. doi: 10.1177/0022022106292079
- McCarthy, A., Lee, K., Itakura, S., and Muir, D. W. (2008). Gaze display when thinking depends on culture and context. *J. Cross. Cult. Psychol.* 39, 716–729. doi: 10.1177/0022022108323807
- McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *J. Pragmat.* 32, 855–878. doi: 10.1016/S0378-2166(99)00079-X
- McCray, G. (2013). “Assessing inter-rater agreement for nominal judgement variables,” in *Language Testing Forum* (Nottingham), 15–17.
- Moons, W. G., Spoor, J. R., Kalomiris, A. E., and Rizk, M. K. (2013). Certainty broadcasts risk preferences: verbal and nonverbal cues to risk-taking. *J. Nonverbal Behav.* 37, 79–89. doi: 10.1007/s10919-013-0146-0
- Muehlenhard, C. L., Koralewski, M. A., Andrews, S. L., Burdick, C. A., Collinge, D., Bowling, B., et al. (1986). Verbal and nonverbal cues that convey interest in dating: two studies. *Behav. Ther.* 17, 404–419. doi: 10.1016/S0005-7894(86)80071-5
- Murphy, N. A. (2007). Appearing smart: the impression management of intelligence, person perception accuracy, and behavior in social interaction. *Personal. Soc. Psychol. Bull.* 33, 325–339. doi: 10.1177/0146167206294871
- Neely, K. K. (1956). Effect of visual factors on the intelligibility of speech. *J. Acoust. Soc. Am.* 28, 1275–1275. doi: 10.1121/1.1908620
- Nelson, N., and Russell, J. (2011). When dynamic, the head and face alone can express pride. *Emotion* 11, 990–993. doi: 10.1037/a0022576
- Nelson, T. O., and Narens, L. (1990). “Metamemory: a theoretical framework and some new findings,” in *The Psychology of Learning and Motivation*, ed G. H. Bower (San Diego, CA: Academic Press), 125–173. doi: 10.1016/S0079-7421(08)60053-5
- Patterson, M. L. (1976). An arousal model of interpersonal intimacy. *Psychol. Rev.* 83, 235–245. doi: 10.1037/0033-295X.83.3.235
- Patterson, M. L. (1982). A sequential functional model of nonverbal exchange. *Psychol. Rev.* 89, 231–249. doi: 10.1037/0033-295X.89.3.231
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Core Team.
- Rimé, B., and Schiaratura, L. (1991). “Gesture and speech,” in *Studies in Emotion & Social Interaction. Fundamentals of Nonverbal Behavior*, eds R. S. Feldman and B. Rimé (New York, NY: Cambridge University Press), 239–281.
- Ruch, W. (1993). “Exhilaration and Humor,” in *The Handbook of Emotion*, eds M. Lewis and J. M. Haviland (New York, NY: Guilford Publications), 605–616.
- Sato, W., and Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* 104, 1–18. doi: 10.1016/j.cognition.2006.05.001
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bull.* 2, 110–114. doi: 10.2307/3002019
- Scherer, K. R., London, H., and Wolf, J. J. (1973). The voice of confidence: paralinguistic cues and audience evaluation. *J. Res. Pers.* 7, 31–44. doi: 10.1016/0092-6566(73)90030-5
- Schmid Mast, M., Murphy, N. A., and Hall, J. A. (2006). “A brief review of interpersonal sensitivity: measuring accuracy in perceiving others,” in *Current*

- Themes in Social Psychology*, eds D. Chadee and J. Young (Kingston: University of the West Indies Press), 163–185.
- Smith, V., and Clark, H. (1993). On the course of answering questions. *J. Mem. Lang.* 32, 25–38. doi: 10.1006/jmla.1993.1002
- Sueyoshi, A., and Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Lang. Learn.* 55, 661–699. doi: 10.1111/j.0023-8333.2005.00320.x
- Sullins, E. S. (1989). Perceptual salience as a function of nonverbal expressiveness. *Personal. Soc. Psychol. Bull.* 15, 584–595. doi: 10.1177/0146167289154011
- Swerts, M., and Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *J. Mem. Lang.* 53, 81–94. doi: 10.1016/j.jml.2005.02.003
- Swerts, M., and Krahmer, E. (2010). Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *J. Phon.* 38, 197–206. doi: 10.1016/j.wocn.2009.10.002
- Tenney, E. R., Spellman, B. A., and Maccoun, R. J. (2008). The benefits of knowing what you know (and what you don't): how calibration affects credibility. *J. Exp. Soc. Psychol.* 44, 1368–1375. doi: 10.1016/j.jesp.2008.04.006
- Tracy, J. L., Robins, R. W., and Schriber, R. A. (2009). Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion* 9, 554–559. doi: 10.1037/a0015766
- Visser, M., Krahmer, E., and Swerts, M. (2014). Children's expression of uncertainty in collaborative and competitive contexts. *Lang. Speech* 57, 86–107. doi: 10.1177/0023830913479117
- Wongpakaran, N., Wongpakaran, T., Wedding, D., and Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med. Res. Methodol.* 13:61. doi: 10.1186/1471-2288-13-61
- Zaki, J., Bolger, N., and Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion* 9, 478–487. doi: 10.1037/a0016551

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mori and Pell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.