Children's Integration of Verbal and Emotional Cues for Communicative Inferences

Marija Bolic (marija.bolic@mail.utoronto.ca) & Yang Wu (yangm.wu@utoronto.ca)
Department of Psychology, University of Toronto Scarborough
1265 Military Trail, Scarborough, ON M1C 1A4 Canada

Abstract

Words and emotional expressions each convey information, but together they can reveal more than either cue alone. Adults integrate these cues to make nuanced inferences, but when does this capacity emerge in development? Children aged 5–10 (N=210) and adults (N=35) watched videos in which a person tasted a friend's food, smiled or frowned, and stated that it "tastes good" or "tastes bad." Participants were asked to infer the food's taste and the speaker's communicative goals. For food taste, adults prioritized negative emotions when cues conflicted; for communicative goals, they used both cues to infer a goal to be honest and relied primarily on speech to infer a goal to make others feel good. Children exhibited substantial developmental progression between ages 5 and 7, although subtle differences from adults persisted at age 10. These results suggest middle childhood is a pivotal period for learning to make sophisticated inferences from multiple cues.

Keywords: pragmatics; social cognition; emotional expressions; speech comprehension; cognitive development

Introduction

Imagine a child watching two adults, Alex and Sam, having dinner. Alex tastes the food Sam made, frowns, and says, "It tastes good." What might the child infer? By integrating facial expressions and speech, the child could conclude that the food tastes bad and that Alex intends to be kind rather than honest. This ability to infer both ground truth (e.g., the food's taste) and communicative goals (e.g., honesty and kindness) is central to everyday social interaction. How does this ability to integrate multiple, sometimes conflicting, cues develop over childhood?

The Understanding of Speech

Although verbal cues can be taken literally, they often require pragmatic interpretation that accounts for the speaker's communicative goals. Grice (1975) proposed a foundational account of speech as cooperative, in which speakers aim to be truthful, relevant, clear, and informative without redundancy. This view has been formalized in the Rational Speech Act (RSA) framework (Goodman & Frank, 2016), which models communication as recursive reasoning in which listeners infer speakers' beliefs under the assumption that speakers aim to be informative while avoiding redundancy.

However, speakers do not always aim to be fully informative. In the opening example, Alex may tell a white lie to protect Sam's feelings, departing from the informational goal assumed in cooperative communication and highlighting the social functions of speech (Brown & Levinson, 1987). Building on the RSA framework, Yoon et al. (2020) model politeness as a tradeoff between informational and social goals, showing that different weightings of these goals yield varying degrees of politeness. While these accounts characterize speakers' communicative goals, they do not address how such understanding develops over childhood.

Children's ability to infer communicative goals from speech develops gradually. Jaswal et al. (2010) found that 3-year-olds often trust verbal testimony even after prior dishonesty, suggesting an early default assumption of honesty. Between ages 3–5, children increasingly distinguish accurate from inaccurate informants and selectively trust truthful speakers (Brosseau-Liard et al., 2014; Jaswal & Neely, 2006; Koenig et al., 2004; Li et al., 2014; Vanderbilt et al., 2011), indicating growing sensitivity to speakers' informativeness. In parallel, children's understanding of the social motives behind speech strengthens: they increasingly recognize that white lies can be used to protect others' feelings and infer prosocial motives in politeness contexts (Broomfield et al., 2002; Vendetti et al., 2019; Warneken & Orlins, 2015; Xu et al., 2010). For instance, Heyman et al. (2009) found that children aged 7–11 evaluated lying and truth-telling differently, treating polite lies as more acceptable than lies that cover transgressions. They also progressively justified polite lies by their impact on others' feelings, indicating a growing understanding of prosocial motives. Overall, development involves increasingly flexible reasoning about both informational and social goals underlying speech.

The Understanding of Emotional Expressions

Communication extends beyond words: nonverbal cues, especially emotional expressions, play a central role in social interaction. Several perspectives address the function of emotional expressions. While Basic Emotions Theory characterizes expressions as automatic reflections of distinct emotions (Cowen et al., 2019; Darwin, 1872; Ekman & Friesen, 1971; Izard, 1994), expressions can also serve communicative functions. These functions may be informational or social, and people can strategically regulate expressions to achieve different goals. Informationally, emotional expressions can function like language in conveying ground truth, shaping others' behavior, and signaling future actions (Scarantino, 2017). Socially, expressions such as smiling may signal politeness

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

or affiliation rather than genuine affect (Kraut & Johnston, 1979). These functional accounts invite consideration about how emotional signals are interpreted via people's intuitive reasoning about emotions.

There are a variety of intuitive inferences adults can make from emotional cues. Prior work has formalized adults' abilities to use emotional cues to infer emotional or mental states, with the assumption that emotional expressions are reflections of those states (Houlihan et al., 2022; Ong et al., 2015; Saxe & Houlihan, 2017; Wu et al., 2018). However, emotional cues can also be produced strategically for communicative purposes and, therefore, do not always reflect underlying states. Thus, without modeling communicative goals, existing approaches cannot fully account for the communicative functions of emotional expressions.

Developmental work has looked at how young children make inferences from observed emotional cues (for review, see: Wu & Gweon, 2021). Even in infancy, children use others' facial expressions to guide their behavior, such as whether to cross a cliff (Moses et al., 2001; Repacholi, 1998; Sorce et al., 1985). They do this not only when emotions are directed at them, but also when observing emotional exchanges between other people (Repacholi & Meltzoff, 2007). When watching an agent show a happy or disgusted reaction to some food, 18-month-olds were able to use those emotional expressions to infer the agent's preference, even when it conflicted with their own (Doan et al., 2015; Repacholi & Gopnik, 1997). These findings suggest a basic understanding of the informational role of emotional expressions. Around ages 6–8, children increasingly reason about the social goals behind emotional concealment, such as to use concealed emotions to determine the desires of the person concealing their emotion and their social partner (Naito & Seki, 2009; Wu & Schulz, 2020). By age 8–9, children demonstrate a more complex understanding, recognizing in which contexts people selectively show or hide negative feelings depending on whether their goal is to be informative or prosocial (Hayashi & Shiomi, 2015).

The Integration of Speech and Emotional Cues

In everyday communication, people rarely rely on a single modality; instead, integrating multiple cues yields richer inferences than any cue alone. Although prior research has often examined speech and emotional expressions in isolation, far less is known about how these cues are integrated.

Studies with adults investigated how people use both speech and emotional cues to infer ground truth and communicative goals (Wu et al., 2021). While adults consider both informational and social goals for speech and emotional cues, they recognize that emotional expressions may be automatic. Thus, when cues conflict, emotional expressions are more likely to be taken as revealing the true state. These inferences have been formalized with a computational model (Wu et al., 2021).

A few studies have examined how children integrate both cues. Children increasingly use the consistency between verbal and emotional cues to judge if someone is telling the truth. Although 5-year-olds largely fail to apply this principle, it emerges by age 7 and is robust by age 9 (Rotenberg et al., 1989). Similarly, Gillis and Nilsen (2017) found that while 4–5-year-olds simply prefer positive statements, 7–8-year-olds prefer information from consistent over inconsistent speakers.

When cues are inconsistent, children must decide which to prioritize. When inferring others' true feelings or preferences, toddlers show no consistent bias toward verbal or prosodic emotional cues (Brady et al., 2024). Children aged 3–8 primarily rely on speech (Friend & Bryant, 2000; Morton & Trehub, 2001; except when emotional expressions are exaggerated, Eskritt & Lee, 2003), while children aged 9–10 adopt an adult-like reliance on emotional expressions (Friend & Bryant, 2000; Morton & Trehub, 2001). These studies highlight that with age, children increasingly favor emotional cues when in conflict with speech cues to make inferences about the speaker's preferences or true feelings.

Relatively little work has addressed whether children can infer the communicative goals underlying a speaker's speech and emotional expressions. To the best of our knowledge, only Rothermich et al. (2020) has touched upon it, showing that children aged 8–12 can infer speakers' social goals from multimodal cues in naturalistic interactions. They rated speakers' intent to be friendly after hearing utterances produced in different delivery styles, judging literal positive statements, teasing, and prosocial lies as more friendly than sarcasm and literal negative statements. However, the stimuli consisted of naturalistic interactions between two people using different delivery styles, incorporating cues such as laughing, eye rolling, and nudging. Emotional cues were not systematically manipulated and the other character's reactions were not held constant, limiting conclusions about how children integrate speech and emotional cues when inferring social goals. Second, they did not examine informational goals, which would provide a more complete picture of the speaker's intent. The present study will address these limitations.

Current study

The current study systematically examined how children aged 5–10 integrated verbal and emotional cues to infer the ground truth of a situation, as well as a speaker's informational goal (i.e., wanting to be honest) and social goal (i.e., wanting to make others feel good). An adult comparison group was included to replicate prior findings (Wu et al., 2021) and to provide a reference point for interpreting children's performance.

Both children and adults watched videos in which a person tasted food prepared by a friend, displayed an emotion, and produced a verbal utterance. Emotional cues (smile vs. frown) were crossed with utterances ("it tastes good" vs. "it tastes bad"), yielding four conditions: smile-good, smile-bad, frown-good, and frown-bad. Participants judged the food's taste (ground truth), whether the speaker wanted to be honest (informational goal), and whether the speaker wanted to make their friend feel good (social goal).

For each question, participants chose yes, no, or not sure, and then rated confidence for yes/no responses (a little sure, sure, very sure), yielding a 7-point scale from -3 (no–very sure) to +3 (yes–very sure). Thus, higher ratings reflected higher certainty about the yes answer, lower ratings indicated higher certainty about the no answer, and the midpoint indicated uncertainty. Prior work has mostly assessed children's responses using binary options (Eskritt & Lee, 2003; Friend & Bryant, 2000; Morton & Trehub, 2001; Rotenberg et al., 1989; except for Gillis & Nilsen, 2017, using simple evaluation scales). Our confidence scale captures variation in certainty among participants who may reach the same judgment, addressing ambiguity in social inferences.

For adults, we predicted replication of prior work (Wu et al., 2021). Regarding ground truth judgments, we expected high certainty for consistent cues, with smile-good rated highest and frown-bad lowest. For inconsistent cues, frowning but saying the food tasted good should be interpreted as a white lie, resulting in low ratings comparable to the frown-bad condition. In contrast, smiling and saying the food tasted bad allows multiple interpretations (e.g. teasing or sarcasm), leading to intermediate, more uncertain ratings. The predictions are summarized as Prediction 1:

Food taste: smile-good > smile-bad > frown-good ≥ frown-bad

For informational goals, we expected consistent cues (smile-good, frown-bad) to signal a goal of honesty more than inconsistent cues (frown-good, smile-bad), with further differences within each type. Between consistent conditions, we predicted greater confidence when a speaker frowned and said the food tasted bad than when they smiled and said it tasted good, as negative statements may signal an honest intent despite social discomfort, whereas positive statements may reflect either an intent to be polite or honest. Between inconsistent conditions, we predicted lower ratings when a speaker frowned and said it tasted good than when they smiled and stated it tasted bad, since the motivation for lying is clearer in frown-good (to preserve feelings). Prediction 2 summarizes the order.

Informational goal: frown-bad > smile-good > smile-bad > frown-good

For social goals, we expected positive speech ("it tastes good") to signal a stronger intent to make others feel good than negative speech ("it tastes bad"), reflecting the assumption that speech is typically intentional, whereas emotional expressions may be intentional or automatic. We further predicted minor differences within each speech type. When positive speech is paired with a frown, it should receive higher social goal ratings because it reflects a white lie, compared to when it is paired with a smile, which may convey either honesty or kindness. When negative speech is paired with a smile, it should receive higher social goal ratings as smiling may mitigate the negativity of the verbal statement, compared to when it is paired with a frown, which reinforces the negative statement. The predictions are outlined in Prediction 3.

Social goal: frown-good > smile-good > smile-bad > frown-bad

Developmentally, we predicted that with age, children would increasingly use multiple cues to make nuanced, adult-like inferences. The study is preregistered: https://aspredicted.org/28f7de.pdf.

Methods

Participants

We tested 210 children aged 5–10 (M = 7.96 years; range = 5.01–11.01; 105 females) via *Children Helping Science* (formerly *Lookit*; Scott & Schulz, 2017), with 35 children per age group (5-year-olds: M = 5.45 years, range = 5.01–5.92; 6-year-olds: M = 6.43 years, range = 6.00–6.99; 7-year-olds: M = 7.50 years, range = 7.07–7.97; 8-year-olds: M = 8.45 years, range = 8.05–8.96; 9-year-olds, M = 9.43 years, range = 9.04–9.95; 10-year-olds, M = 10.48 years, range = 10.03–11.01). Parents reported their children's race/ethnicity as White (113), Asian (38), African American/Black (2), Hispanic/Latino/Spanish (6), Middle Eastern/North African (4), multiracial (38), or did not respond (9).

Following our preregistered exclusion criteria, an additional 23 participants were tested but excluded due to completing less than half of the test trials (4), less than 0.8 mean score on training trials (3), technical issues (3), and self-report of any disability or disorder (12). In addition, one child was excluded for having observed a sibling participate; although not preregistered, this exclusion minimized potential confounds from prior task exposure.

We also recruited 35 adults (M_{Age} = 40.14 years; 16 females, 17 males, 2 other) through *Prolific*. They reported their race/ethnicity as White (28), Asian American/Asian (3), African American/Black (1), and multiracial (3). None were excluded per our preregistered exclusion criteria.

Stimuli

We created video stimuli of actors tasting food made by their friends. Each video began with an actor that displayed a neutral facial expression and stated in a neutral tone: "I will try the food now." The actor then took a bite offscreen and reappeared pretending to chew. Afterward, they displayed either a smile or a frown (target expression), and then said either "it tastes good" or "it tastes bad" (target utterance). The expression was constrained to mainly the face and upper body. The smile expression could have involved lips curved upward, eyebrow raising, shoulder raising or the eye aperture narrowing. The frown expression could have involved making expressions of disgust, such as lips curled downwards, nose wrinkling, pulling their head back, or brow furrowing. The utterance always followed the expression, and both intonation and facial expression during speech were kept relatively neutral to ensure consistency across the conditions. Four sets of videos were recorded, each featuring a different actor producing the expressions and utterances for the four experimental conditions. The video stimuli is available here: https://osf.io/wguqs/overview?view_only=7cd56dba80b0406dad001328f7bdd9d6.

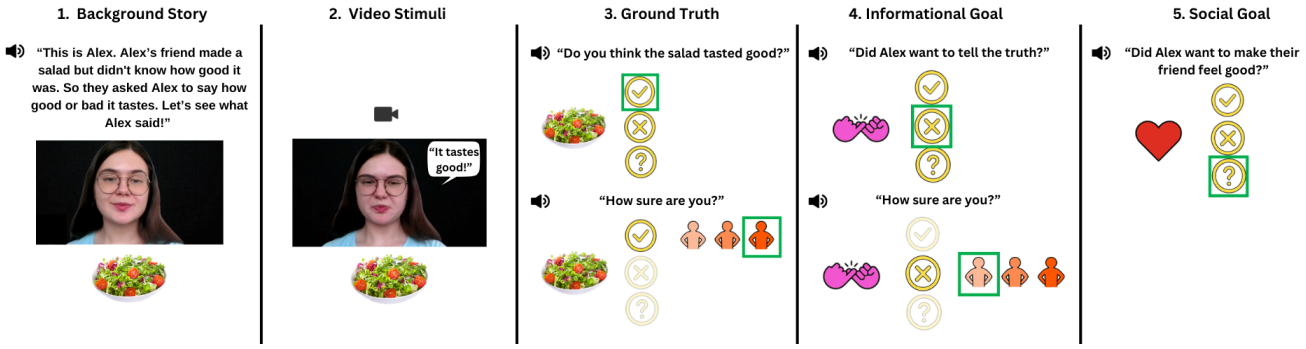

Figure 1: Each trial began with a background story (1), followed by a video (2) showing a target expression and utterance. Children then inferred the ground truth (3), the speaker's informational goal (4), and social goal (5) by selecting a check mark, cross, or question mark. Selecting the former two prompted a confidence rating. Green boxes indicate potential selections.

Procedure and design

Children completed training trials to familiarize them with the response format. They heard a simple story and selected options as instructed, with incorrect responses prompting retries. Training scores were coded as 1 (correct on first attempt), 0.5 (second attempt), or 0 (two or more incorrect attempts). Both children and adults performed near ceiling (mean accuracy scores ≥0.95 after exclusion; see Participants).

Each child then completed four test trials covering all expression-utterance combinations in randomized order. For each condition, one actor (Sam, Taylor, Alex, or Avery) and one food type (pasta, sandwich, pizza, or salad) were randomly assigned, with each appearing exactly once per participant.

On each trial, children heard a background story, watched a video of a person tasting and commenting on food made by a friend, and then answered three questions with the option to replay the video (see Figure 1). For ground-truth inference, participants were shown an image of the food and asked: "Do you think the [food] tasted good?" Participants selected a check mark (✓), cross (×), or question mark (?), corresponding to "yes," "no," and "you're not sure," respectively. Selecting "yes" or "no" prompted confidence ratings ("a little sure," "sure," "very sure") depicted by progressively darker human silhouettes; selecting "not sure" advanced participants to the next question. Informational goals were assessed using a pinky-promise icon with the question: "Did [character] want to tell the truth?" Social goals were assessed using a heart icon with the question: "Did [character] want to make their friend feel good?" The response format was identical to that used for the ground-truth question.

Adult participants completed the study online via *Qualtrics*. They were told that their responses would be compared with those of child participants to examine the development of multi-model inference. The procedure matched that used for children, except that instructions and prompts were presented in audio format for children and in written form for adults. The adult experiment can be accessed here: https://utorontopsych.az1.qualtrics.com/jfe/form/SV_3rbLhspNA2CxJKC.

Results

We analyzed data using preregistered mixed-effects models implemented in *lme4* in *R*. For child data, we fit a model for each test question, predicting ratings from condition, age (centered and continuous), and their interaction, as well as a model for each age group predicting ratings from condition alone. Adult data were analyzed with one model per question, predicting ratings from condition.

Because models with random slopes failed to converge, all analyses included random intercepts only; when singularity occurred, the random intercept was removed and the model was fit as a linear model. Pairwise comparisons were conducted using estimated marginal means, with Tukey adjustments for multiple comparisons.

For adults, responses differed significantly by condition for each question (all all ps $< .001$). For children, age interacted significantly with condition for each question (all ps $< .001$). Detailed results are presented below.

Ground Truth

Consistent with Prediction 1, adults rated smile-good highest (all ps $< .001$), smile-bad intermediate (all ps $< .001$), and both frown-good and frown-bad lowest (all ps $< .001$), with no difference between the two ($p = .924$).

Children's performance largely resembled adults', with the largest developmental change in the frown-good condition (i.e., white lie). Five-year-olds interpreted the taste as good based on the positive statement, rating this condition no differently from smile-good ($p = .104$) and significantly higher than frown-bad ($p < .001$). With age, ratings decreased such that by age 7 they were significantly lower than both smiling conditions (all ps $< .013$) and no longer differed from frown-bad (all ps $> .156$), suggesting that children aged 7 and older interpret positive speech paired with negative expressions as white lies. See Figure 2A.

Together, these findings indicate that by age 7, children integrate speech and emotional cues in an adult-like manner when inferring ground truth. Consistent with prior work (Es-

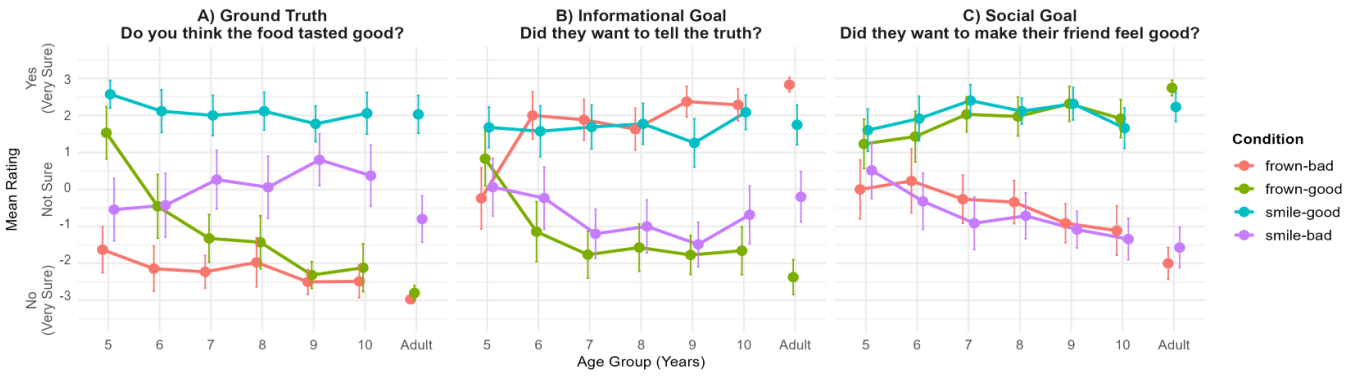

Figure 2: Results. Mean confidence ratings for (A) ground truth, (B) informational goal, and (C) social goal across the four conditions, shown separately for each child age group (5–10 years) and for adults. Error bars represent 95% confidence intervals.

kritt & Lee, 2003; Friend & Bryant, 2000; Morton & Trehub, 2001), this developmental shift reflects growing sensitivity to emotional cues in the interpretation of positive statements.

Informational Goal

Adults' responses aligned with Prediction 2 when inferring the speaker's goal to tell the truth (i.e., an informational goal). They rated consistent cues higher than inconsistent cues overall (all ps $< .001$) and further differentiated within each type. Among consistent conditions, participants were more confident that frown-bad reflected an informational goal than smile-good ($p = .009$). Among inconsistent conditions, smile-bad elicited greater confidence than frown-good ($p < .001$).

Children by age 6 also rated consistent cues higher than inconsistent cues overall (all ps $< .003$). However, they did not further differentiate within each type (all ps $> .096$, except $p = .016$ for 9-year-olds in the consistent conditions), as shown in Figure 2B.

Compared to adults, children may use a simple heuristic—treating consistent cues as indicating honest intent and inconsistent cues as indicating dishonest intent. The ability to integrate emotional and verbal information in a more nuanced, adult-like manner continue to develop beyond age 10.

Social Goal

Consistent with Prediction 3, adults showed greater certainty that the speaker's goal was to make their friend feel good (i.e., a social goal) in conditions with positive speech (smile-good and frown-good) than in conditions with negative speech (smile-bad and frown-bad; ps $< .001$). This pattern became clear in children by age 6 (all ps $< .047$).

For further differentiation within each speech type, adults showed effects in the predicted direction but did not reach significance between the two positive-speech conditions ($p = .253$) or between the two negative-speech conditions ($p = .413$). Children did not even show any predicted trends (all ps $> .359$). See Figure 2C.

These results suggest that although prior work shows that adults and children understand that emotional expressions can be used strategically to make others feel good (Hayashi & Shiomi, 2015; Kraut & Johnston, 1979; Naito & Seki, 2009), they do not robustly rely on emotional cues to infer social goals when speech cues are present. Instead, both adults and children by age 6 prioritize speech content when inferring whether the speaker intends to make others feel good.

Discussion

The present study investigated developmental changes in how individuals integrate speech and emotional cues to infer both ground truth and a speaker's communicative goals. Adults flexibly weighted both cues depending on the type of inference. For instance, when frowning and stating the food tastes good, they prioritized emotional cues for ground truth inferences. They also relied on a more sophisticated principle than cue consistency when inferring informational goals, and weighted speech more heavily for social goal inferences across all conditions. Children showed substantial developmental progression between ages 5 and 7; by age 10, their reasoning was largely adult-like, with subtle residual differences.

This study advances prior work both empirically and methodologically. Empirically, previous studies either manipulated speech and emotional cues without examining communicative goals (Eskritt & Lee, 2003; Friend & Bryant, 2000; Gillis & Nilsen, 2017; Morton & Trehub, 2001; Rotenberg et al., 1989), or examined communicative goals without manipulating both cues (Rothermich et al., 2020). Our study bridges this gap by systematically manipulating these cues and examining inferences about both communicative goals and ground truth. Methodologically, to capture more nuanced inferences arising from these manipulations, we used a confidence scale. This finer-grained measure allowed us to detect distinctions that would be missed with coarser measures used in prior work (Eskritt & Lee, 2003; Friend & Bryant, 2000; Gillis & Nilsen, 2017; Morton & Trehub, 2001; Rotenberg et al., 1989). For example, it revealed a gap between children and adults in inferring informational goals: children largely relied on cue consistency, whereas adults showed graded certainty

within each consistency type, such as being more certain that a speaker wanted to be honest when frowning and saying the food tasted bad than when smiling and saying it tasted good.

Overall, the adult results replicated prior work by Wu et al. (2021) and largely confirmed our predictions, with the exception of some nuanced distinctions in social goal inference. As expected, adults judged positive speech cues as more indicative of social goals than negative speech cues. However, they did not show significant differentiation within speech conditions, suggesting limited use of emotional expressions to further disambiguate these inferences. At first glance, this pattern appears to contrast with Wu et al. (2021), which found that social goal inferences draw on both speech and emotional cues. This discrepancy likely reflects methodological differences: Whereas Wu et al. (2021) used computational modeling that focused on overall pattern fitting to detect small effects, our hypothesis-testing approach requires larger effects or higher statistical power to reach significance. Notably, our adult results were *numerically* consistent with both the predicted patterns and the results of Wu et al. (2021) across all four conditions (see Figure 2C, Adult), suggesting that emotional cues may contribute to social goal inferences but with effects too small for our methods to detect. Future work could combine experimental manipulations with model-based analyses or increase sensitivity (e.g., more trials or participants) to better characterize the role of emotional expressions in social goal inference. That said, regardless of whether emotional cues have small or null effects, the present findings point to adults placing substantially greater weight on speech than on emotional cues when inferring social goals.

The child results reveal a substantial developmental change in middle childhood. For ground truth inferences, a shift occurred between ages 5 and 7, driven primarily by improved understanding of the frown-good condition: younger children relied on speech content, whereas older children prioritized the emotional cue. This is consistent with prior work showing that younger children tend to rely on the literal meaning of speech, but gradually come to understand the prosocial motives behind white lies and increasingly use nonverbal cues to infer ground truth (Eskritt & Lee, 2003; Friend & Bryant, 2000; Heyman et al., 2009; Morton & Trehub, 2001; Xu et al., 2010). This shift is likely supported by developments in theory of mind (Broomfield et al., 2002; Perner & Wimmer, 1985; Sullivan et al., 1994). For communicative goal inferences, while we observed a similar progression toward adult-like performance between ages 5 and 6, more nuanced understanding continued to develop even beyond age 10. Specifically, although children aged 6 and older could use cue consistency to infer honest intent (Gillis & Nilsen, 2017; Rotenberg et al., 1989), they did not reliably differentiate within each consistency category, which required more sophisticated reasoning about the balance between informational and social goals. For example, smile-good could reflect either a desire to be truthful or to be kind, whereas frown-bad more uniquely signals truthfulness. Likewise, children aged 6 and older relied primarily on speech to infer social goals and did not further consider different emotional cues; unlike adults, they did not even show numerical trends outlined in Prediction 3. Together, these patterns suggest that children's inferences increasingly approximate adults' flexible, goal-based integration of speech and emotional cues, but that reasoning requiring fine-grained weighing of informational and social goals continues to develop beyond age 10.

This work suggests several future directions. One promising avenue is to examine how prior knowledge shapes social inferences. In our study, participants observed strangers with minimal contextual information, whereas in real life children often have prior experience with speakers that informs expectations about their goals. For example, knowledge about a speaker's past behavior—such as whether they are typically honest or people-pleasing, or whether this is their first time making the food—may influence inferences about ground truth as well as informational and social goals.

Another direction for future research is to vary properties of emotional expressions to test the generalizability of our findings. One approach is to manipulate the subtlety of expressions: exaggerated expressions may appear more intentional and signal a communicative goal, whereas subtler cues (e.g., a slight eyebrow raise) may be perceived as more automatic and reflective of underlying affect. Future work could examine whether people interpret subtle versus exaggerated cues differently, and how this sensitivity develops across childhood. Another approach is to move beyond posed expressions in controlled settings and test whether these findings generalize to more naturalistic expressions and contexts.

Cross-cultural comparisons represent another promising direction. Individuals from collectivistic cultures tend to suppress emotions more than those from individualistic cultures (Huwaë & Schaafsma, 2018; Kitayama & Markus, 1991), which may influence whether emotional expressions are perceived as automatic or communicative. In individualistic cultures, emotional expressions are likely to reflect underlying affect or signal informational goals, whereas in collectivistic cultures, emotion suppression to maintain social harmony may heighten sensitivity to social goals. Comparing cultures could therefore illuminate how children learn to interpret emotional expressions in multimodal communication and how cultural norms shape this development.

To conclude, we found that across middle childhood, children increasingly integrate multiple cues to make adult-like inferences, although subtle differences persist beyond age 10. Our results show that this development extends beyond sensitivity to individual cues, requiring flexible reasoning about both ground truth and speakers' underlying communicative goals. By revealing how this capacity emerges and where it remains constrained, our work advances developmental theories of social reasoning and underscores the importance of studying communication as an integrated, goal-driven process rather than as isolated cues.

Acknowledgments

We thank the members of the Affective Cognition Lab, our research assistant Zelda Gu, Children Helping Science, and all participating families. The first author is supported by the Faculty of Arts & Science Top Fellowship at the University of Toronto.

References

Brady, S. M., Ogren, M., & Johnson, S. P. (2024). Effects of conflicting emotional cues on toddlers' emotion perception. *British Journal of Developmental Psychology*, *42*(3), 376–391. https://doi.org/10.1111/bjdp.12501

Broomfield, K. A., Robinson, E. J., & Robinson, W. P. (2002). Children's understanding about white lies. *British Journal of Developmental Psychology*, *20*(1), 47–65. https://doi.org/10.1348/026151002166316

Brosseau-Liard, P., Cassels, T., & Birch, S. (2014). You seem certain but you were wrong before: Developmental change in preschoolers' relative trust in accurate versus confident speakers. *PloS one*, *9*(9), e108308. https://doi.org/10.1371/journal.pone.0108308

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge university press.

Cowen, A., Sauter, D., Tracy, J. L., & Keltner, D. (2019). Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression. *Psychological Science in the Public Interest*, *20*(1), 69–90. https://doi.org/10.1177/1529100619850176

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray.

Doan, T., Denison, S., Lucas, C. G., & Gopnik, A. (2015). Learning to reason about desires: An infant training study. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *37*. https://escholarship.org/uc/item/8xt657ph

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124. https://doi.org/10.1037/h0030377

Eskritt, M., & Lee, K. (2003). Do actions speak louder than words? Preschool children's use of the verbal-nonverbal consistency principle during inconsistent communications. *Journal of Nonverbal Behavior*, *27*(1), 25–41. https://doi.org/10.1023/A:1023614123081

Friend, M., & Bryant, J. B. (2000). A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly*, *46*(2), 342.

Gillis, R. L., & Nilsen, E. S. (2017). Consistency between verbal and non-verbal affective cues: A clue to speaker credibility. *Cognition and emotion*, *31*(4), 645–656. https://doi.org/10.1080/02699931.2016.1147422

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (pp. 43–58, Vol. 3). Academic Press.

Hayashi, H., & Shiomi, Y. (2015). Do children understand that people selectively conceal or express emotion? *International Journal of Behavioral Development*, *39*(1), 1–8. https://doi.org/10.1177/0165025414548777

Heyman, G. D., Sweet, M. A., & Lee, K. (2009). Children's reasoning about lie-telling and truth-telling in politeness contexts. *Social Development*, *18*(3), 728–746. https://doi.org/10.1111/j.1467-9507.2008.00495.x

Houlihan, S. D., Ong, D., Cusimano, M., & Saxe, R. (2022). Reasoning about the antecedents of emotions: Bayesian causal inference over an intuitive theory of mind. *Proceedings of the annual meeting of the cognitive science society*, *44*(44). https://escholarship.org/uc/item/7sn3w3n2

Huwaë, S., & Schaafsma, J. (2018). Cross-cultural differences in emotion suppression in everyday interactions. *International Journal of Psychology*, *53*(3), 176–183. https://doi.org/10.1002/ijop.12283

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*(2), 288–299. https://doi.org/10.1037/h0030377

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*(9), 757–758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Kitayama, S., & Markus, H. R. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological science*, *15*(10), 694–698. https://doi.org/10.1111/j.0956-7976.2004.00742.x

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of personality and social psychology*, *37*(9), 1539–1553. https://doi.org/10.1037/0022-3514.37.9.1539

Li, Q.-G., Heyman, G. D., Xu, F., & Lee, K. (2014). Young children's use of honesty as a basis for selective trust. *Journal of Experimental Child Psychology*, *117*, 59–72. https://doi.org/10.1016/j.jecp.2013.09.002

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child development*, *72*(3), 834–843. https://doi.org/10.1111/1467-8624.00318

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child devel-*

opment, *72*(3), 718–735. https://doi.org/10.1111/1467-8624.00311

Naito, M., & Seki, Y. (2009). The relationship between second-order false belief and display rules reasoning: The integration of cognitive and affective social understanding. *Developmental Science*, *12*(1), 150–164. https://doi.org/10.1111/j.1467-7687.2008.00748.x

Ong, D. C., Zaki, J., & Goodman, N. D. (2015). Affective cognition: Exploring lay theories of emotion. *Cognition*, *143*, 141–162. https://doi.org/10.1016/j.cognition.2015.06.010

Perner, J., & Wimmer, H. (1985). "john thinks that mary thinks that. . . " attribution of second-order beliefs by 5-to 10-year-old children. *Journal of experimental child psychology*, *39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Repacholi, B. M. (1998). Infants' use of attentional cues to identify the referent of another person's emotional expression. *Developmental psychology*, *34*(5), 1017. https://doi.org/10.1037/0012-1649.34.5.1017

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14-and 18-month-olds. *Developmental psychology*, *33*(1), 12. https://doi.org/10.1037/0012-1649.33.1.12

Repacholi, B. M., & Meltzoff, A. N. (2007). Emotional eavesdropping: Infants selectively respond to indirect emotional signals. *Child development*, *78*(2), 503–521.

Rotenberg, K. J., Simourd, L., & Moore, D. (1989). Children's use of a verbal-nonverbal consistency principle to infer truth and lying. *Child Development*, *60*(2), 309–322. https://doi.org/10.2307/1130978

Rothermich, K., Caivano, O., Knoll, L., & Talwar, V. (2020). Do they really mean it? Children's inference of speaker intentions and the role of age and gender. *Language and Speech*, *63*(4), 689–712. https://doi.org/10.1177/0023830919878742

Saxe, R., & Houlihan, S. D. (2017). Formalizing emotion concepts within a bayesian model of theory of mind. *Current opinion in Psychology*, *17*, 15–21. https://doi.org/10.1016/j.copsyc.2017.04.019

Scarantino, A. (2017). How to do things with emotional expressions: The theory of affective pragmatics. *Psychological Inquiry*, *28*(2-3), 165–185. https://doi.org/10.1080/1047840X.2017.1328951

Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind*, *1*(1), 4–14. https://doi.org/10.1162/OPMI_a_00002

Sorce, J. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental psychology*, *21*(1), 195. https://doi.org/10.1037/0012-1649.21.1.195

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental psychology*, *30*(3), 395–402. https://doi.org/10.1037/0012-1649.30.3.395

Vanderbilt, K. E., Liu, D., & Heyman, G. D. (2011). The development of distrust. *Child development*, *82*(5), 1372–1380. https://doi.org/10.1111/j.1467-8624.2011.01629.x

Vendetti, C., Kamawar, D., & Andrews, K. E. (2019). Theory of mind and preschoolers' understanding of misdeed and politeness lies. *Developmental psychology*, *55*(4), 823–834. https://doi.org/10.1037/dev0000666

Warneken, F., & Orlins, E. (2015). Children tell white lies to make others feel better. *British Journal of Developmental Psychology*, *33*(3), 259–270. https://doi.org/10.1111/bjdp.12083

Wu, Y., Baker, C. L., Tenenbaum, J. B., & Schulz, L. E. (2018). Rational inference of beliefs and desires from emotional expressions. *Cognitive science*, *42*(3), 850–884. https://doi.org/10.1111/cogs.12548

Wu, Y., & Gweon, H. (2021). Preschool-aged children jointly consider others' emotional expressions and prior knowledge to decide when to explore. *Child Development*, *92*(3), 862–870. https://doi.org/10.1111/cdev.13585

Wu, Y., & Schulz, L. E. (2020). Understanding social display rules: Using one person's emotional expressions to infer the desires of another. *Child development*, *91*(5), 1786–1799. https://doi.org/10.1111/cdev.13346

Wu, Y., Tessler, M. H., Asaba, M., Zhu, P., Gweon, H., & Frank, M. C. (2021). Integrating emotional expressions with utterances in pragmatic inference. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*, 2260–2266. https://escholarship.org/uc/item/1tw571n6

Xu, F., Bao, X., Fu, G., Talwar, V., & Lee, K. (2010). Lying and truth-telling in children: From concept to action. *Child development*, *81*(2), 581–596. https://doi.org/10.1111/j.1467-8624.2009.01417.x

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind*, *4*, 71–87. https://doi.org/10.1162/opmi_a_00035