Journal of Experimental Child Psychology 142 (2016) 391–399

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Preschoolers' real-time coordination of vocal and facial emotional information

Jared M.J. Berman [a], Craig G. Chambers [b], Susan A. Graham [a,*]

[a] Department of Psychology, University of Calgary, Calgary, Alberta T2N 1N4, Canada
[b] Department of Psychology, University of Toronto, Mississauga, Ontario L5L 1C6, Canada

ARTICLE INFO

Article history:
Available online 16 October 2015

Keywords:
Vocal affect
Eye-tracking
Facial emotion
Implicit processing
Emotion
Coordination

ABSTRACT

An eye-tracking methodology was used to examine the time course of 3- and 5-year-olds' ability to link speech bearing different acoustic cues to emotion (i.e., happy-sounding, neutral, and sad-sounding intonation) to photographs of faces reflecting different emotional expressions. Analyses of saccadic eye movement patterns indicated that, for both 3- and 5-year-olds, sad-sounding speech triggered gaze shifts to a matching (sad-looking) face from the earliest moments of speech processing. However, it was not until approximately 800 ms into a happy-sounding utterance that preschoolers began to use the emotional cues from speech to identify a matching (happy-looking) face. Complementary analyses based on conscious/controlled behaviors (children's explicit points toward the faces) indicated that 5-year-olds, but not 3-year-olds, could successfully match happy-sounding and sad-sounding vocal affect to a corresponding emotional face. Together, the findings clarify developmental patterns in preschoolers' implicit versus explicit ability to coordinate emotional cues across modalities and highlight preschoolers' greater sensitivity to sad-sounding speech as the auditory signal unfolds in time.

Introduction

Detecting and recognizing another's emotion is critical to communicative success. Consider, for example, the utterance "It's snowing!" The meaning of this statement varies considerably depending

* Corresponding author.
E-mail address: susan.graham@ucalgary.ca (S.A. Graham).

http://dx.doi.org/10.1016/j.jecp.2015.09.014

upon whether the speaker is excited (perhaps a skier) or upset (perhaps a stranded commuter). In many cases, the emotional disposition of the speaker is communicated alongside the linguistic information in the speech stream via nonlinguistic modulations of the voice (e.g., variations in pitch level, pitch contours, and speech rate; see Banse & Scherer, 1996; Frick, 1985), which together are often referred to as "emotional prosody" or "vocal affect." Although there is a growing literature examining children's attention to vocal affect (e.g., Berman, Chambers, & Graham, 2010; Quam & Swingley, 2012; Sauter, Panattoni, & Happé, 2013), the question of how children at different ages dynamically coordinate vocal affect with the other perceptual information source that commonly signals an individual's emotional disposition, namely facial expression, is not well understood. Here, we examined 3- and 5-year-olds' ability to link vocal affect in unfolding speech with a corresponding emotional face, focusing on differences related to emotional valence and the time course of information processing. Our assessment of this ability is based on implicit/unconscious physiological measures (namely saccadic eye movements to displayed faces) as well as overt/explicit measures (children's points toward what they consciously judge to be the "matching" face for an unfolding utterance).

The building blocks for children's effectiveness and flexibility in coordinating the cues to emotion that are provided by different sensory–perceptual modalities are evident early in development (see Walker-Andrews, 1997, for a review). For example, 3-month-olds will match vocal affect with one of two dynamic emotional faces, but only when the individual depicted is their mother (Kahana-Kalman & Walker-Andrews, 2001). By 7 months of age, however, infants can use unfamiliar faces and voices to discriminate across emotional categories (Flom & Bahrick, 2007; Walker-Andrews, 1986). Furthermore, event-related potential (ERP) studies have found patterns supporting the view that there is more consistent discrimination of and recognition for multimodal versus unimodal cues to emotional state prior to 7 months of age (Grossmann, Striano, & Friederici, 2006, 2007).

When cues to emotion are conveyed via a single modality (face or voice), we see a gradual development of emotion recognition abilities over the childhood years. For example, although neonates can discriminate between different facial expressions (Field, Woodson, Greenberg, & Cohen, 1982), children's ability to associate the label *happy* with happy facial expressions and the label *sad* with sad facial expressions emerges just before 4 years of age. The ability to identify other facial emotions, such as fear and surprise, develops more slowly during childhood (Widen, 2013). In the auditory domain, children can accurately judge a speaker's emotional state based on vocal affect by age 4 if the emotional content of the speech itself is either neutral or filtered out (e.g., Morton & Trehub, 2001; Nelson & Russell, 2011). However, when there are competing cues regarding the relevant emotion conveyed via either the meaning of a sentence (e.g., Morton & Trehub, 2001) or the situational context (Gil, Aguert, Le Bigot, Lacroix, & Laval, 2014), preschoolers' ability to recognize vocal affect is compromised. Research has also shown that preschoolers' success at identifying vocal affect varies across emotional categories, with greater success in recognizing sad vocal affect compared with happiness or fear (Nelson & Russell, 2011), and that the ability to recognize emotion categories from auditory information continues to develop well beyond the age of 5 years (Sauter et al., 2013).

Our goal in the current study was to examine aspects of the multimodal coordination of auditory and visual cues to emotion during the preschool years. However, rather than focusing on the added benefit of multimodal input or the integrative processing of simultaneously perceived facial and speech-based cues, we instead considered how cues provided by one modality (speech information) allow a child to identify a relevant percept in another modality (an emotional face). This "orienting" ability is a core component of emotion recognition in many day-to-day scenarios involving a context with multiple individuals and where the perceiver is not already focused on an individual producing emotional cues via facial expression or vocal affect. This arises in part because auditory information can be easily perceived without directed attention due to the way in which sound travels. As a result, vocal affect is often detected and interpreted by a perceiver first, leading the perceiver to redirect visual attention to establish the identity of the "source" individual.

An important aspect of our study is the inclusion of both *implicit* and *explicit* measures. Much of the research on vocal emotion has focused primarily on young children's detection and recognition of emotional cues using explicit measures that are less able to capture the dynamic nature of the auditory signals that convey emotion, namely the fact that spoken utterances unfold in time. This issue is essential to consider in view of evidence from studies of adults demonstrating that the time required to recognize vocally conveyed emotions varies according to emotional valence. For example, Pell and Kotz (2011) found that, on average, it took adults approximately 400 ms longer to recognize happy-sounding speech than sad-sounding speech. Similarly, 5-year-olds appear to be slower to use sad-sounding speech to identify the objects to which a speaker might be referring (Berman, Graham, & Chambers, 2013b). Importantly, differences of even less than a second are relevant for various aspects of performance in linguistic communication. For example, given a normal speaking rate in English, the duration of the average word is less than a second, as is the typical interval between the "turns" of speakers engaged in a conversation (e.g., Casillas & Frank, 2012). One important question on the issue of timing is whether preschoolers' slower reactions to happy-sounding speech cues in referential tasks can be attributed to timing differences in the uptake of the relevant prosodic cues across emotion categories (mirroring the findings of Pell & Kotz, 2011) or whether this pattern simply reflects greater difficulty in linking a referential object with vocal affect cues in view of the different task used. In the latter case, the explanation may rest on the fact that it could be more obvious to link sad-sounding speech to things like broken objects than it is to link happy-sounding speech to intact or entertaining objects. However, if preschoolers are comparatively slower to use happy-sounding speech to identify a visual percept that has an *inherent* rather than indirectly inferred emotional valence (e.g., a happy face), it is more plausible that timing factors in the perceptual uptake of different vocal affect cues are responsible.

In view of our research question and the fact that the time course of processing is an important consideration, we use real-time eye gaze behavior to explore potential valence and timing effects in children's ability to map emotion in the voice to faces. In contrast to intentionally controlled behaviors such as pointing and verbal responses, eye movements are known to reflect moment-to-moment aspects of processing as speech unfolds in time (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), in part because of their unconscious nature and low metabolic demands (e.g., Abrams & Jonides, 1988). Furthermore, in studies of children's language processing, eye movement measures are known to capture effects that are not reflected in children's explicit behavioral responses (e.g., Fernald, Zangl, Portillo, & Marchman, 2008). To assess the extent to which implicit sensitivities reflected in gaze behavior are present or absent in children's conscious and controlled processing, we also measured children's overt judgments about the face that matches vocal affect in speech, as assessed by the image to which they point.

Our procedure was similar to that used by Paulmann, Titone, and Pell (2012) with adults and was adapted to test whether 3- and 5-year-olds could match sentences recorded with happy-sounding, sad-sounding, or neutral vocal affect to one of three photos (happy, sad, or neutral) of a given individual. We focused on 3- and 5-year-olds because previous research has demonstrated significant progress in children's ability to use vocal affect during this developmental period (e.g., Berman, Graham, Callaway, & Chambers, 2013a; Berman et al., 2010; Quam & Swingley, 2012). The content of the sentences presented was emotionally neutral, and each consisted of an instruction to look at an absent object (e.g., "Look at the doll/mido"). The sentences were drawn from previous studies of vocal affect in children (Berman et al., 2010, 2013a) and had been previously validated as reflecting the relevant vocal affect cues. Furthermore, the sentences contained either novel (nonce) nouns or familiar nouns to control for potentially inherent valance biases (e.g., the noun *doll* might plausibly convey positive associations on its own). For each utterance, children were explicitly asked to *consider what the speaker's face looked like as she was speaking*. Thus, the task required children to envisage the speaker on the basis of the vocal affect cues rather than make explicit judgments about the speaker's emotional state.

Our predictions were as follows. First, we expected that both 3- and 5-year-olds' ability to match vocal affect to visual information across emotions would be evidenced by their visual attention to the correct face, as reflected in the eye gaze record. Second, in light of previous research demonstrating that the preschool years are a time of significant developmental change in children's explicit sensitivity to vocal affect, we expected that 5-year-olds, but not 3-year-olds, would be successful in consciously/overtly identifying the face that matched the vocal affect. Finally, if the timing effects observed in earlier studies (e.g., Berman et al., 2013b) are due to differences in the use of affect cues across the happy/sad categories (rather than factors related to the identification of referential objects), we would predict that both 3- and 5-year-olds, like adults, would be significantly faster at recognizing faces corresponding to sad-sounding vocal affect than to positive-sounding vocal affect.

Method

Participants

The final sample consisted of 32 English-speaking preschoolers: 16 3-year-olds (9 boys, $M = 3.52$ years, $SD = 0.26$) and 16 5-year-olds (8 boys, $M = 5.64$ years, $SD = 0.29$). Data from an additional 5 5-year-olds and 12 3-year-olds were excluded from analyses due to insufficient gaze data resulting from excessive head or body movements.

Stimuli

On each trial, the same three photographs of the face of a young woman were presented on a large display screen. The three photographs varied in terms of the emotional expression portrayed (happy, sad, or neutral) and were drawn from the Karolinska Directed Emotional Faces database (Lundqvist, Flykt, & Öhman, 1998). The relative position of each emotional face in the display was systematically varied across trials.

The auditory stimuli consisted of 24 prerecorded sentences of the format, "Look at the X. Point to the X," where X was either a familiar or unfamiliar noun (e.g., *doll* vs. *mido*). Three versions of each utterance were recorded by a female native English speaker, varying only in the emotional affect expressed by the speaker's voice. The *happy* and *sad* versions were recorded using distinctly happy-sounding and sad-sounding voices, whereas the *neutral* sentences were recorded with neutral speech intonation. The auditory stimuli were drawn from earlier studies in which the relevant affect differences had been statistically validated using an independent sample of adults (Berman et al., 2010, 2013a). For a given participant, each sentence occurred in only one of the three vocal affect types, and the pairing of vocal affect to sentences was counterbalanced across participants.

Apparatus

Children's eye movements were monitored using a Tobii ×50 eye-tracker located below a 46-inch computer monitor. The ×50 is accurate between 0.5 and 0.7 degrees of visual angle with a sampling rate of 50 Hz. We identified three areas of interest (AOIs) for each trial, defined by the boundaries of the three faces, to establish the screen region the child was fixating at successive time points. The recording software logged gaze data every 20 ms, and a fixation was registered if the child gazed at the same AOI for more than 100 ms (the default fixation filter value in Tobii analysis tools, including ClearView and E-Prime Tobii extensions). A set of speakers located on each side of the monitor played the auditory stimuli. To record pointing behavior, a high-definition (HD) camera was positioned behind the child.

Procedure

Each child was seated on a small chair facing a 46-inch computer monitor. The child's eye gaze was calibrated using ClearView software. Only data from those children who showed accurate calibration on three of five test fixation points were included, although for 69% of children calibration was accurate for all five fixation points.

After calibration, each child was familiarized with the photographs of emotional faces (i.e., happy, sad, and neutral). For this phase, each face was presented separately, accompanied by a recorded utterance identifying the relevant emotion (e.g., "This is what I look like when I feel happy"). Following this familiarization phase, the child was presented with 24 trials (8 for each vocal affect type) and was told, "Listen to what [the speaker] says and then point to what her face looked like when she said it." On each trial, the three faces were presented on the screen simultaneously while the recorded sentence played. The faces remained on-screen until the child pointed at one of them.

Results

Eye fixation patterns

For the purpose of analysis, we examined eye movements beginning 200 ms following utterance onset. The 200-ms margin reflects the time interval required for the planning and execution of saccadic eye movements in reaction to a stimulus (cf. Hallett, 1986). Fig. 1 represents the probability of fixating the different faces beginning at this margin and continuing for 1600 ms (the average utterance duration was 1586 ms). Because fixation patterns did not differ according to age group, gender, or whether the instructions contained novel or familiar labels (see below), the data are shown collapsed across these factors. Qualitatively, the results for the happy-sounding vocal affect condition (top panel) show a preference to consider the happy face beginning to emerge approximately 600 ms after the onset of the utterance. This preference continued in the remainder of the analysis window. In the neutral affect condition (middle panel), the fixation profile suggests more or less equal consideration of all three faces. Finally, when the utterances were spoken with sad-sounding vocal affect (bottom panel), children displayed an increasing preference to fixate the sad face beginning approximately 200 ms after the onset of the utterance, and this preference remained until the 800-ms mark.

To evaluate these patterns statistically, we first calculated a difference score by subtracting looks to the happy face from looks to the sad face at each time point for each participant. The exclusion of the neutral face from this score is motivated by children's apparent reluctance to consider this face in their

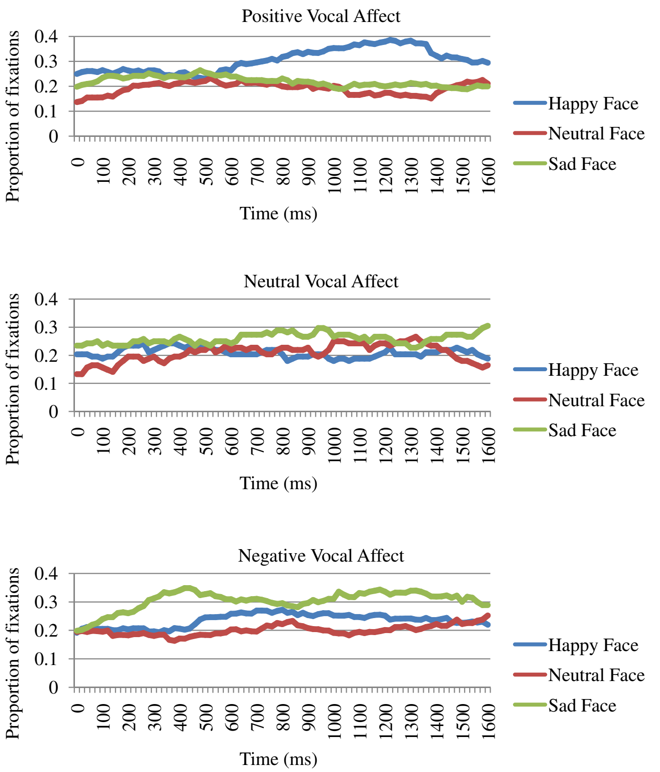

Fig. 1. Proportion of fixations to faces over the course of the utterance, collapsed across age.

pointing behavior (see next section) and because the resulting measure will provide a more sensitive index of potential valence differences in the analyses that follow. This measure provides a straightforward test of the extent to which happy-sounding and sad-sounding vocal affect cues are mapped to faces as the auditory information unfolds in time. Using averaged difference scores, timing effects were examined using four consecutive 400-ms intervals within the 1600-ms window to provide a means to statistically distinguish early-occurring versus late-occurring effects. We conducted a separate analysis of variance (ANOVA) for each interval with affect type as a within-participant factor and age group as a between-participants factor. For the final time interval, we added noun type (familiar or novel) into the ANOVA model because the noun is produced only at this point in the speech stream. For all of the first three time intervals, there was a main effect of vocal affect across conditions (all Fs > 3.4, all ps < .001), consistent with the different patterns of voice-to-face mapping apparent in the figures, and there was no effect of age group. In the fourth interval, there was no significant effect of affect, noun type, or age group (ps > .10).

To explore the apparent valence differences in terms of timing, we conducted follow-up comparisons against chance behavior (i.e., a difference score of 0) for each time interval. These analyses showed that in the sad-sounding affect condition, children looked more to the sad face than the happy face during the first two 400-ms time intervals (ps < .01), whereas in the same time windows in the happy-sounding affect condition, the tendency to consider the happy face did not differ from chance (ps > .18). During the second half of the utterance (i.e., 800–1200 and 1200–1600 ms), children looked more to the happy face than the sad face in the happy-sounding affect condition (ps < .01). However, children's looks to the sad face in the sad-sounding affect condition did not differ significantly from chance in these time windows (ps > .36). Finally, when the utterances were produced with neutral vocal affect, there were no significant differences in the attention deployed to happy versus sad faces in any time window (ps > .19).

To summarize, the eye fixation data indicate that both 3- and 5-year-olds use happy-sounding and sad-sounding vocal affect cues to locate the matching emotional face. However, children use the sad-sounding vocal affect significantly earlier in the utterance to find the corresponding face. The matching of happy-sounding affect with a compatible face was not apparent until the final portion of the utterance (which again was approximately 1600 ms in duration). Interestingly, neutral vocal affect led to consideration of all three faces for both age groups.

Pointing behaviors

Children's pointing behavior was coded as a measure of their explicit ability to judge the emotional face that produced the accompanying speech. Points were coded from videotape by assistants who were unaware of the experimental hypotheses and the specific vocal affect presented on a given trial. Inter-rater reliability for 20% of the data ($n = 6$) was perfect (Cohen's kappa = 1.00, $p < .001$).

We calculated the percentage of points to the face matching the vocal affect in the accompanying speech. Fig. 2 shows the percentage of points to the target face as a function of vocal affect type and age group because preliminary analyses revealed no effect of noun type. Data were submitted to an ANOVA with affect type (happy, neutral, or sad) as a within-participant factor and age (3-year-olds or 5-year-olds) as a between-participants factor. The analysis revealed a significant Age × Affect Type interaction, $F(2,60) = 18.59$, $p < .01$, $\eta_p^2 = .38$. Planned comparisons indicated that 5-year-olds were more likely to point to the correct face in the happy-sounding and sad-sounding affect conditions relative to the neutral condition, $t(15) = 7.78$, $p < .001$, $d = 2.11$ and $t(15) = 7.05$, $p < .001$, $d = 2.20$, respectively, and there was no difference in correct selections across the happy-sounding and sad-sounding affect conditions ($p = .89$). Note that 5-year-olds' selection of the neutral face in the neutral condition corresponded to chance-level behavior ($M = 31\%$, $SE = .043$, where chance is 33%), but their correct selections of the happy and sad faces in the corresponding happy and sad speech conditions were significantly greater than chance (ps < .001). Unlike 5-year-olds, 3-year-olds did not match vocal affect to the correct face at above-chance levels in any condition (ps > .40). Neither gender nor noun type entered into any significant effects.

To summarize, the pointing data indicate that 5-year-olds, but not 3-year-olds, appropriately link a speaker's vocal affect to the corresponding emotional face, although the effect for 5-year-olds is clearly

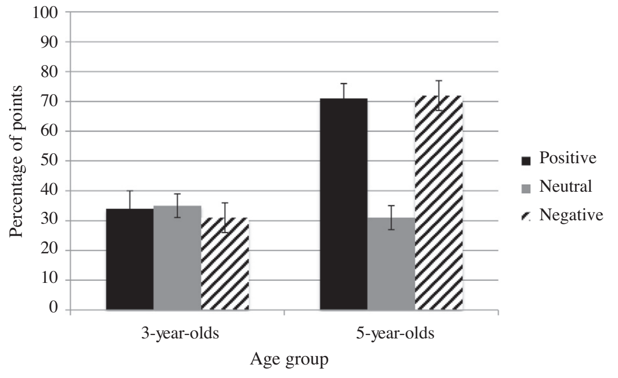

Fig. 2. Percentage of points to corresponding face as a function of vocal affect type.

carried by performance in the happy-sounding and sad-sounding speech conditions. With neutral-sounding speech, there was no apparent tendency to select the neutral face. Furthermore, although 3-year-olds showed a latent ability to map vocal affect to a corresponding face (as shown in unconscious gaze behavior), their pointing behaviors suggest that these mapping abilities are still at an emergent stage at this age; the conscious selection of a matching face did not reflect the accurate use of vocal affect information, consistent with earlier work using other kinds of tasks (e.g., Berman et al., 2010). In contrast, 5-year-olds successfully used positive and negative vocal affect to explicitly identify the corresponding face via pointing, reflecting a greater congruence between implicit processing and explicit judgments at this age.

Discussion

The results of this experiment offer new insights into preschoolers' ability to link auditory and visual cues to a speaker's emotional disposition. First, we found that children as young as 3 years can match a speaker's vocal affect to her corresponding facial affect, as evidenced through their gaze behaviors. As content-neutral utterances unfolded, preschoolers used the nonlinguistic affect cues in the speech stream to locate the happy face in the happy-sounding affect trials and the sad face in the sad-sounding affect trials. This is an impressive outcome given that children of this age and older often show difficulties in using vocal affect in other kinds of tasks used to evaluate sensitivity to vocal affect (e.g., Berman et al., 2010; Sauter et al., 2013). The different experimental scenario used in the current study is likely one reason for this difference. For example, in Berman and colleagues' (2010) study, the task required 3-year-olds to link a speaker's vocal affect with her intention to refer to objects that were depicted as either intact or broken. The current results, therefore, suggest that 3-year-olds' difficulties are isolated to the process of linking vocal affect with the intent to refer to objects and do not simply reflect an inability to correctly recognize and classify a speaker's emotional disposition.

Second, our results highlight the challenge that young children face in interpreting neutral emotions; neither the 3-year-olds nor the 5-year-olds showed evidence of matching neutral vocal affect with the neutral face. This finding is consistent with the proposal that children's emotion categories are acquired gradually, first interpreted broadly in terms of valence and then differentiated into specific categories (Widen, 2013). Neutral emotions, then, are understandably more challenging for children to interpret because they are not easily classified in terms of valence or salient emotion categories.

Third, our results add to the growing literature that highlights differences in the sensitivity to negative versus positive affective information and the specific ways in which this sensitivity is linked to the time course of implicit perceptual processing. Although valence differences were not apparent in children's overt decisions about the match between an emotional face and emotional paralanguage, they were reflected in less conscious behaviors, namely eye movements. Specifically, although the gaze patterns of both 3- and 5-year-olds showed that sad-sounding vocal affect was used very quickly to identify a sad face, it was not until approximately 800 ms into the utterance that a comparable matching effect was found with happy-sounding speech. As noted earlier, this difference in the timing of recognition corresponds to a nontrivial interval of time in the course of real-time language processing. Furthermore, this valence difference in the mapping of voice information to faces is consistent with other findings suggesting that children and adults are faster to detect negative vocal affect than positive vocal affect (e.g., Berman et al., 2013a; Paulmann et al., 2012; Rigoulot & Pell, 2012). The finding that preschoolers detect negative vocal affect more rapidly than positive vocal affect may be broadly consistent with the well-documented finding of a negativity bias in children's social–cognitive processing. In particular, Vaish, Grossmann, and Woodward (2008) proposed that children differentially attend to negative versus positive information when making sense of the world around them. Furthermore, negative stimuli have been hypothesized to carry a greater amount of information, which may require a greater amount of processing (Peeters & Czapinski, 1990) and, therefore, may benefit from more attentional resources.

A final insight provided by our results concerns the developmental differences in children's sensitivity to vocal affect and the critical importance of including implicit measures. Although 5-year-olds' overt pointing decisions and their eye movements reflected their ability to match vocal affect with the correct emotional face, 3-year-olds reflected this ability only in their eye movements. What might account for this discrepancy between the 3-year-olds' implicit and explicit behaviors? One explanation is that the conscious integration of vocal information with the matching facial expression necessary to enact an explicit behavioral response may have overwhelmed 3-year-olds' cognitive resources. That is, pointing requires children to program and then enact complex motor movements to signal a conscious choice. In contrast, information-seeking behaviors such as gaze shifts are unconscious and have lower metabolic demands (e.g., Abrams & Jonides, 1988).

A related explanation is that this dissociation reflects an emergent ability to coordinate vocal affect with emotional faces in 3-year-olds in contrast to the more developed abilities of 5-year-olds. Indeed, the developmental trajectory seen in our results is consistent with other recent research demonstrating significant 3- and 5-year-olds' abilities to use vocal affect to make referential decisions (Berman et al., 2010, 2013a, 2013b; Quam & Swingley, 2012). That is, the implicit effects observed in 3-year-olds may be best understood as a precursor to the more sophisticated and explicit ways in which emotional cues are perceived and used by older children and adults (Berman et al., 2013a, 2013b; Morton & Trehub, 2001). This proposal is consistent with the finding that many emerging cognitive abilities can be detected earlier in development with implicit measures than explicit measures, including resolving spatial ambiguity (Plumert, 1996), false belief understanding (Garnham & Ruffman, 2001; Onishi & Baillargeon, 2005), problem-solving strategies (Goldin-Meadow, 2003; Siegler, 2000; Winsler & Naglieri, 2003), object physics (Hood, Cole-Davis, & Dias, 2003), and communicative perspective-taking (Nilsen & Graham, 2012; Nilsen, Graham, Smith, & Chambers, 2008).

Acknowledgments

This work was supported by funds from the Canada Foundation for Innovation, the Canada Research Chairs program, and the University of Calgary awarded to S.G. and by an operating grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) awarded to S.G. and C.C. J.B. was supported by a graduate fellowship from SSHRC and funding from Alberta Innovates - Health Solutions. We thank Heather MacKenzie, Vanessa Schell, and Charlene Parker for their assistance with this research. Some of the data presented were included in J.B.'s dissertation submitted to the University of Calgary.

References

Abrams, R. A., & Jonides, J. (1988). Programming saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 428–443.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.

Berman, J. M. J., Chambers, C., & Graham, S. A. (2010). Preschoolers' appreciation of speaker vocal affect as a cue to referential intent. *Journal of Experimental Child Psychology, 107*, 87–99.

Berman, J. M., Graham, S. A., Callaway, D., & Chambers, C. G. (2013a). Preschoolers use emotion in speech to learn new words. *Child Development, 84*, 1791–1805.
Berman, J. M., Graham, S. A., & Chambers, C. G. (2013b). Contextual influences on children's use of vocal affect cues during referential interpretation. *Quarterly Journal of Experimental Psychology, 66*, 705–726.
Casillas, M., & Frank, M. C. (2012). Cues to turn boundary prediction in adults and preschoolers. In S. Brown-Schmidt, J. Ginzburg, & S. Larsson (Eds.), *Proceedings of SemDial 2012 (SeineDial): The 16th workshop on the semantics and pragmatics of dialogue* (pp. 61–69). Paris: Université Paris-Diderot.
Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). Amsterdam: John Benjamins.
Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expression by neonates. *Science, 218*, 179–181.
Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology, 43*, 238–252.
Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97*, 412–429.
Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding or belief? *Developmental Science, 4*, 94–100.
Gil, S., Aguert, M., Le Bigot, L., Lacroix, A., & Laval, V. (2014). Children's understanding of others' emotional states: Inferences from extralinguistic or paralinguistic cues? *International Journal of Behavioral Development*. http://dx.doi.org/10.1177/0165025414535123.
Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.
Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science, 9*, 309–315.
Grossmann, T., Striano, T., & Friederici, A. D. (2007). Developmental changes in infants' processing of happy and angry facial expressions: A neurobehavioral study. *Brain and Cognition, 64*, 30–41.
Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York: John Wiley.
Hood, B. M., Cole-Davis, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental Psychology, 39*, 61–70.
Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development, 72*, 352–369.
Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces–KDEF [CD–ROM]*. Stockholm, Sweden: Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet.
Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development, 72*, 834–843.
Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology, 110*, 52–61.
Nilsen, E. S., & Graham, S. A. (2012). The development of preschoolers' appreciation of communicative ambiguity. *Child Development, 83*, 1400–1415.
Nilsen, E. S., Graham, S. A., Smith, S., & Chambers, C. G. (2008). Preschoolers' sensitivity to referential ambiguity: Evidence for a dissociation between implicit understanding and explicit behavior. *Developmental Science, 11*, 556–562.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.
Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: Evidence from eye movements. *Speech Communication, 54*, 92–107.
Peeters, G., & Czapinski, J. (1990). Positive–negative asymmetry in evaluations: The distinction between affective and informational negativity effects. In W. Stroebe & M. Hewstone (Eds.). *European review of social psychology* (Vol. 1, pp. 33–60). Chichester, UK: Wiley.
Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE, 6*(11), e27256.
Plumert, J. M. (1996). Young children's ability to detect ambiguity in descriptions of location. *Cognitive Development, 11*, 375–396.
Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development, 83*, 236–250.
Rigoulot, S., & Pell, M. D. (2012). Seeing emotion with your ears: Emotional prosody implicitly guides visual attention to faces. *PLoS ONE, 7*(1), e30740.
Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31*, 97–113.
Siegler, R. S. (2000). The rebirth of children's learning. *Child Development, 71*, 26–35.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.
Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social–emotional development. *Psychological Bulletin, 134*, 383–403.
Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology, 22*, 373–377.
Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*, 437–456.
Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review, 5*, 72–77.
Winsler, A., & Naglieri, J. (2003). Overt and covert verbal problem-solving strategies: Developmental trends in use, awareness, and relations with task performance in children aged 5 to 17. *Child Development, 74*, 659–678.