Metacognition of Emotion Recognition

Karen J. Kelly

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013
ABSTRACT

Metacognition of Emotion Recognition

Karen J. Kelly

Are people able to determine when they are correct or incorrect in their interpretation of another’s emotional state? This question of whether or not individuals are capable of making accurate judgments about this ability was briefly explored in a handful of studies that concluded that individuals could not make such judgments. This finding did not seem to be consistent with our high-level social abilities. It is difficult to image that individuals are capable of fluently moving though social interactions, emotional exchanges, and interpersonal relationships absent any ability to determine if they are indeed correctly interpreting other’s emotions. In an effort to revisit this question it was necessary to take a deeper look at the methodology used in the original studies. The procedure used to establish metacognitive accuracy, although not incorrect, was not the appropriate choice. Instead of relying on the global measures of metacognition that previous research used, we shifted the focus to relative measures of metacognition that allow individuals to make item-by-item decisions about their perceived accuracy on each stimulus. This methodology has been used in studies involving both static (posed facial expressions and cartoon images) and dynamic (body gait and verbal prosody) stimuli. In each experiment, for each type of stimulus, individuals are able to distinguish those items that they know from those that they do not know – demonstrating metacognition of emotion recognition. This knowledge is not limited to adults, but appears to be developing in the 3rd grade and fully developed by the 5th grade. These findings are discussed with respect to the importance of emotion
recognition in social interactions, the variety of cues that might be useful during the process of emotion recognition, and cognitive development in general.
Table of Contents

Lists of Charts and Graphs……………………………………………………………………..ii
Acknowledgments………………………………………………………………………………iii
Dedication………………………………………………………………………………………iv
Preface………………………………………………………………………………………….v
Chapter 1: Metacognition of emotional face recognition……………………………………1
Chapter 2: Metacognition of dynamical emotion recognition……………………………31
Chapter 3: Metacognition of emotion recognition in children……………………………52
Chapter 4: Conclusions and future directions………………………………………………63
References……………………………………………………………………………………66
Appendices……………………………………………………………………………………73
Lists of Charts and Graphs

I. Chapter 1: Metacognition of emotional face recognition
   
a) Figure 1: Global Pearson and relative gamma correlations for Experiment 1.....15
b) Figure 2: Global Pearson and relative gamma correlations for Experiment 2.....24
c) Table 1: Ekman morphs - Performance according to item difficulty...............25

II. Chapter 2: Metacognition of dynamical emotion recognition
   
a) Figure 3: Relative gamma correlations for Experiments 3 and 4..................48

III. Chapter 3: Metacognition of emotion recognition in children
   
a) Figure 4: Task performance across groups: 3rd, 5th, and CU students..........58
b) Figure 5: Task accuracy plotted against metacognitive accuracy across groups...59
Acknowledgments

This research was supported, in part, by Grant 220020166 from the James S. McDonnell Foundation. I would like to thank Natalie Sebanz, Rachel Burris, Barbie Huelser, Lisa Son, Bridgid Finn, Jamil Zaki, Patrick Kennedy, Maggie Shiffrar, John Franchak, and David Miele. I would also like to acknowledge the support of my committee: Lisa Son, Kevin Ochsner, Niall Bolger, and Marianne Lloyd. Most importantly, I would like to thank Janet Metcalfe for her support, contributions, knowledge, mentoring, and dedication to this project. Her guidance and encouragement throughout my time at Columbia University have been endless.
Dedication

To those in my life who have supported and inspired me throughout the process of completing my degree, I thank my father, David Kelly, and my Uncle Eddie Kelly who served as inspirations to me during this process and, through their own scholarly achievements, set a wonderful example for me. I would like to further acknowledge and thank my father, my husband, my siblings (both by blood and by good fortune), and my friends for their continuous support. Finally, I would like to thank my mother, for everything.
Preface

While humans are adept at recognizing emotional states conveyed by facial expressions, the current literature suggests that they lack accurate metacognitions about their performance in this domain. This conclusion concerning metacognition within the emotional domain, though, comes from global trait-based questionnaires that assess the extent to which an individual perceives him or herself as empathic, as compared to other people. Those who rate themselves as empathically accurate are no better than others at recognizing emotions. Metacognition of emotion recognition can also be assessed using relative measures that evaluate how well a person thinks s/he has understood the emotion in a particular facial display as compared to other displays. While this is the most common method of metacognitive assessment of people's judgments of learning or their feelings of knowing, this kind of metacognition-- 'relative meta-accuracy'-- has not been studied within the domain of emotion.

In chapter one, as well as asking for global metacognitive judgments, people were asked to provide relative, trial-by-trial prospective and retrospective judgments concerning whether they would be right or wrong in recognizing the expressions conveyed in particular facial displays. The question was: Do people know when they will be correct in knowing what expression is conveyed, and do they know when they do not know? Although the results replicated the finding that that global meta-accuracy was unpredictable of performance, relative meta-accuracy, given by the correlation between participants' trial-by-trial metacognitive judgments and performance on each item, were significantly above zero on both the 'Mind in the Eyes' task (Experiment 1) and on the 'Ekman Emotional Expression Multimorph' task (in Experiment 2).
In chapter two this finding was expanded to include metacognition of emotion recognition using dynamic stimuli: emotion expressed in body language and spoken language. In Experiment 3, participants viewed video images of point light walkers expressing basic emotions. In Experiment 4, participants listened to audio clips of sentences with neutral content but spoken with a specific emotional prosody. After each video or audio clip the participants provided trial-by-trial predictions about how well they would be able to recognize the emotion when subsequently given a multiple-choice test. Participants later saw or heard the same clip again and chose which emotion they thought the actors were expressing followed by a retrospective confidence judgment. Participants made prospective metacognitive judgments as well as retrospective metacognitive judgments that were significantly above zero. These findings are discussed with respect to metacognition as an important component of social cognition.

In chapter three, the ability of children to make metacognitive judgments about emotion recognition was investigated. Additionally, a set of age appropriate stimuli was created for use with children and were in need of proper validation. The stimuli consisted of still images taken from well-known cartoon or animated movies that were normed using Columbia University students. As with adults, 3rd and 5th graders were able to make judgments about their ability to recognize emotional expressions that were significantly above zero. While 3rd and 5th graders and Columbia students performed equally well on the task, 3rd graders showed a deficit in their metacognition in that they were not as good at identifying expressions that they would correctly recognize in the future. Fifth graders and Columbia students did not show any metacognitive differences.
These findings are discussed within the domain of children’s metacognition as well as deficits in emotion recognition that contribute to social dysfunction.
Chapter 1

Metacognition of Emotional Face Recognition
Metacognition of Emotion Recognition

This chapter addresses the question of whether people know when they understand others’ emotional expressions. Metacognition--knowing when one knows and when one does not know-- is a skill that is both separate from the more basic level emotion recognition itself, and that has the potential to confer a number of benefits. If one is accurate in knowing that one knows, for certain, what another person is feeling, then it is not necessary to devote time and effort to further determining what the other person's emotion is. One can act, or choose not to act, on one's knowledge of the other person's emotional state directly. In contrast, there is a distinct value to being able to correctly assess that one does not know what the other person's emotional expression is conveying. Rather than acting precipitously, an individual--with good metacognitions--can, when they know that they don't know, take the time to further explore the person's expression or directly ask the person what he or she is feeling, rather than jumping to a wrong conclusion that could have negative consequences for the relationship.

Appropriate uncertainty about one's own lack of understanding can be valuable. Furthermore, one of the well-known benefits of accurate metacognition, in any domain, is that it can be used to promote learning. The individual can choose to allocate attention and time appropriately (see, Finn, 2008; Metcalfe & Finn, 2008; Thiede, Anderson, & Therriault, 2003), and learn to hone their emotion recognition skills effectively. If an individual does not know what they don't know (i.e., they have poor metacognitions) they are not in a position to remedy the situation (Benjamin, Bjork, & Schwartz, 1998). As in other domains, then, meta-accuracy in the domain of emotional recognition should be
helpful in promoting task accuracy, minimizing error, ensuring flexibility in a dynamic social context, and in promoting learning over time (Nelson & Narens, 1990).

Metacognition is broadly defined as the capacity to actively monitor and reflect upon one’s own performance and abilities (Dunlosky & Metcalfe, 2009; Flavell, 1979). It can be measured in several ways, however, and these methods of assessing metacognition address different questions and allow different inferences. The global meta-accuracy and relative meta-accuracy of one's judgments about one's knowledge are the two most important and prominent measures used in the field of metacognition. Global measures usually use a Pearson correlation to calculate the relationship between people's responses on questionnaires about their abilities, preferences, predispositions, or, in this case, emotional sensitivity, and their ultimate overall performance on a relevant task, as compared to that of other people. They require the participant to assess his or her performance in a domain, in comparison with other members of the sample, and also to provide an objective measure--usually by taking a test--of performance on the task. Then the participants' overall metacognitive judgments about their skills or propensities as compared to others are compared to the objective assessments of how they performed on the criterion task as compared to others. For example, a questionnaire might ask how good a student a person thinks he or she is, and then relate that rating to GPA, plotting this correspondence across a sample of people. If people who rate themselves as good students, are, in fact, the students with high GPAs, then this global meta-accuracy correlation will be high: people who think they are good students, are good students.

In contrast to these global measures, relative measures look at the metacognition within the individual and provide information on whether the person knows what he or
she knows and does not know. Such measures require the individual to discriminate among items on a task by providing judgments about whether they will get particular items right or not, on an upcoming test. Then the participant is given the test, and the predictions that were made about the particular items are correlated with whether those items were right or wrong on the test, usually using a gamma correlation (Nelson and Narens, 1990, etc, for discussion of this statistic; c.f., Mason & Rotello, 2008). A gamma correlation is simply a non-parametric correlation coefficient that relates the individual’s trial-by-trial judgments to whether they were right or wrong on each item on the criterion test. A correlation of 1.0 means that the person thought they would be correct on those items that he or she got right, and thought they would be wrong on those items they got wrong. A correlation of zero means that the person's judgments were unrelated to what they got right or wrong on the test. A correlation of -1.0 means that the person thought he or she would get right the answers he or she got wrong, and the wrong answers he or she got right. (One rarely sees this pattern in real data). If the sample mean over people--where gamma is the dependent measure-- is significantly above zero, then one can conclude that people have good metacognition, meaning that, on average, they know what they know and don't know. While relative meta-accuracy is most commonly used in the metacognitive literature on learning and memory (Dunlosky & Metcalfe, 2009) both of these measures--those measuring global meta-accuracy and those measuring relative meta-accuracy-- are commonly, and correctly, called metacognition. Even so, their meaning is different. The former assesses whether people know how good they are with respect to other people. The latter measures whether or not people know what they know.
Within the context of emotion recognition and the ultimate goal of social and interpersonal connectivity and understanding, there is little doubt that metacognition would seem to be a valuable skill. The application of metaknowledge to future social-emotional encounters should result in the opportunity to learn to correct one's emotional errors and promote the future likelihood of an accurate situational response while preserving and nurturing the relevant relationship. Accurate metacognitions, at a relative meta-accuracy level, should promote correct and confident interpretations: an angry expression, confidently understood as such, should elicit an immediate and appropriate response. In contrast, uncertainty in the interpretation of an ambiguous facial expression may prompt the perceiver to seek additional information until a confident assessment can be made, maximizing the likelihood of an adaptive response. Emotional displays signal important information about an individual’s internal states (Darwin, 1872), essential for smooth social functioning and ultimately for survival. Insofar as most individuals can accurately recognize nonverbal emotional expressions (Adolphs, 2006; Ekman, 1982; Tracy & Robins, 2008), and, insofar as metacognition is thought to have a central role in acquiring and maintaining this skill, one would expect the literature to show that people have good metacognition of emotion.

A number of studies have examined people’s metacognitive judgments concerning emotion recognition. Surprisingly, the consensus as exhaustively reviewed below¹ and as noted by Ickes (1993), is that people appear to almost entirely lack

---

¹ There is one additional study that might be relevant that is not included in our literature review— an unpublished masters thesis by Mortimer (1996, cited in Ickes, 2003), in which people were shown video segments and asked if they would be able to answer questions about these segments later. They were then asked inferential questions such as "Who is the child of the two adults?" or "What is the relationship between the man and the woman?". Participants' level of metacognition regarding their trial-by-trial empathic accuracy, was measured using a Pearson correlation. Mortimer found that "the correlations were large enough to be statistically meaningful for only 8 of the 72 perceivers—about 11% of them."
metaknowledge in the domain of emotion recognition. For example, Ickes, Stinson, Bissonnette, and Garcia (1990) designed a study to assess empathic accuracy in naturalistic dyads. The researchers recorded two participants interacting and later asked the participants to review the tapes and note what they and their partner had been feeling. Following this, the participants were asked to complete a self-report questionnaire including personality and self-monitoring measures. While the researchers demonstrated that the participants varied in their ability to recognize the emotions exchanged during the unstructured dyadic interaction, the global metacognitive measures failed to predict performance on this emotion recognition task. Individuals who reported high levels of empathy, skill at deciphering other’s intentions, and emotional intelligence were no more accurate at identifying emotions than those who reported less competence.

Similarly, Levenson and Ruef (1992) adopted a related paradigm to measure empathic accuracy and the effect of physiological synchrony between the target and the perceiver. Participants viewed a dyadic marital interaction and rated the emotional intensity of the observed target person. These ratings were compared with the target’s self-ratings. The accuracy of detecting negative emotions improved when the target and perceiver were synchronized in their autonomic and somatic responses. However, the traditional global empathy scales and the participants' ratings of self-perceived task accuracy and task difficulty did not predict task performance.

In keeping with this line of research Marangoni, Ickes, Garcia, and Teng (1995) used the dyadic interaction paradigm to assess empathic accuracy in simulated psychotherapy sessions. They were interested in individual differences in empathy and whether empathic accuracy improved with increased exposure to the target and with
feedback concerning the target’s feelings. Each participant watched three videotaped interactions between a target and a clinician. The videotapes were paused 30 times each while the participant evaluated the target’s emotional state. At posttest the participants were asked to provide a global judgment of their overall accuracy for each target. The authors found individual differences and general improvement in empathic accuracy following increased exposure to the target and feedback. However, they found no relationship between self-assessment and task accuracy. They concluded that while task accuracy was improved, the meta-accuracy was not improved.

Realo, Allik, Nolvak, Valk, Ruus, Schmidt, and Eilola (2003) created the Mind-reading Belief Scale to more specifically assess metacognition about one’s own emotion recognition ability. The scale included items designed to assess one’s ability to judge another’s traits, mind states, intentions, and emotions. The questionnaire requires the participant to use a 5-point Likert scale to respond to questions such as, “I can read people’s intentions in their faces”. While the scale was psychometrically sound, self-reported empathy on the Mind-reading Belief Scale failed to predict performance on a face or voice emotional expression recognition task.

Ames and Kammrath (2004) asked participants to complete a number of questionnaires to assess individual differences in extraversion, narcissism, self-esteem, and social skills. Participants completed The Interpersonal Perception Task (IPT) (Archer & Costanzo, 1988) in which one must answer questions about video clips that depict various interpersonal interactions. In keeping with previous research, there was no relation between questionnaire responses and performance on the IPT. Those who reported higher levels of social fluency were no more accurate at interpreting the
emotional content of the videos. After completing the IPT, participants were paired together and asked to perform a simulated negotiation task in which they adopted either the role of an entrepreneur selling a family business or that of an executive seeking to purchase the business. The negotiation task was followed by a report of one’s own and one’s partner’s intentions and feelings during the negotiation task. Again, those who reported higher levels of social fluency were no better at deciphering their partner’s intentions and feelings during the negotiation. The authors did, however, find a strong positive relationship between an individual’s predicted task performance and self-reported levels of narcissism. The finding that the more narcissistic the individual the more likely s/he was to predict but not exhibit good performance on emotion recognition tasks renders the utility of these self-reported global metacognitive measures questionable.

Recent work has attempted to discern in more detail the conditions, if any, under which global metacognitive judgments might be predictive. Zaki, Bolger, and Ochsner (2008), using video clips of emotional expressions concerning personal episodes, did find a situation in which self-reported empathic accuracy predicted people’s emotional accuracy. This positive relation, showing that people who thought they were highly sensitive actually performed better, occurred only when the most highly emotionally expressive clips were used. This is the only case in the published literature in which people have shown any above chance metacognitive accuracy on emotion recognition, but even this occurred with only a subset of the materials used. Despite this single provocative finding, then, the consensus remains that people lack metacognition about emotional expression understanding.
According to this review, then, Ickes’ (1993) early summary of the evidence that individuals lack metacognitive insight into emotion recognition--that self-assessment measures show no predictive value with respect to emotion recognition--seems to still hold today. The self-assessment questionnaires, while unable to predict performance on various emotion recognition tasks, do sometimes relate to levels of narcissism, however, rendering their interpretation particularly dubious. Finally, some questionnaires have even demonstrated a negative correlation between perceived and actual ability (Ickes, Stinson, Bissonnette, and Garcia, 1990).

Ickes (1993) suggested a number of reasons why individuals may lack metaknowledge in this realm. He argued that while individuals differ in their ability to identify emotions, people generally have very little insight into this ability because of privacy norms. As a result of these norms, individuals rarely solicit or receive accurate feedback concerning mistakes in emotional interpretation. Further, individuals may intentionally mask their emotions or intentionally provide misleading feedback that makes it difficult to assess one’s accuracy.

While the conclusion suggested by these studies seems straightforward, a complete absence of metaknowledge, nevertheless, seems incompatible with the general notion that humans are emotionally intelligent and socially adapted. Humphrey (1984) compared the complexity of social interaction to a chess game in which individuals must maintain and monitor their own goals while continually monitoring other’s behaviors, building alliances, deceiving, cooperating, strategizing, and manipulating. Successfully navigating such a socially complex world seems unlikely absent any metacognitive insight. Adolphs (2006) argued that one’s perception of emotion is active and inferential
in that one actively seeks out relevant cues and infers meaning based on observations and past experiences. This, too, suggests that individuals monitor their knowledge state, implementing a very basic metacognitive strategy in social interactions. People's obvious skill in negotiating the social world and their apparent lack of metacognition, reflected in the results reviewed above, seem irreconcilable.

Before accepting the conclusion that people have such extremely poor metacognition in the domain of emotional recognition, it is important to note that none of the above-cited studies on emotional metacognition used the relative measures of metacognition discussed earlier. All of the studies described above used questionnaires to assess interpersonal sensitivity based on the individual’s tendency to agree with statements such as, “I am generally sensitive to others' feelings”--a global measure. However, as noted earlier, global meta-accuracy is only one sort of metacognitive accuracy. It measures whether people know where they stand with respect to one another; it does not measure whether or not they know what they know. Without evidence concerning whether people's relative meta-accuracy is above chance or not, it could be premature to conclude that people actually have poor emotional metacognitive capabilities.

The present study investigated whether individuals can accurately predict and assess their performance on two distinct emotion recognition tasks, while exploring both relative and global measures of meta-accuracy and the relation between them. The study further sought to investigate the sensitivity of relative measures across differing levels of emotional content by including stimuli that ranged from relatively neutral to the more dramatically obvious expressions. Experiment 1 investigated people's metacognition of
emotion recognition using the Mind in the Eye's task--a task that assesses one’s ability to determine another’s mental state by viewing just the eyes (Baron-Cohen et al., 2001).

Experiment 2 used the Emotional Expression Multimorph Task (Blair, et al., 2001) that presents an individual’s full face in differing stages of transition from neutral to fully expressed emotions.

Experiment 1

In Experiment 1, participants completed a global metacognitive questionnaire and also performed a relative metacognitive task. The global questionnaire assessed self-perceived levels of empathy, perspective taking, and theory of mind. To evaluate relative meta-accuracy, the participants were then presented stimuli from the Mind in the Eyes test in which individuals viewed sets of eyes and selected the expressed mental state from various response options (Baron-Cohen et al., 2001). During the presentation of each item, participants provided a prospective judgment of whether they would be able to recognize the correct emotion. After making judgments on all items, they then answered each question. After answering the question each participant then made a retrospective judgment of whether they had answered correctly.

Given previous findings, it was expected that the current study would replicate the finding that global metacognitions were unpredictive of performance. The predictions extended to the notion that relative meta-accuracy would be good and that individuals would be able to make highly accurate item-by-item estimates of their performance. It was expected that both the prospective and retrospective relative metacognition judgments would show above zero meta-accuracy. Usually, though, retrospective metacognitive judgments show higher relative accuracy than do prospective
judgments because people, when making prospective judgments, do not know what the alternatives will be (see Glenberg, Sanocki, Epstein, & Morris, 1987). Although it was expected that both prospective and retrospective relative meta-accuracy judgments would be above zero, it was also expected that the retrospective relative meta-accuracy would be higher than the prospective relative meta-accuracy.

Method

Participants

Ninety-nine Columbia University (44 males, 51 females, 4 unknown, $M_{age} = 21.94$ years, $SD_{age} = 4.69$) students participated for course credit or cash. All participants in both experiments were treated in accordance with APA ethical guidelines.

Materials

The questionnaire that assessed global meta-accuracy contained 24 items selected from the Mind-reading Belief Scale (Realo et al., 2003), and the Interpersonal Reactivity Index Perspective-Taking (IRI–P) and Empathic-Concern subscales (IRI-E) (Davis, 1980). The items represented abilities related to understanding another’s mental states. Sample items from each included: “I can read people’s intentions in their faces” (Mind-Reading Belief Scale), “I sometimes try to understand my friends better by imagining how things look from their perspective” (IRI-P), and “I am often quite touched by things that I see happen” (IRI-E). Also included was a 2 task-specific questions, “Relative to other Columbia students I am able to determine what a person is feeling by looking into their eyes” and “Relative to other Columbia students I am not able to determine what a person is feeling by looking into their eyes”. This was termed this the Columbia Empathy Measure (CEM). For our combined samples, the internal consistencies were
measured using Cronbach’s alpha and were: Mind-reading Belief Scale $\alpha = .70$, IRI-P $\alpha = .78$, IRI-E $\alpha = .80$, CEM $\alpha = .76$. All responses were made on a 5-point Likert scale with higher scores indicating that the statement was extremely characteristic of them.

The stimuli for Experiment 1 were derived from the Mind in the Eyes task (see Figure 1 for an example) in which individuals view 36 sets of eyes and select the expressed mental state from 4 response options (Baron-Cohen et al., 2001) which, for reasons outlined below, were increased to 6 options. The stimuli consisted of various individuals posing complex mental states (for example: contemplative, desirous, and aghast). Each image was presented in black and white and cropped to include only the actor’s eye region. To decrease the probability of chance guessing that spuriously decreases gamma scores--making it difficult to detect a real correlation (see Schwartz & Metcalfe, 1994) – 2 distractors were added to the standard response options. Additional distractors were selected from the complete list of response options. Trials in which the added distractors were chosen more often than chance were eliminated, yielding 25 stimuli.

**Procedure**

The testing session began with the questionnaire, which was immediately followed by the Mind in the Eyes task. The stimuli were divided into 4 blocks (9 eyes per block). The displayed emotions and the response options were quite varied and the participants were told they would be given a 6 alternative multiple-choice test. Each block contained a prospective judgment phase and a retrospective judgment phase. During the prospective phase, participants viewed each successive set of eyes and after seeing each were asked, “How confident are you that you will be able to correctly
identify the emotion?” This confidence level was the prospective judgment. During the retrospective phase, participants viewed the same stimuli from the previous phase now accompanied by the test that included the response options and the instruction to select the expressed mental state. After choosing a response, they were then asked to provide a retrospective confidence judgment expressing the likelihood that the given response was correct. Both prospective and retrospective judgments were made using a slider that was anchored at 0% and 100% at each end. These selections were then converted these percentages to proportions.

Results

Performance

Mean accuracy on the Mind In the Eyes task was .73 (SD=.13). Chance performance on the task was .17. A one-sample t-test against the value of chance, showed that our sample performed well above chance \[ t(98) = 41.68, p < .00 \]. In keeping with the literature, this indicated that participants were good at identifying emotional expressions (Tracy & Robins, 2008).

To determine whether or not prospective and retrospective confidence judgments differed, overall, a paired samples t-test was performed. Mean prospective confidence judgments (M = .57, SD = .11) were significantly lower than mean retrospective confidence judgments (M = .65, SD = .12) \[ t(98) = 11.06, p < .001, \, d = 1.58 \]. This is in keeping with our expectations that confidence judgments would change once the response choices were presented, and with the fact that people had seen the stimuli twice by the time they made the retrospective judgments.

Relative Meta-accuracy
Two gamma correlations were computed for each participant. The first was between prospective confidence ratings on each item and whether each item was correct or incorrect; the second was between the retrospective confidence ratings on each item, and whether the items were correct or incorrect. These gamma correlations were taken as dependent measures for each participant from which the sample means of gammas were computed and compared to the null hypothesis of zero, which would indicate no relative meta-accuracy. Both the prospective and retrospective gamma correlations \((G)\) between confidence judgments and accuracy were significantly greater than zero [prospective \(G = .07, \text{SD} = .32, t(98) = 2.21, p = 0.03\); retrospective \(G = .40, \text{SD} = .33, t(98) = 12.20, p < 0.001\)]. As predicted, retrospective relative meta-accuracy was greater than the prospective relative meta-accuracy [\(t(98) = 8.72, p < .001, d = 1.25\)].

*Figure 1.* Global (Pearson correlation between questionnaire response and proportion correct on the emotion recognition task) and relative (gamma correlation, means of within participant correlations between confidence ratings and correct versus incorrect responses on the emotion recognition task) meta-accuracy in Experiment 1 (Mind in the Eyes). The Global questionnaires were the Mind-Reading Belief Scale (MRBS), the Columbia Emotion Measure (CEM), the Interpersonal Reactivity Index Empathy (IRIE) and the Perspective-taking (IRIP) subscales. Gamma correlations are given for Prospective (pro) and retrospective (retro) confidence judgments as related to item accuracy.
Relation of Relative Meta-Accuracy to Performance

Relative meta-accuracy was greater than zero, then, indicating that people do have knowledge about what they know and do not know concerning emotional expressions. The gammas themselves, for each participant, were used to see whether the magnitude of this correlation was, itself, predictive of performance on the Mind in the Eyes task. As noted by Nelson (1984), the gamma correlation (unlike the Pearson correlation) is robust under different levels of task performance. So if a correlation between gammas and performance on the Mind in the Eyes task were found, it would not be spuriously attributed to the mere level of performance, a priori, but rather would indicate that good metacognizers about their own emotion recognition were also good emotion recognizers. However, no relation between the two was found in this experiment. Participants who had particularly high relative meta-accuracy scores did not show particularly high performance on the Mind in the Eyes task, either when prospective gammas were related to proportion correct $[r_{between\ individuals'\ gamma\ scores\ and\ their\ proportions\ correct} = -.01, p = .92]$ or when retrospective gammas were related to proportion correct $[r = .08, p = .39]$.

Global Meta-accuracy

None of the methods of assessing global meta-accuracy—whether people knew that they were good or bad at doing the emotion recognition task—were related to their actual performance on the Mind in the Eyes task. A Pearson correlation between responses on the questionnaires and accuracy on the task was computed. Higher scores on the questionnaires indicate high self-rated levels of mind-reading ability, empathy, and interpersonal sensitivity. Mind-Reading Belief Scale was not correlated with task performance, $r = -.13, p = .21$, n.s.; Interpersonal Reactivity Index (Perspective taking
subscale) was not correlated with task performance, \( r = .08, p = .45, \text{n.s.} \); Interpersonal Reactivity Index (Empathic Concern subscale) was not correlated with task performance \( r = .05, p = .63, \text{n.s.} \); and Columbia Empathy Measure was not correlated with task performance \( r = -.08, p = .442, \text{n.s.} \). These findings replicate the previous literature.

Of further interest was whether global self-assessment on any scale was predictive of any measures other than task accuracy. Perhaps not surprisingly, responses on the Mind-Reading Belief Scale (a global measure) were marginally predictive of the overall magnitude of prospective confidence judgments \( (r = .19, p = .07) \) such that there was a trend towards individuals who thought they were proficient at mind-reading being more confident, overall, in their prospective confidence judgments than individuals who reported lower mind-reading competence. This finding harkens back to the earlier reported finding that narcissism is correlated with questionnaire responses and overall confidence.

The means of the prospective confidence judgments for each participant were not correlated with any other global metacognitive measure, however: the Interpersonal Reactivity Index (Perspective taking subscale) was not correlated with mean confidence judgments \( r = .00, p = .98, \text{n.s.} \); the Interpersonal Reactivity Index (Empathic Concern subscale) was not correlated with confidence judgments \( r = .11, p = .28, \text{n.s.} \); the Columbia Empathy Measure was not correlated with confidence judgments \( r = .14, p = .16, \text{n.s.} \). Mean retrospective confidence judgments were also not correlated with any global metacognitive measures: the Mind-Reading Belief Scale was not correlated with confidence judgments \( r = .16, p = .11, \text{n.s.} \); scores on the Interpersonal Reactivity Index (Perspective taking subscale) were not correlated with confidence judgments \( r = -.07, p \)
scores on the Interpersonal Reactivity Index (Empathic Concern subscale) were not correlated with confidence judgments $r = .02, p = .85, \text{n.s.}$; the Columbia Empathy Measure was not correlated with confidence judgments $r = .12, p = .25, \text{n.s.}$.

And, none of the above global measures of self-assessed metacognition as measured by the personality scales were correlated with relative meta-accuracy gamma scores (all $p$’s $> .05$).

**Discussion**

The most important finding of this experiment was that people were able to make relative metacognitive assessments concerning their emotion recognition performance. They discriminated between items that they answered correctly and incorrectly, particularly when the response options had been presented, and they had just made their emotion recognition judgment. This indicates that people do have some knowledge of whether they know or do not know—they have metacognition—in the domain of emotion recognition, in contrast to the conclusion that would have been drawn from previous studies on this topic.

As expected, this task also replicated the finding that global questionnaire responses did not predict performance on the task. People who reported interpersonal sensitivity, empathy, and the ability to read and interpret mental states and intentions were no better at identifying the facial expressions as depicted in the Mind in the Eyes than were people who reported that they were less competent.

While these results are intriguing, there are a few limitations of the Mind in the Eyes test that might warrant further exploration with a different task. The stimuli are ambiguous and the response options represent complex mental states that are based on
subtle distinctions. In addition, using just the eyes to identify emotions may not be readily comparable to the previous studies of emotion recognition, or to situations that people encounter in their daily lives. Furthermore, in the Mind in the Eyes task, there is no a priori way to distinguish between easy and difficult items, and Zaki et al’s (2008) data indicate that this could be an interesting variable to investigate. Finally, Glenberg, Sanocki, Epstein, and Morris (1987) have noted that accurate metacognitive judgments depend upon the participant knowing the nature of the criterion test. But the test response options were unknown at time of judgment in the Mind in the Eyes task, a factor that, as Glenberg et al. noted, should have negatively impacted on prospective relative meta-accuracy (which gamma scores, while significantly greater than zero, were small). This lack of knowledge of the test may also have accounted for the observed differences in the prospective and retrospective gamma scores.

In spite of these limitations, individuals did reliably assess their meta-accuracy on this task, suggesting a level of metacognitive insight not previously recognized. Even so, it seemed prudent to replicate with a task that circumvented the difficulties of the Mind in the Eyes test.

**Experiment 2**

The results of Experiment 1 indicated, for the first time, that people do know what they know emotionally; they have good relative meta-accuracy. Experiment 2 was created in order to replicate this finding using different and better stimuli. Accordingly, the Ekman Emotional Expression Multimorph Task (Blair, et al., 2001) was used. This stimulus set uses the Pictures of Facial Affect Series (Ekman & Freisen, 1976) to create a set of stimuli that depicts faces that gradually transition from neutral to fully expressed
emotion in 10 stages. The original Ekman picture set contains both males and females expressing the basic emotions that Ekman argued are universally recognizable: happiness, fear, anger, sadness and disgust. (While surprise is often included, our set did not contain pictures expressing this emotion). The set also contains each actor posing a neutral expression. Blair et al. (2001) used these pictures to create a series of morphed images that systematically vary the amount of emotional content available in each image. As the images gradually morph from a neutral expression to a fully expressed emotion, more emotional content becomes available and the posed emotion is slowly revealed. The graded emotional expressions allow objective control of emotional content and categorization of stimuli according to difficulty. This is an important feature of these stimuli in so far as measures of relative meta-accuracy are quite sensitive to restricted range. If the images do not represent a broad enough range of difficulty, the ability to detect metacognitive accuracy is hampered, as the participant is not able to reliably differentiate known items from unknown items.

In contrast to the Mind in the Eyes test, which only displayed a fragment of the face, the morphs display the entire face. Presenting the entire face is also more in keeping with the typical way in which facial expressions are processed in more naturalistic circumstances that allow the viewer to use all available facial cues and information (Adolphs, 2006; Tracy & Robins, 2008). This full-face presentation should provide more ecological validity than seen in the previous experiment.

Further, the Ekman morphs limit the response options such that the same basic emotional choices are presented on each trial, controlling for the problems of subtlety and ambiguity associated with the Mind in the Eyes task. This feature also addresses the
problem of the unknown nature of the test, in the previous experiment. It was expected that presenting the same basic emotion response options on each trial would decrease or possibly entirely eliminate the discrepancy between prospective and retrospective confidence judgments found in Experiment 1. Further, it was also hypothesized that the relative meta-accuracy gammas for the prospective judgments would be higher, given that the test was more transparent.

The expectation was to replicate the overall findings from Experiment 1-- that global metacognitive judgments would not predict task performance but that relative metacognitive judgments would show that people did have the metaknowledge of what they knew and did not know in the emotional face recognition domain. Further, to find that individuals would be sensitive to the amount of emotional content presented in the faces, adjusting their judgments according to stimulus difficulty such that they would be more likely to say they knew the emotion in the easy, expressive faces, and that they did not know when the emotions expressed in the more neutral faces in which the emotions were more difficult to discern and ambiguous.

Method

Participants

One hundred Columbia University (38 males; 60 females; 2 unknown, $M_{age} = 21.4$ years, $SD_{age} = 5.36$) students participated for course credit or cash.

Materials

The previously described questionnaires and scoring—that is the Mind-Reading Belief Scale, the Interpersonal Reactivity Index – Perspective Taking, the Interpersonal Reactivity Index – Empathic Concern, and the Columbia Empathy Measure—were used
with one small exception. The task-specific question read “Relative to other Columbia students, I am able to determine what a person is feeling by looking at their face”.

Ninety faces were selected from the Emotional Expression Multimorph Task: 6 actors (3 male; 3 female) posing 5 emotions (sad, happy, disgust, anger, and fear) (Blair, et al., 2001). The morphs were taken from a previously created set consisting of 10 gradations from neutral to fully expressed emotion. The 2nd, 4th, and 6th levels of each emotion were selected to serve as difficult, medium, and easy items respectively. Each face was presented in black and white on a black backdrop (see Figure 2).

**Procedure**

Global metacognitive judgments were collected using the previously described questionnaires. Relative item-by-item metacognitive judgments were also obtained according to the procedure outlined in Experiment 1. Participants provided prospective and retrospective confidence judgments of their predicted accuracy on each item. There were 6 blocks of 15 trials each. Block arrangement and presentation followed the format described in Experiment 1. Each block was divided into two sections: prospective confidence judgments and test accompanied by retrospective confidence judgments.

Participants were told that each model was expressing sadness, happiness, disgust, anger, or fear. Therefore, during prospective confidence judgments, participants knew that the emotion response options would remain constant and that only these options would be presented during the test.

**Results**

**Performance**
Mean proportion correct on the Ekman Emotional Expression Multimorph Task was .62 (SD = .06). Participants performed significantly better than chance (.20) on this task as indicated by a one-sample t-test against the chance value \( t(99) = 67.95, p < .00 \). This once again confirmed that individuals are generally quite adept at identifying emotional expressions (Tracy & Robins, 2008).

In spite of the fact that emotion response alternatives remained constant throughout the experiment, a paired samples t-test revealed that mean retrospective judgments were significantly higher than mean prospective judgments (prospective M = 71.45, SD = 11.90, retrospective M = 74.28, SD = 11.74) \( t(99) = 5.86, p < .001, d = .85 \), though the difference was not as great as in Experiment 1. Although the difference was very small, people’s mean confidence about what they would get right, a few minutes hence, was slightly lower than their posttest confidence, upon having seen the stimulus a second time and choosing the response alternative. This slight increase in confidence might be due to mere repetition of the question, and was also seen in the first experiment.

**Relative Meta-accuracy**

The main finding of this study was that both the prospective and retrospective gamma correlations \( G \) between judgments and test performance were significantly greater than zero \( \text{prospective } G = .38, \text{SD} = .22, t(99) = 17.14, p < 0.001; \text{retrospective } G = .45, \text{SD} = .45, t(99) = 21.63, p < 0.001 \). This finding demonstrated that once again, relative meta-accuracy was good, indicating that people have metacognitive insight into what they know and do not know on an emotion recognition task. Retrospective gamma correlations were slightly greater than prospective gamma correlations \( t(99) = 2.97, p < .004, d = .31 \).
Figure 2. Global (Pearson correlation between questionnaire response and proportion correct on the emotion recognition task) and relative (gamma correlation, means of within participant correlations between confidence ratings and correct versus incorrect responses on the emotion recognition task) meta-accuracy in Experiment 2 (The Ekman Multi-Morph Task). The Global questionnaires were the Mind-Reading Belief Scale (MRBS), the Columbia Emotion Measure (CEM), the Interpersonal Reactivity Index Empathy (IRIE) and the Perspective-taking (IRIP) subscales. Gamma correlations are given for Prospective (pro) and retrospective (retro) confidence judgments as related to item accuracy.

Relation of Relative Meta-Accuracy to Performance

Interestingly, and unlike in Experiment 1, task accuracy—the proportion correct on the emotion recognition test-- was related to both prospective ($r = .32, p = .001$) and retrospective ($r = .43, p = 0.00$) gammas. People who were better at the emotion recognition task also tended to be more accurate in their relative metacognitive assessments. This is an intriguing finding, suggesting that those with good metacognitive insight into what they know and do not know, in the emotional domain, are also better able to read other’s emotional expressions.

Global Meta-accuracy
As in Experiment 1, global questionnaire responses were not predictive of task performance: The Mind Reading Belief Scale was not correlated with task performance $r = .05$, $p = .65$, n.s.; Interpersonal Reactivity Index (Perspective Taking subscale) was not correlated with task performance $r = .12$, $p = .25$, n.s.; Interpersonal Reactivity Index (Empathic Concern subscale) was not correlated with task performance $r = .14$, $p = .18$, n.s.; the Columbia Empathy Measure was not correlated with task performance $r = -.12$, $p = .25$, n.s.. The questionnaire responses (specifically the Mind Reading Belief Scale) predicted magnitude of mean overall confidence judgments (prospective $r = .23$, $p = .02$ and retrospective $r = .3$, $p < .001$), that is, if people were overconfident on their global judgments they tended to be overconfident overall, on the individual judgments as well.

**Item Difficulty**

As expected, participants were sensitive to the range of item difficulty reflected in the morphs. This sensitivity is captured in their mean accuracy, confidence judgments, and response time. Each pair-wise comparison across item difficulty was significant at the $p < .001$ level.

**Table 1** *Experiment 2: Ekman Morphs - Performance According to Item Difficulty.*

Mean performance (standard deviations) according to item difficulty.

<table>
<thead>
<tr>
<th>Morph Difficulty</th>
<th>Mean Accuracy</th>
<th>Mean Prospective Judgment (0-1)</th>
<th>Mean Retrospective Judgment (0-1)</th>
<th>Mean test RT (in ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>M = .80 (.09)</td>
<td>M = .82 (.11)</td>
<td>M = .84 (.10)</td>
<td>M = 4782.61 (1324.77)</td>
</tr>
<tr>
<td>Medium</td>
<td>M = .67 (.09)</td>
<td>M = .73 (.12)</td>
<td>M = .75 (.12)</td>
<td>M = 6024.02 (2013.28)</td>
</tr>
<tr>
<td>Difficult</td>
<td>M = .39 (.08)</td>
<td>M = .60 (.16)</td>
<td>M = .64 (.16)</td>
<td>M = 7432.98 (2483.73)</td>
</tr>
</tbody>
</table>
These experiments attempted to further investigate and corroborate the finding by Zaki, Bolger, and Ochsner (2008) that global metacognitive measures were accurate when used to predict performance on highly expressive items. However, performance on only the easiest morphs was considered, questionnaire responses did not predict task accuracy (Mind-Reading Belief Scale $r = .07$, $p = .48$, n.s., Interpersonal Reactivity Index (Perspective Taking subscale) $r = .03$, $p = .76$, n.s., Interpersonal Reactivity Index (Empathic Concern subscale) $r = .07$, $p = .49$, n.s.; Columbia Empathy Measure $r = -.21$, $p = .45$, n.s.). Interestingly the Columbia Empathy Measure was negatively correlated with performance on the most difficult items. People who said they were most proficient at identifying emotions had the worst performance on the task (when only the most difficult items were used) $r = -.21$, $p = .04$. Further, the Empathic concern subscale of the Interpersonal Reactivity Index predicted performance on the medium difficulty items $r = .20$, $p = .05$.

In order to further investigate item difficulty, a separate gamma correlation was computed for items at each level of difficulty using both prospective and retrospective confidence judgments correlated with accuracy. All prospective and retrospective gammas for easy, medium, and difficult items were significantly greater than zero at $p < .001$, (Easy $G_{pro} = .28$ ($S.D. = .39$); $G_{retro} = .38$ ($S.D. = .34$); Medium $G_{pro} = .31$ ($S.D. = .27$); $G_{retro} = .39$ ($S.D. = .27$); Difficult $G_{pro} = .16$ ($S.D. = .26$); $G_{retro} = .23$ ($S.D. = .29$). A 2 x 3 ANOVA was performed to further investigate the effects of judgment phase (prospective or retrospective) and item difficulty (easy, medium, difficult). No significant interactions were found. However, as expected there was a significant effect of judgment phase such that gamma correlations between prospective judgments and trial accuracy
were significantly lower than gamma correlations between retrospective judgments and trial accuracy $F(1) = 12.53$, $p=.001$, $d = .12$. There was also a main effect of item difficulty, when the gammas were computed individually on only those subsets of items within participant, such that the most difficult items had significantly lower gamma correlations $F(2) = 29.49$, $p = .00$, $d = .39$.

**Discussion**

Experiment 2 replicated the results of Experiment 1 with respect to both relative and global meta-accuracy. Individuals were able to predict their performance on an item-by-item basis, prospectively, and evaluate it, retrospectively. This relative meta-accuracy indicates that individuals have the ability to both predict and evaluate their performance on this emotion recognition task.

In Experiment 1, test response choices had differed across trials and had not been available during prospective judgments, a fact that probably rendered the prospective judgments less accurate than when participants knew what the test alternatives would be, as was the case in Experiment 2. The similarity of prospective and retrospective judgment accuracy in Experiment 2 is likely a function of the equivalent information being available during the prospective and retrospective phases.

Global meta-accuracy, as captured by the questionnaires, was not predictive of task performance. Participants' estimations of their emotional and interpersonal sensitivity were unrelated to their actual ability to identify the expressed emotions. Unlike Zaki, Bolger and Ochsner (2008), global metacognitive self-assessments predicted performance for highly expressive target stimuli. However, having used static faces
rather than dynamic video clips may have affected this finding. The static faces may have been processed differently than emotionally charged exchanges captured on video.

Participants were sensitive to the difficulty of items as reflected in their prospective and retrospective item-by-item judgments. Easy items were given higher mean judgments than items of medium difficulty, which, in turn were assigned higher mean judgments than very difficult items. Interestingly, people who were good at discriminating what they knew and did not know, in the relative accuracy task, were also the people who performed best on the task—at least in Experiment 2 (though not in Experiment 1). Thus, this particular metacognitive index, in which people made specific judgments about whether they could correctly assess the emotion in particular faces, related to overall performance on the emotion recognition task itself, while none of the global questionnaire measures did.

**Summary and Concluding Discussion**

These studies found, for the first time, that, while individuals have great difficulty making accurate global trait judgments of their own empathy, they are able to make relative metacognitive assessments, correctly identifying the particular emotional exemplars that are more difficult for them to recognize. The failure of global metacognitive self-report measures to predict performance on emotion recognition tests had led to the assumption that individuals lack metacognitive awareness of their interpersonal sensitivity, but the high relative accuracy found here belies that conclusion. Our findings are consistent with the notion that human beings possess a highly developed social mind and are, therefore, socially adept and skilled at recognizing emotions (Adolphs, 2006; Ekman, 1982; Tracy & Robins, 2008). It is likely that reaching this
advanced level of social functioning is accompanied by some metacognitive awareness, as our results indicate. Interpersonal emotional interactions, in the real world, are highly variable and success often requires that the individuals engaged recognize both their accuracy and inaccuracy.

While these data demonstrate that individuals have good relative meta-accuracy when viewing static faces, it would be interesting to know whether these results generalize to more dynamic displays. Most of the previous work on metacognition of emotion recognition has used video taped interactions between live actors engaged in emotional exchanges. Assessing relative meta-accuracy while viewing such naturalistic interactions would extend and add ecological validity to the current findings.

Also, while the current findings demonstrate that people can reliably make accurate relative metacognitive judgments, it is not known how or if they use this metacognitive knowledge. However, accurately assessing one’s performance on a given test of emotion recognition might result in improved performance on other emotion recognition tasks if the person is able to use this metacognitive knowledge to appropriately seek new information in the face of uncertainty. Such a feedback loop might, in the real world, give rise to a positive relation between metacognitive ability and improved emotional recognition. This possibility is offered with caution, since no relation between relative meta-accuracy and task performance was observed in Experiment 1. However, in Experiment 2—an experiment with stronger metacognitive results for reasons mentioned above-- those people with more accurate metacognitive judgments were also better at the emotion recognition task itself. A link between metacognitive insight and task performance may be of particular importance within the realm of
emotion recognition. Both abilities, emotion recognition and metacognition, require that the participant reflect upon and understand mental states.

Establishing whether or not people’s metacognition could be leveraged to improve emotion recognition may be particularly important when working with people who lack proficiency at emotion recognition, such as people with autism or with Asperger’s syndrome. Interventions are often designed to teach people how to correctly identify emotions based on facial and other non-verbal cues. If the individual lacks metacognitive insight into his or her ability to recognize emotions, he or she may not know when to use these newly learned strategies. In order to effectively implement strategies designed to teach emotion recognition, a reasonable strategy may be to first determine if a metacognitive deficit exists, or not. If it does, then that might be the starting point for therapy. If not, then this metaknowledge might be implemented in appropriately directing the emotional learning to where it is needed most.

In summary, global judgments are unlikely to capture the specific intricacies that categorize individual emotional encounters. However, when measured on an item-by-item basis, people are very good at making metacognitive assessments of their emotional knowledge, and some people are very good at it. And this kind of metacognition could be very useful in the real world—allowing people to seek more information when they need it, and to learn from the consequences of such seeking. While both perspectives on metacognition—the global and the relative—are valid, the relative accuracy perspective, in which the person evaluates whether he knows or does not know in each instance according to the dynamics of each specific encounter, may be more representative of people’s nuanced ability to successfully navigate real world interpersonal encounters.
Chapter 2

Metacognition of dynamical emotion recognition
Metacognition of dynamical emotion recognition

Metacognition is an advanced cognitive ability that allows individuals to distinguish what they know from what they do not know (Flavell, 1979). Metacognition is generally studied within the domains of learning, study habits, and general educational practices. In these domains, it is widely accepted that good metacognition is advantageous in that it allows individuals to understand when to seek additional information if something is not fully understood and in a larger sense, establish when and how to allocate cognitive resources (Finn, 2008; Metcalfe & Finn, 2008; Thiede, Anderson, & Therriault, 2003). Metacognitive accuracy in the domain of emotion recognition is demonstrated by the ability to identify when you are correctly or incorrectly recognizing another’s emotional state. Good metacognition of emotion recognition should, therefore, allow individuals to experience more successful social interactions and appropriate emotional responses.

Very little research has been done on metacognition of emotion recognition. While it is known that individuals are good at identifying emotions, previous research has suggested that they may have poor metacognition (Ickes, 1993; Ickes, Stinson, Bissonnette, & Garcia, 1990; Levenson & Ruef, 1992; see Zaki, Ochsner, & Bolger, 2008 for an exception). This early research concluded that people do not have good metacognition about their recognition of emotional expression. However, the operational definition of metacognition of emotion recognition that these studies used did not allow for the evaluation of whether or not individuals can determine when they are accurate and when they are not on an emotion recognition task. Instead, the question that was actually posed was: Could individuals assess the goodness of their own emotion recognition
ability with respect to that of others? The measures of global metacognition--that is, how good a person is at placing him or herself with respect to others on the goodness of emotion recognition accuracy--that were used in these studies were questionnaires that were designed to assess an individual’s beliefs about their own empathic ability, emotional sensitivity, emotional intelligence, and compassion. Participants filled out these self-assessment questionnaires and researchers then attempted to correlate scores on the questionnaires with later performance on an emotion recognition task. A significant correlation between questionnaire scores and task performance would require the participant’s score on the questionnaire to predict his or her performance on the emotion recognition task. In addition, each participant’s score must relate to the scores of the entire sample such that those who performed well on the emotion recognition task must score higher on the questionnaire than those who performed poorly on the emotion recognition task. They found no relationship between the two – people’s global or overall judgments about their own empathic abilities did not predict their accuracy on emotion recognition tasks.

These negative findings indicate that people do not know how good they are at identifying emotions relative to other people (Ickes, 1993, Levenson & Ruef, 1992, Ickes, Stinson, Bissonnette, & Garcia, 1990). But they do not address the issue of whether or not people are able to determine if they will later correctly or incorrectly identify an emotional expression or state. Thus, they do not say whether people have or do not have accurate item-by-item metaknowledge of their own emotion recognition. They do not provide any information about whether people have second order knowledge of when they have understood an emotional expression, and they know what the other person is
feeling, and when they do not know what the other person is feeling. These studies do not utilize a methodology that would allow the researcher to answer this question.

To address this question, Kelly and Metcalfe (2011, please see Chapter 1, which is based on that research) have recently shown that individuals are able to distinguish which particular static facial expressions they will correctly identify from those that they will not be able to identify. They did so by adopting methods used in traditional metacognitive research (Nelson & Narens, 1990). Instead of asking individuals how empathic they are or whether they are able to read other’s thoughts, they asked the participants to make a prospective judgment for each emotionally expressive face stimulus that they viewed concerning whether they would later get the emotion designation correct. They assessed relative accuracy, or the relationship between these metacognitive judgments and actual emotion recognition accuracy on each specified trial for each participant. This procedure resulted in within-participant correlations that were significantly greater than zero –indicating that the participants had metacognition. Thus, on an individual item level, people knew when they were liable to later be correct or incorrect on their emotional recognition-- their relative accuracy was good. Kelly and Metcalfe's (2011) study also showed that at the same time people did not know whether they were better or worse at emotion recognition than other people--their global accuracy was at chance. While not disputing the null results on global accuracy --people's ability to say whether they are good or bad at emotion recognition--we argue that the metacognitive ability tapped by the relative accuracy measure-- knowing when one has understood a particular emotional expression and when one has not yet understood--is crucial for social functioning.
Global and relative measures, then, answer two different questions. Global measures answer the question: “Do I know how well I will perform overall on an upcoming task in relation to the other people who will be performing the same task?”. Relative measures answer the question: “Do I know if I’m likely to be right or wrong about the emotion I think specific stimuli are displaying?”. The answer from the research literature on metacognition of emotion recognition to the first question is no: people cannot make these global self-assessments. The answer to the second question appears, tentatively, to be yes. However, only one study has investigated it. According to Kelly and Metcalfe’s (2011) data, relative measures demonstrated that individuals do display good metacognition with respect to emotion recognition – specifically the recognition of posed facial expressions. But perhaps the positive results would be different with other emotional stimuli.

Emotional displays of expression are not limited to the static arrangement of facial features, even though such static displays are most frequently the subject of research (Ekman, 1982; Sabatinelli, et al., 2010). Several groups of researchers have forwarded arguments against the use of static images, maintaining that recognition of posed facial expressions does not constitute or allow for true and complete emotion recognition (Fiorentini & Viviani, 2011; Zaki & Ochsner, 2011). Additional cues are often used to convey emotional states (Yoshikawa & Sato, 2008). The position and movement of the body, dynamic aspects of emotional expression, and the temporal aspect of emotions that allows them to evolve and change over time all serve as additional cues for accurate emotion recognition. According to these critiques, emotional displays have a
fluid, complex, and dynamic nature that cannot be captured in simple, static images (Wehrle et al., 2000).

According to these critiques, the subtleties involved in understanding other’s mental states and depth of emotion requires the interpretation of more than just static facial expressions. Aviezer, Trope, and Todorov (2012) have even found that in some cases, static facial expressions do not convey sufficient information for the observer to accurately interpret the emotional response. Their research demonstrates that in the case of extreme emotional displays, presented at the peak of intensity, individuals find it difficult to distinguish even the valence of emotions. Perhaps surprisingly, the static facial expressions associated with extreme pain and anguish or pleasure and joy and overlap substantially, rendering a distinction based solely on facial displays extremely difficult. However, when presented with the body postures associated with the emotions, a separate group of individuals were considerably more accurate. Body posture enabled individuals to disambiguate the emotional experience and provided sufficient context to allow them to successfully identify the expressed emotion.

Zaki, Ochsner, and Bolger (2008) have done pioneering research in the dynamics of emotional expression and understanding. They used video clips that depicted a person describing an actual emotional event that they experienced. Participants were asked to judge the intensity and valence of the emotion expressed in the video clips. They found that individual self-reported levels of empathy were only predictive of their task performance when the target was highly expressive. But while the introduction of dynamic emotional situations, such as those used by Zaki et al. (2008) greatly advances our understanding of emotion recognition, the possibility of identification of the specific
emotions involved, rather than just their valence and intensity, is also important in allowing evaluation of claims concerning people's metacognition of emotion recognition. Kelly and Metcalfe's (2011) findings of relatively accurate metacognition of emotion directed at posed facial expressions provide a good starting point, but one that can be criticized for the lack of realistic dynamic information. The present studies, like our earlier studies, investigate metacognition of emotion for specific emotions, but here, rather than using static emotional displays the emotions are expressed dynamically.

Body language and the prosody of spoken language are both situations in which emotion is expressed dynamically. Both forms of expression provide important cues that may be used in the process of emotional identification that, in part, address the three critiques outlined above. Both are dynamic and fluid in nature and represent the ability to process information over time - either as the intonation of the spoken sentence waxes and wanes in intensity or the body gait or posture shifts, recoils, or increases in speed. Attention must be sustained and updated throughout the interaction and body gait necessarily requires the individual to attend to the entire body. Demonstrating metacognition in the realm of spoken and body language is an important step in expanding our understanding of the way that individuals actively monitor their ongoing social interactions. The two experiments that follow will investigate people's metacognition of emotion in the dynamical domains of emotional body language and emotional speech prosody.

**Emotional Body Language - Gait**

The study of expressive body movement and body language was significantly advanced with the advent of the use of point light displays (Cutting & Kozlowski, 1976;
Johanson, 1973). These displays are created by attaching small, reflective surfaces (lights) to the major joints of an actor’s body and filming the actor in the dark while in motion so that only the patterns of the movements of the lights are apparent. Point light displays capitalize on the sensitivity of the visual system to the perception of biological motion. The resultant video image conveys a very limited amount of visual information that contains only the basic movement of the actor, absent a full body image (Atkinson, Dittrich, Gemmell, & Young, 2004). Individuals are so sensitive to these point light displays that they can identify the gender of the actor (Troje, 2002) and the actor’s intention (Runeson & Frykolm, 1983).

Atkinson, Dittrich, Gemmell, and Young (2004) created a set of point light displays that demonstrated that individuals are highly accurate at identifying the biological motion associated with basic emotions. Chouchourelou, Matsuka, Harber, and Shiffrar (2006) expanded on this basic finding by creating and validating a similar series of point light display video clips depicting the following basic emotions: happiness, sadness, neutral, anger, and fear. These video clips included masks that were superimposed over the original video. The masks were created using additional reflective lights (identical in appearance to those comprising the actor’s actual biological motion). These lights were scrambled in order to create a more complicated composition of lights that would distract the viewer from the configuration of the original actor’s lights, rendering the task more difficult by requiring the participant to identify the biological motion through the mask of lights. They were able to show that individuals are quite accurate in identifying the expressed emotion even when a mask is present. The ease with which individuals are able to identify even these highly degraded images of
biological motion underscores the salience of this feature within the realm of emotion recognition. Nothing, however, is known about people’s metacognition about their recognition of emotional gait. Experiment 3 will investigate this question.

**Emotional Spoken Language - Prosody**

Emotional prosody is the intonation and inflexion used in spoken language to communicate emotional content (Frick, 1985; Ross, 1981). It varies along 3 dimensions: fundamental frequency, duration, and intensity (Murray & Arnott, 1993). For example, stressed syllables tend to be higher in frequency, longer in duration, and possess greater amplitude. While these dimensions are used to classify the sound waves themselves, the perception of prosodic speech is generally based on a different set of dimensions: pitch, loudness, pauses, and the length of the utterances (Werner & Keller, 1994). Expressions of each of the basic emotions correspond with specific modulation patterns that vary along these parameters. Again, however, nothing is known about people’s metacognition in making these decisions.

Expanding our understanding of metacognition of emotion recognition is an important component in the study of individual cognition and social fluency. The ability to successfully reflect on and monitor our interpretations of other’s emotional states can potentially influence our ongoing social and personal interactions. The finding that facial expressions, body language, and emotional language are all highly recognizable to most typically developed adults is as important as the breakdown of this ability under certain situations. For instance, an individual must also interpret deceptive, confusing or ambiguous, difficult, or degraded emotional displays.
As well as identifying emotions, people need to be able to distinguish when they are successfully interpreting another’s emotional state and when they are not successful. Accurate metacognitive reflection, therefore, has the potential to encourage the individual to seek additional emotional cues, solicit explicit information about another’s emotional state, or leave a potentially threatening situation.

Experiment 3 will investigate the accuracy of metacognition of emotion with respect to biological motion. Experiment 4 will investigate metacognition with a different dynamic cue - expressivity of spoken language.

**Experiment 3**

Experiment 3 was designed to assess people's relative metacognitive accuracy while viewing videos of point light displays that depicted actors displaying various basic emotions. While it has been established that people can successfully identify the expressed emotions in the selected clips (Atkinson, Dittrich, Gemmell, & Young, 2004; Chouchourelou, Matsuka, Harber, & Shiffrar, 2006), the primary interest for this experiment is in their ability to distinguish those clips they would accurately identify from those that they would not. Body language is an important cue that individuals use to interpret other’s emotional states. Therefore, relative metacognitive accuracy on this task would enable an individual to respond appropriately to another’s anger, fear, happiness, sadness, or neutral emotional state if the individual is capable of using their own judgments to seek additional information in order to achieve a better understanding of the emotional situation.
Method

Participants. Thirty-one Columbia University students (14 males, 17 females, \(M_{\text{age}} = 19.74, \text{SD}_{\text{age}} = 1.77\)) participated in the Emotion in Body Language experiment in exchange for course credit. All participants were treated in accordance with APA ethical guidelines for conducting research.

Procedure. Before the start of the task, participants were provided with instructions and a practice trial. They were explicitly told that every video in the task could be classified as: angry, fearful, happy, sad, or neutral. During the prospective phase, participants were asked to view 20 different video clips. At the conclusion of each video clip, a slider appeared on the screen along with the prompt, “How confident are you that you will be able to correctly identify the emotion being expressed?” The ends of the slider were labeled 0% on the left and 100% on the right. This response constituted the prospective judgment.

The retrospective phase began after the participant had viewed 20 video clips. During the retrospective phase, participants saw the same 20 video clips again, and were asked to identify the expressed emotion as: angry, fearful, happy, sad, or neutral. After a response was selected, they made a retrospective judgment of how sure they were that they were correct on that trial by using the same slider described above.

Materials. Chouchourelou, Matsuka, Harber, and Shiffrar (2006) created the stimuli for the Emotion in Body Language task. They consisted of 60 point light display video clips. Each 3-second video clip depicted a point light walker expressing one of the following affective states: anger, fear, happiness, sadness, and neutrality. Each emotion was presented 12 times - 6 times with the actor walking from left to right and 6 times
with the actor walking from right to left. Half of the clips depicted a male while the other half depicted a female walker.

Each video consisted of 13 point lights on a walker and 13 point lights that were scrambled to create a mask that was superimposed on top of the point light walker. More specifically, a point light walker was present in each video and each video contained a mask of scrambled lights that was created using the same 13 lights that were originally generated by the point light walker. Every mask was tailored to the specific point light walker in order to match the velocity of the walker’s lights with the mask’s lights. Each walker light, therefore, had a corresponding mask light. The mask lights were placed very near to the walker’s lights and the first frame of the walker’s original position determined the original starting location of the mask light. The mask’s lights were scrambled in such a way that their distribution across the image never exceeded the size of the original walker’s position in co-ordinate space. The mask lights were matched for the velocity, luminance, and color of the walker’s lights. The use of these masks renders the task more difficult (Bertenthal & Pinto, 1994; Chouchourelou, Matsuka, Harber, & Shiffrar, 2006).

**Design.** The task was divided into 3 blocks of 20 trials each. Each block contained 4 trials depicting each of the 5 affective states. The presentation of the video clips was randomized within this framework. Blocks were divided into 2 phases: prospective and retrospective.

**Results**

**General Performance.** Participants performed well in correctly identifying the emotions expressed in the Emotion in Body Language task ($M = 84.14\%, \ SD = 8.80\%$).
This level of performance is consistent with previously published findings using the same stimuli (Chouchourelou, Matsuka, Harber, & Shiffrar, 2006).

**Confidence.** Mean confidence ratings were lower for the prospective (M = .69, SD = .10) than the retrospective (M = .74, SD = .90) phases [t(30)=5.89, p=.000].

**Relative Accuracy.** A gamma correlation was performed between metacognitive judgments and trial specific accuracy. Gamma correlations are a non-parametric statistic that allows the researcher to determine if a participant’s perceived accuracy on an individual trial matches up with their actual accuracy on that trial (Nelson & Narens, 1990). Prospective gamma correlations (G) (M = .21, SD = .31) and retrospective G (M = .45, SD = .27) were both greater than zero [prospective G t(30)=3.71, p=.001; retrospective G t(30)=9.40, p=.000].

**Relation of Metacognition to Task Accuracy.** Participant’s accuracy on the Emotion in Body Language task accuracy was not correlated with either prospective gammas (r = .34, p = .06) or retrospective gammas (r = .29, p = .12). People who performed better on the emotion recognition task did not demonstrate higher levels of metacognition.

**Experiment 4**

Experiment 4 was designed to assess metacognition of emotion recognition on emotionally expressive audio clips of sentences spoken in one of five basic emotions. It is well established that individuals are quite good at correctly identifying emotional

---

2 The finding that both prospective and retrospective gamma correlations were significantly greater than zero was replicated in 3 additional studies. Two samples included Columbia University (CU) students (n = 24, prospective gammas p = .001, retrospective gammas p = .000) and (n = 46, prospective gammas p = .000, retrospective gammas p = .000). An additional study was conducted using CU students, non-CU summer students, and members of the public replicated the same finding (n = 26, prospective gammas p = .050, retrospective gammas p = .000).
prosody (norming data published by Blonder, Bowers, & Heilman, 1999). Verbal cues play an important role in interpreting another’s emotional state, rendering the ability to determine if you are correctly interpreting them potentially useful. The experiment that follows is the first to address the question of people's metacognition of emotion recognition based on the prosody of speech.

**Method**

**Participants.** The same 31 participants that had been tested in Experiment 3 completed the Emotion in Spoken Language experiment. Both tasks were completed during one session with the order of administration of these tasks was randomized across participants.

**Procedure.** The procedure was identical to that of Experiment 3, with the exception that participants listened to 30 audio clips during the prospective and retrospective phases instead of viewing video clips.

**Materials.** The stimuli for the Emotion in Spoken Language task were derived from the Florida Affect Battery (FAB) sub-test 8a (Blonder, Bowers, & Heilman, 1999). Administration of this FAB sub-test requires the listener to identify the affective tone of semantically neutral sentences. There were 4 sentences (‘The boy went to the store.’, ‘The lamp is on the table.’, ‘The shoes are in the closet.’, and ‘The chairs are made of wood.’). Every sentence was spoken 5 times by the same speaker? in each of the following affective states: anger, fear, happiness, sadness, and neutrality, resulting in a total of 20 sentences.

As indicated by the accuracy norms established during the construction of the FAB, young adults typically perform very well on this task (M = 96.7, SD = 5.1)
(Blonder, Bowers, & Heilman, 1999). To minimize ceiling effects and render the task comparable to the Emotion in Body Language task, 2 masks were created that were superimposed on the affectively spoken sentences. The masks were constructed using 2 other neutral sentences from the FAB sub-test 8b. Using GarageBand 3.0.5 software, the mask sentences were played backwards (obscuring their semantic meaning) and superimposed on the 20 target sentences from sub-test 8a. Each mask was trimmed so that the starting point of both the sentence and the mask were synchronized. The final stimuli were constructed by randomly pairing 1 of the 2 masks with the affectively spoken sentences.

Three volume levels were used for the masks, resulting in easy, medium, and difficult audio clips. The volume of the affectively spoken sentence was kept constant while the volume of the masks was adjusted to create easy, medium, and difficult stimuli (stimuli are available upon request). The final compiled audio file was played through a set of ablePlanet Linx audio headphones with the computer’s volume set to mid-way in output/sound/system_preferences.

Procedure. The procedure matched that of Experiment 3 with the following exception: There were 2 blocks of 30 trials with each block containing 2 easy, 2 medium, and 2 difficult trials of each of the 5 emotions.

Results

General Performance. Mean accuracy on the emotion recognition task was .76, SD = .93. Accuracy for easy displays was M = .85, SD = .10; for medium was M = .83, SD = .10; and for difficult was M = .65, SD = .11. All 3 levels differed from one another
$F(2, 62) = 81.18, p = .00$ (easy vs. medium $t(31) = 2.15, p = .04$; medium vs. difficult $t(31) = 9.05, p = .00$; easy vs. difficult $t(31) = 11.53, p = .00$).

**Confidence.** Prospective confidence judgments ($M = .67, SD = .90$) and retrospective confidence judgments ($M = .73, SD = .92$) differed significantly from one another, with prospective judgments lower than retrospective judgments [$t(30) = 6.25, p = .000$].

**Relative Accuracy.** Collapsing over difficulty level, both the prospective gamma correlations ($M = .37, SD = .24$) and the retrospective $G$ ($M = .51, SD = .20$) differed from zero [prospective $G t(30) = 8.48, p = .000$; retrospective $G t(30) = 13.80, p = .000$]. (see Figure 3 below).

Prospective gamma correlations for easy ($M = .21, SD = .51$), medium ($M = .27, SD = .45$), and difficult materials ($M = .23, SD = .28$) did not differ according to item difficulty [$F(2, 60) = .151, p = .86$]. Similarly, retrospective gamma correlations did not differ according to item difficulty: easy ($M = .21, SD = .51$), medium ($M = .27, SD = .45$), and difficult ($M = .23, SD = .28$) [$F(2, 60) = .34, p = .70$]. In all cases, the gamma correlations were significantly greater than zero (with $p$ values of: .033, .03, .00 for prospective easy, medium and difficult gammas respectively and .00 for all retrospective gammas).

---

*The finding that both prospective and retrospective gamma correlations were significantly greater than zero was replicated in 2 additional studies with 2 of the samples described above – CU students ($n = 46$, prospective gammas $p = .000$, retrospective gammas $p = .000$) and non-CU summer students, and members of the public replicated the same finding ($n = 26$, prospective gammas $p = .000$, retrospective gammas $p = .000$).*
**Relation of Metacognition to Task Accuracy.** In contrast with the Emotion in Body Language task, in this emotion recognition task accuracy was correlated with metacognitive relative accuracy, both for prospective ($r = .49, p = .005$) and retrospective metacognition ($r = .67, p = .000$). Individuals who performed well on emotion recognition were likely to also perform well on the metacognitive portion of the task while those who had trouble identifying the spoken emotions showed poorer metacognition.

**Comparisons between Experiment 3 and Experiment 4**

**Task performance.** Task accuracy across the Emotion in Body Language and the Emotion in Spoken Language tasks was not correlated ($r = .26, p = .15$): good performance on one task was not predictive of good performance on the other.

**Metacognition.** Metacognition, however, showed a different pattern. Prospective gammas were correlated across both tasks ($r = .46, p = .01$), such that those who made more accurate prospective metacognitive judgments did so on both tasks.\(^4\) Retrospective gammas were not correlated across tasks ($r = .11, p = .57$).

\(^4\) We found that there was a relationship between metacognition across tasks such that if you had good metacognition on the Emotion in Body Language you also had good metacognition on the Emotion in Spoken Language task. While this was initially intriguing, we failed to replicate it in 2 follow-up studies - one with the sample of CU students ($n = 46$) and one with a different population comprised of Columbia students, non-Columbia summer students, and members of local public ($n = 26$) described above. We now believe that the initial finding was spurious.
General Discussion

Individuals in these experiments were able to make metacognitive judgments about whether, on an item-by-item basis, they would be able to identify the specific emotions displayed in both body movement and spoken language. They were also able to make retrospective metacognitive judgments. Our past finding that individuals can make such judgments while viewing static faces (Kelly & Metcalfe, 2011) is now expanded to include two different types of dynamic emotional stimuli. Both tasks, Emotion in Body Language (emotional gait) and Emotion in Spoken Language (emotional prosody), demonstrate that the complexity of emotional expression and cues does not prevent individuals from making metacognitive judgments about their perceived accuracy. This necessarily begs the questions, are there cues that cannot be evaluated metacognitively? How complex can an emotional encounter be and still allow the individual to maintain metacognitive accuracy?

The importance of metacognition has been demonstrated in other domains (Metcalfe & Shimamura, 1994). Accurate metacognition during learning and studying
has been shown to improve learning and help guide effective studying. By understanding what you know and what you do not know, you can make appropriate judgments and behavioral adjustments to enhance learning (Finn, 2008; Metcalfe & Finn, 2008; Thiede, Anderson & Therriault, 2003). Similar benefits of metacognition can also be predicted in the domain of emotion recognition. Individuals who are able to determine when they are correctly assessing another’s emotional state and when they are not would presumably be more socially fluent, adjust better to dynamic social interactions, and react appropriately to other’s emotional states.

Many cues that an individual provides during an emotional exchange are relevant to the ultimate understanding of that person’s emotional state (Zaki & Ochsner, 2011). Social and emotional exchanges occur over a period of time that can be quite short or very long and these cues change as the interaction progresses. It is known from past research that individuals are very good at identifying the emotions in posed facial expressions (Adolphs, 2006; Ekman, 1982; Tracy & Robins, 2008), body language (Atkinson, Dittrich, Gemmell, & Young, 2004; Chouchourelou, Matsuka, Harber, and Shiffrar, 2006), and spoken language (Blonder, Bowers, & Heilman, rev. 1999), even when these stimuli are degraded (as with the masks used in the present experiments). Beyond identifying the valence and intensity of emotions, individuals are capable of explicitly selecting the actual expressed emotion. Our previous finding that individuals display metacognition in the recognition of emotional facial expressions acts as a base on which to start building. The current findings of metacognition of emotion in body language, and spoken language continue to build on this base, creating a scaffold to further expand this line of research.
Having established that individuals demonstrate metacognition for dynamic stimuli in the form of body and spoken language, future work in this domain is needed to further extend our knowledge to even more naturalistic dynamical emotional situations. The tasks presented here require the individual to monitor an ongoing affective state over a short period of time. It is also important to determine the extent to which individuals can monitor affective states over longer periods of affective displays, over more complex emotional displays, and ultimately, over dynamic, multi-actor social interactions. As research continues to expand and push the boundaries of when individuals are able to make metacognitive judgments in the domain of emotion recognition the social implications of such findings must be considered.

The utility of metacognitive accuracy has been established in the educational domain. Being able to reflect on his or her own knowledge allows the learner to appropriately allocate resources to maximize learning and performance (Finn, 2008; Metcalfe & Finn, 2008; Thiede, Anderson, & Therriault, 2003). This logic can be extended to suggest the benefits that may also be conveyed within the emotional domain. Presumably, individuals should be able to use their metacognitive judgments to assess, adjust, and regulate their own interpersonal behavior—to seek out more information, for example, when they do not understand the other's emotions rather than blundering ahead without clarification. This ability to know what one knows and does not know in the emotional domain should, therefore, result in more sophisticated, complex, and clear social exchanges, but whether it, in fact, does so is an open empirical question.

As outlined in the introduction, past research on metacognition of emotion recognition concluded that individuals could not make accurate metacognitive judgments
about their emotion recognition ability. The explanations that were provided to explain this finding were convincing enough to halt further exploration. Shifting the focus from global questionnaires that were designed to measure perception of the goodness of empathic ability to relative measures of metacognition that measure item-by-item judgments of accuracy sheds a new light on the previously dormant topic. It is difficult to imagine the social ease with which individuals generally function, the intimate emotional relationships people engage in, and the acts of kindness and compassion that are regularly performed absent any awareness of our understanding of other’s emotions. As opposed to the notion that individuals cannot make metacognitive judgments about emotion recognition, the present findings indicate that they can actually distinguish when they are correctly or incorrectly identifying another’s emotional state. This seems to more accurately reflect the complexity of our inter-personal exchanges and seems to more clearly reflect our emotional competence and social fluency.
Chapter 3

Metacognition of emotion recognition in children
Metacognition of emotion recognition in children

Much of the research in the field of metacognition is devoted to learning and studying. Within this domain, researchers have focused considerable attention on young children and the extent to which they are capable of making accurate metacognitive judgments (Metcalfe & Finn, 2012, in press; Metcalfe & Finn, 2012; Metcalfe & Kornell, 2005; Metcalfe, Kornell, & Finn, 2009; Miele, Son, Metcalfe, in press; Son, 2010). The findings in this area suggest a developmental difference that may occur between the 3rd and 5th grades. Most 3rd graders and 5th graders demonstrate good metacognition on many tasks, however, 3rd graders show a deficit in their ability to effectively use their metacognitive judgments. This is referred to as an implementation deficit (Metcalfe & Finn, 2012). More specifically, while both 3rd and 5th graders are able to distinguish items they know from items they do not know, when given the opportunity to study, 5th graders base their study decisions on their metacognitive knowledge while 3rd graders chose study items randomly. Furthermore, Finn and Metcalfe (personal communication) have shown that while 5th graders use subtle metacognitive heuristics like the memory for past test (MPT) heuristic, use of such heuristics is often not observed with children in 3rd grade. Thus, a full blown metacognitive capability may not yet be fully developed in children in early middle childhood.

Metacognitive accuracy, when successfully implemented has been linked with a number of benefits that include appropriately allocating study time, adjusting learning strategies, and pursuing knowledge that is appropriate and accessible given an individual’s limitations and abilities (Finn, 2008, Metcalfe & Finn, 2008; Metcalfe & Kornell, 2005; Thiede, Anderson, & Therriault, 2003). In contrast with the established
benefits of metacognitive accuracy within the domains of learning and studying (Nelson & Narens, 1990; Nelson & Dunlosky, 1991; Thiede, 1999), no research has been done that speaks to the effects of metacognition on emotion recognition tasks and social fluency. Such benefits may be especially important to young children who are still in the early developmental stages of learning to regulate their own emotions as well as interact with others in an emotionally productive way. This cannot yet be established, however, as it is not known whether or not children even have good metacognition on emotion recognition tasks.

While research on metacognition of emotion recognition in healthy adult populations has shown some level of emotional metacognitive accuracy using both static and dynamic stimuli (Kelly & Metcalfe, 2011; Kelly & Metcalfe, in preparation), no research has been done on this topic with children of any age. Past research has shown that infants that are only a few days old are capable of generating emotional expressions and within a few brief months can respond to certain emotional cues (Tanguay, 1990). By the age of four, children are able to automatically and accurately identify the basic human emotions (Izard, 1971; Camras & Allison, 1985). This ability is in place very early in development and is highly relevant within the child’s emerging social world.

Determining whether or not and at what age children are able to know when they are correct or incorrect at interpreting emotions must first be established before any conclusions can be made about the possible benefits of good metacognition in this domain. Further, exploring a different form of metacognition – in this case within the domain of emotion recognition – may provide additional information in considering the developmental differences that have been seen in traditional metacognitive research. The
current study does not seek to address the possibility of an implementation deficit but must instead seek to address the most basic question. Do 3rd and 5th grade children demonstrate metacognition on emotion recognition tasks and are there any differences between the two groups?

**Experiment 5**

This experiment was designed to explore whether or not young children (3rd and 5th graders) are able to make metacognitive judgments about their perceived accuracy on an emotion recognition task and to validate a new set of emotional stimuli composed of cartoon images intended for use with children. The inclusion of a group of Columbia University (CU) also allowed for the comparison between children and adults on the same emotion recognition task. If all groups are able to show relative metacognitive accuracy – distinguish items that they will answer correctly from those that they will not – than this is an aspect of metacognition that is in place early in life. If the groups do not perform comparably, however, then developmental differences within this domain might affect social fluency across the age spans.

**Method**

**Participants.** Twenty-three 3rd graders and twenty-three 5th graders participated in the experiment. Both groups were recruited from a NYC elementary school and from the surrounding neighborhood via flyer and word of mouth. The parents of the children who enrolled via flyer or word of mouth were paid $10 for their child’s participation. Two 3rd graders and two 5th graders were removed because they always selected 100% confidence when asked about their perceived accuracy on each trial. One 5th grader was
removed because the session terminated early due to an unforeseen school scheduling change. Thirty-seven CU students participated in exchange for class credit.

**Procedure.** All participants in all 3 groups were told that they would be shown cartoon images expressing different emotions and the emotion choices would always be: happy, sad, neutral, surprised, angry, or fearful. Participants were given a sheet of paper with the response options in order to remember the choices.

Third and 5th grade children went through a short training procedure on what it means to be sure or unsure of something. Children were asked what color shirt they were wearing and how sure they were. They were then asked what color shirt the experimenter was wearing, what color shirt they were wearing 3 days ago, what color shirt the experiment was wearing yesterday…etc. This was done while the children viewed a sample slider that ranged in confidence from “not sure” to “very sure”. When the children were able to identify situations in which they felt not sure, very sure, and varying degrees in between, the experiment began.

All participants completed a practice trial to familiarize themselves with the design of the experiment. During the experiment, participants viewed images of cartoons expressing different emotions and were asked to use a slider anchored at “not sure” to “very sure” to identify how sure they were that they would be able to select the correct emotion when shown the image later in the task. After viewing these images, the same images were shown with the 6 emotions presented on the screen. The participant was asked to select the correct emotion.

**Materials.** The stimuli used in this experiment consisted of still images taken from the following films and television show: The Little Mermaid, Sponge Bob Square
Pants, Finding Nemo, Spirited Away, Monsters Inc., Ratatouille, Lilo and Stitch, Shrek, and Pixar Shorts. Research assistants watched the films or television clips and selected still images based on what they perceived to range from relatively neutral to fully expressed emotions. The emotions that were selected were: happy, sad, neutral, surprised, angry, or fearful. The context within which the emotion was expressed during the film, as well as the facial expression of the cartoon character itself, served as cues for the research assistant to select each image as belonging to one of the six pre-selected emotions.

In order to norm the stimuli, 77 CU students viewed the entire set of 120 images that were chosen. While viewing each image, the participant was asked to complete a forced choice selection, indicating which of the six emotions was being expressed. For each image, the number of times the participants selected each emotion was summed. The “correct” answer was selected based on the most frequently chosen emotion, relying on consensuality to act as a proxy for accuracy (which will be discussed later). This also ensured a range of difficulty across the stimuli to eliminate the possibility of ceiling effects. A rating was assigned for each picture based on the highest frequency of responses – for example, if 40 people chose happy for image 13, the correct response was considered to be happy and this created a rating of 40 – which allowed for the identification of “easy” “medium” and “difficult” images. The images were converted to a standard size and displayed in RealBasic 5.5.5 on Macintosh computers.

**Design.** During the experiment, there were 3 blocks of 20 images each. Each block consisted of a confidence judgment phase (How sure are you that you know the emotion?) and then a response selection phase in which the participant selected a
response from the six options presented to them. The 60 images were chosen randomly from the entire set of images.

**Results**

**Task Accuracy.** Average task accuracy for all 3 groups was above chance: 3rd graders (M = .70, SD = .08), 5th graders (M = .72, SD = .07), and CU students (M = .71, SD = .07). There were no accuracy differences across the 3 groups \( F(2,80) = .56, p = .576 \). Planned individual \( t \)-tests confirmed this across groups (3rd and 5th graders \( p = .949 \); 3rd and CU \( p = 1.00 \); 5th and CU \( p = 1.00 \)).

*Figure 4.* Task performance across all 3 groups – 3rd and 5th graders and CU students. There were no differences in task accuracy across the 3 groups.

**Confidence.** Confidence ratings differed according to group membership: 3rd graders (M = .80, SD = .12), 5th graders (M = .75, SD = .12), and CU students (M = .70, SD = .10) \( F(2,80) = 6.02, p = .004 \). Planned individual \( t \)-tests confirmed that 3rd and 5th graders did not differ in their confidence levels but that both were significantly more confident than CU students (3rd and CU students \( p = .001 \)) and that 5th graders approached a higher level of confidence than CU students (\( p = .070 \)).
Relative Accuracy. Gamma correlations for 3rd graders (M = .17, SD = .35) \[ t(22) = 2.26, p = .034 \], for 5th graders (M = .37, SD = .19) \[ t(22) = 9.26, p = .000 \], and for CU students (M = .38, SD = .20) \[ t(36) = 11.57, p = .000 \] differed from zero. The gammas did differ from one another across groups \[ F(2,80) = 6.28, p = .003 \]. Planned individual t-tests confirmed that 3rd grade gammas were lower than both 5th grade gammas \( (p = .015) \) and CU gammas \( (p = .003) \). Fifth grade and CU student gammas did not differ from one another \( (p = .872) \).

Figure 5. Task accuracy plotted against metacognitive accuracy. While all groups differed from zero, 3rd grade gammas were lower than 5th graders and CU students.

Discussion

Third, 5th, and CU students all performed equally well on the cartoon emotion recognition task. Third graders were more confident on the task than CU students, showing a bias toward thinking they are more accurate than they actually are. All three groups also showed some level of metacognition on the task. Most importantly, however, the 3rd graders, while demonstrating metacognition, were not as good as the 5th graders and the CU students at making metacognitive judgments. Because accuracy on the task
was comparable across groups, the poorer metacognitive accuracy exhibited by the 3rd graders cannot be accounted for by differences in task accuracy. The finding that 5th graders and CU students do not differ metacognitively suggests that once you have got it, you have it.

For 3rd graders, however, the story is not so clear. While metacognitive accuracy statistically is above chance, it is lower than the level of accuracy in 5th graders and CU students: a deficit does exist which may have important developmental and social consequences. The current study replicated the finding that 3rd graders do have metacognition, but also uncovered a different deficit in that their metacognition was not as good as 5th graders. They were less able than 5th graders and CU students to distinguish which emotional faces they were correctly interpreting.

Metacognition is an important tool in terms of monitoring current behavior and abilities, guiding future behavior, and evaluating overall performance when studying and learning (Nelson & Dunlosky, 1991; Nelson & Narens, 1990; Thiede, 1999). The benefits that metacognition confer in learning situations should, therefore, extend to emotional situations in which metacognitive judgments are accurate. The difference between past studies of metacognition in children and the current study is that 3rd grader’s metacognition is not as good as 5th graders or CU students. Therefore, one would expect the benefits of metacognition of emotion recognition to be more limited for 3rd graders.

Making judgments about learning and studying may be quite different from making judgments about emotion recognition. Having found that individuals can make accurate metacognitive judgments using these images served to validate this set of images as appropriate for use within the domain of emotion recognition. There is, however, an
aspect of this and other emotion recognition tasks that must be acknowledged. A large portion of metacognitive research is focused on what most would consider factual information, including mathematical problems, trivia questions, and comprehension exercises (Maki & Berry, 1984; Metcalfe, Kornell, & Finn; Son, 2010). Emotions, however, do not necessarily fit into the category of factual, universally agreed upon information. Rather, the identification of emotional expressions in general, and particularly, difficult or ambiguous ones, is based on consensuality (Barrett, Ochsner, & Gross, 2005). Individuals agree upon what constitutes a “happy” face or a “sad” face and, as such, expressions are categorized according to the most frequently cited emotion. Koriat (2008) has noted that metacognitive accuracy may be related to the level of consensuality specific to a particular item or stimulus.

This emotion recognition task used the same stimuli for 3rd, 5th, and CU students and all groups performed equally well on the recognition task. The difference was seen in metacognitive accuracy. It is possible that 3rd and 5th graders differ on their metacognitive accuracy on this task as a function of consensuality and mere experience. Most 3rd graders have necessarily had fewer opportunities than 5th graders to engage in emotional interactions and interpret emotional stimuli based on their age differences. While their emotion recognition ability seems to be somewhat in place at this age, lesser experience might affect their metacognitive accuracy.

Perhaps 3rd graders have not had enough life experience to monitor their own accuracy in this domain. In support of this notion, Ickes (1993) has proposed that deficits in metacognitive accuracy on emotion recognition tasks may occur because, (1) individuals sometimes try to hide their true feelings by displaying facial expressions
consistent with other emotional states, (2) individuals fail to solicit information during social interactions that might confirm or disconfirm their interpretation of ambiguous displays, and (3) individuals often deliberately mask specific emotions to maintain societal norms about emotional expressivity. Third graders may lack the maturity, sophistication, and experience necessary to overcome the complexities and nuances associated with emotional expressions in order to parse what they know from what they do not know.

Learning to successfully navigate the social world involves learning to control one’s behavior and emotions, recognizing and interpreting other’s emotions, resolving conflicts, creating and maintaining relationships, etc. The ability to do so likely requires the ability to interpret what others are feeling and the ability to determine if you are correct or not in your interpretation. While 3rd and 5th graders and CU students are able to make metacognitive judgments about emotion recognition, 3rd graders are impaired in their ability. Understanding how metacognitive development occurs in all domains remains an open question. Recognizing the deficit exhibited by 3rd graders in metacognition of emotion recognition shifts the focus from a purely cognitive discussion to include social and experiential aspects that may influence metacognitive accuracy.
Chapter 4

Conclusions and future directions
Conclusions and future directions

Human beings are socially adept and skilled at recognizing emotions. It is unlikely that such a highly developed social mind could have emerged absent any metacognitive awareness of one’s ability to recognize emotions. Within the context of emotion recognition and the ultimate goal of social competence, metacognitive awareness of one’s accuracy and inaccuracy may be invaluable.

Based on the evidence obtained from measures of absolute predictive accuracy, it has long been assumed that individuals simply cannot assess or predict their ability to interpret emotional expressions. The present work investigated the efficacy of an alternate measure of metacognitive awareness. Metacognitive awareness was measured using relative predictive accuracy during an emotion recognition task. In addition, the relationship between relative and global measures of metacognitive awareness was explored. Further, the sensitivity of these relative measures was investigated using two different types of stimuli: static and dynamic and across different age groups.

All five tasks that were used: posed facial expressions in the Ekman Multi-morph and the Mind in the Eyes tasks, body gait in the point light displays, emotional prosody as expressed in the FAB, and the posed facial expressions in the cartoon images generated the same result. Columbia students and young children possess the ability to distinguish those emotions that they know they are correctly interpreting from those that they are not. Facial expressions, body language, and spoken language each possess cues that individuals are able to effectively use to both identify the expressed emotion and render relative metacognitive judgments for each emotion.
As discussed above, the study of metacognition in other domains has led to the finding that metacognitive accuracy has important implications in monitoring ongoing and guiding future behavior. The most pressing question that remains is whether or not the benefits conveyed by metacognitive accuracy in learning and studying extend to situations involving metacognition of emotion recognition. In addition to varying the stimuli used in these emotion recognition tasks in order to determine which elements of the emotional expression are relevant and salient to the individual processing the situation, the extent to which this information can be used to modify the perceiver’s behavior must be investigated.

This may be of particular importance for individuals with deficits in emotion recognition ability. For instance, individuals on the autism spectrum often need to have specific emotional cues pointed out for them in order to develop this skill (Howlin, Baron-Cohen, & Hadwin, 1999; Silver & Oakes, 2001). The social impairment that is seen in autism stands in stark contrast to that observed in typically developed individuals. If adults and children are able to identify when they know they are correctly interpreting an emotion, this should presumably enhance their ability to function socially. If individuals lack this ability, it may disrupt their social interactions. Once this has been investigated in typically developed individuals, it might then inform future interventions for those with difficulty recognizing emotional expressions. Understanding whether or not people have good metacognition of emotion recognition and at what age it occurs is an important step in understanding the typical development of emotion recognition throughout the lifespan.
Finally, as our means of communication with others continue to evolve and change and increase in complexity, it is important to understand the limitations that exist for individuals making metacognitive judgments when the social situation is limited. For example, the increased use of electronic communication (text messaging and email) and social media (twitter and facebook) has made deciphering emotional content more difficult. As the cues that we have traditionally and successfully relied on are stripped away, we are left with a type of interpersonal interaction that is much degraded. Understanding when we are correct or incorrect in interpreting emotions in these situations will necessarily rely more heavily on context.

This is simply one example of how future research must emphasize the need for contextualized information. Having established a number of isolated cues that contribute to our ability to identify and evaluate our accuracy in the domain of emotion recognition, we must pull back from the details and begin to investigate the individual’s ability to use context and integrate all available cues to determine their metacognitive accuracy within the domain of emotion recognition.
References

Current Directions in Psychological Science, 15 222-226. doi: 10.1111/j.1467-8721.2006.00440.x


Press.


Kelly, K. J., & Metcalfe, (in preparation). Walking the walk and talking the talk: Metacognition of dynamical emotion recognition.


Years?’, *Brain Dysfunction*, 3, 197–207.


Appendices

Stimuli for Experiment 1 (Mind in the Eyes). These particular images represent the emotions: playful, desire, and worried.
Stimuli for Experiment 2 (Emotional Expression Multimorph Task). a) These images represent one actor posing fear at morphed levels 2, 4, and 6. b) These images represent another actor posing happiness at morphed levels 2, 4, and 6.
Stimuli for Experiment 3 (Point Light Walkers). These particular images represent still shots from masked videos depicting a) angry and b) happy.
Stimuli for Experiment 5 (Cartoon Images). These particular images represent still shots from masked videos depicting a) happy b) sad c) angry d) surprised.