

Running head: Development of emotion comprehension and abstraction

**Charting the development of emotion comprehension and abstraction  
from childhood to adulthood using observer-rated and linguistic measures**

Erik C. Nook<sup>1</sup>, Caitlin M. Stavish<sup>1</sup>, Stephanie F. Sasse<sup>1</sup>, Hilary K. Lambert<sup>2</sup>,

Patrick Mair<sup>1</sup>, Katie A. McLaughlin<sup>2</sup>, & Leah H. Somerville<sup>1</sup>

1—Department of Psychology, Harvard University

2—Department of Psychology, University of Washington

**In press at *Emotion***

**© 2019, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/emo0000609**

Address correspondence to:  
Erik C. Nook  
Department of Psychology  
Harvard University  
William James Hall  
33 Kirkland St.  
Cambridge, MA 02138  
[enook@g.harvard.edu](mailto:enook@g.harvard.edu)

Abstract word count: 242  
Body word count: 11,829

### ABSTRACT

This study examined two facets of emotion development: emotion word comprehension (knowing the meaning of emotion words such as “anger” or “excitement”) and emotion concept abstraction (representing emotions in terms of internal psychological states that generalize across situations). Using a novel emotion vocabulary assessment, we captured how a cross-sectional sample of participants aged 4-25 ( $N=196$ ) defined 24 emotions. Smoothing spline regression models suggested that emotion comprehension followed an emergent shape: knowledge of emotion words increased across childhood and plateaued around age 11. Human coders rated the abstractness of participants’ responses, and these ratings also followed an emergent shape but plateaued significantly later than comprehension, around age 18. An automated linguistic analysis of abstractness supported coders’ perceptions of increased abstractness across age. Finally, coders assessed the definitional “strategies” participants used to describe emotions. Young children tended to describe emotions using *concrete* strategies such as providing example situations that evoked those emotions or by referring to physiological markers of emotional experiences. Whereas use of these concrete strategies decreased with age, the tendency to use more *abstract* strategies such as providing general definitions that delineated the causes and characteristics of emotions or by providing synonyms of emotion words increased with age. Overall, this work (i) provides a tool for assessing definitions of emotion terms, (ii) demonstrates that emotion concept abstraction increases across age, and (iii) suggests that adolescence is a period in which emotion words are comprehended but their level of abstraction continues to mature.

*Keywords:* Emotion, development, abstraction, language

“Use your words” is a phrase frequently used by parents, teachers, and therapists as they ask children to verbalize and talk about how they feel. Although this phrase may appear to be a cliché, classic developmental theories suggest that language is a key self-regulatory tool (Luria, 1961; Vygotsky, 1962). According to these theories, internalized speech helps children regulate their behavior and emotions (Bretherton, Fritz, Zahn-Waxler, & Ridgeway, 1986; Kopp, 1989; Saarni, 1999). Decades of empirical research now support these notions, as a child’s general and emotional vocabulary are related to their ability to manage distressing situations, as well as their executive functioning, mental health, social likability, and academic outcomes (Cole, Armstrong, & Pemberton, 2010; Fabes, Eisenberg, Hanish, & Spinrad, 2001; Kuhn, Willoughby, Vernon-Feagans, & Blair, 2016; Matthews, Biney, & Abbot-Smith, 2018; Roben, Cole, & Armstrong, 2013; Salmon, O’Kearney, Reese, & Fortune, 2016; Trentacosta & Izard, 2007; Vallotton & Ayoub, 2011). Hence, developing a functional emotion lexicon appears to be an important ingredient to overall well-being.

Given the links between emotion vocabulary and psychosocial functioning, understanding how children develop mastery of emotion words is of critical importance. However, three key limitations to our understanding of emotion language development remain. First, even though creative methods have been developed for assessing emotion language in children and adults (see Castro, Cheng, Halberstadt, & Grünh, 2016 and Zeman, Klimes-Dougan, Cassano, & Adrian, 2007 for reviews), research using an objective assessment of participants’ comprehension of emotion words is lacking. Second, because most research on emotion vocabulary has been conducted in young children, we know very little about developments in emotion language through mid and late adolescence, periods of substantial social and emotional change (Somerville & McLaughlin, 2018). Third, although prior work on emotion language

development focuses on when individuals can recognize and use emotion words, a separate but related question concerns the ways in which the concepts underlying these emotion words differ across age. Here, we specifically examined the tendency to represent emotion concepts abstractly (i.e., as internal psychological states that can generalize across situations).

### **Measuring the Development of Emotion Word Comprehension**

Emotion word comprehension refers to an individual's knowledge of the meaning of words used to label emotional experiences (e.g., angry, excited, calm, or disappointed).

Comprehending an emotion word means being able to connect that word with a culturally-agreed-upon concept of what characterizes or defines that emotion (Bloom, 2000; Yin & Csibra, 2015). In this definition, it is important to remember that words and concepts are distinct. In particular, words are symbols that are thought to help people organize (and potentially learn) conceptual information (Barsalou, 1999; Doyle & Lindquist, 2018; Fugate, Gouzoules, & Barrett, 2010; Gopnik, 2001; Lupyan, 2012b, 2012a; Shablack & Lindquist, 2019).

Comprehension of non-emotional words is typically assessed through well-validated and widely-used tests such as the vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 2011) or the Peabody Picture Vocabulary Test (PPVT; L. M. Dunn & Dunn, 1997). In the WASI, participants are shown a word and asked to generate a definition. If a vague or incorrect answer is provided, the experimenter further probes the participant to ensure they provide as full of a definition as possible. Responses are scored according to a rubric of agreed-upon definitions for each word. To date, there is only one measure of emotion vocabulary that takes a similar approach, the Kusché Affective Inventory—Revised (KAI; Greenberg, Kusché, Cook, & Quamma, 1995; Kusché, Belike, & Greenberg,

1988). The KAI is an experimenter-administered interview that provides a multi-faceted assessment of emotion understanding, and the “feelings vocabulary” component involves an assessment of participants’ abilities to define the emotion terms proud, guilty, jealous, nervous, and lonely.

The KAI has been used to show that emotion comprehension is lower in children with learning disabilities, that the general ability to define emotion terms increases during childhood, and that an intervention aimed at educating children about emotion terms can augment this developmental process (Bauminger, Schorr Edelsztein, & Morash, 2005; Greenberg et al., 1995; Kelly, Longbottom, Potts, & Williamson, 2004). However, many studies employing the KAI do not administer the specific subtest of emotion word comprehension described above (Beck, Kumschick, Eid, & Klann-Delius, 2012; Duncombe, Havighurst, Holland, & Frankling, 2013; Locke, Miller, Seifer, & Heinze, 2015; Southam-Gerow & Kendall, 2000). Thus to date, a measure like the KAI has not been used to broadly examine how emotion word comprehension develops across childhood and adolescence. Assessing how emotion definitions vary with age would (i) provide an objective assessment of which words participants comprehend and (ii) provide insight into *how* participants represent emotion terms. Theoretically, insight into the development of emotion word comprehension is important not only because emotion language is related to wellbeing (Cole et al., 2010; Salmon et al., 2016) but also because linguistic labels have recently been proposed to serve a central role in learning new concepts (Doyle & Lindquist, 2018; Fugate et al., 2010; Lupyan, 2012b, 2012a). In particular, some have suggested that emotion words might help children develop specific and multi-dimensional emotion concepts (Lindquist, MacCormack, & Shablack, 2015; Nook, Sasse, Lambert, McLaughlin, & Somerville, 2017; Shablack & Lindquist, 2019).

Furthermore, most research on the development of emotion word comprehension focuses on childhood and neglects protracted development through adolescence. Indeed, the only methods that have been used to assess emotion word comprehension in adolescence include recordings of unstructured free responses to vignettes, checklists in which participants (or their parents or teachers) report which words they know, age of acquisition recalled in adulthood, and tasks in which participants must match emotion words with scripts describing the causes and consequences of emotions (Baron-Cohen, Golan, Wheelwright, Granader, & Hill, 2010; Li & Yu, 2015; O’Kearney & Dadds, 2004; Ponari, Norbury, & Vigliocco, 2018; Widen, Pochedly, & Russell, 2015). Together, these studies show that self-reported emotion comprehension rises across childhood, that the tendency to spontaneously produce more nuanced emotion terms such as “disappointed” rather than “sad” increases across adolescence, that emotions are more accurately associated with scripts than facial expressions, and that overall comprehension of abstract words increases across childhood and into adolescence. Although these methods provide important first steps to understanding emotion language in adolescence, they are limited by their naturalistic and/or self-assessed (rather than experimenter-assessed) natures. As such, an interview-based assessment could validate these patterns.

Greater clarity concerning emotion development in adolescence is needed, as adolescence is a pivotal time of social and emotional change (Somerville & McLaughlin, 2018), with adolescents reporting stronger negative affect, greater difficulty differentiating co-experienced emotions, and increased risk of psychopathology (Kessler et al., 2005; Larson, Moneta, Richards, & Wilson, 2002; Nook, Sasse, Lambert, McLaughlin, & Somerville, 2018). Consequently, assessing emotion comprehension in adolescence could provide insight into the nature of emotion language and emotion representations during this dynamic developmental period.

### **Developing Abstract Emotion Concepts**

As described above, emotion comprehension involves the ability to connect emotion words to culturally agreed-upon conceptual meanings. A separate but related question concerns how the concepts underlying these emotion terms vary across age. Studying emotion concepts themselves is of great interest because recent theories postulate that conceptual processes play a central role in the experience and perception of emotions (Barrett, 2006, 2017; Lindquist, 2017; Lindquist, MacCormack, et al., 2015; Lindquist, Satpute, & Gendron, 2015). Specifically, the constructionist theory of emotions posits that emotions arise when emotion concepts are used to parse ambiguous “core affect”—one’s internal sensations of valence (positivity and negativity) and arousal (activation and deactivation)—into specific emotion types (Barrett, 2006, 2017). A growing body of work supports this model by showing that priming, activating, or impeding emotion concepts can influence how emotions are experienced or perceived at both behavioral and neural levels of analysis (Doyle & Lindquist, 2018; Gendron, Lindquist, Barsalou, & Barrett, 2012; Lindquist, Barrett, Bliss-Moreau, & Russell, 2006; Nook, Lindquist, & Zaki, 2015; Oosterwijk, Lindquist, Adebayo, & Barrett, 2015; Satpute et al., 2016). In fact, a recent study showed that individual differences in emotion concept representations is related to perceptions of emotional facial expressions (J. A. Brooks & Freeman, 2018). Thus understanding emotion concepts is an important scientific goal, and there is reason to specifically explore how emotion concepts develop, as evidence suggests that the conceptual representations underlying emotion words are not static across age but rather become more multidimensional from childhood to adulthood, as one’s vocabulary increases (Nook, Sasse, et al., 2017).

Emotion theorists operating within the constructionist tradition often take a simulation-based approach to concepts, positing that concepts are ultimately grounded in simulations of prior sensory experiences (Barrett, 2017; Barsalou, 1999, 2003b, 2009; Wilson-Mendenhall, 2017; Wilson-Mendenhall, Barrett, & Barsalou, 2013). According to this theory, when thinking about the concept *apple*, the mind draws upon a network of relevant concepts (e.g., *round*, *red*, *food*, *sweet*) which are themselves tied to memories of prior sensory experiences to create an internal simulation of an apple. This internal simulation can then be used to produce other conceptual representations (e.g., *picnic* or *pie*). When applied to emotions, this theory proposes that people construct an experience of, for example, fear by drawing upon and organizing internal simulations of related concepts (e.g., *threat*, *uncertainty*, *escape*), which are ultimately grounded in perceptual memories. In other words, emotions arise through *situated conceptualizations* in which one's emotional experience is constructed by simulating combinations of concepts relevant to a given situation (Barrett, 2017; Wilson-Mendenhall, 2017). This dynamic process of situated conceptualization is thought to explain why completely different situations could be constructed as instances of the same emotion (e.g., the feelings aroused by being alone in the dark woods or being in front of a large audience in a brightly lit room could both be labeled "fear").

Historically, there has been a strict distinction between *concrete* and *abstract* concepts, with concrete concepts referring to things that have clear physical referents (e.g., objects such as apples and balls), and abstract concepts referring to general ideas that lack such clear physical referents (e.g., justice, truth, or freedom; Caramelli, Setti, & Maurizzi, 2004; Hale, 1988; Wang, Conder, Blitzer, & Shinkareva, 2010). Because emotions are not physically constrained objects, emotion concepts have historically been thought of as abstract concepts (see Altarriba & Bauer,

2004 for discussion). However, two key arguments have been made against a strict concrete vs. abstract dichotomy. First, the historical operational definition of abstraction is underspecified, merely defining it as “not concrete” (Barsalou, 2003a; Barsalou, Dutriaux, & Scheepers, 2018). Second, growing empirical and theoretical work suggest that this distinction is not binary but dimensional, such that *all* concepts appear to have some degree of both abstractness and concreteness (Andrews, Frank, & Vigliocco, 2014; Barsalou et al., 2018; Borghi & Binkofski, 2014; Borghi et al., 2017; Della Rosa, Catricalà, Vigliocco, & Cappa, 2010; Lupyan & Winter, 2018; Mahon & Caramazza, 2008). For example, when thinking of the concept *apple*, both concrete qualities that have clear physical boundaries (e.g., its shape and color) and more abstract qualities that lack such physicality (e.g., the psychological states of desire or beauty) can come to mind. Thus even this historically “concrete” concept blends both concrete and abstract aspects.

Consequently, Barsalou et al. (2018) argue that the historical distinction between concrete and abstract concepts should be abandoned and instead the relative abstractness/concreteness of a representation should be defined in terms of how much it focuses on *external* (i.e., physical, non-psychological) versus *internal* (psychological, non-physical) elements and how strongly it *integrates* these elements. In this paper we draw upon this formulation to postulate that concepts vary continuously in their levels of concreteness—abstractness, and we define relatively concrete representations as those that focus on external (i.e., physical and situationally-bounded) elements and relatively abstract representations as those that focus instead on internal (i.e., psychological and not-situationally-bounded) elements.

Interestingly, the argument that all concepts include concrete and abstract qualities converges with data showing that *emotion* concepts do not fit cleanly into either abstract or

concrete categories but instead show features of both concept types (Altarriba & Bauer, 2004; Caramelli & Setti, 2005; Mazzuca, Barca, & Borghi, 2017). One potential reason for this is that emotions can be conceptualized with either an emphasis on concrete external qualities (i.e., situational details that gave rise to specific instances of an emotion) or more abstract internal qualities that generalize across situations (i.e., the psychological principles that give rise to an emotion). For example, one could represent anger in terms of concrete external situational details (e.g., anger is the feeling caused by getting cut off in traffic) or by more abstract internal qualities that integrate across situations (e.g., anger is the feeling caused by one's goals being blocked). Relatedly, neuroimaging data show that emotion processing routinely activates brain regions thought to process both basic sensation and higher-order conceptualization (Kober et al., 2008; Lindquist et al., 2012; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011). Even though psychologists have long known that a given situation can be represented concretely or abstractly (Trope & Liberman, 2010), how this flexibility of representation relates to emotion conceptualization has received little attention in developmental, affective, or cognitive theory.

From a developmental perspective, it is reasonable to hypothesize that across childhood and adolescence the conceptual representations underlying emotion terms shift from a relative focus on external situational details to more internal psychological principles that generalize across situations (i.e., emotion concepts become more *abstract*). Such a hypothesis is an extension of classic Piagetian theories positing that psychological development proceeds from a concrete sensorimotor focus to a more abstract hypothetico-deductive focus (Demetriou et al., 2018; Inhelder & Piaget, 1958). However, there are additional reasons to hypothesize that emotion concepts become more abstract across age. First, data show that mastery of words corresponding to abstract non-physical internal phenomena increases across childhood into

adolescence (Caramelli et al., 2004; Joelson & Herrmann, 1978; Ponari et al., 2018). Second, the cognitive ability to extract general principles and integrate abstract information across specific situations—even at the non-verbal level—increases across age (Crone et al., 2009; Dumontheil, 2014; Ferrer, O’Hare, & Bunge, 2009; Raven, 2000; Whitaker, Vendetti, Wendelken, & Bunge, 2018). And third, as we age, the number and diversity of situations that produce emotions increases (e.g., only after growing to an age at which we give public speeches do we understand that this situation can evoke fear). This experiential diversity likely facilitates expanding emotion concepts from a focus on a narrow set of situations to deeper principles that generalize across situations. For these reasons, we hypothesize that emotion abstraction increases across development, and we use participants’ emotion definitions in an interview-based emotion vocabulary assessment to test this hypothesis.

### **An Automated Linguistic Measure of Abstractness**

Researchers frequently measure the abstractness of conceptual representations by asking human coders to rate how strongly they conform to a given definition of abstraction (e.g., Brysbaert, Warriner, & Kuperman, 2014; Della Rosa et al., 2010; Forgas, 2007; Gray, Parkinson, & Dunbar, 2015; Joelson & Herrmann, 1978; Paivio, Yuille, & Madigan, 1968). This method has been used to generate abstractness norms for thousands of words that are now widely-used in the literature (Brysbaert et al., 2014; Paivio et al., 1968). It can be employed with high inter-rater reliability, and resulting data are face-valid (e.g., children comprehend fewer words that are rated as abstract compared to adolescents and adults; Ponari et al., 2018). However, this method also suffers from some weaknesses: human raters may be biased by stimulus qualities other than its actual abstractness (e.g., the length of a word or the number of words in a response), the way

that raters understand abstractness can differ within and across studies, and human coding is a laborious and time-consuming process that scales with the size of a study's dataset. These challenges limit the reliability and practicability of measuring abstractness via human raters.

Fortunately, automated psycholinguistic techniques can help curtail these limitations. Quantifying abstractness purely based on the words in a response mitigates human bias, standardizes the mathematical computation of abstractness across studies, and can be automated to substantially reduce the time needed to compute abstractness. Semin and Fiedler (1988, 1991) developed a psycholinguistic approach to measuring abstractness called the Linguistic Category Model (LCM; see also Carnaghi et al., 2008; Seih, et al., 2017; Semin, Görts, Nandram, & Semin-Goossens, 2002 for contemporary updates to this theory). Although this model was developed to quantify how abstractly participants represented other people, extant work suggests that this approach produces face valid measures of abstractness in text more broadly (e.g., asking participants to take a distanced perspective on a situation increased linguistic abstractness scores, as would be expected; Seih et al., 2017; Trope & Liberman, 2010). In this model, five word classes are ranked in their levels of abstractness, with use of three kinds of verbs denoting more concrete representations and use of adjectives and nouns denoting more abstract representations (see **Methods** and Carnaghi et al., 2008 for details). For example, defining sadness as “what you feel when you *cry*” (a descriptive action verb) suggests a more concrete representation than defining it as “feeling *sorrowful* in response to a *loss*” (an adjective and a noun).

Here, we used this psycholinguistic method (as well as human ratings) to measure the abstractness of emotion concept representations. Not only does this linguistic approach provide a method for validating human coders' ratings of emotion abstractness, it also lays the foundation for emotion abstractness to be quantified automatically (i.e., without human input).

Such a tool would reduce the burden of empirical research on emotion concept abstraction and allow for rapid processing of large-scale datasets, an approach that has been fruitful in other areas of research (e.g., Doré, Ort, Braverman, & Ochsner, 2015; Fan et al., 2019; Franz, Nook, Mair, & Nock, 2019).

### **The Current Study**

The current project seeks to address the theoretical gaps outlined above by investigating the developmental changes in emotion comprehension and abstraction from early childhood to early adulthood. A broad, performance-based assessment of emotion comprehension (called the Emotion Vocabulary Assessment) was developed by extending the structure of the WASI vocabulary and KAI assessments (Kusché et al., 1988; Wechsler, 2011). This assessment was administered to a cross-sectional sample of participants spanning childhood, adolescence, and early adulthood (i.e., ages 4-25). Trained examiners asked participants to define 24 emotion words, probed their comprehension of each emotion word, and scored their responses on a 0-2 scale. Performance on this assessment constituted participants' emotion comprehension scores.

To allow for direct comparisons of emotion comprehension and abstraction, these measures were assessed in the same dataset. Following data collection, participants' responses were transcribed, and trained coders rated the abstractness of these responses. Coders also identified the concrete and abstract "strategies" participants used to describe the meaning of emotion terms. Finally, a purely linguistic measure of abstraction based on the LCM was used to corroborate coders' abstractness ratings and provide a first step towards the automatic (i.e., non-human) computation of emotion concept abstraction.

## Methods

### Participants

Two hundred three participants enrolled in the study. Data from seven participants were unusable and thus excluded (one did not complete the emotion vocabulary assessment, five did not understand and/or cooperate with task instructions, and audio recordings were inaudible for one participant). Hence, analyses included data from 196 participants (age range = 4.13-25.91,  $M_{\text{age}} = 14.53$ ,  $SD_{\text{age}} = 5.81$ , 51.02% female, 64.29% Caucasian, three participants did not disclose race, range of income-to-needs ratio = 0.07-13.32,  $M_{\text{income-to-needs}} = 4.93$ ,  $SD_{\text{income-to-needs}} = 3.09$ , 9.18% below poverty line, 25 participants did not report data required for computing income-to-needs ratio<sup>1</sup>). Linguistic abstractness scores were excluded from seven participants due to insufficient linguistic data (i.e., usable definitions did not include enough words to produce valid linguistic abstractness scores, see **Supplemental Materials**). Participants were recruited from areas surrounding Harvard University and the University of Washington. All participants were fluent in English with English as their first language. Participants did not have cognitive impairments that would limit their ability to provide consent/assent or complete tasks. Adult participants and guardians of minor participants provided informed consent prior to participation, and minor participants assented to participation. All methods were approved by the Institutional Review Boards of Harvard University (IRB#15-2214: “Development of fundamental emotion processes”) and the University of Washington (IRB#50239: “Typical emotional development”).

Given the absence of published data on the development of emotion abstraction, an *a priori* power analysis was not possible. Hence, we ensured that the sample size was sufficiently powered to detect small-to-medium effects (i.e.,  $\beta > .20$ ; Cohen, 1988). A power analysis

---

<sup>1</sup> Data concerning race and socioeconomic status were provided by adult participants themselves and from parents of participants who were minors.

suggested that 194 participants would be required to detect an effect of this size at 80% power. A *post hoc* power analysis verified that our final sample ( $N = 196$ ) was sufficiently powered (power = 90%) to detect the study's smallest age-related effect (i.e., the overall linear decrease in physiological marker use across age,  $\beta = .23$ ). Data for this study can be accessed at [https://osf.io/m6ue7/?view\\_only=1a06da0cfd564f55bc80cb7a6054232f](https://osf.io/m6ue7/?view_only=1a06da0cfd564f55bc80cb7a6054232f).

### **Emotion Vocabulary Assessment**

A performance-based assessment of emotion vocabulary was created by adapting the format of the vocabulary test of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 2011). In this assessment (see **Supplemental Materials**), participants were asked to define 27 emotion terms, and their responses were scored for comprehension. For each emotion word, the experimenter stated the term aloud, presented a card displaying the written emotion word to the participant, and asked “what does \_\_\_ mean?” Responses were self-paced. To thoroughly probe participants' understanding of emotion terms, participants were prompted to provide more information if their initial response did not demonstrate full comprehension. Responses were audio-recorded and later transcribed.

Participants were asked to define the emotions amazed, angry, annoyed, bored, calm, disappointed, disgusted, embarrassed, excited, grumpy, happy, hate, jealous, lonely, love, nervous, pleased, proud, relaxed, sad, safe, scared, sorry, surprised, thankful, upset, and worried. These emotions were selected to include (i) basic emotions (Ekman & Friesen, 1971), (ii) emotions that correspond to all four quadrants of the circumplex dimensions of valence and arousal (Russell, 1980), and (iii) a mix of emotions that prior work suggested young children can identify (e.g., angry, happy, sad) as well as some that may emerge later in

development (e.g., disappointed; Baron-Cohen et al., 2010). Data from three words—sorry, thankful, and pleased—were not included in measures of emotion comprehension or abstraction. The decision to exclude these three words was made following data collection but before human coders produced measures of emotion abstraction. This decision was made (i) to reduce burden of human coding and (ii) because these three words were considered less central to classical conceptualizations of what constitutes an emotion (i.e., in contrast to classic labels of emotion states, sorry and thankful refer more to interpersonal expressions of affect, and pleased refers more to one's broad level of positively-valenced affect). Emotion words were presented in random order for each participant.

Trained experimenters scored participants' responses during the testing session. A score of 2 points indicated full comprehension, 1 point indicated partial comprehension, and 0 points indicated no comprehension. To earn 2 points, the participant's response must have included (i) a general definition of the emotion word (i.e., an abstract explanation of the general causes or characteristics of the emotion that was not bound to a specific situation; e.g., "people feel sad when they experience loss"), (ii) a synonym of the emotion word (e.g., "sad means feeling sorrowful"), or (iii) an example situation that would likely give rise to that emotion and not others (e.g., "I felt sad when my pet died"). During the assessment, experimenters could refer to a scoring guide that included example definitions and synonyms for each emotion word (see **Supplemental Materials**). Responses earned 1 point if they were of the correct valence but were overly vague and did not specifically describe the emotion in question. For example, the responses "calm means you think things are good" is too vague to specifically define calmness, and "I felt angry when my sister got something that I wanted" is a situation that could produce either anger or jealousy. Hence, these responses would both receive 1 point. Responses earned 0

points if they described emotions of the incorrect valence or if the participant responded by saying “I don’t know.” An *emotion comprehension score* was computed for each participant by summing their scores across all trials and converting this sum to a percentage of the total possible score. Because all participants included in analyses completed the entire emotion vocabulary assessment, the total possible score for every participant was 48 points (2 points for 24 emotions). We converted participants’ scores to percentages by dividing participants’ scores by this total possible score and multiplying by 100.

### **Human Coder Assessment of Emotion Abstractness**

Following transcription, trained coders rated the abstractness of participants’ responses on the emotion vocabulary assessment. First, coders scored each response on three 10-point scales that provided numerical assessments of abstractness. These three scales drew upon extant definitions of abstraction (Burgoon, Henderson, & Markman, 2013; Joelson & Herrmann, 1978) as the extent to which responses focused on generalizable, internal principles rather than external situational details (see also Barsalou et al., 2018). The scales included the questions: (i) “how abstract is this response” (1 = *extremely concrete*, 10 = *extremely abstract*, where abstract was defined for raters as “the extent to which responses referred to general concepts, thoughts, or ideas, and not bound to specific contexts or stimuli”), (ii) “how much does this response depend on thoughts and ideas about the general principles that produce this emotion?” (1 = *not at all*, 10 = *extremely*), and (iii) “how much does this response focus on concrete situation(s) or characteristics that produce this emotion?” (1 = *not at all*, 10 = *extremely*). The third scale was reverse-scored, and ratings on these three scales were averaged to produce a single *coder abstractness score* for each response.

Second, coders indicated which of four “strategies” the participant used to describe each emotion. Three of these strategies came from the *a priori* list that participants could use to obtain a 2-point response (i.e., providing a general definition, example situation, or synonym). To assess how strongly participants conceptualized emotions in terms of bodily sensations, coders also indicated whether or not each response referenced a “physiological marker” (i.e., it described a bodily sensation or physiological change associated with a given emotion; e.g., “people cry when they are sad”). Responses could employ multiple strategies.

To ensure that ratings were reliable, the second author coded all responses, and two research assistants provided a second rating for all responses (i.e., each provided a second rating of approximately half of the responses). Coders were not informed of participants’ ages during coding, and vocal cues as to a participant’s age could not bias coding because coders only had access to transcriptions of participants’ responses. Coders also underwent extensive training before beginning the coding process to increase reliability. Measures of inter-rater reliability for the three abstraction score ratings demonstrate strong reliability. Across all responses, correlations between coders for the three human rated abstractness scales were  $r = .81$  (“how abstract is this response?”),  $r = .79$  (“how much does this response depend on thoughts and ideas about the general principles that produce this emotion?”), and  $r = .90$  (“how much does this response focus on concrete situation(s) or characteristics that produce this emotion?”). The correlation between coders for the average of these three scales was  $r = .89$ , and Cronbach’s alpha across the three scales was  $\alpha = .93$ . Hence, the *coder abstractness score* for each response was averaged across the two coders to provide a single score for each response. Coders came to full agreement on which strategies were used in each response through discussion.

It was determined *a priori* that only definitions that earned 2 points would be included in abstractness analyses. This ensured that estimates of emotion abstraction only included responses that accurately described each emotion type. Crucially, however, full comprehension could be demonstrated both through abstract (e.g., providing a general definition or synonym) and concrete (e.g., providing an example situation) strategies, so restricting abstraction analyses to emotion words that participants fully comprehended did not inherently bias these analyses. Coder abstractness scores were averaged across all available trials within each participant (i.e., participant-level coder abstractness scores were computed without scores for any emotions on which the participant did not receive a comprehension score of 2). We similarly computed the percentage of trials for which each participant used each of the four strategies. We call these percentages *general definition use*, *synonym use*, *example situation use*, and *physiological marker use*. Because younger participants comprehended fewer emotions than older participants (see **Results**), different emotions could have been included in estimates of emotion abstraction across age. Although we conceptualized this as a part of the phenomenon that this study investigates, we conducted supplemental analyses to ensure that dependent variables that showed emergent or linear patterns underwent significant developmental changes even when the sample was restricted to participants who fully comprehended emotion terms (see **Supplemental Materials**). Hence, age-related differences in emotion comprehension do not confound analyses of emotion abstraction reported in the main text.

### **Linguistic Assessment of Emotion Abstractness**

To corroborate human coded ratings of abstractness, we also produced a linguistic measure of abstractness by drawing upon the Linguistic Category Model (LCM; Carnaghi et al.,

2008; Seih et al., 2017; Semin & Fiedler, 1988, 1991; Semin et al., 2002). In this model, word classes are sorted into 1 of 5 levels of abstraction. Abstraction level 1 includes descriptive action verbs (i.e., verbs describing a specific action that can be constrained to a single instance; e.g., “die” in a definition of sad). Abstraction level 2 includes interpretive action verbs (i.e., verbs describing general behaviors that cross specific instances; e.g., “withdraw” in a definition of sad). Abstraction level 3 includes state verbs (i.e., verbs involving mental states without clear beginning or end; e.g., “mourn” in a definition of sad). Abstraction level 4 includes adjectives (i.e., references to qualities that require considering mental states or abstracting across several instances; e.g., “depressed” in a definition of sad). Abstraction level 5 includes nouns (references to classes of things; e.g., “loss” in a definition of sad). A mathematical formula is used to compute a passage’s abstractness by comparing the proportion of words that fall within each abstraction level (see Seih et al., 2017 for details). Scores range from 1 to 5, with higher values indicating higher levels of abstractness.

Following prior work (Pennebaker, Booth, & Francis, 2007; Schmid, 1994; Seih et al., 2017), we used the Linguistic Inquiry and Word Count (LIWC) and TreeTagger software to extract the number of words in each response that fell within each level of linguistic abstraction. We then computed a *linguistic abstractness score* for each response using the formula provided by Seih et al., (2017). To parallel the other measures of emotion abstractness, we computed LIWC measures for responses that earned 2 points. Linguistic abstractness scores for all qualifying responses were then averaged within participants to produce participant-level linguistic abstractness scores. Only two trials did not have any words in abstraction levels 1-5, so a linguistic abstraction score could not be computed. These trials were treated as missing values and not included in linguistic abstractness scores for those participants.

A series of exploratory analyses suggested that linguistic abstractness scores may not have been valid for very young children in our sample (see **Supplemental Materials**). Coherence between linguistic abstractness scores and coder abstractness scores broke down at young ages, and this lack of coherence was explained by the fact that some young participants had few words on which a linguistic abstractness score could be computed (likely because only words from responses that showed full comprehension could be analyzed, and younger children comprehended fewer emotions). Hence, we excluded linguistic abstractness scores from seven participants who had fewer than 100 total words in their usable verbal responses during the assessment. Excluding these participants ensured that the correlation between coder and linguistic abstractness scores was stable across age.

### **Hypotheses and Analyses**

**Interrelations between dependent variables.** In all, this study produced one measure of emotion comprehension, two measures of emotion abstraction (i.e., coder abstractness ratings and linguistic abstractness scores), and four measures of definitional strategies (i.e., general definition use, synonym use, example situation use, and physiological marker use). We hypothesized that people who comprehended more emotion words would also represent emotions more abstractly. We hypothesized that the two measures of emotion abstractness would converge: coder abstractness scores and linguistic abstractness scores should positively correlate. We hypothesized that general definition use and synonym use would track increased abstractness (i.e., they should correlate positively with coder and linguistic abstractness scores) because providing a general definition for an emotion indicates understanding of the internal principles that give rise to emotions across situations and providing synonyms denotes an

understanding of the taxonomic relations between emotion words (i.e., knowing that two words can be used to refer to the same concept). Finally, we hypothesized that example situation use and physiological marker use would track increased concreteness (i.e., they should correlate negatively with coder and linguistic abstractness scores) because providing example situations and physiological markers involves attending to concrete physical details of emotional experiences.

To test these hypotheses, Spearman's correlations assessed the significance and direction of all pairwise relations between these variables at the participant level. Note that because measures of emotion abstraction were only valid for responses on which the participant showed full comprehension of emotion terms, these analyses involved correlating participant-level emotion comprehension scores (i.e., how well participants comprehended the 24 emotion terms we assessed) with participant-level emotion abstraction scores (i.e., how abstractly participants represented the subset of emotions that they fully comprehended). Spearman's correlations were used for all analyses because of the skewed distribution of all dependent variables other than linguistic abstractness scores. Given the number of correlation analyses conducted, we used the Benjamini & Hochberg (1995) false discovery rate (FDR) method to correct for multiple comparisons.

**Overall developmental changes in emotion comprehension and abstraction.** Based on prior work (e.g., Baron-Cohen et al., 2010; Ponari et al., 2018), we hypothesized that both emotion comprehension and emotion abstraction would follow emergent age-related patterns (i.e., they would increase across early development before reaching a plateau; Somerville et al., 2013). We also hypothesized that emotion comprehension would plateau earlier in development than emotion abstraction. We did not have strong hypotheses concerning potential differences

between age-related patterns of the two measures of emotion abstractness and use of the four strategies except that the two measures in which higher scores indicated less abstraction (i.e., example situation use and physiological marker use) should show “inverse emergent patterns” (or *diminishing* patterns) that decreased across childhood and adolescence before plateauing at their nadir.

To test whether measures of emotion comprehension and abstraction indeed followed emergent or diminishing patterns of change (specific kinds of non-linear shapes), we analyzed each dependent variable following the non-linear analytic methods of Rodman, Powers, & Somerville (2017). This approach tests whether “traditional” polynomial patterns (i.e., linear, quadratic, and cubic models) or more complex nonlinear patterns (i.e., spline-based models) best fit age-related patterns in the data. Each dependent variable was subjected to four models that assessed for (i) linear, (ii) quadratic, (iii) cubic, and (iv) other non-linear relations with age. Linear, quadratic, and cubic analyses were tested within linear regression frameworks using the *poly* function in R’s *stats* package (R Core Team, 2016). The *poly* function transforms age into orthogonalized linear, quadratic, and cubic regressors. Linear, quadratic, and cubic regressions were each tested in separate regression models, but lower-order regressors were included in higher-order models (i.e., the linear regressor was included in the quadratic analysis, and both the linear and quadratic regressors were included in the cubic analysis). Finally, we used generalized additive modeling approaches to test for age-related patterns that did not fit linear, quadratic, or cubic patterns. Specifically, thin plate regression smoothing spline analyses were used to produce regression equations that fit the data using cross-validation procedures but were also penalized for the number of parameters to prevent overfitting. The result of these models is a stable smooth curve that describes the data’s age-related patterns but is not constrained to

stereotyped linear, quadratic, or cubic shapes. Spline analyses were conducted using the *mgcv* package in R (Wood, 2003, 2017).

The four models (linear, quadratic, cubic, and spline) were compared to each other as well as a null model containing no age predictors using Akaike Information Criterion (AIC) values (Akaike, 1974). AIC values provide a measure of goodness of fit that takes into account the number of parameters, and lower AIC values indicate better model fit after penalizing for the number of parameters. Adjusted  $R^2$  values for each model are also reported as additional measures of goodness of fit for each model. These values summarize the proportion of variance explained by each model after adjusting for the number of predictors.

When spline regressions were the best fitting model, we visualized the resulting regression lines and found that they conformed to the “emergent” pattern defined by prior literature (Casey, 2015; Somerville et al., 2013) for four dependent variables. Emotion comprehension scores, coder abstractness scores, general definition use, and example situation use showed rapid change in the early portion of the study’s age range before stabilizing (see **Results**). We used a data-driven method to identify the age at which these variables reached their plateaus (see **Supplemental Materials** for details on method development). Two criteria identified plateaus: (i) reduced rate of change in the dependent variable and (ii) the dependent variable reached the maximum/minimum of its development. First, we extracted the best-fit line summarizing age-related change for each variable and computed its first derivative (Simpson, 2014). Because the first derivative quantifies the slope of a curve, the plateau of an emergent curve can be defined as the earliest age at which the first derivative approaches 0 (i.e., the curve flattens). Hence, we set the thresholds for a plateau as occurring when the change in the dependent variable slowed to less than 0.5%/year (for emotion comprehension scores and

strategy use, which are measured on a 100-point scale) or 0.05/year (for coder abstractness scores, which are measured on a 10-point scale).

Second, to prevent “plateaus” from being incorrectly identified at points where the slope of the curve slowed temporarily, we ensured that plateaus reflected the age at which the dependent variable could be considered fully developed. In other words, we wanted to ensure the plateau occurred at a point where the dependent variable had finished increasing to its maximum (or falling to its minimum) and then remained stable. We consequently added a second criterion such that the value of the dependent variable at the plateau must have been near the maximum value (or minimum value, for example situation use, which fell across age). “Near” the maximum or minimum value was operationalized as falling within the 95% CI of the smoothing spline’s estimate of the maximum (or minimum) value of the curve.

Using coder abstractness scores as an illustration (see **Figure 1b**), the curve slowed to < .5%/year at age 17.69, satisfying the first criterion. To check the second criterion, we next evaluated whether this plateau point reflected the conclusion of age-related change (rather than a temporary flattening). The 95% CI of coder abstractness scores around this point were 8.98-9.56 (i.e., the grey shaded region around the maximal point of the curve). Because the value of the curve at age 17.69 (i.e., 9.27) was within this 95% CI, this age was deemed the plateau point.

We used nonparametric bootstrapping methods to test for significant differences in the ages at which each of these variables plateaued. We conducted 10,000 bootstrapped simulations of our data using the *boot* package in R (Canty & Ripley, 2017; Davison & Hinkley, 1997). Within each simulated sample, we (i) conducted spline analyses of emotion comprehension scores, coder abstractness scores, general definition use, and example situation use; (ii) used the first derivative method described above to identify the age of each variable’s plateau; and (iii)

computed pairwise differences between the 4 plateau ages. The 95% bias-corrected and accelerated ( $BC_a$ ) confidence intervals (CIs) of these differences were computed, and those that did not include 0 were considered significant (DiCiccio & Efron, 1996). We also used these bootstrap simulations to compute estimates of the 95%  $BC_a$  CIs of each plateau point for descriptive purposes.

**Development of emotion comprehension and abstraction for each emotion.** To provide descriptive estimates of emotion comprehension and abstraction for each of the 24 emotion words assessed in this study, we subjected emotion comprehension scores and coder abstractness scores for each emotion to smoothing spline analyses. We extracted the lines of best fit for these spline models and visually presented estimates of these values. These descriptive data are provided to allow for speculative interpretations of age-related changes for each emotion and to offer norms that can guide future research.

**Control analyses.** We conducted two sets of analyses to control for potential confounds. First, it was possible that abstractness measures were biased by the length of participants' responses. To address this concern, we used LIWC to extract the word count of each response and tested (i) whether average word count varied across age and (ii) whether controlling for word count eliminated age-related variation in other dependent variables. Second, it is possible that providing a synonym does not actually reflect comprehension of an emotion term but instead is merely evidence that participants could produce terms semantically associated with that emotion word. In other words, providing synonyms might reflect shallow semantic priming in which participants produced words they knew were associated with the target emotion word but did not actually comprehend the concept underlying either word. To address this concern, we computed

the number of trials on which participants only provided synonyms, and we investigated whether the study's results differed when these trials were treated as incorrect responses.

## Results

### Interrelations Between Dependent Measures

Spearman's correlations indicated that participants who showed higher emotion word comprehension had more abstract emotion representations (**Table 1**). Within emotion abstraction measures, modest convergence emerged between human ratings of emotion abstractness and automated linguistic measures of abstractness (**Table 1**). Additionally, the direction of correlations supported the hypotheses that general definition use and synonym use tracked increased abstractness, whereas example situation use and physiological marker use tracked decreased abstractness. All of these relations were significant and survived FDR correction for multiple comparisons except: (i) emotion comprehension scores showed only a trending relationship with physiological marker use ( $p = .076$ ) that did not survive correction, and (ii) synonym use did not correlate significantly with physiological marker use.

### Age-related Change in Emotion Comprehension and Abstraction

Results of analyses using linear, quadratic, and cubic polynomial models, as well as thin plate regression smoothing spline models are presented in **Table 2**. AICs of null models are provided for comparison.

**Emotion comprehension scores.** The spline model produced the best fit for emotion comprehension scores (**Table 2**), revealing that emotion comprehension scores followed an

emergent pattern across age (**Figure 1a**). The first derivative of the spline curve suggested that development of emotion comprehension plateaued at age 10.95, 95% CI = [8.76, 12.11].

**Coder abstractness scores.** The spline model produced the best fit to coder abstractness scores (**Table 2**). Coder abstractness scores also followed an emergent age-related pattern, but these scores plateaued at age 17.69, 95% CI = [16.06, 19.17] (**Figure 1b**). A nonparametric bootstrapping analysis suggested that this plateau was significantly later than the emotion comprehension plateau, 95% CI = [4.83, 9.25].

**Linguistic abstractness scores.** Both the linear and spline model suggested that linguistic abstractness scores followed a linear age-related pattern (**Table 2, Figure 2**) after removing 7 participants with little linguistic data (see **Supplemental Materials**). Hence, this purely linguistic measure of abstractness also suggested that emotion representations became more abstract from childhood to early adulthood. However, the best-fitting shape of this relation was linear rather than emergent, and the strength of the relation between age and linguistic abstractness scores was smaller than the relation between age and coder abstractness scores.

**General definition use.** General definitions were the most frequent strategy participants used to describe emotions (81.21% of trials). A spline model provided the best fit to these data, revealing that general definition use followed an emergent age-related pattern (**Table 2, Figure 3a**). General definition use plateaued at age 17.19, 95% CI = [13.40, 18.50]. This plateau was significantly later than the plateau of emotion comprehension, 95% CI = [4.08, 9.00], but it did not significantly differ from the plateau of coder abstractness scores, 95% CI = [-1.55, 4.50].

**Synonym use.** Participants used synonyms to describe emotions on 33.63% of trials. Both linear and spline models suggested that synonym use followed a linear increase across development (**Table 2, Figure 3b**).

**Example situation use.** Example situations were used to describe emotions on 26.22% of trials. A spline model provided the best fit to the data, revealing that use of this strategy followed a diminishing pattern (**Table 2**). Example situation use fell across childhood and adolescence before reaching a plateau at age 18.19, 95% CI = [16.67, 22.06]<sup>2</sup> (**Figure 3c**). This plateau was significantly later than the plateau of emotion comprehension, 95% CI = [5.32, 11.78], but it did not significantly differ from the plateaus of coder abstractness scores, 95% CI = [-5.18, 0.59], or general definition use, 95% CI = [-1.17, 6.06]. Thus, all plateau ages for measures of emotion abstraction occurred significantly later than the emotion comprehension plateau but did not differ significantly from each other.

**Physiological marker use.** Physiological markers were used to describe 7.29% of trials. A significant linear regression suggested that use of this strategy decreased across age, but the spline model provided the best fit to the data (**Table 2**). This model suggested that physiological marker use followed a non-linear but also not emergent pattern of age-related change (**Figure 3d**). Use of physiological markers was elevated in mid-childhood but low at other ages. First derivative analyses suggested that physiological marker use increased from age 4.13 to age 10.00, then decreased until age 15.05, and then remained low through the rest of the study's age range.

**Controlling for word count.** A linear regression at the participant level showed that average word count of participants' responses increased with age,  $\beta = .21, p = .004$ . However,

---

<sup>2</sup> A plateau could not be identified in 0.84% of bootstrapped simulations for example situation use. In these simulations, the slope decreased across the entire sample and did not slow to the rate that would qualify as a plateau, suggesting that the plateau age was not reached before age 25.91. Although these simulations were excluded from confidence interval estimates, we ensured that replacing these missing values with ages above 26 did not affect the significance of inferences drawn from confidence interval comparisons (i.e., these exclusions do not affect the conclusion that the plateaus of the three abstraction measures do not differ significantly from each other, and they all differ significantly from the comprehension plateau).

differences in the length of participants' responses across age did not confound measures of emotion abstraction. Linear regressions examining coder abstractness scores, general definition use, and example situation use (from childhood until the coder abstraction score plateau of age 17.69) showed that age remained significantly related to these dependent variables even after controlling for the average word count of participants' responses,  $ps < .001$ . Similarly, age was significantly related to linguistic abstractness scores and synonym use across the entire sample even after controlling for average word count,  $ps < .001$ . A series of parallel analyses at the trial level (i.e., mixed-effects models that nested data within participants and examined whether each definition's abstraction metrics were related to participants' age after controlling for the length of each response) also showed that age remained a significant predictor of these dependent variables after controlling for word count,  $ps < .001$ . As such, the length of participants' responses did not confound any of the study's key findings.

#### **Control analyses investigating trials on which participants only provided synonyms.**

A second set of control analyses ensured that results were not significantly affected by the possibility that only providing a synonym for an emotion word may reflect shallow semantic priming rather than full comprehension of an emotion word's underlying concept. Across all trials, 4.49% of trials involved *only* providing a synonym (4.90% of 2-point responses), and removing these trials did not affect conclusions presented in the main text (see **Supplemental Materials** for details).

#### **Age-related Change for Each Emotion**

Visual representations of fits provided by spline models of emotion comprehension scores and coder abstractness scores are shown in **Figure 4**. These analyses are presented for

descriptive purposes. Observable trends suggest that emotion comprehension at age 4 varied substantially across emotions (e.g., emotion comprehension scores were > 80% for love but < 10% for amazed). The age at which emotion comprehension approached ceiling (i.e., exceeded 90%) also varied across emotions (e.g., around age 6 for scared and safe but around age 10 for calm). However, inter-emotion variability was even greater for age-related patterns of emotion abstraction. Whereas coder abstractness scores approached ceiling (i.e., exceeded 9 of 10) as early as age 13 for hate, disappointed, and love, comparable levels only emerged around age 20 for proud and annoyed. This threshold was never reached for nervous.

### **Discussion**

The current study assessed the development of emotion word comprehension and emotion concept abstraction from childhood to adulthood. Emotion comprehension and abstraction both followed emergent age-related patterns of change that increased with age before reaching a plateau. However, emotion comprehension plateaued significantly earlier (i.e., around age 11) than emotion abstraction (i.e., around age 18). A linguistic measure of abstractness was computed from the words participants used to describe emotions, and this measure also suggested that emotion concepts became increasingly abstract across age. Finally, age was related to the strategies participants used to describe emotions. The tendency to use abstract strategies increased across age, with general definition use following an emergent pattern and synonym use following a linear pattern. Conversely, use of concrete strategies decreased across age, with example situation use following a diminishing pattern and physiological marker use showing a non-linear pattern that was slightly elevated in childhood. We discuss the implications of these findings below.

### **An Interview-Based Method for Measuring Emotion Comprehension**

This study offers a novel tool for assessing emotion comprehension, the emotion vocabulary assessment. This interview adapts techniques from the WASI vocabulary assessment (Wechsler, 2011) and updates the approach used by the emotion definition task of the KAI (Kusché et al., 1988). Other than the KAI, several innovative methods have been used to study emotion language in childhood and adolescence. These methods include (i) coding conversations between children and their parents or their peers (Aznar & Tenenbaum, 2013; Bretherton & Beeghly, 1982; Cervantes & Callanan, 1998; J. Dunn, Bretherton, & Munn, 1987; Lagattuta & Wellman, 2002; MacWhinney & Snow, 1990), (ii) asking participants to verbalize how they or other people would feel in emotional situations described by vignettes (O’Kearney & Dadds, 2004, 2005), (iii) asking participants to name as many emotion terms as they can think of until they have exhausted all known terms (Beck et al., 2012; Duncombe et al., 2013; Kusché et al., 1988; Locke et al., 2015), (iv) asking children, adolescents, or their parents to self-report on their emotion word comprehension (Baron-Cohen et al., 2010; Li & Yu, 2015), (v) asking adults to recall the age at which they acquired words (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Stadthagen-Gonzalez & Davis, 2006), (vi) using written tests in which participants identify which emotion would most likely be produced by specific situations (MacCann & Roberts, 2008), (vii) using lexical decision tasks to test whether participants can identify emotion words vs. non-words (Ponari et al., 2018), (viii) asking participants to match emotion words with scripts describing situational causes and consequences of emotions (Balconi & Carrera, 2007; Reichenbach & Masters, 1983; Russell & Widen, 2002; Smith & Walden,

2001; Widen et al., 2015; Widen & Russell, 2010a, 2010b), and (ix) asking participants to define non-emotional and emotional terms in writing (Caramelli & Setti, 2005).

There are pros and cons to each these methods. For instance, some provide naturalistic assessments of emotion word use (e.g., methods i and ii) whereas others offer more focused examination of whether or not participants comprehend a specific set of emotion words (e.g., method vi). Some rely on participants' own self-assessment of whether they know emotion words (e.g., method iv), and others involve experimenter-assessed tests of comprehension (e.g., methods vi-ix). Finally, some of them offer insight into how participants define emotions (e.g., method ix) whereas others do not (e.g., methods i-viii).

The emotion vocabulary assessment developed here expands upon these prior methods to assess comprehension of emotion words through an interview format, which allows experimenters to ask follow-up questions in real time and thereby fully probe participants' comprehension. The breadth and thoroughness of this assessment offers several strengths. For example, prior work on emotion word comprehension across childhood and adolescence relied on participants' self-assessment of their emotion comprehension, which is susceptible to demand characteristics (Baron-Cohen et al., 2010; Li & Yu, 2015). Hence, the present study used an experimenter-assessed method to validate that emotion word comprehension emerges across early childhood and plateaus around age 11. Additionally, verbally probing and recording participants' emotion definitions allows for deeper insight into how people represent the concepts underlying emotion words. Although we studied the abstraction of these concepts, this method could be used to study other facets of emotion concept development.

When comprehension patterns for individual emotions were examined, we found that comprehension of all 24 emotions assessed in this project reached 90% comprehension within a

4-year window (ages 6-10). Hence, these years represent a particularly active period of maturation in emotion word comprehension. However, we also found variability across emotions both in levels of comprehension at age 4 and in the age at which they approached full comprehension. As such, the overall plateau point we identified in this study was influenced by the specific emotions included in this assessment, and this value should not be taken as absolute.

Although we caution against over-interpreting patterns for each emotion, they suggest that some emotion words (e.g., love, scared, excited, angry, sad, happy) may be comprehended before more “social” (e.g., proud, embarrassed, jealous) or “nuanced” emotion words (e.g., nervous, annoyed, amazed). These patterns accord with the notion that a core set of emotion concepts arise early in development and differentiate into more nuanced categories across age (Bridges, 1930; Nook, Sasse, et al., 2017; Widen, 2013; Wu, Muentener, & Schulz, 2017). Interestingly, this pattern aligns with Rosch’s overarching cognitive theory that “basic” levels of categories (i.e., those that are most representative of a category and most frequently used by caregivers; e.g., “fear”) are learned early in development, and subordinate category members (i.e., those that describe more nuanced levels of a category; e.g., “nervousness”) are mastered with greater experience across age (Mervis & Rosch, 1981; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Additionally, the fact that classic “social” emotions were later-emerging than non-social emotions also converges with work showing that theory of mind shows protracted maturation in social emotional development (Blakemore, 2008; Burnett, Bird, Moll, Frith, & Blakemore, 2009; Choudhury, Blakemore, & Charman, 2006; Sebastian et al., 2012). Future research that specifically manipulates these factors (e.g., by comparing social and non-social emotions) using lists of emotion words that are matched for word length and other salient dimensions could further test this notion.

Although emotion vocabulary assessment provided several benefits beyond prior methods, two weaknesses of this interview-based approach should be noted. First, we originally counted synonyms as demonstrating full comprehension because (i) Wechsler vocabulary assessments (on which this test was modeled) accept synonyms as correct responses (Wechsler, 1999) and (ii) providing synonyms demonstrates at least some understanding of how words within a taxonomy are organized (i.e., that two words refer to the same concept and not different concepts). However, it is possible to provide a synonym but not fully comprehend the concept underlying the target word. Even though control analyses suggested that this possibility did not invalidate this study's results, researchers may nonetheless wish to remove this criterion in future work. Second, success on interview-based assessments likely relies on several cognitive and social factors beyond emotion comprehension and abstraction. These factors include developmental increases in general vocabulary and working memory (which could potentially allow older participants to more easily organize and verbally express their thoughts; Farkas & Beron, 2004; Siegel & Ryan, 1989), as well as increased awareness of the "social script" that one should provide abstract definitions during vocabulary assessments. Although these weaknesses exist for all developmental studies that use interview-based methods, future work could control for these factors by measuring them all in one study or using non-emotion terms as control items.

### **Emotion Abstraction: Implications for Affective, Developmental, and Cognitive Theory**

Because the emotion vocabulary assessment collected participants' verbal definitions of emotions, data from this study could also be used to investigate the phenomenon of emotion concept abstraction. Scholars have long known that non-verbal and verbal abstraction skills grow across development and that multiple components of emotion understanding also develop

(Ferrer et al., 2009; Nook, Sasse, et al., 2017; Pons, Harris, & de Rosnay, 2004; Widen, 2013). However, insight into whether these two faculties interact has been lacking (though see Waggoner & Palermo, 1989). The current findings suggest that these processes indeed converge: emotion concepts become more abstract from childhood to adolescence. We observed that emotion representations became more abstract through childhood and adolescence. In particular, emotion word definitions shifted from understanding emotions in terms of external concrete situations to understanding them in terms of internal principles that generalized across situations (Barsalou et al., 2018). Further, we found that emotion abstraction was a later-emerging phenomenon than emotion word comprehension. Control analyses showed that these results were not driven by differences in response length across age.

The notion that emotion words can be associated with relatively more concrete or abstract representations extends theories of emotion. Burgeoning theoretical and empirical work argues that emotion concepts shape how individuals understand their own and others' emotions (Doyle & Lindquist, 2018; Gendron et al., 2012; Lindquist, MacCormack, et al., 2015; Nook et al., 2015; Satpute et al., 2016). However, these theories have not considered how the abstraction of one's emotion concepts might vary. Based on simulation-based theories of concept representation (Barrett, 2017; Barsalou, 1999, 2009; Wilson-Mendenhall, 2017), these data suggest that when defining emotion words, children and adolescents draw upon more "external" simulations (i.e., simulations directly connected to perceptual memories; e.g., an episode of crying), whereas adults draw upon simulations that are more "internal" and integrated (i.e., simulations involving other conceptual representations; e.g., the abstract concept of loss). This reveals that development of emotion concepts involves building an inner network of conceptual representations that progressively become more abstract with age. Future research could extend

this work by charting the fine-grained development of this network and by examining if and how the abstraction of emotion concepts impacts actual emotional experiences. Theories of situated conceptualization (Barrett, 2017; Wilson-Mendenhall et al., 2013, 2011) suggest that this would be the case, but because this study focused on conceptual understanding of emotion words outside of actual emotional experiences, this remains an open question.

Developmentally, these results grant additional insight into how emotions vary across childhood and adolescence. Researchers have found that several aspects of emotion understanding emerge across childhood (e.g., Pons et al., 2004), and the current findings reveal that building abstract emotion representations is another task for the developing mind. These findings also extend understanding of the unique affective processes that occur during adolescence. These data suggest that adolescence is a period in which many emotion words are fully comprehended but the concepts underlying these words have not yet reached mature levels of abstraction. Recent research has shown that adolescents less readily differentiate negative affect into specific types and that multi-dimensional emotion representations continue to develop through adolescence into young adulthood (Nook, Sasse, et al., 2017; Nook et al., 2018). The current results extend these findings to also include the notion that emotion abstractness develops during this period. It remains an open question whether the gap between these two milestones might be related to the elevated risk of mental illness in adolescence (Kessler et al., 2005), but this possibility should motivate future research testing whether and how emotion abstraction relates to emotion regulation and well-being across development.

These results also have implications for cognitive theories of abstract concepts. Some of these theories argue that emotions “ground” abstract concepts (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Ponari et al., 2018; Vigliocco et al., 2014; but see Borghi et al.,

2017). These theories posit that affective experiences may be the first cues that non-physical phenomena exist, leading individuals to build concepts for other non-physical abstract phenomena. As such, these scholars argue that emotion might provide a crucial role in building abstract concepts altogether. This argument calls for increased attention to the intersection between affective science and cognitive theories of concept representations, and our findings provide an initial foray. Because we did not compare age-related patterns of change for emotional and non-emotional concepts, data from the current study cannot directly test if or how emotion concept development relates to more general developments in concept abstraction. Nonetheless, future research could use the emotion vocabulary assessment to examine whether increased emotion concept abstraction (or attention to the affective nature of abstract experiences—like the pleasing tone of justice) facilitate the development of abstract non-emotional concepts.

Although age-related variation in the strategies participants used to define emotions overall supported the notion that emotion representations became more abstract across age, use of these strategies did not all share the same age-related pattern. Whereas example situation and general definition use followed emergent patterns, synonym use increased linearly across age, and physiological marker use overall fell across age but was slightly elevated in childhood. Because identifying synonyms requires the taxonomic understanding that two words share roughly equivalent conceptual referents, increased synonym use across age accords with general increases in emotion abstraction. However, the lack of an emergent pattern within our observed age window suggests that synonym use may arise younger than 4 years of age. Future research could investigate this possibility. Interestingly, the physiological marker pattern suggests that the external simulations children use to understand emotions are tightly linked to their

physiological experiences of emotion. A speculative interpretation for this finding is that children are taught about emotion words via specific physiological cues (e.g., smiles, tears, racing heart), and so these cues are salient when they are asked to define emotion words. However, this unexpected result merits further research.

Although we refrain from overinterpreting data from individual emotions, descriptive data suggest that age-related patterns of emotion concept abstraction are more heterogeneous across emotions than emotion word comprehension. This observation shows that emotion comprehension and abstraction demonstrate interesting qualitative differences across emotions. However, this finding also raises the possibility that representations of some emotions may inherently be more abstract than others (e.g., definitions of nervousness appear to remain more concrete than definitions of other emotions even in young adulthood). If this finding replicates in future work, researchers should investigate what qualities of different emotion concepts lead participants to represent them as more or less abstract.

To connect this and the preceding sections, it is worth noting that including assessments of emotion comprehension and abstraction in one study provided several important benefits. First, this design allowed for tight control of emotion abstraction measures. It is meaningless to analyze the abstractness of participants' emotion definitions if they are not in fact providing a valid definition of the emotion in question. Thus, measuring emotion comprehension and abstraction in one study allows for inaccurate definitions to be discarded. Second, this design allowed for direct comparisons between measures of emotion comprehension and emotion abstraction. In particular, comparing these phenomena within-subject allowed us to empirically demonstrate that abstraction plateaus significantly later than comprehension. This comparison underlies some of the study's key conclusions, including (i) that the concepts underlying emotion

words undergo continued abstraction even after they are comprehended, (ii) that children primarily demonstrate comprehension of emotion words by providing example situations (rather than general definitions), and (iii) that adolescence is a period in which emotion terms are comprehended but there is ongoing development in abstractness.

### **Toward an Automated Linguistic Measure of Abstraction**

A final methodological contribution of the current work is the use of a purely linguistic approach to measure abstraction of emotion definitions. We found that linguistic abstractness scores defined by the LCM showed moderate correlations with abstractness ratings provided by human coders and increased linearly across age. These results support the notions that linguistic and human methods can converge as measures of abstractness and that emotion concept abstraction increases across age. As such, these data provide a first step towards developing a fully automated linguistic measure of emotion concept abstraction.

However, four concerns emerge when considering the strength of the relationship between these two approaches: (i) the two measures did not reveal the same patterns of developmental change (i.e., one was linear and the other emergent), (ii) the strength of the correlation between coder and linguistic measures was modest (i.e.,  $r = .45$ ), (iii) age was more strongly associated with coder abstractness scores than linguistic abstractness scores, and (iv) coherence between linguistic abstractness scores and coder abstractness scores broke down at young ages. Although lack of coherence was explained by the fact that some young participants had few words on which a linguistic abstractness score could be computed (see **Supplemental Materials**) and excluding these participants stabilized the correlation between coder and

linguistic abstractness scores across age, this suggests that the current psycholinguistic approach was particularly noisy in computing the abstractness of young children's speech.

These discrepancies could either imply that these measures tap different constructs or that the measures suffer from too much measurement error to reliably assess the underlying construct of "abstractness." It is our intuition that these discrepancies arose because the linguistic measure was less sensitive to detecting and quantifying abstract features than human coders (i.e., the measurement error of the LCM approach reduced its correlation with the human measure). However, this is purely speculative. Future work could take a data-driven approach to this problem and use machine-learning methods to determine the optimal mathematical formula for determining emotion abstraction from linguistic data. Yet this work will need to contend with the difficult question of how to benchmark linguistic measures (i.e., what should be considered the "true" measure of abstractness against which linguistic measures are compared?). Many researchers use human ratings to measure abstractness, but these perceptions may be subject to bias. Hence, a more thorough treatment of construct validity in abstractness measures—including a definition of how to define "true" measures of abstractness—is needed for the field to refine this automated linguistic measure.

### **Limitations and Future Directions**

In addition to the limitations and future directions outlined throughout the discussion, further research can address three additional limitations of the current work. First, the cross-sectional nature of the current design limits inferences about the causal influence of age on the development of emotion comprehension and abstraction. A longitudinal design could rule out

potential third variables or cohort effects that limit causal inferences concerning relations between age and the dependent variables studied here.

Second, although the current project characterizes age-related changes of emotion comprehension and abstraction, it is limited in that these patterns are not compared to those of non-emotional words and concepts. Hence, the project does not address the question of whether these patterns are unique to emotion words or rather the product of more domain-general developmental shifts. Although prior theoretical and empirical work (as well as the principle of parsimony) suggests that emotion development likely scaffolds on the development of other faculties (like non-emotion vocabulary and non-verbal abstraction; e.g., Nook, Sasse, et al., 2017; Ponari et al., 2018), this question should be empirically tested. Put more broadly, the current study does not identify the mechanisms of increased emotion abstraction across age. This development could arise because of domain-general (i.e., both verbal and non-verbal) increases in the ability to engage in abstract representations (Caramelli et al., 2004; Dumontheil, 2014; Ponari et al., 2018; Whitaker et al., 2018). Thus, the extent to which the phenomenon assessed here scaffolds on broader developments in the ability to engage verbal and non-verbal abstractions remains an important question for future research. However, increased emotion abstraction across age could also arise because of increased diversity in one's affective experiences, allowing for simulations of the same emotion concept to draw upon abstract features that generalize across situations (e.g., feeling afraid both when alone in the woods or when giving a speech because of a common sense of *threat*). This explanatory hypothesis also merits further research.

Third, the current study did not address the downstream consequences of emotion comprehension and abstraction. Several lines of work suggest that having mastery over a large

set of emotion words facilitates adaptive emotion regulation (A. W. Brooks, 2014; Jamieson, Mendes, Blackstock, & Schmader, 2010; Kashdan, Barrett, & McKnight, 2015; Torre & Lieberman, 2018), implying that increased emotion comprehension might help people label their affective states in ways that allow them to effectively down-regulate negative affect. However, it is less clear whether increased emotion abstraction is helpful or harmful for mental wellness. On the one hand, emotion abstraction might facilitate emotion regulation. Prior work demonstrates that (i) psychological distancing facilitates emotion regulation (Ayduk & Kross, 2010; Kross et al., 2014; Moser et al., 2017; Nook, Schleider, & Somerville, 2017) and (ii) psychological distancing involves representing stimuli at higher levels of abstraction (Lieberman & Förster, 2009; Soderberg, Callahan, Kochersberger, Amit, & Ledgerwood, 2015). Combining these ideas suggests that being able to view one's emotional experiences abstractly might allow one to more adaptively engage in behaviors that are in one's long-term best interests when faced with emotionally evocative situations (e.g., Fujita & Carnevale, 2012).

On the other hand, increased abstraction could actually lead to *increased* susceptibility to distress and mental disorders. As Davey, Yücel, & Allen (2008) argue, the rewards that young children pursue tend to be concrete in nature (e.g., food, immediate expressions of care), but these rewards shift to be more abstract in adolescence and adulthood (e.g., social belonging, career success). These more abstract sources of joy and contentment are also more difficult to attain (e.g., subtle forms of social rejection or minor errors during career evaluation can disrupt social belonging and career success). Increased abstraction across development could increase one's feelings of rejection, frustration, and anxiety, ultimately increasing one's vulnerability to depression (Davey et al., 2008). Thus if, when, and how emotion abstraction confers psychological benefits remains an important question for future research.

### **Conclusion**

The current project utilizes a novel method for assessing emotion comprehension and abstraction, takes advantage of cutting-edge statistical and linguistic methods, and reveals that the nature of emotion concept changes across development. These findings extend how we understand emotion concepts and emotion development and may foster additional insight into the implications of abstract emotion representations.

**Acknowledgments**

The authors thank China Rose Marie Bolden, Aridenne Dews, Eliot Fearey, Megan Garrad, Tierra Gogue-Garcia, Kevin Kent, Alka Sareen, Manna Sirak, Taylor Stacy, and Christina Uhrig for assistance with data collection; Charlotte Anrig, Hannah Broos, Sandy Li, Azure Reid-Russell, and José Santiago for assistance with transcriptions; and Sandy Li and Sadhana Ponnaluri for assistance with ratings. This project was supported by a Harvard University startup grant to LHS, a National Institutes of Mental Health grant to KAM (R01-MH103291), and a National Science Foundation Graduate Research Fellowship (DGE1144152) to ECN.

**Author Contributions**

ECN, SFS, HKL, KAM, and LHS collaboratively developed the study design. ECN, SFS, and HKL collected behavioral data, and CMS led rating process. ECN, CMS, PM, and LHS designed analytic plan and analyzed data. ECN, CMS, and LHS interpreted results. ECN, CMS, and LHS drafted the manuscript, and all other authors provided critical revisions. All authors approved the final version of the manuscript for submission.

**Related publications**

Portions of these results have been shared via presentations at the Social Communication Across the Lifespan conference (2018) and the Nebraska Symposium on Motivation (2018).

**Conflicts of Interest**

Authors declare no conflicts of interest.

## REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American Journal of Psychology*, *117*(3), 389. <https://doi.org/10.2307/4149007>
- Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science*, *6*(3), 359–370. <https://doi.org/10.1111/tops.12096>
- Ayduk, Ö., & Kross, E. (2010). From a distance: Implications of spontaneous self-distancing for adaptive self-reflection. *Journal of Personality and Social Psychology*, *98*(5), 809–829. <https://doi.org/10.1037/a0019205>
- Aznar, A., & Tenenbaum, H. R. (2013). Spanish parents' emotion talk and their children's understanding of emotion. *Frontiers in Psychology*, *4*, 1–11. <https://doi.org/10.3389/fpsyg.2013.00670>
- Balconi, M., & Carrera, A. (2007). Emotional representation in facial expression and script. A comparison between normal and autistic children. *Research in Developmental Disabilities*, *28*(4), 409–422. <https://doi.org/10.1016/j.ridd.2006.05.001>
- Baron-Cohen, S., Golan, O., Wheelwright, S., Granader, Y., & Hill, J. (2010). Emotion word comprehension from 4 to 16 years old: A developmental survey. *Frontiers in Evolutionary Neuroscience*, *2*, 109. <https://doi.org/10.3389/fnevo.2010.00109>
- Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46.

[https://doi.org/10.1207/s15327957pspr1001\\_2](https://doi.org/10.1207/s15327957pspr1001_2)

- Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1–23. <https://doi.org/10.1093/scan/nsw154>
- Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences, 22*(4), 577–660.
- Barsalou, L. W. (2003a). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358*(1435), 1177–1187. <https://doi.org/10.1098/rstb.2003.1319>
- Barsalou, L. W. (2003b). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*(5–6), 513–562. <https://doi.org/10.1080/01690960344000026>
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 1281–1289. <https://doi.org/10.1098/rstb.2008.0319>
- Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170144. <https://doi.org/10.1098/rstb.2017.0144>
- Bauminger, N., Schorr Edelsztein, H., & Morash, J. (2005). Social information processing and emotional understanding in children with LD. *Journal of Learning Disabilities, 38*(1), 45–61. <https://doi.org/10.1177/00222194050380010401>
- Beck, L., Kumschick, I. R., Eid, M., & Klann-Delius, G. (2012). Relationship between language competence and emotional competence in middle childhood. *Emotion, 12*(3), 503–514. <https://doi.org/10.1037/a0026320>

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1), 289–300. <https://doi.org/10.2307/2346101>
- Blakemore, S. J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, 9(4), 267–277. <https://doi.org/10.1038/nrn2353>
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
- Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts*. New York, NY: Springer.
- Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, 143(3), 263–292. <https://doi.org/10.1037/bul0000089>
- Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology*, 18(6), 906–921. <https://doi.org/10.1037/0012-1649.18.6.906>
- Bretherton, I., Fritz, J., Zahn-Waxler, C., & Ridgeway, D. (1986). Learning to talk about emotions: A functionalist perspective. *Child Development*, 57(3), 529. <https://doi.org/10.2307/1130334>
- Bridges, K. M. B. (1930). A genetic theory of the emotions. *Pedagogical Seminary and Journal of Genetic Psychology*, 37, 514–527.
- Brooks, A. W. (2014). Get excited: Reappraising pre-performance anxiety as excitement. *Journal of Experimental Psychology: General*, 143(3), 1144–1158. <https://doi.org/10.1037/a0035325>
- Brooks, J. A., & Freeman, J. B. (2018). Conceptual knowledge predicts the representational

structure of facial emotion perception. *Nature Human Behaviour*, 2(8), 581–591.

<https://doi.org/10.1038/s41562-018-0376-6>

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911.

<https://doi.org/10.3758/s13428-013-0403-5>

Burgoon, E. M., Henderson, M. D., & Markman, A. B. (2013). There are many ways to see the forest for the trees. *Perspectives on Psychological Science*, 8(5), 501–520.

<https://doi.org/10.1177/1745691613497964>

Burnett, S., Bird, G., Moll, J., Frith, C. D., & Blakemore, S.-J. (2009). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience*, 21(9), 1736–1750. <https://doi.org/10.1162/jocn.2009.21121>

Canty, A., & Ripley, B. (2017). Boot: Bootstrap R (S-Plus) functions. R package version 1.3-20.

Caramelli, N., & Setti, A. (2005). Different domains in abstract concepts. In B. Bara, L.

Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society* (pp. 1997–2002). Mahwah, NJ: Erlbaum.

Caramelli, N., Setti, A., & Maurizzi, D. D. (2004). Concrete and abstract concepts in school age children. *Psychology of Language and Communication*, 8(2), 19–34.

Carnaghi, A., Maass, A., Gresta, S., Bianchi, M., Cadinu, M., & Arcuri, L. (2008). Nomina sunt omina: On the inductive potential of nouns and adjectives in person perception. *Journal of Personality and Social Psychology*, 94(5), 839–859. [https://doi.org/10.1037/0022-](https://doi.org/10.1037/0022-3514.94.5.839)

[3514.94.5.839](https://doi.org/10.1037/0022-3514.94.5.839)

Casey, B. J. (2015). Beyond simple models of self-control to circuit-based accounts of adolescent behavior. *Annual Review of Psychology*, 66(1), 295–319.

<https://doi.org/10.1146/annurev-psych-010814-015156>

Castro, V. L., Cheng, Y., Halberstadt, A. G., & Grühn, D. (2016). EUREKA! A conceptual model of emotion understanding. *Emotion Review*, *8*(3), 258–268.

<https://doi.org/10.1177/1754073915580601>

Cervantes, C. A., & Callanan, M. A. (1998). Labels and explanations in mother–child emotion talk: Age and gender differentiation. *Developmental Psychology*, *34*(1), 88–98.

<https://doi.org/10.1037/0012-1649.34.1.88>

Choudhury, S., Blakemore, S. J., & Charman, T. (2006). Social cognitive development during adolescence. *Social Cognitive and Affective Neuroscience*, *1*(3), 165–174.

<https://doi.org/10.1093/scan/nsl024>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ:

Lawrence Erlbaum. <https://doi.org/10.4324/9780203771587>

Cole, P. M., Armstrong, L. M., & Pemberton, C. K. (2010). The role of language in the development of emotion regulation. In S. D. Calkins & M. A. Bell (Eds.), *Human brain development. Child development at the intersection of emotion and cognition* (pp. 59–77). Washington, DC: American Psychological Association.

Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S.

A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*(1), 55–66. <https://doi.org/10.1111/j.1467-7687.2008.00743.x>

Davey, C. G., Yücel, M., & Allen, N. B. (2008). The emergence of depression in adolescence:

Development of the prefrontal cortex and the representation of reward. *Neuroscience and Biobehavioral Reviews*, *32*(1), 1–19. <https://doi.org/10.1016/j.neubiorev.2007.04.016>

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their applications*. Cambridge:

Cambridge University Press.

- Della Rosa, P. A., Catricalà, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract—concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods*, *42*(4), 1042–1048. <https://doi.org/10.3758/BRM.42.4.1042>
- Demetriou, A., Makris, N., Spanoudis, G., Kazi, S., Shayer, M., & Kazali, E. (2018). Mapping the dimensions of general intelligence: An integrated differential-developmental theory. *Human Development*, *61*(1), 4–42. <https://doi.org/10.1159/000484450>
- DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*(3), 189–228. <https://doi.org/10.1214/ss/1032280214>
- Doré, B., Ort, L., Braverman, O., & Ochsner, K. N. (2015). Sadness shifts to anxiety over time and distance from the national tragedy in Newtown, Connecticut. *Psychological Science*, *26*(4), 363–373. <https://doi.org/10.1177/0956797614562218>
- Doyle, C. M., & Lindquist, K. A. (2018). When a word is worth a thousand pictures: Language shapes perceptual memory for emotion. *Journal of Experimental Psychology: General*, *147*(1), 62–73. <https://doi.org/10.1037/xge0000361>
- Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostral lateral prefrontal cortex. *Developmental Cognitive Neuroscience*, *10*, 57–76. <https://doi.org/10.1016/j.dcn.2014.07.009>
- Duncombe, M., Havighurst, S. S., Holland, K. A., & Frankling, E. J. (2013). Relations of emotional competence and effortful control to child disruptive behavior problems. *Early Education & Development*, *24*(5), 599–615. <https://doi.org/10.1080/10409289.2012.701536>
- Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers

and their young children. *Developmental Psychology*, 23(1), 132–139.

<https://doi.org/10.1037/0012-1649.23.1.132>

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129. <https://doi.org/10.1037/h0030377>

Fabes, R. A., Eisenberg, N., Hanish, L. D., & Spinrad, T. L. (2001). Preschoolers' spontaneous emotion vocabulary: Relations to likability. *Early Education & Development*, 12(1), 11–27.

[https://doi.org/10.1207/s15566935eed1201\\_2](https://doi.org/10.1207/s15566935eed1201_2)

Fan, R., Varol, O., Varamesh, A., Barron, A., van de Leemput, I. A., Scheffer, M., & Bollen, J. (2019). The minute-scale dynamics of online emotions reveal the effects of affect labeling.

*Nature Human Behaviour*, 3(1), 92–100. <https://doi.org/10.1038/s41562-018-0490-5>

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge:

Differences by class and race. *Social Science Research*, 33(3), 464–497.

<https://doi.org/10.1016/j.ssresearch.2003.08.001>

Ferrer, E., O'Hare, E. D., & Bunge, S. A. (2009). Fluid reasoning and the developing brain.

*Frontiers in Neuroscience*, 3(1), 46–51. <https://doi.org/10.3389/neuro.01.003.2009>

Forgas, J. P. (2007). When sad is better than happy: Negative affect can improve the quality and effectiveness of persuasive messages and social influence strategies. *Journal of*

*Experimental Social Psychology*, 43(4), 513–528. <https://doi.org/10.1016/j.jesp.2006.05.006>

Franz, P., Nook, E. C., Mair, P., & Nock, M. K. (2019). Using topic modeling to detect and

describe self-injurious and related content on a large-scale digital platform. *Revision under Review*.

- Fugate, J. M. B., Gouzoules, H., & Barrett, L. F. (2010). Reading chimpanzee faces: Evidence for the role of verbal labels in categorical perception of emotion. *Emotion, 10*(4), 544–554. <https://doi.org/10.1037/a0019017>
- Fujita, K., & Carnevale, J. J. (2012). Transcending temptation through abstraction. *Current Directions in Psychological Science, 21*(4), 248–252. <https://doi.org/10.1177/0963721412449169>
- Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion, 12*(2), 314–325. <https://doi.org/10.1037/a0026007>
- Gopnik, A. (2001). Theories, language, and culture: Whorf without wincing. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 45–69). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511620669.004>
- Gray, A. W., Parkinson, B., & Dunbar, R. I. (2015). Laughter's influence on the intimacy of self-disclosure. *Human Nature, 26*(1), 28–43. <https://doi.org/10.1007/s12110-015-9225-8>
- Greenberg, M. T., Kusche, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. *Development and Psychopathology, 7*(01), 117. <https://doi.org/10.1017/S0954579400006374>
- Hale, S. C. (1988). Spacetime and the abstract / concrete distinction. *Philosophical Studies, 53*(1), 85–102.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures*. New York, NY: Basic Books.
- Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal*

*of Experimental Social Psychology*, 46(1), 208–212.

<https://doi.org/10.1016/j.jesp.2009.08.015>

Joelson, J. M., & Herrmann, D. J. (1978). Properties of categories in semantic memory. *The American Journal of Psychology*, 91(1), 101. <https://doi.org/10.2307/1421826>

Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science*, 24(1), 10–16.

<https://doi.org/10.1177/0963721414550708>

Kelly, B., Longbottom, J., Potts, F., & Williamson, J. (2004). Applying emotional intelligence: Exploring the Promoting Alternative Thinking Strategies curriculum. *Educational Psychology in Practice*, 20(3), 221–240. <https://doi.org/10.1080/0266736042000251808>

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6), 593.

<https://doi.org/10.1001/archpsyc.62.6.593>

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, 42(2), 998–1031.

<https://doi.org/10.1016/j.neuroimage.2008.03.059>

Kopp, C. B. (1989). Regulation of distress and negative emotions: A developmental view.

*Developmental Psychology*, 25(3), 343–354. <https://doi.org/10.1037//0012-1649.25.3.343>

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental*

*Psychology: General*, 140(1), 14–34. <https://doi.org/10.1037/a0021446>

Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., ... Ayduk,

O. (2014). Self-talk as a regulatory mechanism: How you do it matters. *Journal of*

*Personality and Social Psychology*, 106(2), 304–324. <https://doi.org/10.1037/a0035173>

Kuhn, L. J., Willoughby, M. T., Vernon-Feagans, L., & Blair, C. B. (2016). The contribution of

children's time-specific and longitudinal expressive language skills on developmental

trajectories of executive function. *Journal of Experimental Child Psychology*, 148, 20–34.

<https://doi.org/10.1016/j.jecp.2016.03.008>

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for

30,000 English words. *Behavior Research Methods*, 44(4), 978–990.

<https://doi.org/10.3758/s13428-012-0210-4>

Kusché, C. A., Belike, R. L., & Greenberg, M. T. (1988). *Kusché Affective Inventory-Revised*.

*Unpublished measure*. Seattle, WA: University of Washington.

Lagattuta, K. H., & Wellman, H. M. (2002). Differences in early parent-child conversations

about negative versus positive emotions: Implications for the development of psychological

understanding. *Developmental Psychology*, 38(4), 564–580. [https://doi.org/10.1037/0012-](https://doi.org/10.1037/0012-1649.38.4.564)

[1649.38.4.564](https://doi.org/10.1037/0012-1649.38.4.564)

Larson, R. W., Moneta, G., Richards, M. H., & Wilson, S. (2002). Continuity, stability, and

change in daily emotional experience across adolescence. *Child Development*, 73(4), 1151–

1165. <https://doi.org/10.1111/1467-8624.00464>

Li, Y., & Yu, D. (2015). Development of emotion word comprehension in Chinese children from

2 to 13 years old: Relationships with valence and empathy. *PLOS ONE*, 10(12), e0143712.

<https://doi.org/10.1371/journal.pone.0143712>

- Lieberman, N., & Förster, J. (2009). The effect of psychological distance on perceptual level of construal. *Cognitive Science*, *33*(7), 1330–1341. <https://doi.org/10.1111/j.1551-6709.2009.01061.x>
- Lindquist, K. A. (2017). The role of language in emotion: existing evidence and future directions. *Current Opinion in Psychology*, *17*, 135–139. <https://doi.org/10.1016/j.copsyc.2017.07.006>
- Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, *6*(1), 125–138. <https://doi.org/10.1037/1528-3542.6.1.125>
- Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: predictions from psychological constructionism. *Frontiers in Psychology*, *6*(444). <https://doi.org/10.3389/fpsyg.2015.00444>
- Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion? *Current Directions in Psychological Science*, *24*(2), 99–108. <https://doi.org/10.1177/0963721414553440>
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., Feldman Barrett, L., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *The Behavioral and Brain Sciences*, *35*(3), 121–143. <https://doi.org/10.1017/S0140525X11000446>
- Locke, R. L., Miller, A. L., Seifer, R., & Heinze, J. E. (2015). Context-inappropriate anger, emotion knowledge deficits, and negative social experiences in preschool. *Developmental Psychology*, *51*(10), 1450–1463. <https://doi.org/10.1037/a0039528>
- Lupyan, G. (2012a). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*. <https://doi.org/10.3389/fpsyg.2012.00054>
- Lupyan, G. (2012b). What do words do? Toward a theory of language-augmented thought. In B.

- H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 57, pp. 255–297). Waltham, MA: Academic Press. <https://doi.org/10.1016/B978-0-12-394293-7.00007-8>
- Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1752). <https://doi.org/10.1098/rstb.2017.0137>
- Luria, A. (1961). *The role of speech in the regulation of normal and abnormal behaviors*. New York, NY: Liveright.
- MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, 8(4), 540–551. <https://doi.org/10.1037/a0012746>
- MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: an update. *Journal of Child Language*, 17(2), 457–472. <https://doi.org/10.1017/S0305000900013866>
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1–3), 59–70. <https://doi.org/10.1016/j.jphysparis.2008.03.004>
- Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions. *Language Learning and Development*, 14(3), 186–223. <https://doi.org/10.1080/15475441.2018.1455584>
- Mazzuca, C., Barca, L., & Borghi, A. M. (2017). The particularity of emotional words: A grounded approach. *Rivista Internazionale Di Filosofia e Psicologia*, 8(2), 124–133. <https://doi.org/10.4453/rifp.2017.0010>
- Mervis, C., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32(1), 89–115. <https://doi.org/10.1146/annurev.ps.32.020181.000513>

- Moser, J. S., Dougherty, A., Mattson, W. I., Katz, B., Moran, T. P., Guevarra, D., ... Kross, E. (2017). Third-person self-talk facilitates emotion regulation without engaging cognitive control: Converging evidence from ERP and fMRI. *Scientific Reports*, 7(1), 4519. <https://doi.org/10.1038/s41598-017-04047-3>
- Nook, E. C., Lindquist, K. A., & Zaki, J. (2015). A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion*, 15(5), 569–578. <https://doi.org/10.1037/a0039166>
- Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, 1(12), 881–889. <https://doi.org/10.1038/s41562-017-0238-7>
- Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2018). The nonlinear development of emotion differentiation: Granular emotional experience is low in adolescence. *Psychological Science*, 29(8), 1346–1357. <https://doi.org/10.1177/0956797618773357>
- Nook, E. C., Schleider, J. L., & Somerville, L. H. (2017). A linguistic signature of psychological distancing in emotion regulation. *Journal of Experimental Psychology: General*, 146(3), 337–346. <https://doi.org/10.1037/xge0000263>
- O’Kearney, R., & Dadds, M. (2004). Developmental and gender differences in the language for emotions across the adolescent years. *Cognition & Emotion*, 18(7), 913–938. <https://doi.org/10.1080/02699930341000356>
- O’Kearney, R., & Dadds, M. R. (2005). Language for emotions in adolescents with externalizing and internalizing disorders. *Development and Psychopathology*, 17(2), 529–548.

<https://doi.org/10.1017/S095457940505025X>

Oosterwijk, S., Lindquist, K. A., Adebayo, M., & Barrett, L. F. (2015). The neural representation of typical and atypical experiences of negative images: Comparing fear, disgust and morbid fascination. *Social Cognitive and Affective Neuroscience*, (July), nsv088.

<https://doi.org/10.1093/scan/nsv088>

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 76(1), 1–25.

<https://doi.org/10.1037/h0025327>

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count: LIWC [Computer software]. Austin, TX: LIWC.net. Retrieved from LIWC.net

Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, 21(2), e12549.

<https://doi.org/10.1111/desc.12549>

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, 1(2), 127–152. <https://doi.org/10.1080/17405620344000022>

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria:

R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time.

*Cognitive Psychology*, 41(1), 1–48. <https://doi.org/10.1006/cogp.1999.0735>

Reichenbach, L., & Masters, J. C. (1983). Children's use of expressive and contextual cues in judgments of emotion. *Child Development*, 54(4), 993. <https://doi.org/10.2307/1129903>

Roben, C. K. P., Cole, P. M., & Armstrong, L. M. (2013). Longitudinal relations among

- language skills, anger expression, and regulatory strategies in early childhood. *Child Development*, 84(3), 891–905. <https://doi.org/10.1111/cdev.12027>
- Rodman, A. M., Powers, K. E., & Somerville, L. H. (2017). Development of self-protective biases in response to social evaluative feedback. *Proceedings of the National Academy of Sciences*, 114(50), 13158–13163. <https://doi.org/10.1073/pnas.1712398114>
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382–439. [https://doi.org/10.1016/0010-0285\(76\)90013-X](https://doi.org/10.1016/0010-0285(76)90013-X)
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178.
- Russell, J. A., & Widen, S. C. (2002). Words versus faces in evoking preschool children's knowledge of the causes of emotions. *International Journal of Behavioral Development*, 26(2), 97–103. <https://doi.org/10.1080/01650250042000582>
- Saarni, C. (1999). *The development of emotional competence*. New York, NY: Guilford Press.
- Salmon, K., O'Kearney, R., Reese, E., & Fortune, C.-A. (2016). The role of language skill in child psychopathology: Implications for intervention in the early years. *Clinical Child and Family Psychology Review*, 19(4), 352–367. <https://doi.org/10.1007/s10567-016-0214-1>
- Satpute, A. B., Nook, E. C., Narayanan, S., Shu, J., Weber, J., & Ochsner, K. N. (2016). Emotions in “black and white” or shades of gray? How we think about emotion shapes our perception and neural representation of emotion. *Psychological Science*, 27(11), 1428–1442. <https://doi.org/10.1177/09567976166661555>
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In D. B. Jones & H. L. Somers (Eds.), *Proceedings of the International Conference on New Methods in*

*Language Processing* (pp. 44–49). Manchester, England: University of Manchester Institute of Science and Technology.

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., De Brito, S. A., McCrory, E. J. P., & Viding, E. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, *7*(1), 53–63. <https://doi.org/10.1093/scan/nsr023>

Seih, Y.-T., Beier, S., & Pennebaker, J. W. (2017). Development and examination of the Linguistic Category Model in a computerized text analysis method. *Journal of Language and Social Psychology*, *36*(3), 343–355. <https://doi.org/10.1177/0261927X16657855>

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, *54*(4), 558–568. <https://doi.org/10.1037/0022-3514.54.4.558>

Semin, G. R., & Fiedler, K. (1991). The Linguistic Category Model, its bases, applications and range. *European Review of Social Psychology*, *2*(1), 1–30. <https://doi.org/10.1080/14792779143000006>

Semin, G. R., Görts, C. A., Nandram, S., & Semin-Goossens, A. (2002). Cultural perspectives on the linguistic representation of emotion and emotion events. *Cognition & Emotion*, *16*(1), 11–28. <https://doi.org/10.1080/02699930143000112>

Shablack, H., & Lindquist, K. A. (2019). The role of language in emotional development. In V. LoBue, K. Perez-Edgar, & K. Buss (Eds.), *Handbook of Emotion Development*. New York, NY: Springer.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, *60*(4), 973.

<https://doi.org/10.2307/1131037>

Simpson, G. (2014). First derivative of a spline term. *GitHub Repository*. Retrieved from

<https://gist.github.com/gavinsimpson/e73f011fdaaab4bb5a30>

Smith, M., & Walden, T. (2001). Understanding feelings and coping with emotional situations:

A comparison of maltreated and nonmaltreated preschoolers. *Social Development*, 8(1), 93–116. <https://doi.org/10.1111/1467-9507.00082>

Soderberg, C. K., Callahan, S. P., Kochersberger, A. O., Amit, E., & Ledgerwood, A. (2015).

The effects of psychological distance on abstraction: Two meta-analyses. *Psychological Bulletin*, 141(3), 525–548. <https://doi.org/10.1037/bul0000005>

Somerville, L. H., Jones, R. M., Ruberry, E. J., Dyke, J. P., Glover, G., & Casey, B. J. (2013).

The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. *Psychological Science*, 24(8), 1554–1562. <https://doi.org/10.1177/0956797613475633>

Somerville, L. H., & McLaughlin, K. A. (2018). What develops during emotional development?

Normative trajectories and sources of psychopathology risk in adolescence. In Andrew S. Fox, Regina C. Lapate, Alexander J. Shackman, & Richard J. Davidson (Eds.), *The Nature of Emotion: Fundamental Questions* (2nd ed.).

Southam-Gerow, M. A., & Kendall, P. C. (2000). A preliminary study of the emotion

understanding of youths referred for treatment of anxiety disorders. *Journal of Clinical Child Psychology*, 29(3), 319–327. [https://doi.org/10.1207/S15374424JCCP2903\\_3](https://doi.org/10.1207/S15374424JCCP2903_3)

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition,

imageability, and familiarity. *Behavior Research Methods*, 38(4), 598–605.

<https://doi.org/10.3758/BF03193891>

Torre, J. B., & Lieberman, M. D. (2018). Putting feelings into words: Affect labeling as implicit

emotion regulation. *Emotion Review*, 10(2), 116–124.

<https://doi.org/10.1177/1754073917742706>

Trentacosta, C. J., & Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion*, 7(1), 77–88.

<https://doi.org/10.1037/1528-3542.7.1.77>

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance.

*Psychological Review*, 117(2), 440–463. <https://doi.org/10.1037/a0018963>

Vallotton, C., & Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly*, 26(2), 169–181.

<https://doi.org/10.1016/j.ecresq.2010.09.002>

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., &

Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion.

*Cerebral Cortex*, 24(7), 1767–1777. <https://doi.org/10.1093/cercor/bht025>

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Waggoner, J. E., & Palermo, D. S. (1989). Betty is a bouncing bubble: Children's

comprehension of emotion-descriptive metaphors. *Developmental Psychology*, 25(1), 152–163. <https://doi.org/10.1037/0012-1649.25.1.152>

Wang, J., Conder, J. A., Blitzer, D. N., & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping*, 31(10), 1459–1468. <https://doi.org/10.1002/hbm.20950>

*Mapping*, 31(10), 1459–1468. <https://doi.org/10.1002/hbm.20950>

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX:

Psychological Corporation.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence—Second Edition Manual*.

Bloomington, MN: Pearson.

Whitaker, K. J., Vendetti, M. S., Wendelken, C., & Bunge, S. A. (2018). Neuroscientific insights into the development of analogical reasoning. *Developmental Science*, *21*(2), e12531.

<https://doi.org/10.1111/desc.12531>

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, *5*(1), 72–77.

<https://doi.org/10.1177/1754073912451492>

Widen, S. C., Pochedly, J. T., & Russell, J. A. (2015). The development of emotion concepts: A story superiority effect in older children and adolescents. *Journal of Experimental Child Psychology*, *131*, 186–192. <https://doi.org/10.1016/j.jecp.2014.10.009>

Widen, S. C., & Russell, J. A. (2010a). Children's scripts for social emotions: Causes and consequences are more central than are facial expressions. *British Journal of Developmental Psychology*, *28*(3), 565–581. <https://doi.org/10.1348/026151009X457550d>

Widen, S. C., & Russell, J. A. (2010b). Differentiation in preschooler's categories of emotion. *Emotion*, *10*(5), 651–661. <https://doi.org/10.1037/a0019005>

Wilson-Mendenhall, C. D. (2017). Constructing emotion through simulation. *Current Opinion in Psychology*, *17*, 189–194. <https://doi.org/10.1016/j.copsyc.2017.07.015>

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Situating emotional experience. *Frontiers in Human Neuroscience*, *7*, 764.

<https://doi.org/10.3389/fnhum.2013.00764>

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011).

Grounding emotion in situated conceptualization. *Neuropsychologia*, *49*(5), 1105–1127.

<https://doi.org/10.1016/j.neuropsychologia.2010.12.032>

- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society B*, 65(1), 95–114.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). CRC Press.
- Wu, Y., Muentener, P., & Schulz, L. E. (2017). One- to four-year-olds connect diverse positive emotional vocalizations to their probable causes. *Proceedings of the National Academy of Sciences*, 114(45), 11896–11901. <https://doi.org/10.1073/pnas.1707715114>
- Yin, J., & Csibra, G. (2015). Concept-based word learning in human infants. *Psychological Science*, 26(8), 1316–1324. <https://doi.org/10.1177/0956797615588753>
- Zeman, J., Klimes-Dougan, B., Cassano, M., & Adrian, M. (2007). Measurement issues in emotion research with children and adolescents. *Clinical Psychology: Science and Practice*, 14(4), 377–401. <https://doi.org/10.1111/j.1468-2850.2007.00098.x>

## TABLES

**Table 1.** Spearman's correlations showing interrelations between dependent measures.

	1.	2.	3.	4.	5.	6.
1. Emotion comprehension scores	-					
2. Coder abstractness scores	.52***	-				
3. Linguistic abstractness scores	.20**	.45***	-			
4. General definition strategy use	.50***	.74***	.25***	-		
5. Synonym strategy use	.23**	.33***	.29***	.17*	-	
6. Situation strategy use	-.49***	-.90***	-.50***	-.63***	-.32***	-
7. Physiological marker strategy use	-.13#	-.35***	-.40***	-.16*	-.11	.36***

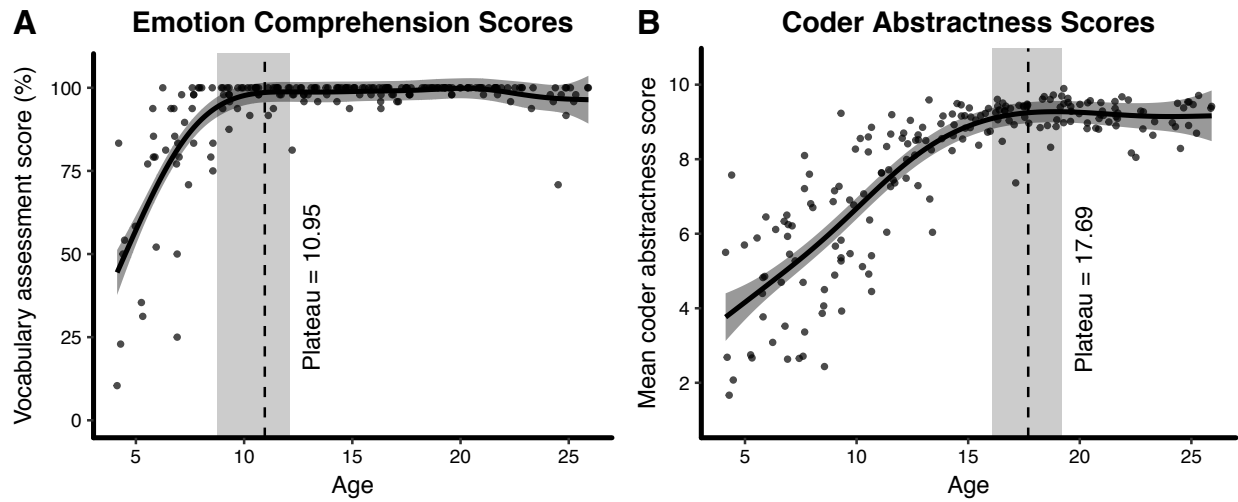
Notes: See main text for tests of linear and non-linear relations between age and other variables. All significant relations (i.e.,  $ps < .05$ ) survive FDR correction for multiple comparisons, but other relations (i.e.,  $ps > .05$ ) do not. \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ , # $p < .10$ .

**Table 2.** Results of null, polynomial, and thin plate smoothing spline models for each dependent variable.

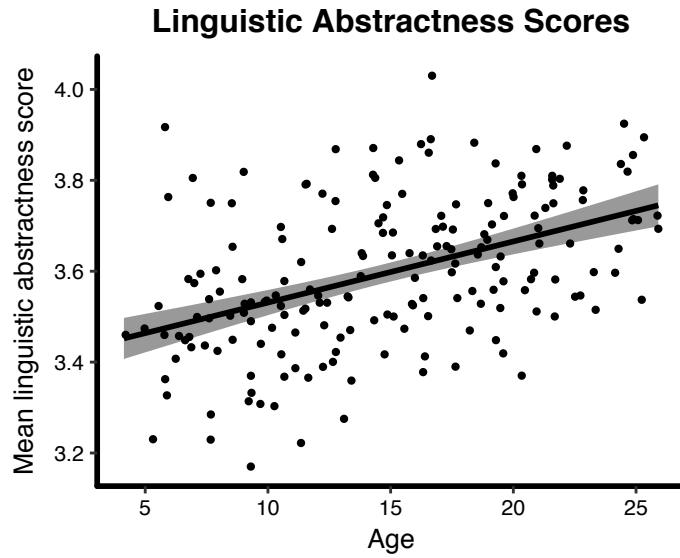
Dependent variable	Model	$\beta$	$p$	Adj $R^2$	AIC
Emotion comprehension scores	Null				1608.10
	Linear	.48	< .001***	.23	1558.51
	Quadratic	-.50	< .001***	.47	1484.31
	Cubic	.32	< .001***	.58	1442.01
	<b>Spline</b>			<b>.62</b>	<b>1426.36</b>
Coder abstractness scores	Null				833.48
	Linear	.76	< .001***	.58	663.07
	Quadratic	-.37	< .001***	.72	586.51
	Cubic	-.01	.745	.72	588.40
	<b>Spline</b>			<b>.73</b>	<b>580.05</b>
Linguistic abstractness scores	Null				-142.50
	<b>Linear</b>	<b>.46</b>	<b>&lt; .001***</b>	<b>.21</b>	<b>-185.25</b>
	Quadratic	-.001	.991	.20	-183.25
	Cubic	-.05	.467	.20	-181.79
	Spline			.21	-185.25 <sup>+</sup>
General definition use	Null				1859.18
	Linear	.70	< .001***	.49	1728.89
	Quadratic	-.46	< .001***	.70	1627.36
	Cubic	.07	.057	.70	1625.66
	<b>Spline</b>			<b>.70</b>	<b>1623.66</b>
Synonym use	Null				1738.33
	<b>Linear</b>	<b>.32</b>	<b>&lt; .001***</b>	<b>.10</b>	<b>1719.66</b>
	Quadratic	-.05	.458	.09	1721.10
	Cubic	-.01	.848	.09	1723.06
	Spline			.10	1719.66 <sup>+</sup>
Example situation use	Null				1866.19
	Linear	-.67	< .001***	.45	1750.57
	Quadratic	.32	< .001***	.55	1712.61
	Cubic	.02	.687	.55	1714.44
	<b>Spline</b>			<b>.56</b>	<b>1710.20</b>
Physiological marker use	Null				1392.78
	Linear	-.23	.001**	.05	1384.27
	Quadratic	-.01	.882	.04	1386.25
	Cubic	.13	.062	.05	1384.68
	<b>Spline</b>			<b>.10</b>	<b>1378.66</b>

Note: Bold text indicates best fitting model for each dependent variable, as determined by having the lowest AIC. <sup>+</sup> In these two cases, spline model algorithms revealed that the best fits were identical to linear models; hence we interpreted the linear models.  $\beta$  = standardized beta, Adj  $R^2$  = adjusted  $R^2$ , AIC = Akaike Information Criterion, \*\*\*  $p < .001$ , \*\*  $p < .01$

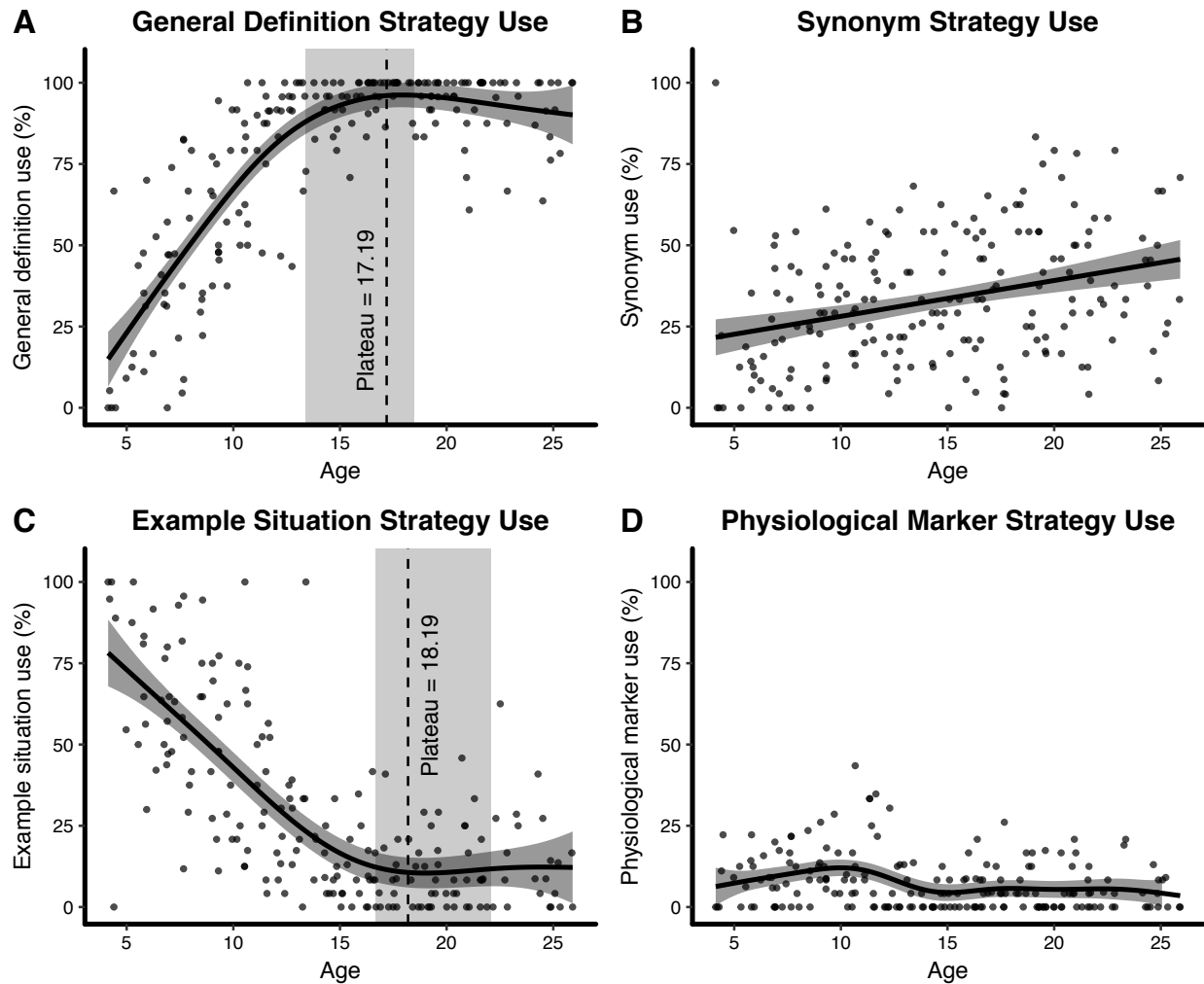
## FIGURES



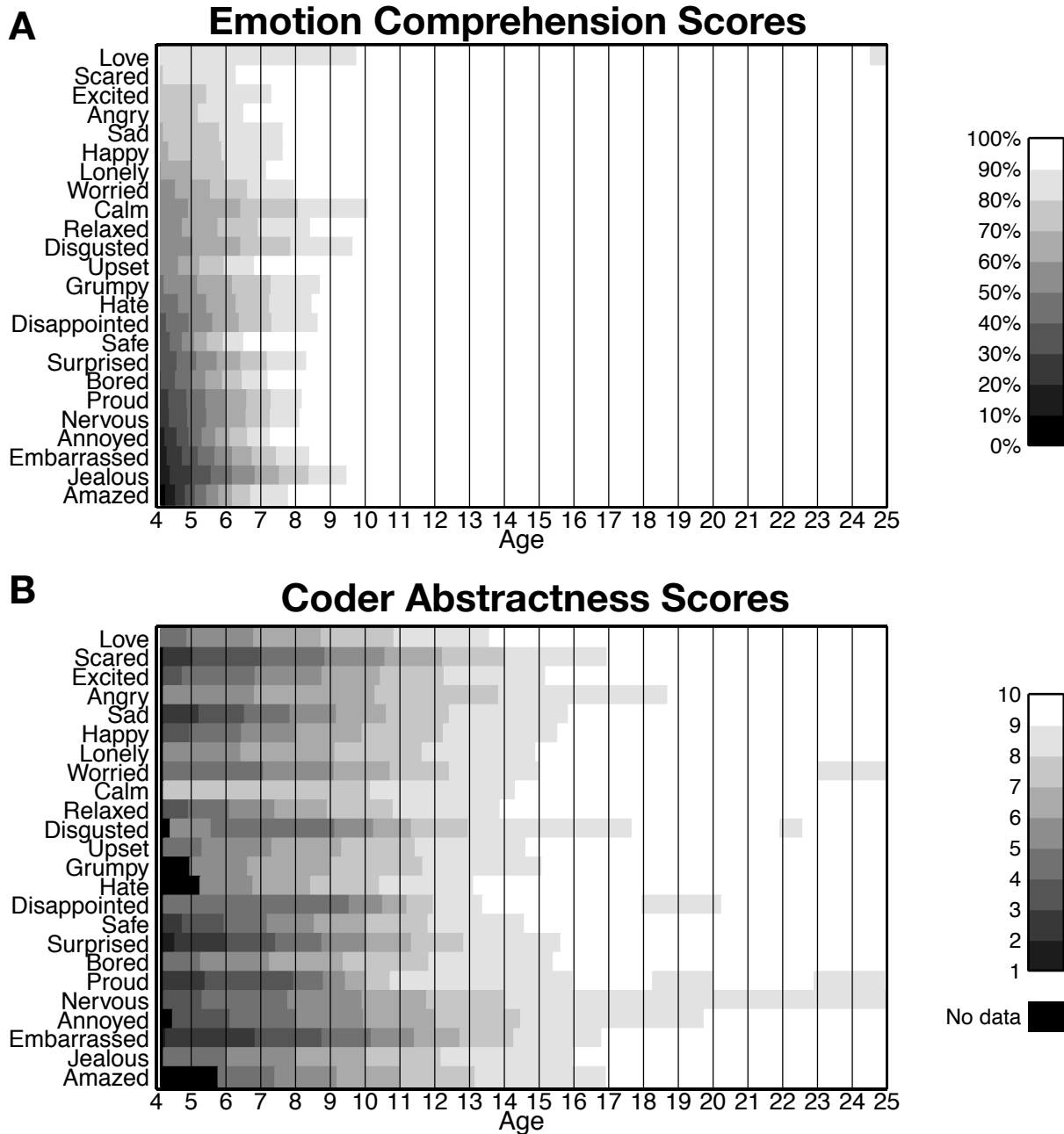
**Figure 1.** Age-related pattern of emotion comprehension and one measure of emotion abstraction. Thin plate regression smoothing splines revealed that A) emotion comprehension scores and B) coder abstractness scores followed emergent age-related patterns. However, emotion comprehension scores plateaued significantly earlier than coder abstractness scores. Solid black lines depict spline model fits (dark grey regions represent 95% CIs from spline models). Vertical dashed lines show developmental plateaus computed using first derivatives of spline fits (vertical light grey regions represent 95% CIs from bootstrap simulations).



**Figure 2.** Linguistic abstractness scores, a purely linguistic measure of abstractness computed following Seih et al., (2017), increased linearly with increasing age. This result corroborates abstractness scores provided by human coders. Dark grey regions represent 95% CI.



**Figure 3.** Age-related patterns of 4 different strategies for describing emotions. A) Use of general definitions followed an emergent pattern that increased across childhood and adolescence before plateauing. B) Use of synonyms increased linearly with increasing age. C) Use of example situations followed a diminishing pattern that decreased across childhood and adolescence before plateauing. D) Use of physiological markers followed a non-linear pattern that was elevated in childhood. Solid black lines depict spline model fits (dark grey regions represent 95% CI from spline models). Vertical dashed lines show developmental plateaus computed using first derivatives of spline fits (vertical light grey regions represent 95% CIs from bootstrap simulations).



**Figure 4.** Spline-based estimates of A) emotion comprehension and B) emotion abstraction (i.e., coder abstractness scores) for each emotion across age. Lighter colors represent higher levels of comprehension or abstraction. Emotions are ordered in terms of initial comprehension scores at age 4. In panel B, black regions indicate ages at which no participants fully comprehended the emotion, so estimates of emotion abstractness could not be computed.

## Supplemental Materials

To accompany

Charting the development of emotion comprehension and abstraction  
from childhood to adulthood using observer-rated and linguistic measures

### Contents

- I. Justification of linguistic exclusions
- II. Justification of plateau definition
- III. Analyses only including participants who fully comprehended all emotions
- IV. Analyses excluding trials on which participants only provided synonyms
- V. Emotion vocabulary assessment

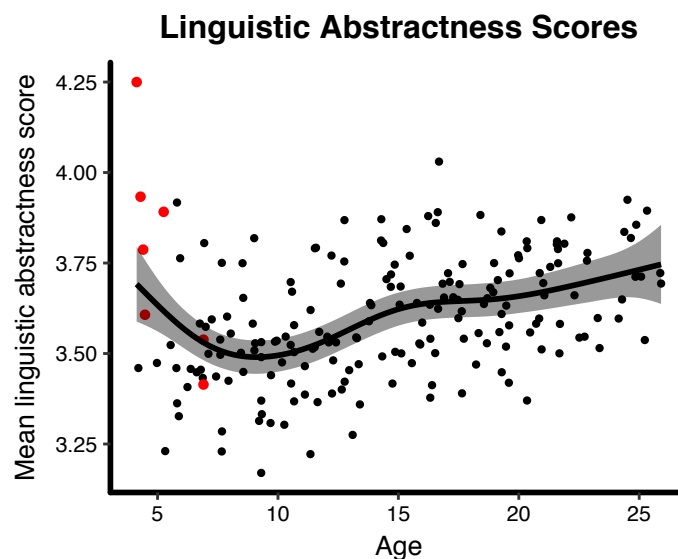
## I. Justification of linguistic exclusions

Initial analyses of linguistic abstractness scores including all participants revealed that the thin plate smoothing spline model,  $R^2 = .21$ ,  $AIC = -167.63$ , provided a better fit than the null,  $AIC = -130.09$ , linear,  $\beta = .34$ ,  $p < .001$ ,  $R^2 = .11$ ,  $AIC = -152.44$ , quadratic,  $\beta = .14$ ,  $p = .034$ ,  $R^2 = .13$ ,  $AIC = -155.03$ , and cubic,  $\beta = -.20$ ,  $p = .002$ ,  $R^2 = .17$ ,  $AIC = -162.59$ , models (**Figure S1**, log likelihood test comparing cubic and spline models:  $p = .012$ ). However, further analyses suggested that linguistic abstractness scores for very young participants might not be valid. The relationship between linguistic abstractness scores and coder abstractness scores varied across age: there was a significant interaction between age and coder abstractness scores in predicting linguistic abstractness scores,  $\beta = .45$ ,  $p < .001$ . Thus the concordance between coder and linguistic measures of abstractness were not equivalent across age when all participants were included. Specifically, concordance was weaker at younger ages and stronger at older ages.

We hypothesized that this breakdown in concordance at young ages may have arisen because there was little usable linguistic data on which linguistic abstractness measures could be computed for young children. Young participants both (i) spoke fewer words on average than older participants (see **Results**) and (ii) they had fewer usable trials to include in linguistic analyses (given that they were less likely to fully comprehend emotions and abstraction scores were only valid for emotions that participants fully comprehended). To test whether a paucity of linguistic data in young children could explain low concordance between human and linguistic abstractness scores at young ages, we computed the total word count for all usable emotion definitions for each participant. Indeed, total usable word count increased across age,  $\beta = .28$ ,  $p < .001$ . However, controlling for usable word count reduced the interaction between age and coder abstractness scores in predicting linguistic abstractness scores to a statistical trend,  $\beta = .23$ ,

$p = .052$ . Hence, accounting for age-related differences in usable word count helped stabilize concordance between coder and linguistic measures of abstractness across age.

Consequently, we excluded 7 participants aged 4.13-6.91 whose usable word count was < 100 usable words (3.57% of participants). Following this exclusion, the concordance between coder and linguistic measures of abstractness was equivalent across age: there was no longer a significant interaction between age and linguistic abstractness scores in predicting coder abstractness scores,  $\beta = .19, p = .157$ . Researchers should note that having sufficient linguistic data appears to be important for computing stable linguistic abstractness scores.



**Figure S1.** Linguistic abstractness scores across age when including data from all participants. The 7 red points represent participants who were excluded from primary analyses due to having overall low levels of usable linguistic data. Excluding these participants stabilized the correlation between coder abstractness scores and linguistic abstractness scores across age. Line of best fit and 95% CIs (grey shading) represent results from the thin plate smoothing spline model.

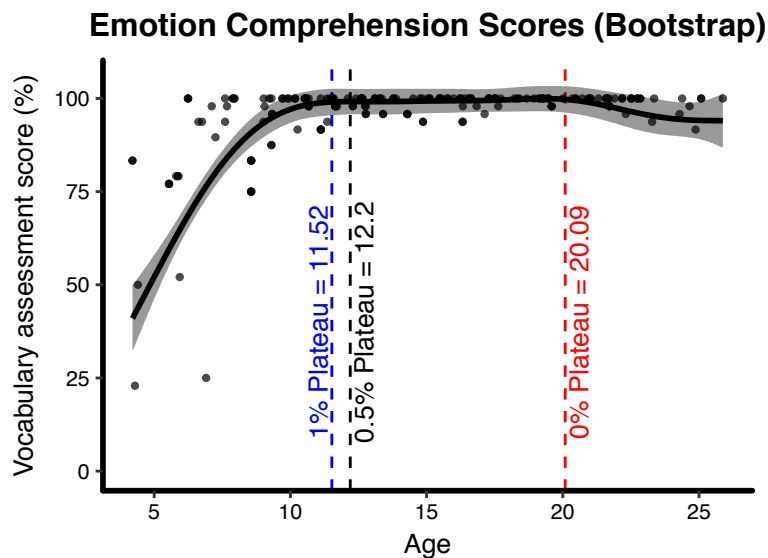
## II. Justification of plateau definition

Because the field has not developed standard thresholds for defining plateaus of nonlinear fit lines, we created a definition of a plateau by adapting mathematical principles for use with empirical data. As stated in the main text, a plateau can be mathematically defined as occurring when (i) there is reduced rate of change in the dependent variable and (ii) the dependent variable reaches the maximum/minimum of its development. However, it was important that this mathematical definition could be flexibly applied across datasets, especially given that we applied it to bootstrapped simulated datasets. Thus we began with “mathematical ideals” and ended our search when we had discovered a definition that could accurately detect plateaus across datasets.

An ideal mathematical definition of a plateau would be the point at which the first derivative of the curve reaches 0. However, applying this definition to empirical data—which are far noisier than mathematical ideals—led to clearly inaccurate definitions of the plateau in some bootstrapped simulations. One such simulated bootstrap sample is presented in **Figure S2** below. As you can see, the slope of the curve clearly slows and flattens in childhood, but the 0%/year threshold (red line) is only reached in young adulthood. Thus this threshold can be set too strictly to produce an accurate definition. As further evidence that this threshold is too strict, around 7% of bootstrapped simulations of example situation use supposedly did not reach a plateau within the measured age window, but an examination of these simulated samples suggested otherwise: the curve clearly plateaued, but it did not reach the 0%/year threshold.

Consequently, we adjusted this threshold to reasonable values (e.g., 0.05%/year, 0.10%/year, 0.50%/year, 1%/year) and tested how well these thresholds fit the data. We found that the 1%/year threshold was slightly too lenient, as the plateau was identified before the slope

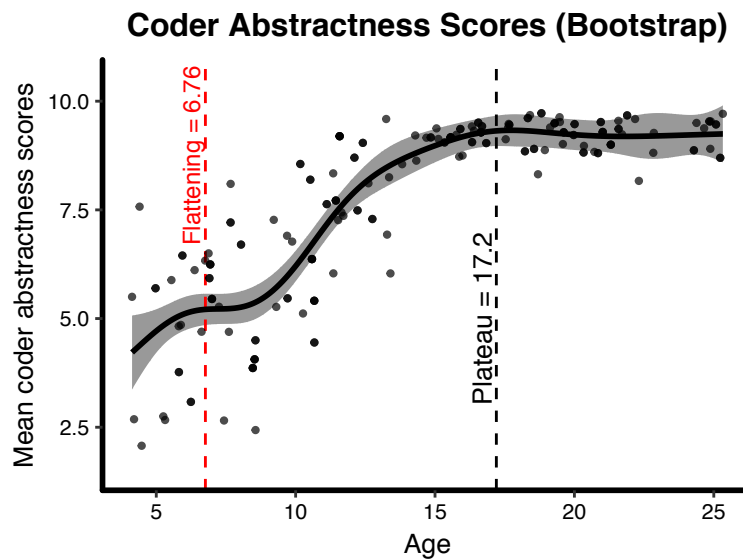
appeared to fully flatten (see blue line in **Figure S2**). Thus we selected 0.5%/year as the threshold because it provided a good balance between these two demands (see black line **Figure S2**). This not only removed inaccuracies in the emotion comprehension score bootstraps (as depicted in **Figure S2**), it also reduced “unidentified” plateaus in bootstrap simulations from 5.84% to 0.84% for example situation use, from 1.26% to 0% for coder abstraction scores, and from 0.11% to 0% for general definition use.



**Fig S2.** One bootstrapped sample of emotion comprehension scores (i.e., a randomly drawn sample from the study’s data with replacement that was used to make bootstrapped confidence intervals of plateaus) that illustrates the influence of the first-derivative threshold on how plateaus are identified. Setting the threshold to 0%/year causes the plateau to be identified almost 8 years after the curve appears to flatten (red dashed line), and setting the threshold to 1%/year causes the plateau to be identified slightly before the curve appears to flatten (blue dashed line). Hence, 0.5%/year (black dashed line) was chosen to provide a well-fitting balance between these demands.

However, this criterion alone was not sufficient for accurately defining plateaus. In particular cases, the dependent variable had a slope that showed only a *temporary* flattening far before the variable hit a final plateau point (e.g., at age 6.76, see red line in **Figure S3** below). Even though change in the dependent variable slowed to < 0.5%/year at this point, scores increased again after this temporary flattening and then eventually plateaued at the end of

adolescence. Thus we added a second criterion such that plateaus were only characterized once the age-related change had completed. We defined this as within the 95% CI of the smoothing spline's estimate of the maximum value of the curve (i.e., within the grey window surrounding the highest point of the black line). This ensured that a “plateau” wasn't when the curve flattened *momentarily* but when it flattened at the peak of development. We observed that adding this second definition ensured that all plateaus were defined accurately (black dashed line in **Figure S3**).



**Figure S3.** One bootstrapped sample of coder abstractness scores (i.e., a randomly drawn sample from the study's data with replacement that was used to make bootstrapped confidence intervals of plateaus) that illustrates the influence of the additional criterion for identifying plateaus (the value of the spline regression must be within the 95% CI of the maximal value of the spline) on how plateaus are identified. Without this criterion, any temporary flattening of the spline curve could be inaccurately labeled a plateau (red dashed line), but with this criterion, the accurate plateau is identified (black dashed line).

This is the process by which we created a specific statistical definition from the general mathematical principles of plateaus. Because the bootstrapping approach forced us to test the stability of this definition across 10,000 samples of 4 different dependent variables, there is reason to believe it can locate plateaus with high accuracy, stability, and generalizability.

### III. Analyses only including participants who fully comprehended all emotion words

Measures of emotion abstraction presented in the manuscript only include data from emotions that participants fully comprehended. Because younger participants comprehended fewer emotion terms than older participants (see **Results**), abstraction scores may reflect different sets of emotions for different ages. This could result in lower abstraction measures in younger ages because the emotions they understood were less abstract. Although we conceptualized this possibility as a part of the phenomenon this study investigates, we conducted supplemental analyses to ensure that age-related increases in emotion abstraction were not *solely* due to age-related differences in emotion comprehension. We used linear regressions to test whether each dependent variable was significantly related to age when restricting the sample to participants who (i) fully comprehended all 24 emotions and (ii) were within the active developmental window for that dependent variable.

For dependent variables that followed emergent developmental patterns (i.e., coder abstractness scores, general definition use, and example situation use), we restricted the sample to the 64 participants aged 6.24-17.54 who fully comprehended all 24 emotion words and were younger than the plateau of coder abstractness scores ( $M = 12.85$ ,  $SD = 2.98$ , 53.13% female). We used an age ceiling to restrict this sample to the active window of increasing emotion abstraction. Linear regressions revealed that age was significantly related to coder abstractness scores,  $\beta = .79$ ,  $p < .001$ , general definition use,  $\beta = .72$ ,  $p < .001$ , and example situation use,  $\beta = -.68$ ,  $p < .001$ , even within this restricted sample.

For dependent variables that followed linear developmental patterns (i.e., linguistic abstractness scores and synonym use), we restricted the sample to the 113 participants aged 6.24-25.91 who fully comprehended all 24 emotion words ( $M = 16.36$ ,  $SD = 4.81$ , 53.10% female).

No age ceiling was placed on this subset of participants, as these dependent variables increased linearly across the full sample (see **Results**). Again, linear regressions revealed that both linguistic abstractness scores,  $\beta = .45$ ,  $p < .001$ , and synonym use,  $\beta = .26$ ,  $p = .006$ , significantly increased across age even within this restricted sample.

These supplementary analyses suggest that age-related differences in emotion comprehension do not confound analyses of emotion abstraction presented in the main text. Age-related changes in emotion abstraction arose even when emotion comprehension was held constant.

#### IV. Analyses excluding trials on which participants only provided synonyms

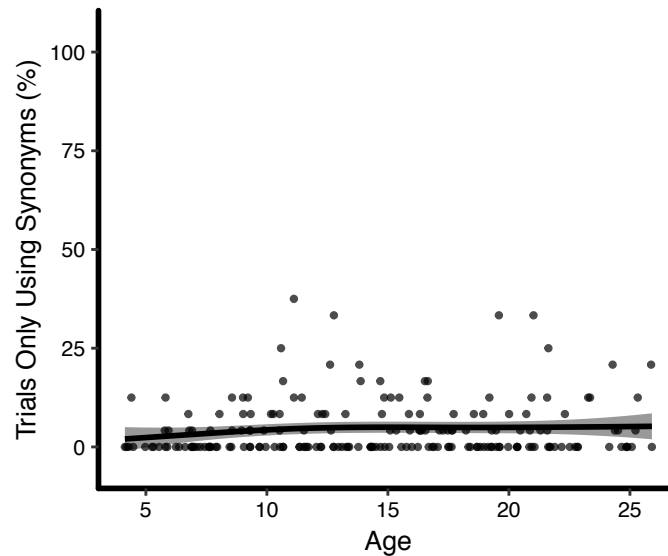
We originally counted synonyms as demonstrating full comprehension because i) Wechsler vocabulary assessments (on which this test was modeled) grant synonyms full credit for several items (Wechsler, 1999) and ii) providing synonyms demonstrates at least some understanding of how words within a taxonomy should be organized (i.e., that two words refer to the same concept and not different concepts). However, only providing a synonym might reflect shallow semantic priming rather than full comprehension. Hence, we conducted control analyses to protect against the possibility that synonym use did not reflect full comprehension.

Overall, the percentage of trials on which participants only provided synonyms was 4.49% (4.90% of 2-point responses). We analyzed relations between age and the tendency to only provide synonyms (quantified as the number of trials on which participants only provided a synonym) as we did for all other variables. AICs suggested that the smoothing-spline model provided a slightly better fit than the null model (**Table S1**). However, three points of evidence suggest that there is not significant variation in the tendency to only provide synonyms across age: i) there was no significant linear relationship with age ( $p = .193$ ), ii) the  $R^2$  of this model was only .01, and iii) no plateau could be defined because the rate of change of this variable was always lower than the plateau threshold of .5%/year (i.e., according to the plateau definition set for other variables, this variable doesn't change at all over the age window we sampled).

**Table S1.** Results of null, polynomial, and thin plate smoothing spline analyses of the tendency to only provide synonyms.

Dependent variable	Model	$\beta$	$p$	Adj $R^2$	AIC
Tendency to only provide synonyms (% of trials)	Null				1327.46
	Linear	.09	.193	.004	1327.74
	Quadratic	-.08	.243	.006	1328.35
	Cubic	.10	.163	.01	1328.36
	<b>Spline</b>			<b>.01</b>	<b>1327.28</b>

Note: Bold text indicates best fitting model for each dependent variable, as determined by having the lowest AIC.  $\beta$  = standardized beta, Adj  $R^2$  = adjusted  $R^2$ , AIC = Akaike Information Criterion



**Figure S4.** The percentage of trials on which participants only provided synonyms presented across age. The black line and grey shaded region present the fit of the smoothing spline regression and 95% CI, respectively.

We then conducted a set of analyses to test how dramatically results would shift if we assumed that trials on which participants only provided synonyms actually did not reflect comprehension of the emotion words. We recoded all trials on which participants only provided synonyms as 0-point responses (which necessitated removing all abstraction measures from these trials) and recomputed all dependent variables. Overall, we find that patterns of results are very similar to what is presented in the main text, and all major conclusions remain unchanged.

The best fitting age models do not change for all dependent variables (note that synonym use was not analyzed as a dependent variable for these analyses; **Table S2**). Emotion comprehension scores, coder abstractness scores, general definition use, and example situation use followed emergent patterns; linguistic abstractness scores increased linearly across development; and physiological marker use was elevated in childhood. Although rescored synonym-only responses as incorrect shifted the plateau of emotion comprehension scores slightly earlier (from age 10.95 to age 10.17), abstraction plateaus remained stable: the coder

abstractness score plateau remained at age 17.69, general definition use plateau shifted from age 17.19 to age 17.08, and the example situation use plateau remained at age 18.19.

Additionally, the significance of bootstrapped comparisons between plateaus remained identical. The emotion comprehension plateau remained significantly earlier than the plateaus of coder abstractness scores, 95% CI = [3.69, 11.00], general definition use, 95% CI = [3.90, 11.52], and example situation use, 95% CI = [4.83, 15.00]. Additionally, bootstrapped comparisons suggested no significant differences between plateaus for measures of emotion abstraction: coder abstractness scores vs. general definition use, 95% CI = [-1.79, 3.56], coder abstractness scores vs. example situation use, 95% CI = [-4.43, 0.59], and general definition use vs. example situation use, 95% CI = [-1.25, 5.48]. Note that these bootstrap comparisons were calculated after removing 0.65% of example situation simulations in which a plateau was not detected.

These results suggest that conclusions presented in the main text are not affected by the possibility that providing a synonym might not reflect full emotion word comprehension. Nonetheless, it may be wise for future researchers who use the emotion vocabulary assessment to only grant 2 points to responses that include example situations or general definitions.

**Table S2.** Results of null, polynomial, and thin plate smoothing spline models for each dependent variable after removing trials on which participants only provided synonyms.

Dependent variable	Model	$\beta$	$p$	Adj $R^2$	AIC
Emotion comprehension scores	Null				1631.82
	Linear	.41	< .001***	.16	1597.78
	Quadratic	-.43	< .001***	.35	1550.54
	Cubic	.26	< .001***	.41	1530.98
	<b>Spline</b>			<b>.45</b>	<b>1519.10</b>
Coder abstractness scores	Null				847.61
	Linear	.77	< .001***	.59	672.65
	Quadratic	-.38	< .001***	.73	590.42
	Cubic	-.01	.785	.73	592.34
	<b>Spline</b>			<b>.75</b>	<b>583.25</b>
Linguistic abstractness scores	Null				-133.81
	<b>Linear</b>	<b>.48</b>	<b>&lt; .001***</b>	<b>.23</b>	<b>-180.81</b>
	Quadratic	-.0005	.994	.22	-178.81
	Cubic	-.07	.278	.22	-178.01
	Spline			.23	-180.81 <sup>+</sup>
General definition use	Null				1860.86
	Linear	.73	< .001***	.54	1710.86
	Quadratic	-.44	< .001***	.73	1608.55
	Cubic	.09	.019*	.73	1604.90
	<b>Spline</b>			<b>.74</b>	<b>1601.39</b>
Example situation use	Null				1881.05
	Linear	-.69	< .001***	.47	1758.78
	Quadratic	.33	< .001***	.57	1716.12
	Cubic	.02	.642	.57	1717.90
	<b>Spline</b>			<b>.58</b>	<b>1713.33</b>
Physiological marker use	Null				1403.99
	Linear	-.24	< .001***	.05	1394.74
	Quadratic	-.006	.933	.05	1396.73
	Cubic	.12	.084	.06	1395.67
	<b>Spline</b>			<b>.09</b>	<b>1390.08</b>

Note: Bold text indicates best fitting model for each dependent variable, as determined by having the lowest AIC. <sup>+</sup>In these two cases, spline model algorithms revealed that the best fits were identical to linear models; hence we interpreted the linear models.  $\beta$  = standardized beta, Adj  $R^2$  = adjusted  $R^2$ , AIC = Akaike Information Criterion, \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$

## V. Emotion vocabulary assessment

**Materials.** 27 laminated index cards with emotions written on them, scoring manual, scoring sheet, and computer for audio recording. Emotions: amazed, angry, annoyed, bored, calm, disappointed, disgusted, embarrassed, excited, grumpy, happy, hate, jealous, lonely, love, nervous, pleased, proud, relaxed, sad, safe, scared, sorry, surprised, thankful, upset, and worried.

**General recommendations.** Make sure that the deck of emotion cards has been shuffled multiple times prior to starting the task so that emotion order is random. Avoid providing corrective or praising feedback in this task. Attempt to keep them motivated but not bias their answers. Use encouraging but neutral feedback. For example, you might say things like “ok”, “let’s try another”, or “you’re doing good work” rather than “that’s correct.” Try to only look at the scoring sheet while they are speaking (i.e., don’t give them feedback via eye contact), and try to resist nodding or giving other forms of feedback. Try not to cut them off after they provide a suitable response, as they should share whatever they think about each emotion.

**Procedure.** Begin by providing them with the following instructions: “This game is called the Word Definitions Game. In this game, I’m going to say a word, and I want you to do your best to tell me what the word means. I won’t be able to talk to you about the game until we’re all done with the study. Also, we should only spend a few minutes on this game, so try to give each definition in about 30 seconds or less. Are you ready to start?” Start the audio recorder. On each trial, pull a card from the deck of emotion terms, show it to them, and say: “What does \_\_\_\_\_ mean?” Follow the scoring guidelines below for each trial and probe them if their responses do not initially earn 2 points.

**Overall scoring guidelines.** Two-point answers include: (i) a reasonable description of the definition, even if the definition is not *exact* (see example definitions below), (ii) a synonym

(even if colloquial) or closely related emotional term (e.g., “pissed off” or “mad” for angry), or (iii) an example situation that would reasonably evoke the emotion in question and not other emotions (e.g., “I felt sad when my pet died” for sad). If the participant provides an example situation that could evoke multiple emotions (e.g., “I felt angry when my sister got something that I wanted,” which could evoke jealousy or anger), query them to ask for a more specific definition or other situations that can confirm that they know what the word means. **Note:** It is possible that synonym use actually represents shallow semantic priming rather than true comprehension of an emotion word’s underlying concept. Although we found that trials on which participants only provided synonyms were rare (4.49% of all trials), we encourage future researchers to consider not counting synonyms as 2-point responses to avoid the possibility of mistakenly accepting priming as comprehension.

One-point answers include a response that is of the correct valence but overly vague. No features are provided to distinguish this emotion from other similar emotions (e.g., “good,” “bad,” “positive,” “negative,” etc.). This includes responses that involve situations that could give rise to several emotions of a similar valence as the target emotion. If participants’ responses only earn 1 point, respond with a query (e.g., “can you tell me more about [emotion]?” or “can you tell me more about that situation?”). Repeat queries as needed to probe participant’s understanding of the emotion term. Score the participant’s response based on their responses to queries (i.e., if a 2-point response is given in response to queries, they earn 2 points for this emotion, even if their initial response earned 1 point or 0 points).

Zero-points are awarded when it is clear that the participant does not know what the emotion word means, even after probing. These responses include: (i) an incorrect definition of the term, (ii) an example of a situation that is unlikely to evoke the emotion in question (e.g., “I

feel happy when my pet dies”), or (iii) the response “I don’t know.” For all zero-point answers, query to thoroughly assess the participant’s understanding of the emotion term. Score their response based on their responses to queries (i.e., they can earn 1 or 2 points for the emotion if later responses clarify their understanding of the emotion term). Some example queries include saying (i) “it’s ok if you don’t know all of these words, but let’s just make sure. What do you think [emotion] means?”, (ii) “What are some other feelings that might be like [emotion]?”, (iii) “What are some other words people use to describe feeling [emotion]?”, (iv) “What are some things that might make someone feel [emotion]?”, (v) “Have you ever heard of someone feeling [emotion] in a book or movie? What made them feel that way?” However, use these latter forms of encouragement (probes for synonyms/situations) sparingly. One goal of the assessment is to measure how much people generate these types of strategies on their own, so only use them if thorough probing suggests they are not able to spontaneously generate a full definition.

**Definition and synonym guidelines.** Definitions below are based on: Merriam-Webster Dictionary for Adults, Merriam-Webster Dictionary for Children, Oxford American Dictionary, and Google Dictionary. Synonyms are taken from a variety of sources. These definitions are only to serve as a rough guide in your scoring.

***Angry.*** Definition: strong feeling of being upset, annoyed, displeased, or hostile.  
Synonyms: irate, mad, annoyed, cross, vexed, irritated, indignant, irked, furious, enraged, infuriated, in a temper, displeasure, fury, aggravated, livid; ticked off, pissed off; losing one’s temper.

***Annoyed.*** Definition: slightly angry; irritated. Synonyms: irritated, vexed, angry, exasperated, irked, get on someone’s nerves, ruffle someone’s feathers.

***Amazed.*** Definition: surprised greatly; confused; filled with wonder; filled with

astonishment. Synonyms: astonished, astounded, surprised, stun, stagger, shock, stupefy, awe, stop someone in their tracks, leave open-mouthed, leave aghast, take someone's breath away...

**Bored.** Definition: weary or restless because one is unoccupied or lacks interest in one's current activity. Synonyms: disinterested, dull, sick and tired, fatigued, blasé, spiritless, tired, inattentive, turned off.

**Calm.** Definition: in a quiet and peaceful state or condition; not feeling or showing nervousness, anger or other emotions. Synonyms: serene, tranquil, relaxed, unruffled, unperturbed, unflustered, untroubled...

**Disappointed.** Definition: sad, unhappy, or displeased because someone or something has failed to fulfill one's hopes or expectations. Synonyms: upset, saddened, let down, cast down, disheartened, downhearted, downcast, depressed, dispirited, discouraged, despondent, dismayed, distressed.

**Disgusted.** Definition: a strong feeling of dislike for something that has a very unpleasant appearance, taste, smell, etc.; annoyance and anger that you feel toward some behavior that is not good, fair, appropriate; revulsion or profound disapproval. Synonyms: revolt, repel, repulse, sicken, nauseate, turn someone's stomach.

**Embarrassed.** Definition: confused and foolish in front of other people; self-consciousness, shame, or awkwardness. Synonyms: mortified, red-faced, blushing, abashed, shamed, ashamed, humiliated, awkward, self-conscious, uncomfortable...

**Excited.** Definition: eager enthusiasm and interest. Synonyms: thrilled, exhilarated, animated, enlivened, electrified.

**Grumpy.** Definition: easily annoyed or angered; having a bad temper or complaining often; sulky. Synonyms: crabby, ill-tempered, short-tempered, crotchety, testy, irritable, prickly,

grouchy, snappy, cranky, irritable.

**Happy.** Definition: pleasure and enjoyment because of life, situation, etc.; contentment. Synonyms: cheerful, cheery, merry, joyful, jovial, jolly, jocular, gleeful, delighted, untroubled, smiling, beaming, grinning, in good spirits, in a good mood, lighthearted, pleased, content, satisfied, gratified, sunny, joyous.

**Hate.** Definition: intense or passionate dislike (for someone); have a strong aversion to (something). Synonyms: loathe, detest, despise, dislike, abhor, repelled by...

**Jealous.** Definition: intolerant of rivalry or unfaithfulness; envy of someone or their achievements and advantages. Synonyms: envious, covetous, desirous.

**Lonely.** Definition: sad from being apart from other people; sad because one has no friends or company; solitary. Synonyms: isolated, alone, lonesome, friendless, with no one to turn to, abandoned, rejected, unloved, unwanted, outcast, gloomy, sad, depressed, desolate, forsaken.

**Love.** Definition: strong or constant affection for a person; feel a deep romantic or sexual attachment to someone. Synonyms: deep affection, fondness, tenderness, warmth, intimacy, attachment, endearment, care very much for, hold very dear, adore, think the world of, be devoted to, dote on, worship.

**Nervous.** Definition: worried and afraid about what might happen; easily agitated or alarmed; tending to be anxious; highly strung. Synonyms: high-strung, anxious, edgy, tense, excitable, jumpy, skittish, brittle, neurotic.

**Pleased.** Definition: happy or satisfied. Synonyms: happy, glad, delighted, grateful, thankful, content, satisfied.

**Proud.** Definition: deep pleasure, satisfaction, or happiness as a result of one's own

achievements, qualities, or possessions or those of someone with whom one is closely associated; attitude or people who think that they are better or more important than others.

Synonyms: pleased, glad, happy, delighted, joyful overjoyed, thrilled, satisfied, gratified, content.

**Relaxed.** Definition: calm and free from stress, worry, or anxiety; free from tension and anxiety; at ease. Synonyms: comfy, cozy, relaxed, content, satisfied, peaceful, resting, easygoing, undisturbed.

**Sad.** Definition: grief or unhappiness; sorrow. Synonyms: unhappy, sorrowful, dejected, depressed, downcast, miserable, down, blue, down in the dumps, blah.

**Safe.** Definition: free from harm or risk; secure from threat of danger, harm, or loss. Synonyms: secure, protected, shielded, sheltered, guarded, out of harm's way.

**Scared.** Definition: afraid of something; nervous, frightened, fearful. Synonyms: afraid, startled, nervous, fearful, panicky, alarmed, intimidated, terrified, petrified, terrorized, spooked.

**Sorry.** Definition: sorrow or regret; distress, especially through sympathy with someone else's misfortune. Synonyms: sad, unhappy, sorrowful, distressed, upset, downcast, downhearted, despondent.

**Surprised.** Definition: feelings caused by something that is unexpected or unusual. Synonyms: amaze, astonish, astound, dumbfounded, shocked, startled.

**Thankful.** Definition: glad that something has happened or not happened, that something or someone exists, etc. pleased and relieved; gratitude. Synonyms: grateful, appreciative, filled with gratitude, relieved.

**Upset.** Definition: unhappy, worried; disappointed. Synonyms: distress, trouble, unsettled, worried, bothered, agitated, hurt, sadden, grieve, perturbed, disquieted.

***Worried.*** Definition: fear or concern because you think that something bad has happened or could happen; anxious, upset, or troubled about actual or potential problems. Synonyms: anxious, perturbed, troubles, bothered, concerned, upset, distressed, uneasy, agitated, nervous, edgy, tense, keyed up, jumpy, stressed, strung out.