UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DOCTORAL DISSERTATION

Head Office: Università degli Studi di Padova

Department of Developmental Psychology and Socialization

Ph.D. COURSE IN: PSYCHOLOGICAL SCIENCE

SERIES: XXXVIII

SEEING AND HEARING EMOTIONS: MULTIMODAL AFFECTIVE PERCEPTION IN PREVERBAL INFANTS

Coordinator: Ch.ma Prof.ssa Lucia Regolin

Supervisor: Ch.ma Prof.ssa Teresa Farroni

Co-Supervisor: Ch.ma Prof.ssa Judit Gervain

Ph.D. student: Chuchu Jia

Table of Contents

ABSTRACT

The ability to perceive and integrate affective information across sensory modalities is fundamental to infants' emerging socioemotional development. In the first year of life, preverbal infants rely heavily on nonverbal cues, such as facial expressions and vocal prosody, to interpret emotions and navigate social interactions. Understanding how infants process these multimodal emotional signals provides critical insight into the developmental foundations of emotion perception and social communication. This thesis investigates the developmental trajectory of multimodal emotion perception in infancy through three complementary studies that incorporate evidence from a systematic review, behavioral experiments, and neuroimaging.

Study 1 presents a systematic review that synthesizes empirical evidence on infants' multimodal perception of affective information across facial and vocal modalities during the first year of life. Findings from 47 experiments indicate that by 5 to 7 months of age, infants demonstrate sensitivity to the congruency between facial and vocal expressions, often exhibiting a preference for happy stimuli. The review further highlights that dynamic and socially relevant stimuli enhance intermodal matching performance, and the brain exhibits differential patterns in multimodal emotion processing.

Study 2 reports a behavioral experiment involving 35 infants aged 6 to 8 months, which employed a visual preference-based intermodal matching paradigm to assess their ability to match emotions across sensory modalities. Infants were presented with face–voice emotional pairings under three conditions: happy–angry, happy–neutral, and angry–neutral. Results indicate that infants reliably matched congruent happy expressions, exhibiting a visual preference for happy faces when paired with a corresponding vocal emotion. Additionally, emotion-matching performance was significantly correlated with both age and family expressiveness, as measured by parent-reported questionnaires, suggesting that both developmental stage and environmental factors contribute to the early perception of emotion.

Study 3 investigates the neural mechanisms underlying infants' sensitivity to multimodal emotional congruency using functional near-infrared spectroscopy (fNIRS). Infants aged 6 to 7 months are presented with congruent (e.g., happy face + happy voice) and incongruent (e.g., happy face + angry voice) emotional stimuli, while hemodynamic responses are recorded from

bilateral temporal and prefrontal regions. Results are expected to show differential neural activation between congruent and incongruent emotional conditions, suggesting that infants exhibit distinct neural responses to multimodal affective cues. This study aims to provide neurophysiological evidence of early affective integration and to identify the cortical brain regions involved in processing multimodal emotional information during infancy.

Taken together, these studies offer converging evidence for the early emergence of multimodal emotion perception in infancy and its modulation by both developmental and social environment factors. The findings contribute to a growing understanding of how infants begin to interpret complex affective information and may inform future research into the neural and experiential mechanisms that support early socioemotional development.

CHAPTER I.

INTRODUCTION AND CONCEPTUAL FRAMEWORK

1.1 Early Emotion Cognition in Infancy

1.1.1 Definition and Early Characteristics

In recent years, emotions have attracted increasing scholarly attention in infant development research and have gradually become a central focus in conceptualizations of early development. From the earliest stages of life, infants not only display distinct emotional expressions and temperamental tendencies but also begin to form initial emotional bonds with caregivers through their affective responses. A systematic examination of the functions and developmental trajectories of emotions in infancy can advance our understanding of the nature of emotions and highlight the essential role they play in shaping later socio-emotional development.

Psychologists have proposed multiple perspectives on the definition and classification of emotions. In developmental psychology, emotions are commonly defined as complex response patterns that arise from the evaluation of internal and external events, encompassing subjective experience, physiological arousal, expressive behavior, and action tendencies (Campos et al., 1989; Izard, 1993). Emotions serve both adaptive functions, enabling individuals to cope with environmental challenges, and social functions, facilitating the regulation of interactions with others, particularly when they affect individual well-being (Campos et al., 1994). Although the meaning of emotions extends beyond communication per se, during infancy, emotions primarily function as a communicative tool (E. Z. Tronick & Cohn, 1989). Through emotional signals, infants establish connections with their environment and gradually develop the foundations of social interaction and self-regulation.

Some researchers argue that infant emotions constitute an innate biopsychological system, which in early life primarily manifests as basic emotions such as joy, surprise, anger, fear, disgust, and sadness. These emotions gradually differentiate during the process of socialization and become increasingly intertwined with cognition, language, and social experience (Lewis et al., 2010; Saarni, 1999). Developmental psychology places particular emphasis on the social nature of emotions: they are not only reflections of internal states but also serve as communicative signals. Through emotional expressions, infants guide caregivers' responses to meet their needs, a process that forms both a core mode of interaction with the environment and a critical foundation for socialization and cognitive growth.

Regarding the classification of emotions, while existing theories differ in emphasis, most frameworks distinguish between positive emotions (e.g., joy, enthusiasm) and negative emotions (e.g., anger, sadness) (Izard et al., 1991; J. A. Russell, 1980; Watson et al., 1988). Moreover, two classical approaches have had a profound impact on research in emotional development: the basic emotions theory and the discrete emotions theory. The basic emotions theory suggests that certain emotions are biologically based and universally present across cultures, including happiness, surprise, anger, fear, disgust, and sadness (Ekman, 1992a). Building on this perspective, the discrete emotions theory further suggests that these emotional systems are relatively distinct from birth and progressively differentiate and mature throughout development (Izard, 1994). Empirical evidence indicates that infants can already display positive and negative basic emotions within the first months of life and begin to differentiate discrete emotions such as fear, anger, and sadness around 6–8 months, whereas self-conscious emotions (e.g., pride, jealousy) depend on more advanced cognitive and social development.

1.1.2 Influential Factors

From the perspective of developmental psychology, emotional development is a dynamic process shaped by the interplay of multiple factors. These include biological characteristics, such as neural maturation and temperament, as well as environmental influences, including parent–infant interactions, social relationships, and cultural contexts. From an evolutionary standpoint, emotions are grounded in innate biological foundations while also being shaped by experiential factors (Campos et al., 1994; Darwin & Darwin, 1872; Sroufe, 1997). The biological roots of emotion are closely linked to early neural development, particularly involving structures such as the limbic system and brainstem (LeDoux, 1996; Panksepp, 2004). Evidence suggests that even in the earliest months of life, infants display basic emotions such as joy, anger, and distress, indicating that the underlying neural mechanisms are functional from the outset.

At the same time, parent–infant interactions and broader social relationships provide crucial support for the expansion of infants' emotional experiences (Feldman, 2007b; E. Z. Tronick & Cohn, 1989). Caregivers play a central role in emotion regulation; for instance, when infants cry or exhibit distress, soothing behaviors not only alleviate their negative affect but also reduce stress hormone levels (Kirby, 2024). In daily life, parents may elicit positive affect by engaging in playful interactions or, conversely, expose infants to distress through family conflict, illustrating the concrete impact of social interaction on emotion regulation.

Cultural contexts also exert a profound influence on how emotions are expressed. While basic emotions are thought to be universal across human cultures, their modes and intensities of expression vary. For instance, Asian infants generally exhibit lower frequency and intensity of both positive and negative emotions compared to non-Latino White infants (Camras et al., 2006), partly because Asian parents are more likely to encourage emotional restraint rather than overt expression (Cole & Tan, 2007). Further evidence shows that Japanese parents tend to suppress children's negative emotional displays, whereas non-Latino White mothers are more likely to intervene actively when their children experience distress (Cole & Tan, 2007). In sum, emotional development is rooted in human biological evolution, as well as profoundly shaped by cultural environments and interpersonal relationships. Biological mechanisms provide the foundational architecture for emotions, while social interactions and cultural practices offer emotional experience with its diversity and complexity (Song et al., 2014).

Some developmental psychologists conceptualize emotions as products of adaptation to the environment (Saarni, 2008), emphasizing that infants' emotional responses are always embedded within the specific contexts in which they occur, and in early life, are primarily defined by interpersonal interactions. Within this framework, emotional expression and the recognition of others' emotions serve to convey internal states, regulate infants' own behavior, and shape the responses of others. From a functionalist perspective, emotions are understood as processes unfolding within dynamic social interactions rather than as isolated phenomena (Lamb et al., 2013). For instance, when an infant cries, a caregiver's sensitive and appropriate response can facilitate the establishment of early emotional bonds. Empirical evidence demonstrates that caregivers' gaze direction significantly influences infants' allocation of attention (Farroni et al., 2002). During such interactions, parents often adjust their responses in accordance with infants' emotional expressions, while infants, in turn, modify their emotional displays based on parental feedback, illustrating the bidirectional nature of regulation (Bridgett et al., 2009). Under conditions of sensitive and timely responsiveness, caregivers' appropriate feedback can foster infants' emotional development regardless of whether the baby expresses positive or negative affect (Ainsworth, 1979).

Moreover, functional theories of emotion highlight the close association between emotions and goal-directed behavior (Campos et al., 1989, 1994). Individuals typically experience positive emotions when goals are achieved, whereas blocked or frustrated needs are often expressed as anger or sadness. The nature of the goal shapes the type of emotional experience; for instance, during the first year of life, infants display heightened sensitivity and avoidance toward threat-related cues, which is closely linked to the emergence of fear responses. As development progresses, the functions of emotion gradually expand. Emotions not only enable individuals to monitor their internal states and perceive the emotions of others but also provide a foundation for social learning. Once language skills are acquired, individuals can articulate their emotional experiences within culturally specific contexts and begin to develop empathy and compassion. With the maturation of social functioning, they also come to recognize the distinction between internal emotional experience and external expression, learning to employ regulatory strategies. Emotional functioning thus becomes critical for the acquisition of multiple social competencies, such as emotion management, resilience in stressful contexts, and the establishment of positive interpersonal relationships.

1.1.3 Developmental Functions

The early course of emotional development in infancy can be understood through Lewis's (2000, 2007) influential distinction between primary emotions and self-conscious emotions. Primary emotions, which are shared with nonhuman animals, emerge within the first six months of life and include surprise, interest, joy, anger, sadness, fear, and disgust. By contrast, self-conscious emotions, such as guilt, pride, shame, empathy, and jealousy, are intrinsically linked to self-concept and social evaluation, and their developmental onset remains contested. While some studies argue that these emotions do not appear until the second year of life (Lewis et al., 1989), others suggest that precursors of these emotions may be observable earlier (Draghi-Lorenz et al., 2001). LeDoux (1996) contends that, due to the immaturity of neural structures, infants are unlikely to experience self-reflective emotions during the first year.

Despite these debates, there is general agreement that the development of emotion is not an isolated maturational process but is deeply intertwined with attachment styles and broader trajectories of socio-emotional growth. Attachment refers to the close emotional bond formed between individuals, and the relationship between infants and their primary caregivers has long been a central focus of developmental psychology. One of the most influential contributions to attachment theory is Ainsworth and colleagues' (1979) Strange Situation paradigm. In this procedure, infants are exposed to a structured sequence of episodes involving the caregiver and a stranger, including separations and reunions. Observations of infants' behavioral responses across these episodes allow researchers to classify attachment styles. Findings identified two broad categories: secure and insecure attachment, with the latter further subdivided into avoidant and ambivalent types. Each attachment style is associated with distinctive behavioral characteristics. Infants with secure attachment use the caregiver as a "secure base" from which to explore novel environments. Avoidant infants display distancing or avoidance of contact, reflecting underlying insecurity. Ambivalent infants exhibit heightened distress and may respond to the caregiver's return with resistance or anger.

Caregivers' sensitivity and timely responsiveness to infants enhance a sense of security and facilitate the formation of secure attachment. Conversely, when caregivers provide limited or inconsistent emotional feedback or when infants experience difficulties in recognizing and interpreting others' emotions, the likelihood of developing insecure attachment increases (Ainsworth, 1979). Bowlby (1979) emphasized that the quality of attachment is largely determined by infants' ability to interpret caregivers' emotions and behaviors, as well as by the emotional support caregivers provide within social interactions. Because emotional expression functions as a critical communicative signal in attachment relationships, positive interactions not only foster infants' experiences of joy but also provide effective comfort during distress. Taken together, these findings indicate that infants' early capacities for emotion perception and expression are closely tied to the development of attachment patterns, highlighting the importance of investigating the cognitive mechanisms of emotion in developmental research.

Beyond attachment classifications, infants are immersed in the social world from the earliest stages of development. A substantial body of research demonstrates that infants are innately predisposed to attend to social stimuli. For instance, they preferentially orient toward faces and face-like cues (Fantz, 1961; Farroni et al., 2002; Johnson et al., 1991), and are especially responsive to human voices, particularly those of caregivers (DeCasper & Fifer, 1980). Consequently, across diverse cultural contexts, caregiver–infant face-to-face play emerges around 2–3 months of age. Such interactions, which integrate facial expressions, vocal characteristics, and gestures (Brazelton et al., 1974; E. D. Tronick et al., 1977), provide a foundation for the later development of emotion cognition skills. Empirical findings further indicate that during this developmental window, infants also actively anticipate positive emotional signals from others. Research employing the still-face paradigm has shown that infants expect caregivers to respond with smiles or positive vocalizations; when caregivers remain expressionless and unresponsive, infants exhibit increased withdrawal and negative affect (Haley & Stansbury, 2003; Weinberg & Tronick, 1996). A meta-analysis further revealed that the classic still-face effect of reduced positive affect and gaze, and increased negative affect, as well as a partial carry-over effect into the reunion episode, consisting of lower positive and higher negative affect compared to baseline (Mesman et al., 2009).

Another key social-cognitive achievement in infancy is the gradual development of the ability to interpret others' emotions (Walker-Andrews, 1997). Before acquiring language, infants rely primarily on emotional cues such as facial expressions, vocal tone, and eye contact to infer others' intentions and affective states, thereby enabling effective social interaction (E. Z. Tronick & Cohn, 1989). A central manifestation of this capacity is social referencing, whereby infants in uncertain situations observe others' emotional reactions to guide their own behavior (Walden & Ogan, 1988). Through social referencing, infants become better able to interpret ambiguous contexts and select appropriate responses, with emotion understanding serving as its necessary foundation (Campos, 1982). For instance, infants use parental emotional signals to determine whether to approach or avoid strangers (Pelaez et al., 2012). By around 12 months, maternal facial expressions, such as smiling or fear, strongly influence infants' willingness to explore novel environments. In the classic "visual cliff" experiment, infants' crawling behavior was highly influenced by their mothers' emotional expressions (Sorce et al., 1985). Emotional cues not only guide infants' safe exploration but also promote joint attention and the comprehension of communicative intent (Farroni, Massaccesi, et al., 2007). Evidence further shows that caregivers' emotional expressions help infants integrate attention with referential behaviors, thereby supporting vocabulary acquisition and language development (Tomasello, 2014). Neuroimaging studies also reveal distinct brain activation patterns when infants process positive versus negative emotional cues, providing a neural basis for subsequent social communication (Grossmann et al., 2007; Leppänen & Nelson, 2009).

In sum, research consistently demonstrates that during the first year of life, infants develop increasing sensitivity to others' emotional signals. This ability enables adaptive responses in uncertain contexts, also lays a foundation for social interaction, emotion regulation, and the emergence of early language. Accordingly, investigating emotion cognition in infancy holds significant importance for understanding socio-emotional development.

1.2 The 6-8-Month Developmental Window

6-8 months of age is often described as a developmental "turning point" in infants' socio-emotional growth (Grossmann, 2010; Leppänen & Nelson, 2009; Nelson, 2001). At this stage, coordinated advances in perception, attention, learning, and socio-emotional functioning become evident. Infants' responses shift from being driven primarily by low-level stimulus features to showing heightened sensitivity to socially meaningful signals such as faces, vocal tones, and emotional expressions. This reorganization of emotion perception and social processing strategies provides the groundwork for higher-order socio-emotional learning (Grossmann, 2010; Leppänen & Nelson, 2009). These behavioral changes occur in parallel with the rapid maturation of cortical and subcortical networks involved in emotion processing, which begin to exhibit selective responsiveness to social signals. Such neural specialization provides the basis for multimodal integration and more efficient emotion recognition, suggesting the emergence of increasingly complex social interaction abilities (Grossmann, 2010; Leppänen & Nelson, 2009).

By 6-8 months, infants demonstrate greater stability in differentiating emotional expression (Caron et al., 1988; Flom & Whiteley, 2014; Flom & Bahrick, 2007). A common feature of this stage is the positivity bias: stronger responses to positive cues, such as smiling faces and joyful voices, compared with neutral or negative ones (Flom & Whiteley, 2014; Nawrot, 2003). This bias is thought to serve an adaptive function by facilitating caregiver–infant bonding and promoting social interaction (Grossmann, 2010; Leppänen & Nelson, 2009; Nelson, 1987). At the same time, preferential processing of negative emotions such as fear and anger begins to emerge, suggesting a developmental recalibration between positive and negative emotion processing (Grossmann, 2010). This period also marks a transition in face processing from broad sensitivity to species- and race-specific specialization. Six-month-old infants can still discriminate across species and racial groups, but by nine months this ability narrows significantly, reflecting a typical trajectory of perceptual narrowing; seven months represents a critical midpoint in this process (Pascalis et al., 2002).

Another striking developmental change at this age concerns attentional strategies. Compared with earlier stages, seven-month-old infants no longer sustain attention on a single emotional cue but instead flexibly allocate resources across multiple social signals. Although this shift may lead to shorter or less consistent responses to individual cues, it reflects a strategic reorganization of attention to cope with increasingly complex social environments. Neuroimaging and electrophysiological evidence further indicate that processing of socio-emotional signals becomes more functionally specialized during this period (Flom & Whiteley, 2014; Grossmann, 2010; Heck et al., 2016).

Taken together, these converging findings support the view that six to eight months constitutes a sensitive window for investigating infants' emotion perception, discrimination, and cross-modal integration. At this age, infants display sufficient neural and behavioral maturity to yield reliable experimental effects, while not yet having undergone the full perceptual narrowing associated with later stages. This balance of developmental readiness and experimental feasibility makes this age period an especially suitable age for the behavioral and neuroimaging studies reported in the present work.

1.3 Core Concepts and Definitions

1.3.1 Affective Information

In developmental psychology, affective information refers to emotional cues transmitted through multiple channels, such as facial expressions, vocal prosody, and body posture, that observers can use to infer others' emotional states (Zadra & Clore, 2011). For infants, these cues constitute the primary input for emotion perception and social interaction: by extracting, processing, and integrating such signals, infants infer others' emotional states and social intentions (Bahrick & Lickliter, 2000; Grossmann, 2010; Walker-Andrews, 1997). In the present study, affective information is operationalized as the emotional meaning conveyed by visual (facial expressions) and auditory (prosodic/intonational) signals. Developmental evidence shows that infants are sensitive to socio-emotional cues from birth (e.g., preferential responses to maternal faces and voices) and, by 6–8 months, exhibit more advanced processing of multimodal affective information. This includes improved intermodal matching and enhanced abilities for multimodal integration, marking a critical window in the development of emotion understanding that provides a mechanistic foundation for subsequent processes such as social referencing and attachment (Farroni et al., 2007; Soken & Pick, 1999).

The focus on facial and vocal signals is supported by both theoretical and evolutionary considerations. From an evolutionary perspective, faces directly convey intentions and emotions (e.g., gaze direction, facial configurations), whereas voices transmit crucial affective information even at a distance or under visual constraints (e.g., pitch, intensity, rhythm). Together, they form the earlier and more reliable channels of human social communication (DeCasper & Fifer, 1980; Grossmann, 2010). For instance, distinct facial expressions of happiness, anger, or neutrality provide salient visual cues, while variations in vocal prosody communicate parallel emotional meaning in the auditory domain, enabling infants to establish semantic correspondence and cross-modal integration. From a methodological perspective, faces and voices afford high experimental controllability and replicability: researchers can precisely match multimodal stimuli in terms of duration, intensity, and category, thereby minimizing potential confounds. At 6-8 months, a stage marked by limited attentional resources yet heightened social motivation, faces and voices also possess a natural attentional salience, ensuring robust task engagement and increased measurement sensitivity (Walker-Andrews, 1997). Thus, using facial expressions and vocal prosody as core stimuli not only reflects the ecological reality of early socio-emotional processing but also provides an optimal paradigm for testing intermodal matching and integration.

In sum, this study defines affective information as emotion-related signals conveyed through the face–voice channels and focuses on the critical developmental window at 6-8 months. Through behavioral paradigms, we systematically examine infants' intermodal matching and integration of visual–auditory emotional cues. The study design offers a fine-grained developmental account of how infants construct emotional meaning from multimodal social signals, while striking a balance between ecological validity, stimulus control, and attentional engagement. This framework supports interpreting the data and linking the findings to broader theoretical constructs of attachment and social referencing.

1.3.2 Emotion Perception

Emotion perception is typically defined as the process by which individuals detect, discriminate, and interpret others' emotional states through multimodal emotional cues, including facial expressions, vocal prosody, body posture, and gestures (Ekman, 1992a; Scherer, 2005). In the context of infancy, the concept is further specified as infants' sensitivity to emotional cues, making it a core component of early social cognition and interaction (Saarni, 2008).

During the first year of life, emotion perception develops rapidly and systematically. Newborns can already discriminate the affective valence of vocal tones and show a preference for positive facial expressions, such as smiles, indicating an innate sensitivity to emotional signals

(DeCasper & Fifer, 1980; Farroni et al., 2007; Johnson et al., 1991). Between six and eight months, infants display more stable and fine-grained responses to facial and vocal expressions and gradually acquire the ability to integrate emotional information across modalities (Caron et al., 1988; Grossmann et al., 2006; Soken & Pick, 1999). The maturation of this ability enables infants to regulate their own affective states in response to caregivers' feedback, as well as to rely on others' emotional reactions as contextual information to guide behavior in uncertain situations (Pelaez et al., 2012; Sorce et al., 1985). Emotion perception is therefore considered a foundational capacity for socio-emotion development in infancy.

Methodologically, infant emotion perception is typically assessed through indirect behavioral responses. Common paradigms include visual preference, habituation–dishabituation tasks, and intermodal emotion matching, which allow researchers to infer infants' emotional processing capacities in the absence of language (Grossmann, 2010; Ruba & Repacholi, 2020). Within this theoretical and methodological framework, the present study operationally defines emotion perception as infants' ability, before language acquisition, to process and respond to the affective information conveyed by facial expressions and vocal prosody. This definition emphasizes perceptual sensitivity and early multimodal correspondence rather than explicit recognition via linguistic labels, aligning with experimental paradigms feasible in infancy and providing an empirical basis for examining the role of emotion processing in attachment, social referencing, and communicative development.

1.3.3 Intermodal/Multimodal Emotion Perception

Intermodal emotion perception is commonly defined as the ability to establish correspondences across sensory modalities and to match and integrate emotional signals (Walker-Andrews, 1997). The most typical example is visual–auditory matching: when viewing a happy face, infants can associate it with a simultaneously presented happy voice (Flom & Whiteley, 2014; Walker, 1982). Multimodal perception refers more broadly to the simultaneous processing and integration of two or more sensory signals, encompassing not only visual and auditory information but also cues from other modalities such as touch and smell (Bahrick & Lickliter, 2012; Della Longa et al., 2023). In studies of infant emotional development, this term is generally used to describe how infants combine cues from multiple sensory sources to form more complete and robust emotional representations (Marks, 1978).

From a functional perspective, multimodal perception is regarded as a foundational mechanism for socio-emotional development. First, cross-modal integration provides redundancy and stability in emotion recognition, enabling infants to extract others' emotional intentions more reliably in noisy or incomplete environments (Bahrick & Lickliter, 2000). Second, the development of this capacity is closely linked to the increasing reliance on social referencing between 7 and 9 months of age, when infants begin to use caregivers' facial expressions and vocal tones to evaluate the safety of novel situations and guide their own exploratory behavior (Sorce et al., 1985). Thus, multimodal perception represents a critical step in emotion processing.

Developmental evidence further shows that infants exhibit sensitivity to emotional cues within the first months of life. Newborns already display preferences for their mother's voice and face (DeCasper & Fifer, 1980; Farroni et al., 2007). By approximately 3.5 months, infants can detect the emotional congruence between their mother's face and voice, although this sensitivity is less evident when the stimuli are provided by strangers (Walker-Andrews et al., 2011). Around seven months, infants' intermodal matching abilities increase markedly, allowing them to more consistently establish correspondences between visual and auditory emotional signals (Vaillant-Molina & Bahrick, 2012).

In the present study, we adopt the broader term multimodal perception to refer to infants' ability to integrate affective information across facial and vocal modalities, whereas intermodal matching specifically denotes infants' recognition of emotional congruence across sensory modalities within experimental paradigms. This distinction ensures terminological clarity, aligns with methodological conventions in developmental psychology, and provides a consistent conceptual framework for the design and interpretation of the present experiments.

1.4 Theoretical Accounts of Multimodal Emotion Processing

1.4.1 Ecological and Perceptual Learning Accounts

James J. Gibson's (1966) ecological approach to perception emphasizes the principle of direct perception. Within this framework, perception is conceptualized as an active rather than a mediated process: individuals do not reconstruct internal representations of the external world through inferential processing, but instead directly extract stable and meaningful invariants from the environment. Applied to infant research, this perspective suggests that infants are not passive recipients of external stimuli but active explorers who engage with their surroundings and directly obtain emotional intentions conveyed by facial expressions and vocal signals through visual and auditory modalities.

Building on this view, Eleanor J. Gibson (1991) developed the theory of perceptual learning, which highlights perceptual development as a process of progressively differentiating and extracting invariants. Through continuous interaction with the environment, infants gradually increase their sensitivity to salient affective information while reducing reliance on irrelevant or redundant information. Repeated exposure to caregivers' facial expressions, vocal prosody, and bodily actions enables infants to detect multimodal invariant patterns. With accumulated experience, infants become increasingly attuned to emotion-relevant features, such as dynamic facial movements, acoustic intensity, and rhythm, while downplaying irrelevant variability, thereby establishing more reliable correspondences across sensory modalities (Walker-Andrews, 1997).

This invariant-based perspective on differentiation and integration aligns closely with the Intersensory Redundancy Hypothesis (IRH; Bahrick & Lickliter, 2000). The IRH suggests that, in early development, information redundantly specified across multiple sensory channels is more salient and thus more readily detected and learned, facilitating the processing of emotional signals. Together, ecological and perceptual learning theories provide a theoretical basis for understanding how infants extract affective information in naturalistic environments; they also offer a developmental framework that underpins the multimodal integration mechanisms described by the IRH.

1.4.2 Intersensory Redundancy Hypothesis (IRH)

Building on the ecological and perceptual learning traditions, Bahrick and Lickliter (2000) proposed the Intersensory Redundancy Hypothesis, which systematically explains how infants learn and recognize environmental information through simultaneous multimodal input. The central claim of the IRH is that when identical information is redundantly specified across two or more sensory channels, for instance, when facial expressions and vocal prosody simultaneously convey "happiness" or "anger", shared amodal features such as temporal structure, rhythm, and intensity changes are highlighted. These features preferentially capture infants' attention and receive deeper processing (Bahrick & Lickliter, 2000, 2012). Within the context of emotion processing, dynamic facial movements and vocal prosody typically convey congruent affective information. Such multimodal redundancy increases the salience of emotional signals, facilitating intermodal matching and discrimination. By contrast, when the same cues are presented unimodally (i.e., only a face or only a voice), the absence of redundant support makes it more difficult for infants to extract affective features consistently.

Empirical evidence provides direct support for the IRH. Studies show that infants as young as 3–5 months demonstrate enhanced information processing when emotion signals are preserved in multimodal stimulation (Bahrick & Lickliter, 2012; Vaillant-Molina & Bahrick, 2012). In emotional paradigms, when facial expressions and vocal prosody transmit the same affective intent, infants are better able to discriminate emotions and exhibit more consistent preferences than under unimodal conditions, with higher accuracy and stability in intermodal matching (Bahrick & Lickliter, 2012; Flom & Bahrick, 2007). This framework explains why infants can rapidly grasp others' emotional intentions in natural interactions. It also offers clear methodological guidance for laboratory paradigms: by employing audiovisual stimuli, researchers can maximize the attentional and processing advantages afforded by intersensory redundancy, thereby capturing developmental changes in infants' emotion perception with greater sensitivity. Collectively, these theoretical and empirical foundations justify the present study's use of audiovisual paradigms to investigate infants' emotion perception and intermodal matching ability.

1.4.3 Neurocognitive and Developmental Models

One influential neurocognitive account of infant face processing is Johnson's (2011) CONSPEC/CONLERN model. This framework proposes that infants' early preference for faces is supported by two interacting systems. CONSPEC is a subcortical mechanism, present at birth, that orients infants' attention toward face-like patterns via structures such as the superior colliculus and pulvinar (Johnson et al., 1991). The function is to ensure that newborns reliably attend to faces, thereby providing the necessary input for subsequent learning. With increased visual and social experience, the cortical learning system, CONLERN, becomes progressively engaged. This system, involving occipital-temporal regions such as the fusiform gyrus, supports the experience-dependent refinement of face perception and the ability to discriminate between individual faces. Unlike ecological or perceptual learning theories, which stress direct access to environmental information or learning mechanisms, this framework links emotion processing to specific neural structures and functions. Examining brain activation during infants' emotional cognition provides critical insights into the mechanistic question of how emotion perception is realized in the brain: which regions are involved, at what developmental stage, and in what coordinated manner they support infants' detection, discrimination, and integration of social-emotional cues such as faces and voices.

Neuroimaging and electrophysiological evidence indicate that even within the first months of life, the infant brain exhibits sensitivity to emotional stimuli (Nelson, 1987). At the structural level, the superior temporal sulcus (STS) is implicated as a key region for integrating audiovisual information, including emotional expressions (Kreifelts et al., 2007). The amygdala and prefrontal cortex play important roles in evaluating emotional salience and regulating attentional resources (LeDoux, 1996). By approximately seven months, responses in prefrontal and temporal regions to facial expressions and vocal prosody become markedly stronger, reflecting the increasing specialization of neural networks involved in emotion processing (Grossmann, 2010; Grossmann et al., 2007). These neural changes align with behavioral evidence showing improvements in infants' intermodal emotion matching, providing biological support for their emerging ability to flexibly allocate attention across social signals.

Importantly, neurocognitive accounts highlight not only neural maturation but also the contribution of experience-dependent plasticity. Through everyday interactions with caregivers, infants are continuously exposed to multimodal emotional signals. Such experiences shape functional connectivity patterns in the brain, which in turn facilitate the development of emotion perception and integration (Johnson, 2011). Thus, the neurocognitive framework reveals the neural mechanisms underlying infants' emotion perception and emphasizes the role of environmental input and social interaction in the co-development of brain and emotional systems.

Guided by this theoretical perspective, the present study combines behavioral paradigms with fNIRS to examine hemodynamic responses in the temporal–frontal network of 6-8-month-old infants while processing facial expressions and vocal prosody. By incorporating precisely controlled congruent and incongruent conditions, the design allows us to test the selectivity and coordination of temporal–frontal responses to emotional signals and to link individual differences in behavioral performance with neural activation.

1.4.4 Socio-Emotional and Evolutionary Perspectives

Beyond ecological, perceptual learning, and neuro-maturational approaches, social-emotional and evolutionary perspectives further illuminate the adaptive functions and developmental significance of multimodal emotion processing in infancy. From a social-emotional standpoint, the integration of facial expressions and vocal prosody enables infants to more accurately discern caregivers' emotional states during early interactions, thereby supporting the formation and maintenance of attachment bonds. In uncertain contexts, infants rely on others' emotional reactions to regulate their own approach–avoidance and exploratory strategies, a process central to social referencing (Sorce et al., 1985). These exchanges create dynamic emotional dialogues, in which caregivers and infants achieve multimodal affective coupling in rhythm, intensity, and timing, making emotional communication and relational coordination possible (Feldman, 2007b; E. D. Tronick et al., 1977). Sensitivity to multimodal affective information facilitates the emergence of basic emotion regulation and provides a foundation for higher-order socio-cognitive abilities, such as language acquisition and theory of mind, which depend on inferences about interactive intentions and mental states (Bahrick & Lickliter, 2000; Tomasello, 2014; Walker-Andrews, 1997).

From an evolutionary perspective, multimodal emotion processing is understood as an adaptive mechanism. Rapid and accurate recognition of caregivers' emotional signals, for example, fear and anger, elicits proximity seeking, protective responses, and distress reduction, thereby serving survival and safety functions (Bowlby, 1979). At the same time, infants' preferential attention to positive emotions, such as smiles, has been interpreted as an adaptive orientation toward affiliation and cooperation, which strengthens early social bonds and facilitates smooth interaction (Leppänen & Nelson, 2009; Vaish et al., 2008). In natural environments, affective information is typically presented in multimodal synchrony (e.g., angry expressions accompanied by intense, fast-paced prosody). Thus, multimodal integration is associated with immediate adaptive benefits and long-term socio-cognitive outcomes (language, reasoning, social understanding).

In sum, social-emotional and evolutionary perspectives converge in demonstrating that multimodal emotion processing in infancy is not a simple sum of unimodal perceptions, but a functional integration rooted in interactive rhythms and intersensory redundancy. At the micro level, it optimizes attentional allocation and emotion regulation; at the macro level, it shapes the early forms of attachment, social referencing, and communication. This integrative perspective underscores the necessity and ecological validity of examining how coordinated face–voice cues shape socio-emotional development during the first year of life, thereby providing both theoretical and functional insights for the multimodal paradigm adopted in the present study.

1.4.5 Integrative Summary and Relevance to the Present Dissertation

The development of multimodal emotion processing in infancy can be explained within multiple theoretical frameworks, each offering complementary perspectives on how infants gradually perceive and integrate multimodal affective information. Ecological and perceptual learning theories emphasize that emotion processing is deeply embedded in everyday interactions, whereby infants progressively increase their sensitivity to key emotional cues through continuous engagement with their environment (E. J. Gibson, 1991; J. J. Gibson, 1966). Building on this foundation, IRH proposes that when emotional features are redundantly specified across multiple sensory channels, this redundancy preferentially captures infants' attention and facilitates their early recognition and learning of emotions (Bahrick et al., 2004; Bahrick & Lickliter, 2000). In parallel, neurocognitive approaches elucidate the neural mechanisms underlying these capacities, linking infants' behavioral responses to the functional maturation of regions such as the superior temporal sulcus, amygdala, and prefrontal cortex, thereby providing a biological account of emotion perception (Grossmann, 2010; Johnson, 2011; Leppänen & Nelson, 2009). Finally, from social-emotional and evolutionary perspectives, multimodal emotion processing is understood not only as a perceptual ability but also as an adaptive function: it supports the formation of secure attachment in parent–infant interactions, facilitates emotional communication and self-regulation, and provides survival and social advantages at an evolutionary level (Sorce et al., 1985; E. D. Tronick et al., 1977).

According to this integrative theoretical background, the present dissertation systematically investigates the mechanisms of multimodal emotion processing in infancy through three complementary studies. Study 1 (systematic review) evaluates empirical evidence on infants' multimodal emotion perception, highlighting how different methodologies support these theoretical assumptions. Study 2 (behavioral experiment) examines infants' performance in intermodal matching tasks, directly testing predictions derived from the IRH and social-emotional perspectives regarding the congruency sensitivity for multimodal affective information and the sensitivity to positive and negative emotions. Study 3 (fNIRS experiment) explores the neural representations of multimodal emotion processing, extending the neurocognitive framework to preverbal infants. Together, these studies form a logically connected sequence: the review provides a theoretical and empirical overview, the behavioral experiment tests functional mechanisms of emotion matching, and the neuroimaging study identifies the underlying neural implementation. By integrating behavioral and neural evidence, this dissertation aims to provide a systematic and multidimensional framework for understanding how 6-8-month-old infants perceive, match, and integrate affective information.

1.5 Research Paradigms and Methodological Approaches

1.5.1 Behavioral Paradigms

For decades, behavioral paradigms have been the primary methodological tools for studying infants' emotion perception. These paradigms provide reliable and developmentally appropriate means of assessing preverbal infants' sensitivity to affective signals in a non-invasive manner. Different experimental approaches reveal complementary aspects of infants' capacities and limitations in emotion processing, with three paradigms being representative: visual preference, habituation–dishabituation, and intermodal matching.

The visual preference paradigm is a classic method in infant perception and cognition research. Its central assumption is that infants' looking time reflects their preferences and discrimination abilities (Fantz, 1961). In typical experiments, infants are presented with two or more stimuli simultaneously, and researchers infer preference or discrimination by comparing looking times. Longer fixation generally indicates greater interest or recognition of distinctive features (Aslin, 2007). In the domain of emotion research, this paradigm has been widely employed to examine whether infants can differentiate facial expressions or detect face–voice congruency under multimodal conditions, thereby providing direct evidence for the early development of emotion perception (Farroni et al., 2007; Farroni, Massaccesi, et al., 2007; Grossmann, 2010).

The habituation–dishabituation paradigm, in contrast, emphasizes infants' responses to novelty. The logic is that repeated exposure to the same stimulus leads to a gradual decline in looking time ("habituation"); renewed attention to a novel stimulus ("dishabituation") indicates detection of a difference between old and new stimuli (Bornstein, 1985; Fantz, 1964). This method indirectly reveals infants' discrimination abilities, memory representations, and information-processing strategies. In emotion research, habituation paradigms are often used to test whether infants can distinguish emotional expressions or vocal tones. For example, six-month-old infants habituated to one facial expression show increased looking when presented with a different expression, suggesting behavioral discrimination of basic emotions (Caron et al., 1988; Flom & Bahrick, 2007). While highly sensitive, this paradigm can be time-consuming and is susceptible to fatigue and mood fluctuations.

In studies of multimodal emotion processing, the intermodal matching paradigm is widely used. In this task, infants are typically shown two faces depicting different emotions (e.g., happy vs. angry) alongside an auditory cue (e.g., happy or angry vocalization) (Spelke, 1976). If infants establish an intermodal congruency sensitivity, they look longer at the face matching the auditory signal (Walker, 1982). Such preferential looking is taken as evidence of intermodal matching. Together, these behavioral paradigms provide important methods for studying how infants in the first year of life perceive, discriminate, and integrate affective information. They have established the methodological foundation for the present dissertation's use of the intermodal matching paradigm to examine mechanisms of multimodal emotion integration.

1.5.2 Neuroimaging Approaches

In infant research, beyond behavioral paradigms, non-invasive neuroimaging techniques have played an indispensable role in understanding the neural mechanisms underlying emotion perception. The most widely used approaches include electroencephalography/event-related potentials (EEG/ERP), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS), each with distinct strengths and limitations.

EEG/ERP offers millisecond-level temporal resolution, enabling the detection of rapid neural processes associated with infants' responses to emotional stimuli and providing insights into early mechanisms of attention allocation and emotion perception (Nelson & McCleery, 2008). For instance, the roles of the Nc and P400 components in facial and emotional expression processing have been consistently demonstrated (Grossmann et al., 2006; Leppänen, Moulson, et al., 2007). However, EEG is highly sensitive to motion and ocular artifacts, has limited spatial resolution, and is less effective in characterizing neural activity patterns in dynamic, multimodal contexts. By contrast, fMRI provides millimeter-level spatial resolution and whole-brain coverage, allowing precise localization of cortical and subcortical activations (Dehaene-Lambertz et al., 2002). Yet, its noisy scanning environment, stringent postural requirements, and susceptibility to motion artifacts are not well-suited for infant participation during natural sleep or under extreme immobilization, thereby limiting its ecological validity in social and emotional research.

fNIRS has emerged as a balanced and feasible approach for studying emotion processing in infancy (Gervain et al., 2023). By measuring the absorption and scattering of near-infrared light through the scalp and skull, fNIRS estimates changes in oxyhemoglobin and deoxyhemoglobin concentrations in cortical regions (approximately 1–2 cm), indirectly reflecting localized neural activity. Its spatial and temporal resolution lies between EEG and fMRI, while offering unique advantages: it is quiet, non-invasive, free of electromagnetic interference, and compatible with natural infant seating, making it particularly well-suited for presenting dynamic socio-emotional stimuli such as faces and voices (Nakato et al., 2009; Urakawa et al., 2015; Zhang et al., 2017). Moreover, compared to fMRI, fNIRS is more tolerant of minor head movements, significantly improving data yield and ecological suitability (Lloyd-Fox et al., 2010). As a result, fNIRS has become a core technique in the study of infant social-emotional development. In this dissertation, we employ fNIRS as the primary neuroimaging tool to investigate multimodal emotion processing in 6-8-month-old infants. This choice reflects both methodological feasibility and theoretical rationale.

1.5.3 Contextual and Environmental Measures

Although laboratory-based behavioral and neuroimaging paradigms have provided critical evidence for understanding infant emotion processing, a comprehensive account of the factors shaping emotion perception must also incorporate infants' everyday social experiences. The family environment constitutes the primary context for emotional learning and socialization, where infants are continuously exposed to caregivers' emotional signals and accumulate experience with emotional expression through daily interactions. Empirical studies indicate that the frequency, intensity, and valence of emotional expressions within the family directly influence infants' strategies and sensitivity in recognizing, processing, and interpreting emotional information (Denham, 1993; Eisenberg et al., 1998). Thus, quantifying the emotional environment of the family is essential for understanding individual differences in socio-emotional development.

A widely used tool for this purpose is the Family Expressiveness Questionnaire (FEQ) (Halberstadt, 1986). Grounded in family socialization theory, the FEQ conceptualizes the family as the earliest and most influential environment for infants' emotional learning. It relies on parental reports, asking caregivers to rate the frequency of emotional expressions displayed in everyday family situations, thereby capturing the overall pattern of emotional expressiveness. The instrument comprises four subscales: positive dominant, positive submissive, negative dominant, and negative submissive, designed to distinguish different tendencies of emotional expression across positive and negative contexts. Previous research has shown that FEQ scores reliably reflect variations in the family's emotional environment and are significantly associated with children's emotion understanding, regulation, and broader social development (Halberstadt & Eaton, 2002; Ogren et al., 2018).

In the present dissertation, the FEQ is employed to examine whether variability in family expressiveness predicts infants' performance in intermodal emotion-matching tasks. By integrating behavioral indices of emotion perception with contextual measures of the family environment, this study aims to explore how individual developmental factors (e.g., age and attentional strategies) interact with environmental inputs to shape early multimodal emotion processing.

1.5.4 Integrative Considerations

Overall, behavioral paradigms, neuroimaging methods, and contextual measures provide complementary and mutually validating approaches for investigating emotion perception in infancy. Behavioral tasks (e.g., visual preference and intermodal matching) directly capture infants' ability to recognize and discriminate affective information, offering sensitive behavioral indices of perceptual capacities. Neuroimaging methods, particularly fNIRS, record hemodynamic changes associated with multimodal emotion integration, thereby supplying neural explanations and mechanistic evidence for behavioral responses. Meanwhile, contextual measures such as the FEQ reveal the social-environmental origins of individual variability, clarifying how the emotional environment of the family shapes infants' performance in experimental tasks.

This multilevel framework allows research on socio-emotional development to be interpreted across behavioral, neural, and contextual dimensions. Behavioral paradigms demonstrate whether infants can perceive and match emotional cues; neuroimaging identifies how these abilities are instantiated within specific developing brain networks; and contextual measures explain the environmental roots of individual differences.

In this dissertation, this methodological integration constitutes the guiding logic of the overall research design. Chapter 2 presents a systematic review synthesizing cross-paradigm evidence on multimodal emotion perception, establishing the theoretical and methodological foundation for the empirical studies. Chapter 3 employs behavioral experiments to examine intermodal matching abilities in 6–8-month-old infants while incorporating FEQ data to assess the influence of the family expressiveness. Chapter 4 applies fNIRS to investigate the neural representations of multimodal emotion processing, thereby linking behavioral performance with underlying neural mechanisms. Through this multilevel design, the dissertation seeks to provide converging evidence to address the central question: How do infants between 6 and 8 months perceive and integrate emotional information across sensory modalities?

1.6 Dissertation Roadmap

This dissertation comprises three interrelated studies designed to systematically examine how infants aged 6–8 months perceive and integrate affective information across sensory modalities. By combining a systematic review, behavioral experiments, and neuroimaging methods, it aims to establish a multidimensional understanding of early emotion-processing mechanisms at both theoretical and empirical levels.

Chapter 2 (Study 1: Systematic Review) provides a comprehensive synthesis of empirical research on multimodal affective information processing during the first year of life, with a particular focus on evidence from multimodal paradigms. The review summarizes major findings regarding methodological approaches, developmental trajectories, and specific emotional effects, and identifies existing gaps and controversies, thereby offering theoretical and methodological insights for subsequent empirical studies.

Chapter 3 (Study 2: Behavioral Experiment) investigates infants' responses in an intermodal matching task, examining their preferential looking to face–voice congruency and assessing whether systematic matching tendencies vary as a function of age or emotional types. This experiment provides direct behavioral evidence for how infants integrate emotional signals during this critical developmental window.

Chapter 4 (Study 3: fNIRS Experiment) extends the inquiry to the neural level by employing fNIRS to examine hemodynamic responses in the temporal and prefrontal cortices during the processing of multimodal emotional stimuli. By integrating neural activation patterns with behavioral performance, this study aims to provide the neural basis of early multimodal emotion integration.

Chapter 5 (General Discussion) synthesizes findings across the three studies and situates them within broader theoretical frameworks of socio-emotional development and multimodal affective information processing. It highlights the contributions of this dissertation to the understanding of multimodal affective information processing in preverbal infants, outlines theoretical and methodological directions for future research, and discusses the potential implications of family and environmental factors for early emotion processing.

In sum, through a cross-level design, this dissertation provides a unified framework for understanding how infants in the first year of life perceive, discriminate, and integrate affective information across sensory modalities, underscoring the significance of 6–8 months as a critical window for the development of emotion perception.

CHAPTER II.

SEEING AND HEARING AFFECTIVE INFORMATION: A REVIEW OF INFANT MULTIMODAL PERCEPTION IN THE FIRST YEAR OF LIFE

Infants' ability to perceive and integrate affective information across sensory modalities plays a fundamental role in their early social and emotional development. During the first year of life, preverbal infants depend heavily on nonverbal cues, particularly facial expressions and vocal prosody, to interpret others' emotional states and to guide their behavior in social contexts. This review synthesizes empirical research on infants' intermodal perception of affective information, with a focus on their capacity to integrate and match emotional cues across visual and auditory modalities. This work outlined the early development of infants' processing of affective information, with particular attention to the developmental trajectory of intermodal emotion perception and the asymmetries in the emergence of emotion-specific sensitivity.

This study aims to understand how infants perceive and integrate affective information across sensory modalities, which is essential for clarifying the developmental origins of socio-emotional communication. Although previous research has shown that infants are sensitive to facial and vocal expressions from early in life, less is known about how they match and integrate these signals during the first year of life. The present work aims to address these questions by systematically reviewing existing evidence and conducting empirical studies to underscore the crucial role of intermodal emotion perception in infancy and offer valuable insights into the early development of socioemotional cognition.

2.1 Introduction

Affective information refers to signals conveyed by an individual's emotional state, encompassing cues related to its expression, perception, and processing. Such information can be communicated through multiple channels. Among its various components, expressive cues such as facial expressions, vocal tone, bodily movements, and tactile signals (Darwin & Darwin, 1872; Ekman et al., 1997) have long been the core focus of research. This emphasis arises not only from their observability and measurability but also from their cross-cultural universality and evolutionary significance. Expressive cues play a crucial role in enabling the communication of both emotions and intentions during social interactions.

The ability to perceive and interpret affective information forms the foundation of infants' early social development. From birth, infants develop in a socially rich interactive environment in which engagement with others is not only essential for immediate survival but also for shaping the long-term trajectories of socioemotional and behavioral development. Understanding others is therefore considered a fundamental task through which infants begin to comprehend and adapt to the external world. Previous research has demonstrated that newborns exhibit a preference for face-like stimuli (Goren et al., 1975; Johnson et al., 1991) and can discriminate and show preferences for certain facial expressions over others shortly after birth (Farroni et al., 2007). Such early sensitivity to emotional expressions is regarded as an adaptive feature of human development (Darwin & Darwin, 1872; Nelson, 1987), enabling infants to engage effectively with caregivers and other social partners from the earliest stages of life. Even before language acquisition, infants are able to infer caregivers' intentions, needs, and emotional states through nonverbal cues such as facial expressions, vocal prosody, and body movements, and adjust their own social behaviors accordingly. Taken together, these findings indicate that infants are highly attentive to affective information from a very early age, particularly to nonverbal emotional signals such as facial and vocal expressions. Moreover, they are able to extract meaningful information from these cues, which facilitates the discrimination and recognition of emotional expressions and fosters adaptive, dynamic relationships with the social world. A natural question arises when we consider the developmental trajectory of how these emotional cues are understood.

In the visual domain, newborns can discriminate certain facial expressions within days after birth as reflected by a visual preference for a happy compared to a fearful facial expression (Farroni et al., 2007), providing an early basis for the development of the social brain. By 3–5 months of age, infants are able to differentiate between various emotional expressions, such as happiness, anger, fear, and sadness (Caron et al., 1988; Flom & Bahrick, 2007), a capacity thought to be influenced by the valence and intensity of expressions, White et al. (2019) found that 5-month-old infants could distinguish sadness from disgust and happiness from surprise, but struggled to differentiate more similar negative emotions such as anger and disgust. Positive emotions, particularly happiness, tend to be recognized earlier and elicit stronger preferences, likely due to their high social salience and frequent occurrence in caregiver–infant interactions. By 7 months, infants continue to show preferences for positive emotions but also begin to show sensitivity to negative emotions (Grossmann et al., 2007). For instance, Kotsoni et al. (2001) demonstrated that infants could discriminate happy versus fearful expressions and even showed a preference for fearful faces. Similarly, event-related potential (ERP) studies revealed enhanced early attentional Nc components in response to fearful expressions (Leppänen, Moulson, et al., 2007), suggesting the emergence of threat sensitivity. By 8 months, infants show heightened Nc responses to fearful (relative to happy) expressions (Missana, Rajhans, et al., 2014), indicating increasing specialization in processing threat-related emotions. At 12 months, infants continue to distinguish negative emotions. For instance, Lee et al. (2021) found that infants could categorically differentiate anger from happiness but still struggled to distinguish sadness from happiness, highlighting the functional importance of emotional salience and not just a perceptual processing of the facial expression. As development progresses, infants also rely on caregivers' facial expressions to regulate their own behavior in ambiguous contexts, a phenomenon known as social referencing (Klinnert et al., 1986).

In the auditory domain, affective information also plays an important role. Research has shown that newborns are already sensitive to changes in intonational contours (Zhang et al., 2014). An NIRS study demonstrated that, compared to neutral prosody, emotional prosody elicited stronger brain responses (Zhang et al., 2017). Similar findings have been reported in infants aged 3–7 months, who exhibited significant differential activation in the anterior temporal cortex (Blasi et al., 2011). Five-month-old infants were able to discriminate among different emotional prosodies, such as happiness, sad, anger, and neutrality (Walker-Andrews & Grolnick, 1983).

An important consideration is that, unlike static facial expressions, vocal expressions unfold dynamically over time, offering richer emotional information and exerting a powerful influence on infants' attention and behavior (Grossmann, 2010). Notably, infants show heightened sensitivity to positive prosody early in development. This sensitivity may arise from its greater social salience as well as from distinctive acoustic features, such as higher pitch and a wider pitch range, that are characteristic of infant-directed speech (Fernald & Kuhl, 1987; Melanie et al., 2017; Singh et al., 2002). At the same time, infants' ability to allocate greater attentional resources to negative emotional information appears to emerge earlier in the vocal domain than in the facial domain (Grossmann, 2010).

However, in real-world contexts, affective information is rarely conveyed through a single sensory modality. Instead, affective information is typically embedded within rich, multiple sensory inputs, such as the combination of facial expression and vocal prosody. Intermodal perception is therefore essential for constructing a coherent representation of others' emotional states. This ability enables infants to detect redundancies and correspondences across modalities, thereby enhancing the salience of affective information and facilitating more efficient processing. Evidence suggests that intermodal perception gradually emerges during the first year of life, although findings remain inconsistent. Some studies have shown that infants begin to demonstrate intermodal recognition around 6–7 months of age (Flom & Whiteley, 2014; Phillips et al., 1990). For instance, in a cross-modal transfer experiment, Soken and Pick (1999) found that 7-month-old infants looked longer at affectively congruent facial expressions when paired with a happy voice, suggesting that they were sensitive to violations of emotional congruency. Neurophysiological studies provide converging evidence showing that incongruent stimuli elicited larger negative ERP deflections, whereas congruent pairings enhanced positive ERP components, indicating early neural sensitivity to cross-modal emotional congruency (Grossmann et al., 2006). Together, these findings suggest that by 6–7 months of age, infants are able to detect and differentiate the congruency of multimodal emotional signals. Nevertheless, other studies have reported that, when paired with neutral faces, 9-month-old infants still fail to show cross-modal matching among happy, sad, and angry expressions (Ogren et al., 2018).

Nevertheless, the timing of the earliest emergence of infants' emotion discrimination remains a matter of debate (Grossmann, 2010). Classical theories propose that certain "basic" emotions, such as happiness, sadness, fear, anger, surprise, and disgust, are universally recognizable across cultures (Ekman, 1994). In contrast, constructionist theories deny the existence of universal emotional signals (J. A. Russell, 1994) and emphasize the high variability of emotional experience and expression. Research generally suggests that if preverbal infants possess conceptual emotion categories, these are initially organized along a valence dimension (J. A. Russell, 1980), and given the variability in terminology across the literature, the present review follows the conventional distinction between 'positive' and 'negative' emotions. This early evaluative mechanism is hypothesized to operate within the first few months of life and serves adaptive functions, such as guiding approach or avoidance behaviors in response to others' emotional cues. As development progresses, these broad categories gradually refine into discrete emotion concepts (Widen, 2013). Many studies report that positive emotions, particularly happiness, are recognized earlier and more accurately than negative emotions (Farroni et al., 2007; Grossmann et al., 2007; Hiatt et al., 1979). By 7 months, infants demonstrate sustained attention and neural sensitivity to fearful faces (Peltola et al., 2013). For instance, Kotsoni et al. (2001) and Leppänen et al. (2007) found that infants exhibited an attentional preference for fearful expressions, accompanied by an enhanced P400 ERP response, reflecting an early threat-detection mechanism. In contrast, recognition of anger seems to emerge later: Grossmann et al. (2007) and Missana et al. (2014) reported that 8-month-old and 12-month-old infants exhibited stronger ERP responses to angry faces, suggesting a slower developmental trajectory for the processing of socially directed negative emotions. Nevertheless, the developmental sequence of infants' recognition of negative emotions remains contested.

Given the foundational role of intermodal perception in infants' socioemotional development, this review synthesizes current empirical findings on how preverbal infants perceive and integrate affective information across visual and auditory modalities. Previous reviews have rarely provided a systematic synthesis of studies examining infants' perception of multimodal emotional expressions or their sensitivity to cross-modal emotional congruency. Moreover, infants' responses to multimodal emotional cues may shaped by both emotion type and modality salience. For instance, happy expressions are often more preferred than negative ones, and some studies suggest that auditory cues may temporarily dominate under dynamic conditions (Vaillant-Molina et al., 2013). However, the distinct contributions of sensory modalities and emotion types to infants' behavioral and neural responses remain insufficiently understood. Only a limited number of studies have investigated how these factors interact to shape affective processing during the first year of life. These gaps underscore the need for a comprehensive synthesis of current findings on how preverbal infants perceive, integrate, and respond to affective information across sensory modalities.

In synthesis, this review aims to examine the roles of sensory modalities and emotion types in multimodal emotion integration. Specifically, it focuses on studies employing multimodal emotion tasks, particularly intermodal matching paradigms, to assess infants' sensitivity to the congruency between facial and vocal expressions during the first year of life. It also summarizes the developmental trajectory of emotion specificity in infancy. The review begins with an overview of research on infants' emotion processing and a description of the methodological approach adopted. It then synthesizes findings from behavioral and neuroimaging studies, emphasizing the development of intermodal emotion specificity and changes in sensitivity to emotional congruency. Finally, it discusses the limitations of existing research and proposes directions for future studies. By integrating current evidence, this review seeks to clarify the developmental trajectory of infants' intermodal emotion perception and to consider its implications for the early emergence of emotion understanding and social cognition.

2.2 Method

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for conducting and reporting systematic reviews (Page et al., 2021).

2.2.1 Literature Search

2.2.1.1 Search Strategy

A systematic literature search was conducted across three electronic databases—EBSCO (including APA PsycArticles, APA PsycInfo, and Psychology and behavirol science), Elsevier's Scopus®, and PubMed®—to identify empirical studies examining infants' perception of affective information across visual and auditory modalities during the first year of life. The search was completed on July 31, 2025, the search strategy combined four conceptual categories using Boolean operators: (1) infant age (e.g., infant, newborn, preverbal); (2) intermodal or cross-modal processing (e.g., intermodal, multimodal, modality preference, integration, matching); (3) sensory modalities (visual, auditory); (4) affective content (e.g., emotion, affective information, emotional expression, facial expression, vocal expression).

The following Boolean search structure was applied across title, abstract, and keyword fields, where applicable: (infant* OR "young infant*" OR newborn* OR neonate* OR baby* OR preverbal*) AND (intermodal* OR "cross-modal" OR multimodal* OR "modality preference" OR match* OR integration) AND (visual* OR auditory*) AND (emotion* OR "emotional expression" OR "affective information" OR "facial expression*" OR "vocal expression*"). The search was restricted to English-language publications, with no limitation on publication date. Where possible (e.g., in PubMed), filters were applied to include studies focusing on infancy.

2.2.1.2 Study Selection

The initial database search yielded a total of 395 potentially relevant records (PubMed = 69; Scopus = 221; EBSCO = 105). After applying language, age, and discipline filters, 230 relevant records remained (PubMed = 40; Scopus = 138; EBSCO = 52). All references were imported into Zotero reference management software, where duplicates were identified and removed using the built-in duplicate detection function. This process resulted in 177 unique records for further screening. To ensure transparency and traceability, a cleaned set of unique records was generated, and source-specific subfolders (e.g., PubMed, Scopus, PsycInfo) were reconstructed based on database-origin tags (e.g., source: PubMed). Each unique record was assigned a primary database tag, ensuring that it appeared only once in the review dataset. These tags were used to create three non-overlapping final folders, each representing a distinct subset of the included literature. All de-duplicated records were compiled into a temporary screening collection, and titles and abstracts were assessed against the predefined inclusion and exclusion criteria. During screening, each record was tagged with a status label (e.g., status: included, status: excluded, status: unclear) and, if excluded, with the corresponding exclusion reason (e.g., reason: older than 12 months, reason: non-affect). This tagging system facilitated transparent monitoring and systematic classification during the screening phase, enabled efficient tracking, filtering, and categorization of studies, and ensured consistent documentation of inclusion decisions and data extraction progress.

Study screening was conducted in two stages: (1) Title and abstract screening, and (2) Full-text review. Title and abstract screening were independently performed by two researchers (CJ and TL), followed by full-text evaluation of all records deemed potentially relevant. A standardized coding form was employed to document specific reasons for exclusion at each stage of the screening process. Any discrepancies between coders were resolved through discussion until consensus was reached.

During the screening phase, n = 150 records were excluded for the following reasons: reviews or meta-analyses (n = 15), methodological studies (n = 9), atypical development (n = 34), animal studies (n = 6), non-affect studies (n = 13), non-multimodal studies (n = 21), non-audiovisual modalities (n = 13), and studies involving infants older than 12 months (n = 39). Each excluded study was assigned a single primary exclusion reason (e.g., reason: older than 12 months, reason: non-affect) to allow clear frequency calculation and reporting. When multiple criteria were applied, the most direct or critical reason was selected for tagging, following PRISMA guidelines. Additional information was documented in notes for transparency.

The entire screening process was documented following PRISMA guidelines, and a flow diagram (Figure 2.1) illustrating the number of included and excluded records at each stage was prepared accordingly. As a result, n = 27 studies encompassing n = 47 experiments were deemed eligible for inclusion in the present systematic review.

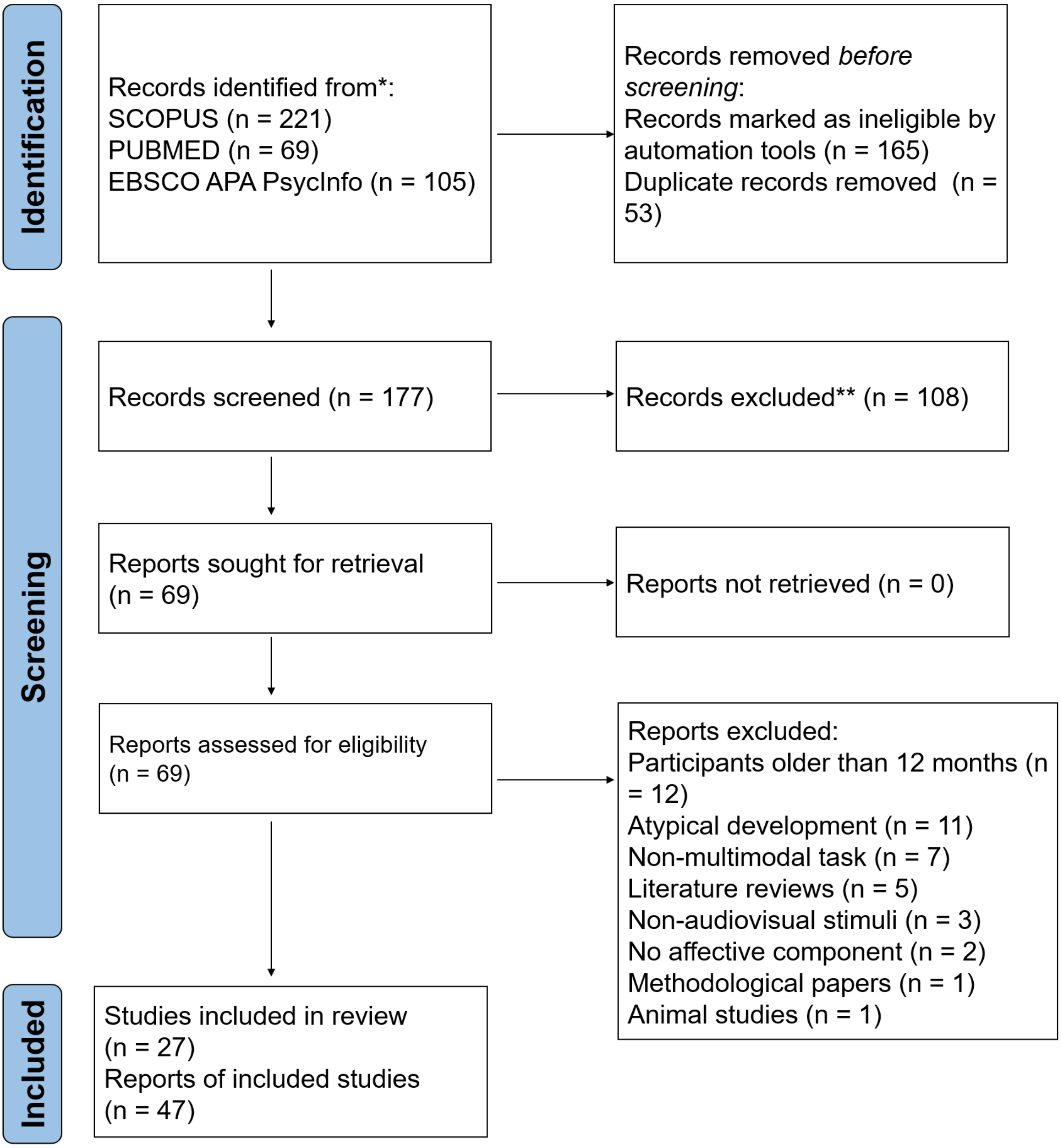


Figure 2.1 Flow diagram (following PRISMA guidelines).

2.2.2 Eligibility Criteria

To be eligible for inclusion, studies were required to be written in English (as filtered during the database search), report original empirical data (e.g., behavioral or neurophysiological measures), and be published in peer-reviewed scientific journals. Eligible studies specifically if they (a) involved human infants under 12 months of age with typical development, (b) assessed perception, matching, or integration of affective information across visual and auditory modalities.

Studies were excluded if they were reviews, meta-analyses, theoretical or methodological papers, commentaries, or conference abstracts lacking empirical data. Further exclusion criteria encompassed studies involving populations with atypical development (e.g., infants with neurodevelopmental disorders or hearing impairment), preterm or prenatal samples, and participants older than 12 months of age. We also excluded studies that did not involve human participants, lacked an intermodal or affective component, or focused on sensory modalities other than facial and vocal cues.

Further details on the inclusion and exclusion criteria applied during both the title and abstract screening and the full-text review phases are presented in Table 2.1, which outlines the decision framework used to ensure consistency and relevance across all selected studies.

Table 2.1 Eligibility criteria for study inclusion and exclusion

	Inclusion criteria	**Exclusion criteria**
Language	Written in English	Other than English
Source and type of publication	Original empirical research published in Scientific Journals	Reviews, meta-analyses, theoretical or methodological papers, commentaries, or conference abstracts without empirical data
Population	Human infants (< 12 months) with typical development	Infants with atypical development (e.g., neurodevelopmental disorders), preterm or prenatal samples, animal models, participants > 12 months
Sensory modality	Assessment of affective information perception, matching, or integration in both visual and auditory modalities	Studies without an intermodal or affective component, or focusing on sensory modalities other than facial and vocal cues
Other	—	Non-human studies

Note. Inclusion and exclusion criteria were applied at both the title–abstract screening and full-text review stages.

2.2.3 Data Extraction

For each paper included, we systematically collected comprehensive information from each eligible study to facilitate structured synthesis and comparison. Publication details included the full list of authors, year of publication, and digital object identifier (DOI) to ensure accurate referencing. Sample characteristics were recorded in detail, including the total sample size, the number of male and female participants (when reported), age range, mean age, and standard deviation of age to capture the developmental context and sample variability.

In addition to publication and sample information, we extracted key methodological features relevant to the review's aims. These included the experimental design (e.g., between-subjects, within-subjects, longitudinal), the specific paradigm employed (e.g., intermodal matching, visual preference, habituation), and the number and nature of experimental conditions tested. We also documented the types of emotions investigated (e.g., happiness, anger, fear), the modality or combination of modalities presented, the type of stimuli used (e.g., static facial images, dynamic videos, naturalistic recordings), and the primary outcome measures adopted (e.g., looking time, neural responses such as Electroencephalogram).

All extracted information was recorded by the first author and checked by the third author. Some of the included papers reported results from multiple age groups; however, only one of these groups met our inclusion criteria. Therefore, we extracted and recorded information only for the experiments involving that specific age group.

In this section, we paid particular attention to the type of visual stimuli used (dynamic facial expressions vs. static facial photographs). Early visual preference studies often employed static images, which may lack the ecological validity of dynamic expressions that more closely approximate real-life social interactions, and may therefore elicit greater sensitivity in infants due to their consistency with everyday experiences, ultimately enhancing the ecological validity of experimental paradigms. Previous research has shown that dynamic stimuli are more effective in capturing and maintaining infants' attention (Grossmann, 2010), as subtle motion changes help sustain their visual engagement with the screen, leading to more stable and reliable data collection. In contrast, static stimuli may reduce the sensitivity of experiments in measuring infants' cross-modal processing abilities, potentially underestimating their true capacities to perceive and integrate affective information across sensory modalities. Notably, in the present review, dynamic stimuli were more frequently used in multimodal studies, potentially influencing the observed patterns of cross-modal emotion perception.

We also recorded the measurement methods employed in each study. Behavioral measures such as looking time are widely used and allow for inferences about perceptual discrimination based on visual attention patterns. However, they are indirect and may be influenced by attentional fluctuations, fatigue, and non-specific preferences. Neurophysiological measures provide more direct evidence of neural processing but differ in temporal and spatial resolution, and can be more sensitive to movement artifacts, which is a particular concern in infant populations.

This structured data extraction approach ensured consistent coding across studies and supported a nuanced interpretation of the findings. By jointly considering experimental paradigms, emotion types, and measurement methods, it also facilitated subsequent analyses aimed at clarifying how these factors contribute to our understanding of affective information perception and integration in infancy.

2.3 Results

2.3.1 Overview of Included Studies

As outlined in the Method section, we systematically compiled detailed experimental information for all studies included in this review, as shown in Table 2.2, presenting a summary of the key characteristics and main findings of all studies included. For each study, we extracted the variables, including the authors and publication year, sample size, and the age of participants, the type of task and stimuli employed, the measurement approach, and the categories of emotions investigated. Particular attention was given to listing the main findings for each age group, allowing for a clear comparison of developmental trends across studies. Where studies contained multiple experiments, each experiment is listed separately to reflect its distinct design and outcome. The final dataset comprises 27 studies, encompassing a total of 47 experiments. Studies are ordered chronologically by the first author's surname to facilitate cross-study comparison. Table 2.2 provides a structured overview that serves as the foundation for the thematic synthesis presented below, in which findings are organized according to specific patterns of infant emotion perception and integration, as well as their developmental progression within the first year of life.

Table 2.2 Overview of the studies included in the systematic review

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
1	Caron et al. (1988)	1	4 months N = 48	Dynamic videos + synchronized voice	Habituation–dishabituation	Looking time	Happy vs Sad	4-month-olds only showed discrimination in the Sad → Happy direction.
		2	5 months N = 48	Dynamic videos + synchronized voice	Habituation–dishabituation	Looking time	Happy vs Sad	5-month-olds successfully discriminated between happy and sad expressions.
		3	5 months N = 24	Dynamic videos + synchronized voice	Habituation–dishabituation	Looking time	Happy vs Angry	5-month-olds did not show evidence of discrimination.
		4	7 months N = 24	Dynamic videos + synchronized voice	Habituation–dishabituation	Looking time	Happy vs Angry	7-month-olds were able to discriminate between happy and angry expressions.
		5	7 months N = 24	Dynamic videos, Voice vs. No-Voice conditions	Habituation–dishabituation	Looking time	Angry vs Happy	7-month-olds in the voice condition successfully discriminated between angry and happy expressions. (Auditory information is crucial for distinguishing expressions that are visually similar.)
2	Flom & Bahrick (2007)	6	3 months N = 18	Dynamic video + synchronized audio	Habituation–recovery	Looking time	Happy, Sad, Angry	3-month-olds showed no discrimination of affect.
			4 months N = 18	Dynamic video + synchronized audio	Habituation–recovery	Looking time	Happy, Sad, Angry	4-month-olds showed clear discrimination of affect; no preference.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
2	Flom & Bahrick (2007)	6	5 months N = 18	Dynamic video + synchronized audio	Habituation–recovery	Looking time	Happy, Sad, Angry	5-month-olds showed clear discrimination of affect; no preference.
			7 months N = 18	Dynamic video + synchronized audio	Habituation–recovery	Looking time	Happy, Sad, Angry	7-month-olds showed clear discrimination of affect; no preference.
		7	4 months N = 18	Static face (neutral expression) + audio	Habituation–recovery	Looking time	Happy, Sad, Angry	4-month-olds showed no discrimination of affect.
			5 months N = 18	Static face (neutral expression) + audio	Habituation–recovery	Looking time	Happy, Sad, Angry	5-month-olds showed clear discrimination of affect; no preference.
			7 months N = 18	Static face (neutral expression) + audio	Habituation–recovery	Looking time	Happy, Sad, Angry	7-month-olds showed clear discrimination of affect; no preference.
		8	4 months N = 18	Dynamic video +delayed audio (2s)	Habituation–recovery	Looking time	Happy, Sad, Angry	4-month-olds showed no discrimination of affect.
			5 months N = 18	Dynamic video +delayed audio (3s)	Habituation–recovery	Looking time	Happy, Sad, Angry	5-month-olds showed clear discrimination of affect; no preference.
		9	4 months N = 18	Alternating dynamic audio and video	Habituation–recovery	Looking time	Happy, Sad, Angry	4-month-olds showed no discrimination of affect.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
3	Flom & Whiteley (2014)	10	7 months N = 32	Static facial expressions + dynamic audio	Intermodal matching	Proportion of Total Looking Time (PTLT)	Happy vs Sad	7-month-olds showed successful intermodal matching, mainly for happy affect, happy facial preference.
			12 months N = 32	Static facial expressions + dynamic audio	Intermodal matching	Proportion of Total Looking Time (PTLT)	Happy vs Sad	12-month-olds showed significant matching, matching both happy and sad affect reliably; happy facial preference.
4	Flom et al. (2018)	11	3 months N = 20	Dynamic audiovisual	Infant-controlled habituation paradigm (50% criterion)	Looking time	Positive (happy-like) vs. Negative (sad-like)	3-month-olds showed no discrimination between their peers' positive and negative affective expressions.
			5 months N = 20	Dynamic audiovisual	Infant-controlled habituation paradigm (50% criterion)	Looking time	Positive (happy-like) vs. Negative (sad-like)	5-month-olds discriminated their peers' positive and negative affective expressions.
		12	3 months N = 20	Dynamic audiovisual	Infant-controlled habituation paradigm (70% criterion)	Looking time	Positive (happy-like) vs. Negative (sad-like)	3-month-olds discriminated between positive and negative affective expressions (with the more stringent 70% habituation criterion).
5	Grossmann et al. (2006)	13	7 months N = 18	Static facial expressions + dynamic audio	Passive crossmodal perception task	ERP	Happy vs Angry	7-month-olds showed significant N400-like negativity for incongruent pairs; absent for congruent pairs.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
6	Heck et al. (2016)	14	10 months N = 22	Multimodal (face + voice); dynamic vs. static visual stimuli	Passive crossmodal perception task	Eye-tracking: fixation duration	Fear, Sad, Happy	10-month-olds showed increased attention (longer fixation, slower disengagement) to dynamic bimodal stimuli.
7	Isomura & Nakano (2016)	15	5 months N = 15	Dynamic audiovisual stimuli	Passive crossmodal perception task	EMG	Laughter (positive), Crying (negative), and Neutral expressions	5-month-olds showed significant facial mimicry response in multimodal conditions, but not in unimodal conditions.
8	Kahana-Kalman & Walker-Andrews (2001)	16	3.5 months N = 72	Dynamic audiovisual stimuli	Intermodal preference	Looking time	Happy vs Sad	3.5-month-olds showed congruency sensitivity when displayed by their mothers, even asynchrony; happy expressions elicited more positive affect.
9	Montague & Walker-Andrews (2002)	17	3.5 months N = 32	Dynamic videos	Intermodal preference	Proportion of Total Looking Time (PTLT)	Happy, Sad, Angry	3.5-month-olds showed intermodal matching preferences only for their mothers' expressions; no preference for fathers or unfamiliar adults.
10	Nawrot (2003)	18	5-9 months N = 26	Dynamic video + emotional music	Intermodal matching	Looking time	Happy vs Sad	5-9-month-olds showed preferences for happy over sad displays regardless of congruency.
11	Ogren & Johnson (2020)	19	9 months N = 40	Dynamic video + asynchronous audio	Intermodal matching	Eye tracking (Looking time)	Angry, Happy, Sad, Neutral	9-month-olds showed no congruency sensitivity and no preference.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
12	Ogren et al. (2018)	20	9 months N = 38	Dynamic video + audio	Intermodal matching with a silent baseline	Proportion of Total Looking Time (PTLT)-Eye tracking	Happy, Sad, Angry (each paired with Neutral)	9-month-olds showed no significant group-level emotion matching; happy emotion matching positively related to total family expressiveness.
13	Otte et al. (2015)	21	9 months N = 82	Static facial expressions + non-verbal audio	Multimodal emotion processing task	EEG/ERP	Happy vs Fearful	9-month-olds exhibited heightened sensitivity to fearful voices, as reflected by increased amplitudes in ERP components.
14	Palama et al. (2018)	22	6 months N = 24	Static facial expressions + non-verbal audio	Cross-modal transfer: auditory familiarization + visual preference	Eye tracking (Looking time)	Happy, Angry, Neutral	6-month-olds can transfer emotional information from auditory to visual modalities.
15	Phillips et al. (1990)	23	7 months N = 77	Static facial expressions + auditory tones	Categorization metaphorical mapping	Looking time habituation-discrimination	Joy, Anger, Surprise, Sadness	7-month-olds looked longer at facial expressions that metaphorically matched auditory cues.
16	Soken & Pick (1992)	24	7 months N = 50	Dynamic audiovisual	Intermodal matching	Looking time	Happy vs Angry	7-month-olds looked longer at affectively concordant displays; happy faces preference.
			7 months N = 50	Dynamic audiovisual	Intermodal matching	Looking time	Happy vs Angry	7-month-olds looked longer at affectively concordant displays; happy faces preference.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
		25	7 months N = 20	Dynamic audiovisual (5s asynchrony)	Looking preference	Looking time	Happy vs Interested	7-month-olds looked longer at affectively congruent displays for both expressions; no preference between expressions.
		26	7 months N = 20	Dynamic audiovisual (6s asynchrony)	Looking preference	Looking time	Happy vs Sad	7-month-olds showed no significant preference for congruent displays; strong visual preference for happy faces regardless of audio.
17	Soken & Pick (1999)	27	7 months N = 20	Dynamic audiovisual (7s asynchrony)	Looking preference	Looking time	Happy vs Angry	7-month-olds showed significantly longer looking at affectively congruent displays; visual preference for happy expressions.
		28	7 months N = 20	Dynamic audiovisual (8s asynchrony)	Looking preference	Looking time	Interested vs Sad	7-month-olds showed no significant preference for congruent displays, visual preference for interested expressions.
		29	7 months N = 20	Dynamic audiovisual (9s asynchrony)	Looking preference	Looking time	Interested vs Angry	7-month-olds showed significant preference for congruent displays; visual preference for interested expressions.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
		30	7 months N = 20	Dynamic audiovisual (10s asynchrony)	Looking preference	Looking time	Sad vs Angry	7-month-olds showed significant preference for congruent displays; visual preference for angry expressions.
18	Tsui et al. (2016)	31	8-10 months N = 15	Dynamic facial expressions + affective voices	Rule learning (AAB rule) via habituation test	Looking time	Happy, Sad, Surprised, Disgust	8-10-month-olds successfully learned the AAB rule only with congruent bimodal affective stimuli.
		32	8-10 months N = 15	Dynamic facial expressions + affective voices	Rule learning (AAB rule) via habituation test	Looking time	Happy, Sad, Surprised, Disgust	8-10-month-olds succeeded in learning the AAB rule with congruent bimodal stimuli.
		33	8-10 months N = 17	Dynamic facial expressions + affective voices	Rule learning (AAB rule) via habituation test	Looking time	Happy, Sad, Surprised, Disgust	8-10-month-olds failed to learn the AAB rule with incongruent bimodal affective stimuli.
19	Vaillant-Molina & Bahrick (2012)	34	5.5 months N= 32	Dynamic video + 3D toys	Habituation + Preference test	Looking time	Happy vs Fearful	5.5-month-olds in bimodal (audiovisual) condition detected the change in affect-object pairing and preferred the toy previously paired with happy affect.
20	Vaillant-Molina et al. (2013)	35	3.5 months N = 20	Dynamic facial expressions + audio (infants)	Intermodal matching	PTLT (proportion total looking time), PFL (proportion of first looks)	Joy vs Anger	3.5-month-olds showed no evidence of matching facial and vocal affective displays.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
			5 months N = 20	Dynamic facial expressions + audio (infants)	Intermodal matching	PTLT (proportion total looking time), PFL (proportion of first looks)	Joy vs Anger	5-month-olds matched the vocal and facial affect of other infants based on general valence.
21	Vogel et al. (2012)	36	5 months N = 15	Static facial expressions + audio	Passive viewing; sound-face congruency ERP task	EEG	Happy vs Sad	5-month-olds showed neural sensitivity to audio-visual emotional congruency (larger Nc for congruent pairs).
			9 months N = 17	Static facial expressions + audio	Passive viewing; sound-face congruency ERP task	EEG	Happy vs Sad	9-month-olds showed sensitivity to emotional congruency only for own-race faces.
22	Walker (1982)	37	7 months N = 16	Dynamic facial expressions + audio	Intermodal matching	Looking time	Happy vs Neutral	7-month-olds looked longer at the sound-specified film vs. silent; overall preference for happy face in soundtrack condition.
		38	5 months N = 16	Dynamic facial expressions + audio	Intermodal matching	Looking time	Happy vs sad	7-month-olds looked longer at sound-specified films; strong preference for happy faces
			7 months N = 16	Dynamic facial expressions + audio	Intermodal matching	Looking time	Happy vs sad	7-month-olds showed a preference for sound-specified films, particularly the happy film.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
		39	7 months N = 16	Dynamic facial expressions + audio	Intermodal matching (asynchrony)	Looking time	Happy vs Neutral	7-month-olds preferred sound-matched films despite asynchrony; happy-face preference.
		40	7 months N = 16	Dynamic facial expressions + audio	Intermodal matching (normal vs. inverted)	Looking time	Happy vs Angry	7-month-olds in the normal orientation group preferred sound-specified film and happy faces.
23	Walker-Andrews (1986)	41	7 months N = 16	Dynamic facial expressions + audio	Intermodal matching	Looking time	Happy vs Angry	7-month-olds looked significantly longer at the congruent facial expression.
			5 months N = 16	Dynamic facial expressions + audio	Intermodal matching	Looking time	Happy vs Angry	5-month-olds did not show significant preference for sound-congruent facial expressions.
24	Walker-Andrews et al. (2011)	42	3.5 months N = 40	Dynamic facial + vocal videos of parents	Habituation / dishabituation	Looking time	Happy vs Sad	3.5-month-olds showed increased looking to changes in emotion or person when habituated to their parents' dynamic facial and vocal expressions.
			3.5 months N = 40	Dynamic facial + vocal videos of strangers	Habituation / dishabituation	Looking time	Happy vs Sad	3.5-month-olds did not show differential looking to changes in emotion or person when habituated to unfamiliar individuals.

Study No.	Authors, Year	Exp. No.	Sample/Age	Stimulus Type	Task Type	Measure	Emotion Type	Key Findings
25	Walker-Andrews & Lennon (1991)	43	5 months N = 88	Static facial images + emotional vocal sounds	Habituation-dishabituation	Looking time	Happy, Angry, Sad	5-month-olds discriminated against vocal emotions when accompanied by facial expression, but not with non-social visual stimuli.
26	Xiao & Emberson (2019)	44	9 months N = 18	Static facial images + emotional vocal sounds	Sequential cueing + visual preference	Eye-tracking (initial latency, looking preference)	Happy vs Angry	9-month-olds showed longer looking time to the emotional face when it was congruent with the preceding vocal sound.
26	Xiao & Emberson (2019)	45	9 months N = 18	Static facial images + emotional vocal sounds	Simultaneous audio-visual presentation	Eye-tracking (initial latency, looking preference)	Happy vs Angry	9-month-olds showed no congruency effect.
		46	6.5 months N = 18	Static facial images + emotional vocal sounds	Sequential cueing + visual preference	Eye-tracking (initial latency, looking preference)	Happy vs Angry	6.5-month-olds showed no congruency effect.
27	Yong & Ruffman (2016)	47	7 months N = 24	Static facial images + emotional vocal sounds	Intermodal matching	Looking time	Happy, Angry, Sad	7-month-olds showed no reliable face–voice matching overall.

2.3.2 Thematic Synthesis of Findings

2.3.2.1 Affective Information Discrimination Across Studies

Across the reviewed literature, findings regarding infants' ability to discriminate emotional expressions across visual and auditory modalities varied. Twelve studies involving infants aged 3–7 months reported clear evidence of successful discrimination, such as between happy and sad, happy and angry, or among multiple emotions, typically using habituation–dishabituation or visual preference paradigms (Caron et al., 1988; Vaillant-Molina & Bahrick, 2012). These studies often employed dynamic facial expressions paired with synchronized vocal cues, enabling even younger infants, such as 4-month-olds, to detect common affective information across modalities. Infants at 3 months of age exhibited discrimination of affective expressions only when provided with a more stringent 70% habituation criterion, as opposed to the conventional 50% criterion (Flom et al., 2018). Results were generally consistent across modalities, although some variation was observed depending on infants' age and the specific emotion pairs tested. For instance, Caron (1988) found that 4-month-olds showed discrimination only in the sad–happy direction, not the reverse. In the same study, a separate experiment indicated that 7-month-olds could discriminate happy from angry expressions, only under multimodal conditions (visual and auditory), not when facial expressions were presented without accompanying vocal cues, suggesting that auditory information plays a crucial role in distinguishing visually similar expressions. These findings align closely with the intersensory redundancy hypothesis (Bahrick & Lickliter, 2000; Bahrick et al., 2004), which suggests that infants are more likely to discriminate emotions when dynamic facial expressions are paired with congruent vocal cues than in unimodal conditions. These findings are also in line with existing evidence that positive emotions tend to be discriminated earlier and more accurately than negative emotions in infancy (Grossmann, 2010; Ruba & Repacholi, 2020).

In contrast, six studies involving infants aged 3-5 months reported no significant evidence of discrimination between emotional expressions (Caron et al., 1988; Flom & Bahrick, 2007). This absence of discrimination was observed across both unimodal and multimodal presentations (when unimodal conditions were tested as controls) (Caron et al., 1988; Flom & Bahrick, 2007). Notably, the absence of significant effects was consistent regardless of the type of emotional contrast tested (e.g., happy vs. sad, happy vs. angry vs. sad). In several cases, infants demonstrated comparable looking times or response patterns to both emotional stimuli

(Flom & Bahrick, 2007), suggesting that perceptual sensitivity to differences in affective information across modalities may not yet be fully established at this age. Furthermore, no systematic differences were found between studies employing habituation–dishabituation and visual preference paradigms, indicating that the null findings were not attributable to methodological factors. Together, these results suggest that within the first half of the first year, the ability to reliably discriminate discrete emotional expressions, whether in unimodal or multimodal contexts, may remain in an early stage of development.

2.3.2.2 Affective Information Congruency Sensitivity Across Modalities

Findings on infants' sensitivity to the congruency of affective information across visual and auditory modalities are mixed (A. R. Heck et al., 2016; Isomura & Nakano, 2016; Soken & Pick, 1992). Evidence from 30 experiments indicates that infants are sensitive to cross-modal emotional congruency. Particularly, among infants aged 5–7 months, 21 experiments, including 3 using neurophysiological measures, reported positive evidence for the congruency sensitivity (Flom & Whiteley, 2014; Kahana-Kalman & Walker-Andrews, 2001; Otte et al., 2015). The most robust and consistent effects were observed for positive emotions, especially happiness, as suggested by 14 experiments that reported a preference for happy expressions in congruent conditions. In these cases, infants tended to look longer at faces that matched the accompanying vocal cues. Such patterns are interpreted as evidence for the emergence of intermodal integration abilities. Methodologically, most studies employed either passive cross-modal perception paradigms or intermodal matching tasks to assess whether infants preferentially attended to visual stimuli that conveyed the same affective information as the auditory cues.

Seven months of age is widely considered a critical period in the emergence of emotion and cognition, and it has become a focal point of research in this field. Among the reviewed studies, 15 experiments focusing specifically on 7-month-old infants consistently reported evidence that, by this age, infants already exhibit sensitivity to emotional congruency across visual and auditory modalities (Flom & Whiteley, 2014; Grossmann et al., 2006). This sensitivity was observed regardless of whether the visual stimuli were dynamic or static. Most of these studies employed intermodal matching or visual preference paradigms, whereas a smaller number used alternative measures, such as facial mimicry or rule-learning tasks, to assess infants' preference for and sensitivity to congruent emotional expressions. Notably, several studies further reported that even when audiovisual cues were asynchronous, infants nevertheless tended to prefer the emotionally congruent component.

Converging neurophysiological evidence further supports these behavioral findings. Event-related potential studies have revealed distinct neural responses to congruent versus incongruent multimodal emotional expressions in 7-month-old infants, providing developmental evidence for the emergence of cross-modal emotion processing (Grossmann et al., 2006). Similar neural patterns have also been reported in 5 and 9 months of age (Otte et al., 2015). Furthermore, recognition of congruent affective information across modalities has been observed as early as 3.5 months, with all such early studies employing dynamic visual stimuli, which possibly enhances infants' sensitivity to congruency (Kahana-Kalman & Walker-Andrews, 2001; Montague & Walker-Andrews, 2002; Walker-Andrews et al., 2011). These studies found that at this early age, congruency sensitivity was evident only for familiar individuals (e.g., mothers) but not for strangers. Overall, sensitivity to cross-modal emotional congruency seems to develop gradually: it can be detected as early as 3.5 months under favorable conditions and becomes more stable by 6–7 months.

In contrast, a smaller subset of 11 experiments reported no significant evidence of cross-modal sensitivity to emotional congruency (Ogren et al., 2018; Ogren & Johnson, 2020; Soken & Pick, 1999). These null findings were observed across samples ranging from 3 to 9 months of age. A closer examination reveals that studies employing asynchronous presentation of visual and auditory information consistently failed to detect congruency sensitivity (Ogren & Johnson, 2020; Soken & Pick, 1999), even when dynamic visual stimuli were used. Similarly, two studies employing static visual stimuli reported null results despite presenting audiovisual cues synchronously (Xiao & Emberson, 2019; Yong & Ruffman, 2016). In some cases, these outcomes may reflect the younger age of participants, which could have limited their ability to detect cross-modal congruency (Vaillant-Molina et al., 2013; Walker-Andrews et al., 2011; Xiao & Emberson, 2019). Taken together, these findings suggest that the presence and detectability of congruency sensitivity in infancy are highly dependent on both stimulus presentation parameters and developmental stage.

2.3.2.3 Patterns of Emotional Preference Across Studies

Regarding infants' emotional preferences, a synthesis of the reviewed literature identified 16 experiments reporting a preference toward positive emotions, particularly happy facial expressions. Most of these studies focused on 7-month-old infants, although evidence of a happiness preference was also observed as early as 3.5 months and up to 12 months of age. Across this body of work, the most common emotion pairings contrasted happiness versus anger or sadness. The preference for happiness was observed in studies using both static and dynamic visual stimuli. Overall, the type of emotion pairing appeared to exert a stronger influence on eliciting a detectable preference than the specific visual stimulus format. Interestingly, this preference for happy expressions frequently co-occurred with sensitivity to cross-modal emotional congruency, suggesting a possible link between the two phenomena.

Only two studies reported a preference for negative emotions in infancy. One ERP study found that 9-month-old infants exhibited heightened sensitivity to fearful voices, as reflected in increased amplitudes of specific ERP components (Otte et al., 2015). The other study, using a looking-preference task with asynchronous dynamic audiovisual stimuli, reported that 7-month-old infants preferred angry facial expressions, possibly due to the high arousal associated with anger (Soken & Pick, 1999). Previous evidence suggests that sensitivity to negative emotions typically emerges around 7 months of age but becomes more frequently observed in the second half of the first year (Otte et al., 2015).

Eight experiments reported no evidence of an emotional preference, with such findings concentrated between 4 and 7 months of age. In 6 of these experiments, infants showed clear discrimination between emotional expressions but no consistent preference for one over another. This may indicate that although infants are able to distinguish emotions, they may allocate comparable levels of attention or looking time to different expressions (Ruba & Repacholi, 2020). In addition, when the contrast between expressions was within valence, such as between happy and interested, no preference was observed (Soken & Pick, 1999). This pattern suggests that the emotional valence of the stimuli may not have been sufficiently strong to elicit a detectable bias under the given task conditions.

2.3.3 Developmental Trajectory of Multimodal Emotion Processing

Evidence from the included studies indicates a gradual emergence and refinement of emotion perception and integration abilities during the first year of life. Research on newborns to approximately 3-month-old infants has yielded mixed findings. Four out of 9 experiments reported that infants in this age range were unable to discriminate between different emotions and showed no sensitivity to cross-modal emotional congruency (Flom & Bahrick, 2007; Vaish et al., 2008; Walker-Andrews et al., 2011). This may be partly due to the ongoing development of the visual system, which limits the processing of visual emotional cues, as well as the fact that this period precedes the critical phase for the development of emotion cognition, resulting in relatively inconsistent performance in emotion-related tasks. Nevertheless, three experiments found that by 3.5 months of age, infants could already exhibit sensitivity to cross-modal emotional congruency, but only when the stimuli were presented by familiar individuals (e.g., mothers) and involved dynamic audiovisual information (Kahana-Kalman & Walker-Andrews, 2001; Montague & Walker-Andrews, 2002; Walker-Andrews et al., 2011). One of these studies further reported that happy expressions elicited greater positive affect (Kahana-Kalman & Walker-Andrews, 2001). Overall, during this early stage, emotional cues in various task contexts may not yet be salient enough to elicit distinct differential responses, and the perceptual systems supporting multimodal integration are still undergoing maturation.

By 4–6 months of age, six experiments reported that infants could successfully discriminate between distinct emotional expressions, most consistently when happiness was contrasted with negative emotions such as anger or sadness. During this period, initial evidence of sensitivity to cross-modal emotional congruency also appeared, suggesting the early development of multimodal affective integration. These findings indicate that by mid-infancy, the perceptual systems supporting the processing of visual and auditory emotional cues may have advanced sufficiently to allow for more coordinated processing across modalities. However, results in this age range remained mixed: several studies still reported either no reliable discrimination between emotions or no evidence of congruency sensitivity. Such inconsistencies may reflect differences in task demands, the salience of emotional cues, or the developmental variability typical of this transitional stage.

By 7–8 months of age, research findings consistently indicate that infants reach a relatively stable and advanced stage of emotion processing. At this point, they not only reliably discriminate between different emotional expressions but also demonstrate clear sensitivity to cross-modal emotional congruency, with effects most pronounced for positive emotions, particularly happiness. Multiple studies have shown that infants at this age tend to allocate more looking time to facial expressions that match the accompanying vocal cues, suggesting significant progress in the integration of visual and auditory affective information. During this period, the preference for positive emotions is frequently reported, whereas preferences for negative emotions remain rare and typically occur only in specific task contexts or under particular stimulus conditions (Otte et al., 2015; Soken & Pick, 1999). Evidence further suggests that the advantage for positive emotions may be related to the greater social appeal and higher salience of happy expressions, whereas the preference for negative emotions may require more threatening cues or higher arousal levels to be elicited. Overall, the 7–8-month period appears to represent a critical consolidation phase in the development of multimodal emotion processing, reflecting both the further maturation of perceptual systems for cross-modal emotion integration and infants' increasing efficiency in using emotional cues during social interaction.

By 9–12 months of age, the available empirical evidence remains relatively limited, and findings are less consistently supportive of further consolidation in emotion processing abilities than might be expected. Within this age range, studies are roughly divided between those reporting sensitivity to cross-modal emotional congruency and those finding no such evidence. Notably, one study indicated that 9-month-old infants exhibited heightened sensitivity to negative emotions, aligning with theoretical accounts of an emerging preference for threat-related affect in late infancy (Otte et al., 2015). The mixed pattern of results may reflect both methodological variations, such as differences in task design, stimulus complexity, emotion categories tested, and individual variability in the developmental trajectory of multimodal emotion processing during this later stage of the first year.

Taken together, findings across the first year of life suggest a gradual yet heterogeneous progression in multimodal emotion processing. In the earliest months (0–3 months), evidence for both emotion discrimination and cross-modal congruency sensitivity remains limited, likely due to the immaturity of perceptual systems and the low salience of emotional cues, though sensitivity may emerge under optimal conditions involving dynamic stimuli from familiar individuals. By 4–6 months, some infants begin to demonstrate reliable discrimination, particularly for positive emotions, and early signs of congruency sensitivity, although results remain inconsistent across studies. A notable shift occurs around 7–8 months, when robust and consistent evidence is found for both discrimination and congruency sensitivity, alongside a pronounced preference for positive emotions, suggesting a likely consolidation phase in multimodal emotion processing. By 9–12 months, findings become more mixed than anticipated: while some infants demonstrate broadened discrimination abilities and stable congruency sensitivity, others show no such effects; importantly, sensitivity to negative emotions begins to appear more frequently, aligning with theoretical accounts of an emerging preference for threat-related affect in late infancy. Overall, these patterns underscore that the development of multimodal emotion processing is shaped by a complex interplay of perceptual maturation, task demands, stimulus salience, and the socio-emotional relevance of the presented cues. To provide an overview of empirical findings on infants' emotion cognition during the first year of life, Figure 2.2 presents a quantitative overview of the distribution of experimental findings. Distinguishing between evidence and non-evidence for emotion discrimination, intermodal congruency sensitivity, and emotion preference. These plots summarize the relative proportion of studies reporting each outcome across different age groups. Complementing this distributional summary, Figure 2.3 presents a visual map of research findings by developmental stage, illustrating not only the types of results reported but also the methodological characteristics (behavioral vs. neuroimaging approaches) and stimulus formats (static vs. dynamic visual stimuli) employed in each age window. These schematic highlights age-related clustering of research evidence and the diversity of paradigms used in infant emotion research.

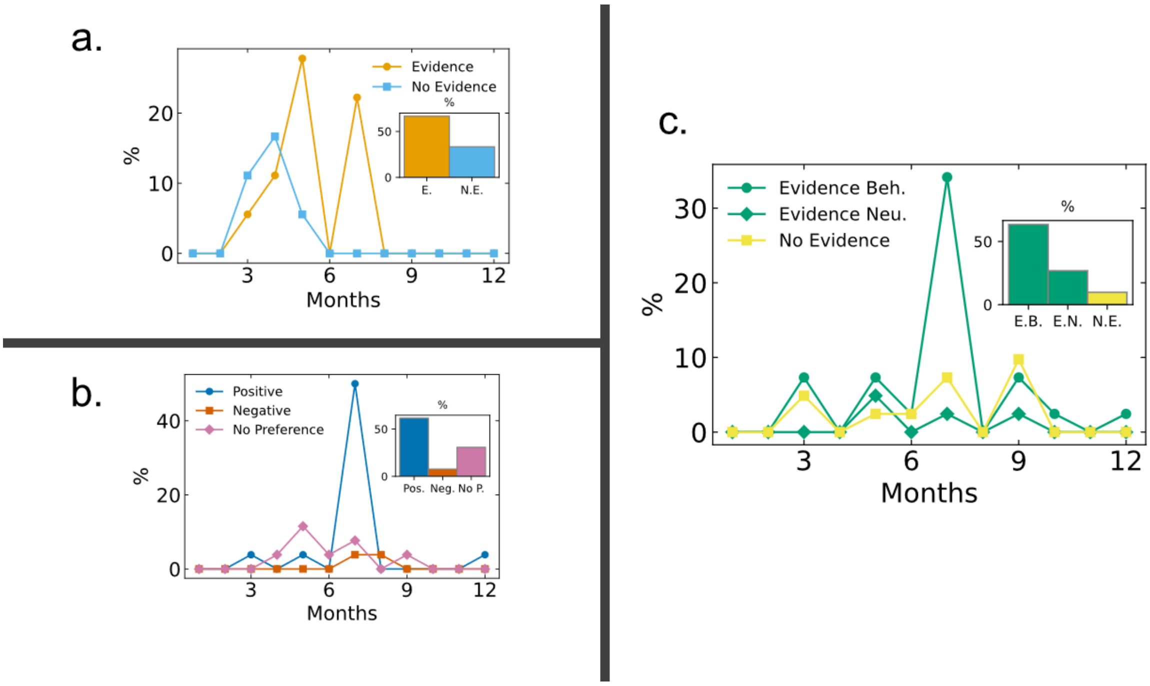


Figure 2.2 Distribution of experimental findings on infant emotion cognition during the first year of life.

Note. (a) Emotion discrimination: evidence versus no evidence of infants' ability to discriminate emotional expressions. (b) Emotion preference: studies reporting positive, negative, or no preference. (c) Intermodal congruency sensitivity: behavioral and neural evidence versus no evidence of sensitivity to congruent audiovisual emotional cues. Insets show overall proportions across studies.

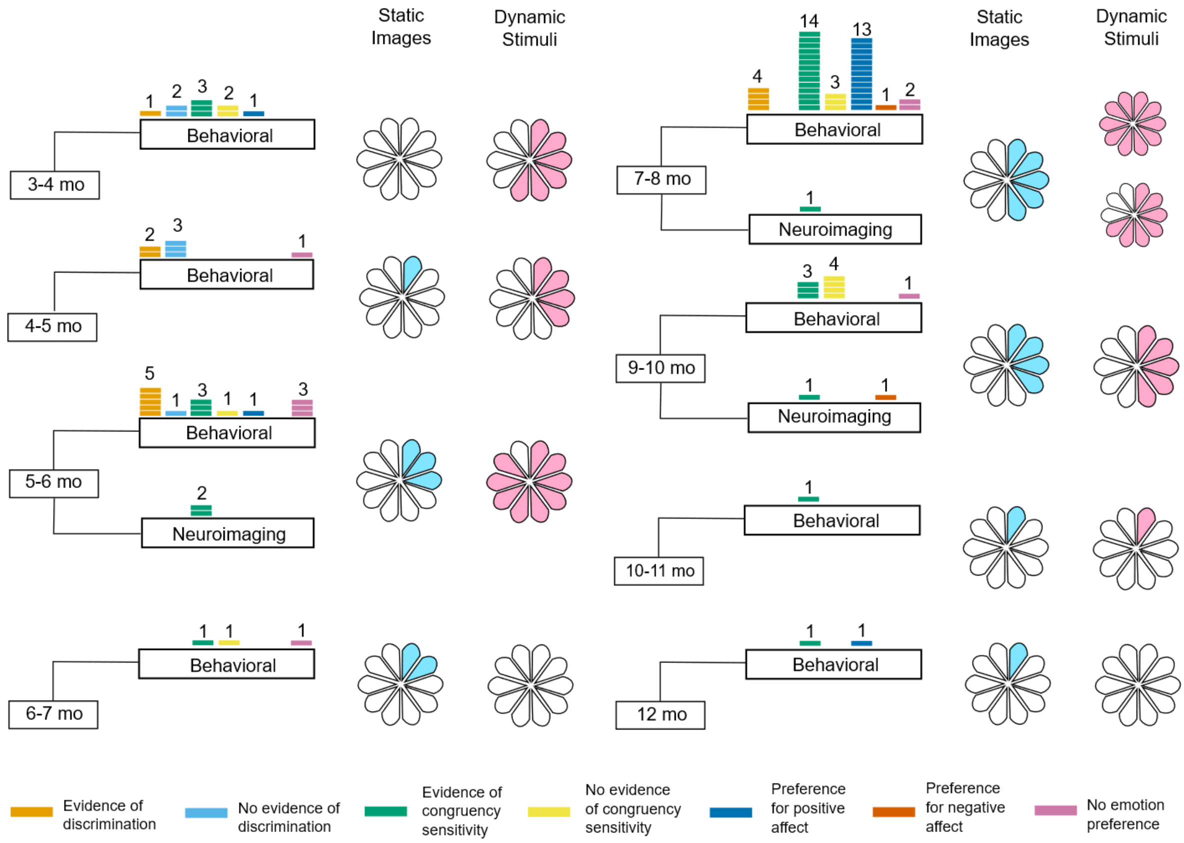


Figure 2.3 Distribution of experimental findings and stimulus types across infant age groups. *Note*. Visual summary of the distribution of experimental findings across infant age groups (3–12 months). For each developmental stage, the figure illustrates the type of evidence reported (discrimination, congruency sensitivity, emotion preference), the research method (behavioral vs. neuroimaging), and the type of visual stimuli used (static vs. dynamic). The schematic highlights both age-specific clusters of findings and the methodological diversity of the literature.

2.4 Discussion

2.4.1 Summary of Main Findings

This systematic review synthesizes evidence from 27 studies examining how infants perceive and integrate visual and auditory affective information during the first year of life. The findings collectively suggest gradual development across several key domains: the ability to discriminate emotional expressions in multimodal contexts, sensitivity to cross-modal emotional congruency, preferences for specific emotional valences, and the broader developmental trajectory of these abilities throughout the first year.

First, regarding infants' ability to discriminate affective information across visual and auditory modalities, the finding revealed a developmental progression during the first half of the first year of life. Evidence from studies of 3–7-month-olds demonstrates emotion discrimination, with particular sensitivity to positive emotions such as happiness. This ability emerges more often when dynamic facial expressions are paired with synchronized vocal cues, suggesting redundant information across modalities facilitates infants' faster recognition and processing of emotional cues. By contrast, studies of younger infants aged 3–5 months more frequently report null findings, with no differences between unimodal and multimodal conditions or across task paradigms. These findings suggest that, while the capacity for cross-modal emotion discrimination may begin to emerge under optimal conditions at around 4 months of age, it is not yet reliably established in the early months of life and becomes more consistently observable closer to mid-infancy.

Second, research on cross-modal emotional congruency perception suggests that this ability begins to stabilize around 5–7 months of age. The effect is more pronounced for positive emotions and is most consistently observed in dynamic, synchronous multimodal contexts. Sensitivity may also be reported under certain conditions in younger infants, for instance, at 3.5 months when stimuli are presented dynamically and provided by familiar caregivers such as the mother. Nevertheless, a smaller subset of studies with infants aged 3–9 months have reported no evidence of congruency sensitivity, particularly when audiovisual cues are asynchronous or when static images were used as visual stimuli. These mixed findings suggest that the emergence of cross-modal congruency sensitivity is shaped not only by infants' developmental age but also by the properties of emotional stimuli and the specific task context.

Third, in terms of preferences for specific emotional valences, a consistent bias toward positive emotions is evident, yet it is important to consider that infants showed considerable (individual) variability. This preference appears as early as 3.5 months and becomes more consistent by mid-infancy, and is observable with both dynamic and static visual stimuli. In contrast, preferences for negative emotions are relatively rare and tend to emerge in the second half of the first year. For instance, under high-arousal conditions, infants allocate more attention to angry or fearful faces, aligning with theoretical accounts of increasing sensitivity to threat-related emotions later in infancy (Otte et al., 2015). In some cases, infants may discriminate between emotions without exhibiting clear preferences, particularly when the contrasts are subtle or involve emotions within the same valence category (e.g., happiness vs. interest). Overall, these findings suggest that emotion preferences are not fixed traits, but dynamic outcomes shaped by developmental stage, emotion type, and situational context.

Considering the current evidence, we can draw the developmental trajectory of multimodal emotion processing during the first year of life, which appears to follow a gradual yet uneven progression. In the earliest months (0–3 months), evidence for both emotion discrimination and cross-modal congruency sensitivity is limited, with occasional signs of sensitivity only under highly salient and socially relevant conditions, such as dynamic audiovisual stimuli from familiar individuals. By 4–6 months, some infants begin to demonstrate more reliable discrimination of positive emotions and early signs of congruency sensitivity, though findings remain inconsistent across studies. By 7–8 months, both abilities are more robust and consistently observed, suggesting a potential consolidation phase in the development of multimodal emotion processing. At the end of the first year (9–12 months), findings are more mixed than expected: while some infants display broadened discrimination abilities and stable congruency sensitivity, others show no such effects. During this period, heightened sensitivity to negative emotions, such as anger or fear, also becomes more frequent, aligning with theories that propose the emergence of attentional biases toward threat-related cues in late infancy. Overall, the current evidence supports a tentative developmental sequence of emotion processing: happiness→anger/sadness→fear. This sequence likely reflects the relative salience and ecological frequency of different emotions in infancy. This trajectory reveals both age-related gains and persistent variability, highlighting the interplay between maturational processes and stimulus characteristics in shaping infants' multimodal emotion processing.

In summary, the developmental patterns observed across these domains offer important insights into the mechanisms underlying multimodal emotion processing in infancy. The progressive emergence and consolidation of emotion discrimination, congruency sensitivity, and emotional preferences, together with shifts in developmental trajectories, align with theoretical accounts of emotion perception, such as the IRH. At the same time, the variability observed across age groups and experimental conditions underscores the importance of contextual factors, including stimulus salience (dynamic vs. static), familiarity (caregiver vs. stranger), and temporal synchrony (synchronous vs. asynchronous audiovisual cues). These factors interact to shape developmental outcomes. Taken together, this body of converging evidence provides a foundation for future research examining how current models of cross-sensory processing and socioemotional development account for the patterns identified in this review.

2.4.2 Implications for Theories of Early Multimodal Emotion Processing

The developmental patterns identified in this review offer important insights for theories seeking to explain how multimodal emotion processing emerges and develops during infancy. The integrated evidence, in terms of emotion discrimination, cross-modal congruency sensitivity, emotional preference, and their developmental timeline, underscores that these capacities do not develop in isolation but are shaped by shared underlying mechanisms.

To explain how infants perceive affective information across sensory modalities, several theoretical frameworks have been proposed. J. J. Gibson's (1966) ecological approach to perception argued that different forms of sensory stimulation are not obstacles to unified event perception but rather its essential foundation. In his view, the senses function as an integrated "perceptual system," with cross-modal, non-modality-specific information representing the core of amodal perception. Building on this perspective, Bahrick (2000; 2004) and colleagues systematized the concept within developmental psychology and introduced the IRH. According to this hypothesis, intersensory redundancy in early infancy promotes the detection of cross-modal information in multimodal events. When information is presented redundantly across multiple sensory channels (e.g., a smiling face accompanied by a joyful voice), infants are more likely to process emotional cues. This early sensitivity to such redundancy cross-modal information provides an efficient pathway for perceptual processing, enhancing attention, directing infants toward meaningful unified events, and supporting their early learning of socially relevant cues, including emotional expressions (Bahrick et al., 2004; Flom & Bahrick, 2007). Evidence from intermodal matching paradigms further indicates that even young infants can detect congruency between facial and vocal expressions, demonstrating a preliminary ability to integrate emotional signals into coherent perceptual representations (Walker-

Andrews et al., 2011). Therefore, investigating the development of cross-modal perception of affective information during the first year is crucial for understanding how infants extract meaningful social information from complex multimodal environments. Consistent with the predictions of the IRH (Bahrick et al., 2004; Bahrick & Lickliter, 2000), the reviewed evidence suggests that infants are more likely to show emotion discrimination and congruency sensitivity when exposed to dynamic audiovisual stimuli presented synchronously.

Second, the findings of this review indicate that infants can discriminate among different emotional expressions, with a notable advantage for positive emotions such as happiness. This advantage can be understood through complementary evolutionary, social, and perceptual perspectives. From an evolutionary perspective, positive emotional expressions (e.g., smiling) convey signals of safety, social bonding, and caregiving availability, thereby carrying high adaptive value and being more easily recognized and processed early in life (Farroni et al., 2007; Leppänen & Hietanen, 2004). The process of socialization further reinforces this bias: caregivers more frequently express positive emotions in daily interactions, thereby providing infants with extensive exposure to happy facial and vocal cues. Such input likely strengthens the advantage for positive emotions (Eisenberg et al., 1998). Perceptual features of happy expressions also contribute to this effect, as their distinctive and prototypical characteristics (e.g., upturned mouth corners, eye constriction, and salient mouth movements) increase visual contrast and facilitate detection (Jessen & Grossmann, 2014; M. White, 1995). Together, these factors help explain why positive emotions, particularly happiness, are consistently discriminated earlier and more reliably than negative emotions in infancy. Overall, these influences contribute to a broader "positivity bias" in early emotion perception. This bias not only enables young infants to recognize and prefer happy expressions but also supports more stable integration of positive emotions across different sensory modalities compared to negative emotions. Moreover, this bias appears to be amplified when emotional cues are provided by familiar individuals, suggesting that familiarity and emotional salience may jointly shape the developmental trajectory of infants' emotion perception (Walker-Andrews et al., 2011). Consistent with the evidence reviewed here, younger infants often succeed in discriminative tasks only under conditions involving dynamic stimuli from familiar individuals. This finding aligns with theoretical models emphasizing the role of social relevance in guiding attentional regulation and shaping early perceptual processes.

Finally, the findings of this review suggest that infants' sensitivity to negative emotions, particularly threat-related expressions such as anger and fear, begins to emerge in mid-infancy and becomes increasingly pronounced during the second half of the first year of life (Jessen & Grossmann, 2016; Nelson & De Haan, 1996; Vaish et al., 2008). This developmental shift can be interpreted from both evolutionary and socioemotional perspectives. From an evolutionary perspective, this development reflects an adaptive adjustment in attentional priorities during the second half of the first year of life. As infants gradually begin to explore their environment more independently (Campos et al., 2000), heightened vigilance to potential threats holds significant survival value (Leppänen & Nelson, 2009; Peltola et al., 2009). At the same time, this sensitivity may also be partly driven by infants' curiosity toward novel stimuli (Malatesta & Haviland, 1982). From a socialization perspective, infants are initially exposed primarily to positive emotions conveyed by caregivers, while negative emotions occur relatively infrequently in early social experiences (Peltola et al., 2008; Rose et al., 1982). Consequently, when infants encounter these less familiar emotions later in development, they tend to allocate more attention in order to better interpret their social significance (De Haan et al., 2013). Overall, these explanations together suggest that the development of multimodal emotion processing is not driven by a single mechanism but emerges from the dynamic interplay between perceptual maturation, changes in attentional allocation, and the social significance of emotional cues.

2.4.3 Implications for the Developmental Trajectory of Multimodal Emotion Processing

The developmental trajectory outlined in this review has important implications for understanding the emergence of multimodal affective information perception and integration across the first year of life. Although some unimodal visual studies have reported that newborns are able to discriminate between happy and fearful or sad expressions (Field et al., 1983; Farroni et al., 2007), these findings remain controversial. In the multimodal domain, the evidence included in this review suggests that reliable results of emotion discrimination or congruency perception have not yet been found during early infancy (1–3 months). This lack of effect may be related to limitations of the early immature visual system, such as reduced contrast sensitivity, scanning ability, and visual acuity (Nelson, 1987), which make facial expressions difficult to discriminate reliably in the first months of life. It may also reflect insufficient developmental readiness for emotion perception at this stage. Prior unimodal studies suggested that infants at 3 to 4 months generally fail to reliably discriminate between positive and negative facial expressions when tested with static images (Barrera & Maurer, 1981). However, there is also evidence that younger infants are able to demonstrate cross-modal processing under specific conditions. For instance, by 3.5 months, infants have been shown to detect emotional congruency when presented with dynamic audiovisual cues from familiar individuals such as their mothers (Kahana-Kalman & Walker-Andrews, 2001).

Between 4 and 6 months of age, infants' ability to discriminate emotional expressions becomes increasingly consistent, particularly for positive emotions such as happiness, and the earliest signs of cross-modal congruency perception begin to appear. However, these abilities remain relatively fragile and highly dependent on specific task contexts during this period. Empirical studies show that 4-month-old infants can discriminate multimodal happy expressions from angry or fearful ones, although findings are inconsistent when comparing happiness with sadness (Flom et al., 2018; Flom & Bahrick, 2007; Montague & Walker-Andrews, 2002). By 5 months, infants generally succeed in distinguishing positive from negative facial expressions (Bornstein & Arterberry, 2003; Caron et al., 1988; Flom & Bahrick, 2007), but fail to do so when anger serves as the familiarization expression (Schwartz et al., 1985). Such asymmetries in familiarization are common in research on infant emotion processing, especially when infants are familiarized with fearful expressions (Parker et al., 2005). This developmental progression suggests that emotion perception in infancy is shaped by the joint influence of neural maturation and experiential input. During the first months of life, neural systems that support the processing of facial and vocal emotions are still undergoing rapid structural and functional development, such as the superior temporal sulcus, amygdala, and prefrontal regions, which may limit infants' ability to detect subtle emotional differences or to integrate affective information across modalities.

By 7–8 months of age, converging evidence indicates that infants demonstrate robust and reliable abilities in both emotion discrimination and cross-modal congruency tasks, with the strongest effects observed for positive emotions, particularly happiness (Flom & Whiteley, 2014; Nawrot, 2003; Soken & Pick, 1992, 1999; Walker, 1982). This developmental window is widely regarded as a critical stage in the development of multimodal emotion processing, as the ability to match facial and vocal expressions may allow infants to decode social signals more efficiently and thereby interpret others' communicative intentions more accurately. Empirical evidence further shows that infants between 5 and 7 months can already match happy and angry vocalizations with the corresponding facial expressions (Flom & Whiteley, 2014;

Grossmann et al., 2006; Palama et al., 2018; Phillips et al., 1990; Soken & Pick, 1992, 1999; Walker, 1982; Walker-Andrews, 1986). Electrophysiological studies also reveal that 7-month-old infants exhibit larger Nc amplitudes when viewing fearful or angry faces compared to happy ones (Jessen & Grossmann, 2015; Xie et al., 2019), suggesting that greater attentional resources begin to be allocated to fearful expressions during this stage. In contrast, most studies have not observed consistent differences in the P400 or N290 components for positive versus negative facial expressions across infancy (Taylor-Colls & Pasco Fearon, 2015). These developmental changes likely reflect the combined influence of sensory–perceptual maturation, increasing attentional selectivity, and accumulating social experience. Notably, infants' performance in cross-modal matching also depends on the ecological validity of auditory cues. For instance, Phillips (1990) found that 7-month-old infants were unable to match loud versus soft tones to angry and sad faces, respectively. However, when the vocal expressions included natural speech, 7-month-olds succeeded in cross-modally matching angry and sad faces (Soken & Pick, 1999). This finding suggests that successful cross-modal matching may rely on auditory cues with higher ecological validity.

By 9–12 months of age, research remains limited, and the available findings do not provide consistent support for the hypothesis that emotion processing undergoes further consolidation during this period. For instance, Ogren et al. (2018) reported that 9-month-old infants failed to form cross-modal matches for happy, sad, and angry expressions when paired with neutral faces. At this developmental stage, outcomes may be more strongly shaped by factors such as task design, the specific emotion categories tested, and the complexity of the stimuli. Importantly, infants show heightened sensitivity to negative emotions during this period, particularly in their responses to threat-related expressions such as anger and fear (Otte et al., 2015). This pattern is consistent with evolutionary–developmental accounts, which propose that as infants acquire greater mobility and begin to explore the environment more independently, their attentional priorities adaptively shift toward monitoring potential threats (Campos et al., 2000; Leppänen & Nelson, 2009). Moreover, some evidence suggests that the fear bias begins to diminish by 11–12 months (Peltola et al., 2013), potentially reflecting the accumulation of greater social experience with fearful expressions (Xie et al., 2019).

Taken together, these findings underscore that the first year of life represents a critical period of dynamic change in multimodal emotion perception. More basic abilities, such as the discrimination of positive emotions, emerge early, whereas more complex skills of cross-modal integration develop progressively later. Mapping this developmental trajectory not only advances understanding of socio-emotional development in infancy but also offers a theoretical basis for early screening and intervention, with potential applications for populations at heightened risk of socio-emotional difficulties.

2.4.4 Methodological Considerations and Limitations of the Present Review

The findings of this review should be interpreted considering the methodological and sample heterogeneity across the included studies. Although the review encompasses a broad developmental range from newborn to 12-month-old infants, the distribution of studies across age groups is markedly uneven. For instance, research on 7-month-old infants is disproportionately represented, whereas empirical studies of infants aged 9–12 months remain relatively sparse. Such imbalances may limit a comprehensive understanding of the developmental trajectory of emotion processing and may overemphasize characteristics observed in certain age groups. Moreover, substantial variability existed in sample characteristics and age definitions across studies. Some studies, for instance, classified a broad age range (e.g., 5–9 months) as a single "7-month-old" group, potentially obscuring more fine-grained developmental differences. Sample sizes were generally small, with some studies including fewer than 20 infants, and several failing to report effect sizes or statistical power. These limitations reduce the robustness of statistical conclusions and constrain the generalizability of the findings.

Second, notable variability also exists in task paradigms and measurement approaches. Most behavioral studies employed visual preference or habituation–dishabituation paradigms, while others used more complex matching tasks. Although such diversity reflects methodological richness, it also complicates cross-study comparisons and synthesis. Differences in trial duration, number of trials, and intermodal presentation formats may all influence infants' task performance. Measurement methods constitute another critical source of variability. Behavioral studies typically relied on looking time as an indirect index of perceptual discrimination, which is susceptible to attentional fluctuations and fatigue. In contrast, neurophysiological techniques such as event-related potentials provide valuable complementary insights into neural processing but differ in temporal and spatial resolution and are highly sensitive to motion artifacts, which is an especially salient concern in infancy studies.

Third, there is also considerable variation in stimulus characteristics across studies. For instance, visual stimuli may consist of static images or dynamic expressions; the types of emotional pairings vary (e.g., happy vs. angry, happy vs. sad, happy vs. interested); the nature of the auditory stimuli also differs (infant-directed speech, artificial speech, or non-speech sounds); audiovisual synchrony (synchronous vs. asynchronous); and the identity of the emotion source (familiar vs. unfamiliar individuals). All these factors may systematically influence infants' performance and introduce potential confounding variables when integrating findings across studies.

Finally, the current evidence is almost entirely based on cross-sectional studies, with a lack of longitudinal research tracking the same infants across multiple developmental stages. This gap constrains the ability to establish the temporal sequence and causal mechanisms underlying the development of emotion processing. In addition, the present review primarily relied on English-language databases, potentially overlooking relevant non-English publications. Publication bias toward positive findings may also lead to an overestimation of the age at which cross-modal emotion processing abilities first emerge.

2.4.5 Future Research Directions

Building on the findings of this review and methodological considerations, future research could address several factors considered. First, there is the need to optimize the distribution of participant age groups. Current empirical evidence on multimodal emotion processing is concentrated on 7-month-old infants, with relatively few investigations targeting the earliest (0–3 months) and later (9–12 months) stages of the first year of life. Previous reviews and related work indicate that the visual system is still immature in the earliest months of life (0–3 months), limiting stable emotion recognition, but by around 5 months of age, infants already display relatively mature recognition abilities, at least for positive emotions (Ruba & Repacholi, 2020). Evidence from infants aged 9–12 months remains inconsistent, and the mechanisms and theoretical accounts underlying this variability are not yet well understood. Achieving more balanced representation across developmental stages will allow future studies to provide a clearer and more comprehensive picture of the developmental trajectory of multimodal emotion processing, while reducing the risk of overemphasizing the distinctiveness of certain stages.

Second, future studies could systematically manipulate and carefully control key stimulus characteristics. For instance, the presentation of visual stimuli (static vs. dynamic), the type of emotional contrast (cross-valence vs. within-valence), audiovisual synchrony, and the familiarity of the emotion source. Clearly reporting these variables will not only enhance the comparability of findings but also establish a stronger basis for cross-study comparisons and meta-analytic synthesis. Moreover, future research could also expand the range of emotional categories under investigation, moving beyond commonly examined cross-valence contrasts (e.g., happy vs. sad) to explore infants' processing of more subtle or complex social emotions, such as interest, surprise, or other mixed emotions.

In addition, advancing methodological improvements is equally important. Integrating multiple levels of measurement within the same experimental design, such as combining visual preference paradigms with EEG/ERP or fNIRS, can provide deeper insights into the neural mechanisms underlying behavioral outcomes. For instance, such designs can help determine whether infants' lack of behavioral preference reflects perceptual insensitivity or constraints imposed by task design. Longitudinal studies following the same group of infants across multiple time points are also essential for clarifying whether emotion perception and integration abilities unfold gradually or emerge in developmental shifts, as well as for identifying stability, developmental patterns, and individual differences. Future research could further increase sample size and cultural diversity by including infants from varied sociocultural contexts, thereby enhancing the generalizability of findings. Finally, enhancing research transparency and reproducibility remains particularly important. Practices such as preregistration, data sharing, and the publication of null results can reduce publication bias and improve the transparency and balance of the evidence base, contributing to a more robust framework for the study of multimodal emotion processing in infancy.

CHAPTER III.

MULTIMODAL INTERGRATION OF AFFECTIVE INFORMATION: BEHAVIORAL EVIDENCE FROM 6-8-MONTH-OLD INFANTS

The previous chapter highlighted the importance of affective information for early socio-emotional development and reviewed evidence showing that infants are sensitive to emotional signals conveyed across multiple modalities. However, relatively little is known about how these sensitivities manifest in real-time behavior, particularly during the critical period between 6 and 8 months of age when socio-emotional competencies undergo rapid change. To address these questions, the present study investigates whether infants demonstrate reliable preferences for congruent over incongruent facial–vocal emotional expressions, and whether the type of emotion influences their intermodal matching performance, thereby providing direct behavioral evidence of multimodal affective integration. In addition, the study explores how individual differences in this ability relate to the expressiveness of the family environment, which represents the primary social context in which infants first encounter and learn to interpret emotional expressions.

3.1 Background and Rationale

Infants enter the world immersed in social interaction and emotional communication from the very moment of birth. The ability to perceive and integrate others' emotional states is not only essential for evolutionary survival but also profoundly shapes the formation and development of emotional cognition and regulation (Campos et al., 1989; Saarni, 2000; Vaish et al., 2008). From an adaptive perspective, early sensitivity to affective information enables infants to interact effectively with the world around them and secure social support (Darwin & Darwin, 1872; Ekman, 1992b). Such sensitivity is regarded as an evolutionary advantage, allowing communication of survival-relevant information (Darwin & Darwin, 1872). Consistent with this view, Ekman (1992b)'s theory of "basic emotions" emphasized that fundamental emotions, such as fear, anger, and happiness, are universally shared across cultures and are directly linked to adaptive functions. The ability to recognize and perceive these emotions is therefore essential for infants' survival and adaptation in the real world. From a developmental perspective, research demonstrates that infants rely on caregivers' affective signals to guide their behavior in uncertain or ambiguous contexts (Campos et al., 1989). A classic example is the visual cliff experiment, in which infants decided whether to cross the deep side based on their mothers' facial expressions, illustrating the role of emotional cues in social referencing (Sorce et al., 1985). Similarly, the still-face paradigm revealed that when mothers maintained face their infants but remained facially unresponsive, infants quickly displayed wariness and withdrawal, suggesting the importance of emotional communication for their sense of security (E. Tronick et al., 1978). Together, such findings highlight that recognizing and interpreting caregivers' emotional state is fundamental not only for social referencing, but also for the establishment of secure attachment, as these signals shape infants' expectations of safety and responsive care. Remarkably, even before language acquisition, infants are already able to infer others' intentions and needs from nonverbal emotional cues such as facial expressions, vocal tones, and body movements to adjust their own behavior accordingly (Farroni et al., 2004, 2007; Sroufe, 1982).

Empirical findings illustrate that newborns are equipped with striking predispositions for social engagement. For instance, neonates preferentially orient toward their mother's voice over that of a stranger (DeCasper & Fifer, 1980), and are able to imitate adult facial gestures such as mouth opening and tongue protrusion within 0.7 to 71 hours of birth (Meltzoff & Moore, 1983). Moreover, newborns show a preference for face-like stimuli within days of birth (Johnson et al., 1991), as well as demonstrating a preference for certain facial expressions (Farroni et al., 2007). These behavioral tendencies are consistent with evolutionary accounts of basic emotions as biologically hardwired communicative signals (Ekman, 1992b). Such foundational skills play an important role in developing more complex socio-emotional competencies, such as attachment styles and social referencing, but also provide a platform for socio-cognitive capabilities, including joint attention, language acquisition, and theory of mind (Tomasello, 2003).

The advancement of cognitive neuroscience methods has provided critical opportunities for clarifying the neural mechanisms underlying the development of emotion cognition in infancy. Recent progress in developmental neuroscience converges with behavioral observations to further support this view. Studies employing neuroimaging techniques such as EEG, fNIRS, and fMRI have shown that even very young infants recruit occipito-temporal and prefrontal regions when processing emotional expressions, suggesting the presence of specialized neural systems for social perception from the very beginning of life (De Haan & Nelson, 1999;

Grossmann et al., 2010; Hoehl & Striano, 2010). Studies such as Nelson (1996) suggest that 7-month-old infants exhibit distinct brain responses when processing different emotional expressions. Nakato et al. (2009) demonstrated that infants' neural responses in face-sensitive regions become view-invariant by around 8 months of age. Moreover, Grossmann et al. (2010) reported that even in early infancy, the brain responds selectively to emotional vocal cues, showing distinct activation patterns for angry and happy prosody. Notably, LeDoux (1996) proposed a theoretical framework for understanding the early development of the infant emotion system and how emotions emerge from the brain. Early specialization aligns with the social brain hypothesis, which proposes that human brain development has been shaped by the demands of navigating complex social environments (Dunbar, 1998).

Existing research has broadly charted the developmental trajectory of infants' emotional processing during the first year of life. Although unimodal visual studies have reported that newborns are able to discriminate between happy and fearful or sad expressions (Field et al., 1983; Farroni et al., 2007), these findings remain controversial. Reliable evidence of emotion discrimination or congruency perception has not been observed in early infancy (1–3 months), likely due to immaturities in the visual system (Nelson, 1987), such as reduced contrast sensitivity, scanning ability, and acuity. By 3-4 months, infants distinguish visual emotions including happiness, anger, sadness, and fear (Caron et al., 1988; Kaneshige & Haryu, 2015), and also discriminate emotions in vocal intonation (Blasi et al., 2011; Zhang et al., 2014). However, evidence for multimodal processing is more complex and heterogeneous. 3.5-month-olds show congruency sensitivity to their mothers' but not strangers' expressions (Kahana-Kalman & Walker-Andrews, 2001; Walker-Andrews et al., 2011), and by 5–6 months, some studies report successful detection of face–voice congruency, though results vary across paradigms such as habituation, preferential looking, and intermodal matching (Poncet et al., 2022; Soken & Pick, 1999; Walker-Andrews, 1986). By 7-9 months, evidence becomes more consistent: infants reliably match facial and vocal emotions and exhibit distinct neural responses to incongruent pairings (Grossmann et al., 2006; Palama et al., 2022; Phillips et al., 1990). ERP and fNIRS studies similarly show different temporal and prefrontal activation patterns to different multimodal emotional expressions, indicating the emergence of specialized neural systems for emotion integration (Grossmann, 2010; Hoehl & Striano, 2010). By contrast, other studies have reported no evidence of such congruency sensitivity until as late as 9 months of age (Ogren et al., 2018; Xiao & Emberson, 2019). Overall, this period is often considered a critical window in emotional–cognitive development. While many studies have focused on infants of a single month (e.g., only 7-month-olds), Recent work further highlights meaningful variation across 6–8 months. For example, Tang et al. (2023) tracked infants monthly from 6 to 9 months and found age-related changes in social responsiveness. Similarly, Pérez-Edgar et al. (2020) emphasize the importance of capturing finer-grained developmental trajectories rather than relying on narrow age cohorts. Given that perceptual and integrative abilities during this developmental window likely unfold gradually and progressively, highlighting the need for comparative analyses across multiple age cohorts to clarify how chronological age shapes the progressive emergence of emotion integration.

Among various emotional expressions, positive emotions, particularly happy facial expressions, play an especially important role at the early development stage in infancy. Although infants show variability, a consistent positivity bias is evident from about 3.5 to 7 months across visual and auditory modalities (Flom & Whiteley, 2014; Leppänen & Hietanen, 2004). This advantage can be explained through evolutionary, social, and perceptual perspectives: smiles signal safety and bonding, carrying high adaptive value and being readily processed early (Farroni et al., 2007; Leppänen & Hietanen, 2004), frequent caregiver expression of positive emotions increases infants' exposure and strengthens this bias (Eisenberg et al., 1998), and the distinctive perceptual features of happy faces facilitate detection (Jessen & Grossmann, 2014; M. White, 1995). Consequently, positive emotions are consistently processed earlier and more reliably across modalities, especially when expressed by familiar individuals (Walker-Andrews et al., 2011). However, this early positivity bias does not diminish the role of negative emotions: sensitivity to anger and fear emerges later, following distinct developmental trajectories (Jessen et al., 2016; Nelson & De Haan, 1996; Vaish et al., 2008). Review by Leppänen et al. (2009), synthesizing ERP evidence, suggests that infant brains are sensitive to positive emotions during the first six months, while after seven months they demonstrate stronger processing of threatening emotions such as anger and fear. This shift reflects adaptive adjustments in attentional priorities as infants gain independence (Campos et al., 2000), and heightened vigilance to potential threats holds significant survival value (Leppänen & Nelson, 2009; Peltola et al., 2009), while also being partly driven by curiosity toward novel stimuli (Malatesta & Haviland, 1982). Socialization processes also contribute, as early environments are dominated by positive caregiver emotions, whereas negative emotions occur less frequently, prompting infants to allocate more attention when they later encounter them (De Haan et al., 2013; Peltola et al., 2008; Rose et al., 1982). Overall, happiness benefits from high frequency and perceptual salience, whereas threat-related emotions gain importance later for survival and social interpretation, emphasizing the need to examine how infants integrate both positive and negative cues across modalities.

Beyond universal developmental trajectories, identifying the factors that contribute to infants' emotion perception offers valuable insights into the developmental pathways of both emotion perception and understanding. Among these factors, the infant's social environment, particularly the level of family emotional expressiveness, emerges as a critical influence that may account for individual differences. Family emotional expressiveness refers to the frequency, intensity, and style in which parents display emotions during daily interactions (Halberstadt, 1986). Research has shown that emotional expressiveness within families not only effectively shapes the family's emotional atmosphere but also provides infants with rich and continuous emotional input. During early infant development, this input can both enhance their emotional perception abilities and potentially constrain their development. Variability in parental expressiveness has been linked to children's social-emotional development outcomes, including infants' emotional understanding, emotional regulation, and social competence (Cole et al., 2013; Halberstadt et al., 1999). For instance, children raised in families with frequent positive expressiveness tend to demonstrate stronger emotional awareness and more adaptive emotional regulation strategies, whereas those in environments characterized by high levels of negative expression are more likely to experience difficulties in socio-emotional adjustment (Spinrad & Eisenberg, 2019). Although much of this evidence comes from studies of preschool and school-aged children, emerging work suggests that these associations may originate in infancy. Infants growing up in more emotionally expressive families are exposed to a richer repertoire of affective cues, which may enhance the predictability, salience, and diversity of emotional signals available for learning (Cheung et al., 2016). Conversely, environments low in expressiveness may limit opportunities for infants to practice recognizing and integrating multimodal affective information. Neurobehavioral studies further suggest that the family emotional atmosphere can shape the functioning of infants' stress-response systems and emotion-processing networks, pointing to the early socialization effects of family expressiveness on both neural and behavioral development (Feldman, 2007a, 2015). Integrating the developmental stage framework with research on environmental factors, such as family emotional expressiveness (Ogren et al., 2018), may help provide a more comprehensive account of why some infants demonstrate stronger or earlier intermodal integration than others.

Empirical evidence connecting family expressiveness with infants' multimodal emotion perception remains scarce. Ogren et al. (2018) have reported a positive correlation between family expressiveness and emotion matching for happy trails in 9-month-old infants. The result provides valuable insight and possibility into how family expressiveness may influence emotion perception early in development. Clarifying how family expressiveness influences infants' emotion matching is essential not only for understanding how environmental variation contributes to individual differences in early emotion processing, but also for examining how developmental trajectories are shaped by the interplay between maturational changes and socialization experiences. The present study therefore aims to test whether the level of family expressiveness predicts the robustness of intermodal emotion matching during a critical developmental window in infancy.

In summary, existing research suggests that infants gradually acquire the ability to integrate affective information across modalities during the first year of life. Nevertheless, several critical questions remain unresolved. First, although 7-month-old infants often demonstrate relatively stable multimodal integration, systematic comparisons across the broader 6–8-month age range are scarce, leaving the developmental trajectory of progressive, age-related change insufficiently understood. Second, the specific contribution of different emotion types to early emotion perception remains unclear. While existing studies have reported both a bias toward positive emotions and heightened sensitivity to threat-related expressions, it is not well established how these differences influence infants' intermodal matching performance. Third, there are studies investigating associations between the family emotional environment and children's socio-emotional development; direct evidence linking family expressiveness to infants' intermodal emotion processing remains limited.

The primary aim of the present study is to advance understanding of early multimodal emotion perception by integrating a laboratory-based behavioral paradigm with parent-reported assessments of family emotional expressiveness. Specifically, the study seeks to address the following questions:

- Are infants between 6 and 8 months of age able to integrate affective information across facial and vocal modalities? Does performance vary across this key developmental window?
- How do specific emotional types shape infants' intermodal emotion matching abilities? Does such integration reflect an advantage for positive emotions?

- Whether infants' intermodal matching ability is influenced by the level of emotional expressiveness in the family environment?

Based on prior theoretical and empirical evidence, we mainly formulated the following hypotheses:

- Developmental trajectory hypothesis: Infants aged 6 to 8 months are able to match affective information across facial and vocal modalities, with performance showing age-related variation in intermodal emotion perception.
- Positivity bias hypothesis: Across conditions, infants will show stronger and more consistent intermodal matching for positive emotion.
- Environmental influence hypothesis: Higher levels of family emotional expressiveness will be positively associated with infants' intermodal matching performance.

3.2 Methods

3.2.1 Participants

A total of 38 healthy, full-term infants were recruited for this study through the BabyLab database at the University of Padua. Three infants were excluded from the final analyses: two due to excessive fussiness during testing and one due to a technical error. Thus, the final sample consisted of 35 infants (19 female, 16 male) ranging in age from 6.00 to 8.52 months ($M = 7.08$, $SD = 0.77$). All included infants were full-term (37-42 weeks of gestation), born without delivery complications, had no history of neurological or developmental disorders, and came from Italian-speaking families. Parents provided written informed consent before participation, and the study was approved by the Ethics Committee of the Department of Developmental Psychology and Socialisation, University of Padua. A sensitivity analysis conducted with G*Power 3.1 showed that, with $n = 35$ and $\alpha = .05$ (two-tailed), the study had 80% power to detect effects of Cohen's $d \geqslant 0.48$ in paired comparisons.

3.2.2 Materials

3.2.2.1 Stimulus Creation

Dynamic audiovisual recordings were created as experimental stimuli, consisting of three emotion conditions: Happy, Angry, and Neutral, presented by two native Italian speakers. To generate the auditory stimuli, 19 semantically neutral scripts were developed (full materials are available in the Appendix). Each script consisted of an 8-second sentence designed to preserve natural vocal characteristics while minimizing linguistic content. Examples include “Vieni qui,” “Perché fai questo,” and “Arrivederci” (“Come here,” “Why are you doing this,” “Goodbye”). The speakers were instructed to produce each sentence with intonations that clearly expressed the three target emotions. All audio files were edited using Ocenaudio and Praat to standardize volume, duration, intensity, and background noise.

For the visual stimuli, the same speakers displayed the target emotions through facial expressions while reading the predefined scripts. Recordings focused on the upper body against a neutral, non-white background, with direct eye contact maintained and only minimal natural head movements permitted. Audio and video were recorded separately and subsequently synchronized using OpenShot video editor; a stimulus example is shown below (Figure 3.1).

Figure 3.1 Example of visual stimuli.
Note. The figure shows one of the models displaying facial expressions of happiness (right) and anger (left), which served as the visual component of the audiovisual stimuli presented to infants.

3.2.2.2 Stimulus Validation

After the editing process, all stimulus materials underwent validation to ensure that the affective information was clearly identifiable. For this purpose, 23 undergraduate students (aged 19–26 years; 16 female) participated in an online survey programmed via the Labvanced platform. Each stimulus was presented independently: auditory stimuli were played against a plain black background with no accompanying facial expressions, while visual stimuli were shown without sound. Participants were asked to classify the emotion conveyed by each stimulus (happy, angry, or neutral) and to rate its perceived intensity on a 7-point Likert scale. Their responses were then used to identify and select the most representative stimuli of each emotional category for inclusion in the experiment.

In the Labvanced validation survey, participants accurately recognized the emotional content of the final set of audio and video stimuli, with an overall accuracy of 83% (angry: 78%, happy: 88%, neutral: 82%). Based on these results, the 12 scripts that received the highest ratings were selected for use in the current study. Each selected stimulus represented a unique pairing of script and emotional condition, ensuring that no verbal content was repeated with the same emotion across stimuli.

Acoustic analyses of the final audio set were conducted using Praat. For each sentence, we extracted mean pitch, maximum pitch, minimum pitch, duration, and intensity. Duration and intensity values were standardized to ensure consistency across stimuli. Although pitch levels varied systematically across emotional conditions, both the happy (Mean difference = 70.13, $SE = 5.07$, $p < .001$) and angry (Mean difference = 50.79, $SE = 5.07$, $p < .001$) conditions exhibited significantly higher mean pitch values than the neutral condition. This variation was considered appropriate, as it reflects natural patterns of emotional prosody (e.g., elevated pitch in happy tones). Therefore, the recordings were regarded as high-quality, preserving ecological validity while reliably conveying the intended emotions.

3.2.2.3 Surveys

Informed consent was obtained from the parents of all participating infants, who also provided demographic information and completed the Family Expressiveness Questionnaire (FEQ; Halberstadt, 1986). The FEQ is a 40-item self-report instrument that assesses patterns of emotional expression and socialization within the family, using scenarios that capture both verbal and nonverbal forms of expression. It measures the extent to which family members typically display emotions in everyday contexts and has been widely used to examine links between family expressiveness, individual emotional expression, nonverbal behavior, and emotional skills. Parents rated each item on a 9-point Likert scale (1 = behavior not at all frequent in my family, 9 = behavior very frequent in my family), indicating the extent to which each item reflects a behavior exhibited within the family.

The FEQ comprises a total score ranging from 40 to 360 and is divided into four subscales of 10 items each that represent the affect dimension crossed by the power dimension: Positive-Dominant (PD), Positive-submissive (PS), Negative-Dominant (ND), and Negative-submissive (NS). Scores for each subscale range from 10 to 90 (Halberstadt et al., 1995). For the present study, the questionnaire was translated into Italian using a double independent translation and back-translation procedure. Two bilingual translators independently translated the original English version into Italian, and then each back-translated the other's version into English. The back-translated versions were compared with the original questionnaire to ensure semantic accuracy and conceptual equivalence.

3.2.3 Procedure

The experiment was programmed using E-Prime 2.0 (Psychology Software Tools). Infants were seated on their caregiver's lap at approximately 60 cm from the monitor. A video camera recorded infants' visual fixation on the emotional stimuli, while both stimulus presentation and infant behavior were simultaneously captured using OBS Studio (Open Broadcaster Software, Version 30.1.2).

The experimental setup consisted of a 70 × 40 cm monitor with a resolution of 1280 × 720 pixels. Two identical speakers were positioned 10 cm from each side of the screen to ensure consistent orientation and sound delivery of the auditory stimuli. Each trial commenced with a 2-second attention-getting clip, followed by the presentation of two affective video clips (8 seconds each) displayed side-by-side. During the 8-second trial, a vocal expression was played, with its emotional content congruent with one of the two video clips. Each video measured 63 × 31.5 cm, corresponding to a visual angle of approximately 55.4° × 30° at a viewing distance of 60 cm. The two videos were separated by a blank interval of 10.5 cm, equivalent to a visual angle of 34.1°. An overview of the experimental flow is provided in Figure 3.2.

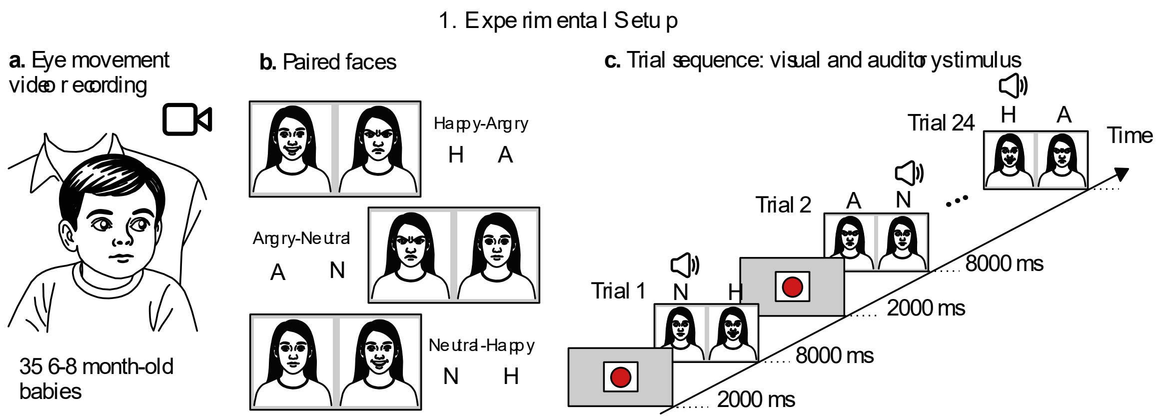


Figure 3.2 Experimental setup and procedure. The schematic illustrates the experimental design. *Note.* (a) Infants' eye movements were recorded with a video camera while seated on their caregiver's lap. (b) Stimulus pairs consisted of dynamic video clips of emotional facial expressions, presented in three combinations: Happy–Angry, Angry–Neutral, and Neutral–Happy. (c) Each trial began with a 2-s attention-grabbing stimulus, followed by an 8-s presentation of two side-by-side facial expressions, accompanied by a vocal expression congruent with one of the videos. In total, 24 trials were presented in a randomized order. The diagram is an illustrative schematic created in Inkscape and does not include images of real participants.

The experiment comprised 24 trials, organized into three blocks of eight trials each. Each trial lasted 8 seconds and was presented in a randomized order. The spatial position of the emotional expressions (left vs. right) was also randomized, with two constraints: (a) no more than three consecutive trials displayed the same emotion, and (b) no more than two consecutive trials presented the same emotion on the same side of the screen. In each trial, the accompanying audio clip corresponded to the emotional expressions of one of the simultaneously displayed videos. The audio was temporally asynchronous with both videos to prevent infants from relying solely on temporal synchrony to identify the emotional match.

Before each 8-second trial, a centrally presented attention-getting stimulus was displayed for 2 seconds to reorient the infant's gaze toward the screen. Trials commenced only once the infant was confirmed to be attending the display. Parents were instructed to hold their infants securely while allowing them to look freely at the screen, without attempting to direct their attention or otherwise interact with them. However, if an infant had completely disengaged from the preceding trial, parents were permitted to gently reorient the child toward the monitor during the attention-getting phase.

3.2.4 Data Coding and Data Analysis

3.2.4.1 Data Coding

All video recordings were coded using Datavyu software. Infant gaze behavior was analyzed on a frame-by-frame basis, with each coding unit corresponding to 30 frames. Two trained researchers independently coded the infants' fixations for each trial, categorizing whether the infant was looking to the left, right, or away from the screen. Discrepancies between coders were reviewed jointly and resolved through discussion, with inconsistent segments re-coded when necessary. Based on the finalized coding, the total looking time directed toward each stimulus was computed for every trial.

3.2.4.2 Descriptive Statistics

We conducted descriptive analyses of participants' demographic characteristics and parental information, as well as responses to the Family Expressiveness Questionnaire (Halberstadt, 1986). Descriptive statistics for looking times across different emotion types are presented in Table 1. FEQ results are summarized in Table 2. Figure 3 provides an illustrative example of trial-by-trial task performance from a single participant in one emotional condition.

3.2.4.3 Data Analysis

1. Looking Time

Infants' gaze behavior was coded frame by frame from the video recordings of each trial. For every trial, we calculated the total looking time directed toward the congruent and incongruent facial expressions. These trial-by-trial measures were then averaged across trials to obtain each infant's mean looking time in the congruent and incongruent conditions.

In addition to this main measure, we further examined infants' overall attentional allocation during the task. Specifically, for each trial, we computed the cumulative looking time directed at each emotional expression (happy, angry, neutral), regardless of congruency. Looking time directed away from the stimuli (i.e., off-screen) was also recorded as an index of overall task engagement. Together, these measures enabled us to capture both infants' selective emotion-matching performance and their broader distribution of attention across emotional categories and non-stimulus areas.

2. Emotion Matching Score

The primary dependent variable is the Emotion Matching Score (EMS), which evaluates infants' ability to allocate attention to facial expressions that were congruent with concurrently presented vocal cues. For each trial, we recorded the total fixation duration on the target face. In the happy condition, for example, this referred to fixations on the happy facial expressions when paired with either congruent vocal cues (also happy) or incongruent vocal cues (expressing a different emotion). We then calculated the mean fixation time on the target face separately for congruent (matching) and incongruent (non-matching) trials. The EMS was computed as follows: EMS = Average fixation time on the target face in matching trials – Average fixation time on the same face in non-matching trials (Bergelson & Swingley, 2012). By definition, a positive EMS reflects a stronger attentional preference for facial expressions that are congruent with vocal emotions, whereas a negative or near-zero EMS suggests a lack of such preference. To determine whether EMS values differed across emotional conditions, we conducted a three-level repeated-measures analysis of variance (ANOVA) with Emotion (Happy, Angry, Neutral) as the within-subject factor. Where the omnibus test revealed significant effects, we carried out pairwise comparisons with Bonferroni correction to control for inflated Type I error. In addition, to evaluate whether infants' EMS values significantly deviated from chance, we performed one-sample t-tests against zero for each emotional condition separately. This allowed us to determine whether the observed attentional preferences for congruent face–voice pairings were statistically reliable within each specific emotion.

To examine potential age-related differences, infants were divided into three groups according to chronological age: 6, 7, and 8 months. We then performed separate simple linear regression analyses for each emotional condition (Happy, Neutral, Angry). This analysis enabled us to evaluate whether age uniquely contributed to variance in EMS across different emotional contexts.

Finally, to investigate whether characteristics of the family expressiveness environment contributed to individual differences in infants' emotion matching abilities, we conducted a series of multiple linear regression analyses. Specifically, FE value, measured via parent-report questionnaires, was entered as the predictor variable, while EMS values for each emotional condition, as well as the overall mean EMS, were entered as dependent variables in separate models. These analyses allowed us to test whether higher levels of family expressiveness predicted stronger congruent face–voice preferences in infants.

Statistical analyses were conducted with R (version 4.3.3), IBM SPSS Statistics (version 28.0), and Python within Jupyter Notebook.

3. Linear Mixed-effects Model (LMMs)

To investigate the effects of multiple factors on infants' emotion matching scores while accounting for individual variability, we constructed a series of linear mixed-effects models. In all models, Subject was included as a random intercept to control for repeated measures and baseline differences in emotional processing.

The initial null model (Model 0) included only the random intercept for Subject to estimate baseline variance in emotion matching scores. Fixed effects were then added in a stepwise fashion: first Emotion, followed by Age, and subsequently their interaction. Gender was then introduced as an additional predictor to test for main effects and potential higher-order interactions. The models were specified as follows:

Model 0 (null model): Specified the hypothesis of no effect of the independent variables and accounted only for individual variability across subjects.

Model 1: Specified the hypothesis of the main effect of emotion condition on the outcome.

Model 2: Specified the hypothesis of additive effects of emotion condition and age, without interaction.

Model 3: Specified the hypothesis of a two-way interaction between emotion condition and age, testing whether the effect of emotion varies as a function of age.

Model 4: Specified the hypothesis of a two-way interaction between emotion and age, with an additional additive effect of gender.

Model 5: Specified the hypothesis of a three-way interaction among emotion, age, and gender, including all corresponding lower-order effects and interactions.

Among the specified models, we selected the model that produces the lowest AIC value (Hooper et al., 2008) and quantified the strengths of evidence supporting this selection using its AIC weight, which can be interpreted as conditional probability for each model of being the most plausible, given the data and the set of candidate models (Wagenmakers & Farrell, 2004). Then we tested the effects predicted by the best model using analysis of deviance (ANOVA, type III Wald chi-square test, “car” package; Fox & Weisberg, 2018), and post-hoc contrasts were conducted using the emmeans package (Lenth, R. V et al., 2024).

3.3 Results

3.3.1 Descriptive Results

A total of 35 infants (16 males, 19 females) were included in the final analyses. Their ages ranged from 6.00 to 8.52 months (*M* = 7.08, *SD* = 0.77). An additional 3 infants were tested but excluded due to excessive fussiness during testing (n = 2) or technical error (n = 1). The overall EMS across conditions was slightly positive (*M* = 0.03, *SD* = 0.11). Descriptive statistics for EMS in the Happy, Angry, and Neutral conditions are summarized in Table 3.1.

Table 3.1 Descriptive statistics for Emotion Matching Scores across conditions

Condition	***M***	***SD***	**Min**	**Max**	**N**
Happy	0.05	0.13	-0.22	0.32	35
Angry	-0.05	0.11	-0.24	0.16	35
Neutral	0.04	0.14	-0.25	0.33	35
Overall	0.03	0.11	-0.20	0.24	35

Note: EMS = Emotion Matching Score. Values represent minimum (Min), maximum (Max), mean (M), and standard deviation (SD) across participants.

In addition, parents completed the Family Expressiveness Questionnaire. FEQ scores ranged from 169 to 269 (*M* = 225.8, *SD* = 26.1), which are comparable to values previously reported for parent-completed FEQ assessments (Ogren et al., 2018), indicating moderate to normative levels of emotional expressiveness within the participating families. The FEQ scores observed in our sample are presented in Table 3.2.

Table 3.2 Descriptive Statistics for Family Expressiveness Questionnaire (FEQ) Subscales

Subscale	**Mean**	**Standard Deviation**	**Observed Range**
Total	225.8	26.1	169-269
Positive-Dominant	74.5	8.3	58-88
Positive-Submissive	74.4	6.3	63-85
Positive	148.9	13.4	124-170
Negative-Dominant	33.7	10.6	13-59
Negative-Submissive	43.2	11.6	17-60
Negative	76.9	19.7	30-119

Note. The FEQ subscales include Total Expressiveness, Positive Expressiveness (subdivided into Positive-Dominant and Positive-Submissive), and Negative Expressiveness (subdivided into Negative-Dominant and Negative-Submissive).

To provide a more detailed illustration of infants' trial-by-trial behavior, we recorded individual performance in both congruent and incongruent conditions for each emotional category. Figure 3.3 shows the performance of a representative participant (Participant 12) under the "Happy" condition, separately for congruent (left panel) and incongruent (right panel) trials. Each horizontal bar represents a single trial, with time on the x-axis and the visual state (looking at the happy face vs. looking at the alternative expressions) on the y-axis. Blue segments indicate periods in which the infant fixated on the happy expression. As shown, the participant exhibited sustained looking at the congruent happy face across most trials, whereas looking patterns in the incongruent condition were more variable. Figure 3.4 and Figure 3.5 present the same participant's performance in the Angry and Neutral conditions, respectively. The Appendix presents the trial-level performance of all participants.

Figure 3.3 Emotional matching performance based on auditory affective categories for participant 12. Based on the coded data, we extracted the duration of infants' gaze directed toward different emotional facial expressions, as well as the time spent looking elsewhere. As illustrated in this figure, the gaze behaviour of participant 12 is presented for the condition in which the target emotion was happiness. The figure displays looking times toward the happy face across multiple trials. In each pair of bars, the upper bar represents the looking time in the congruent condition (i.e., when the happy face matched the vocal emotion), and the lower bar represents the incongruent condition (i.e., when the happy face did not match the vocal emotion).

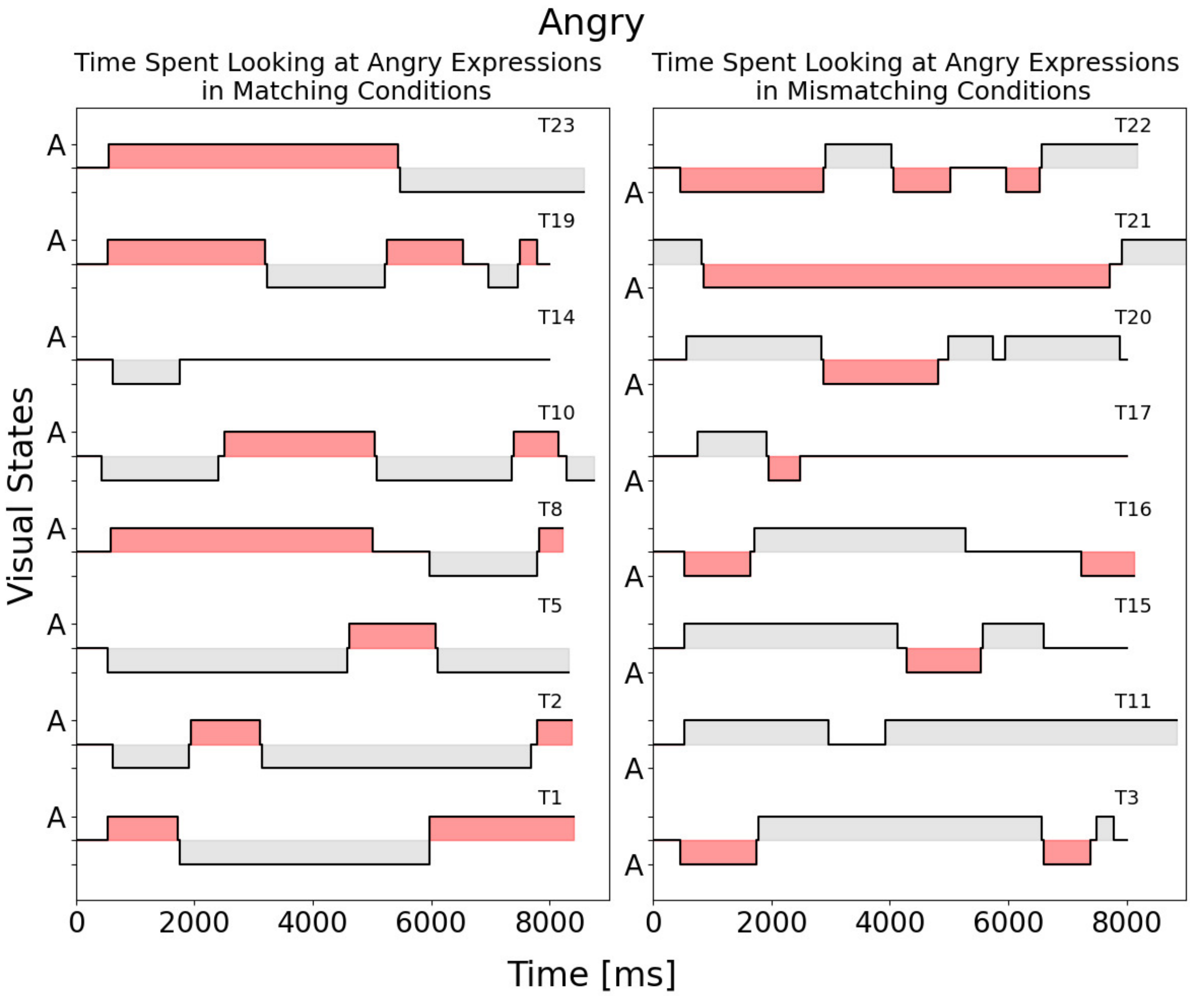


Figure 3.4 Trial-by-trial looking behavior of Participant 12 in the Angry condition.

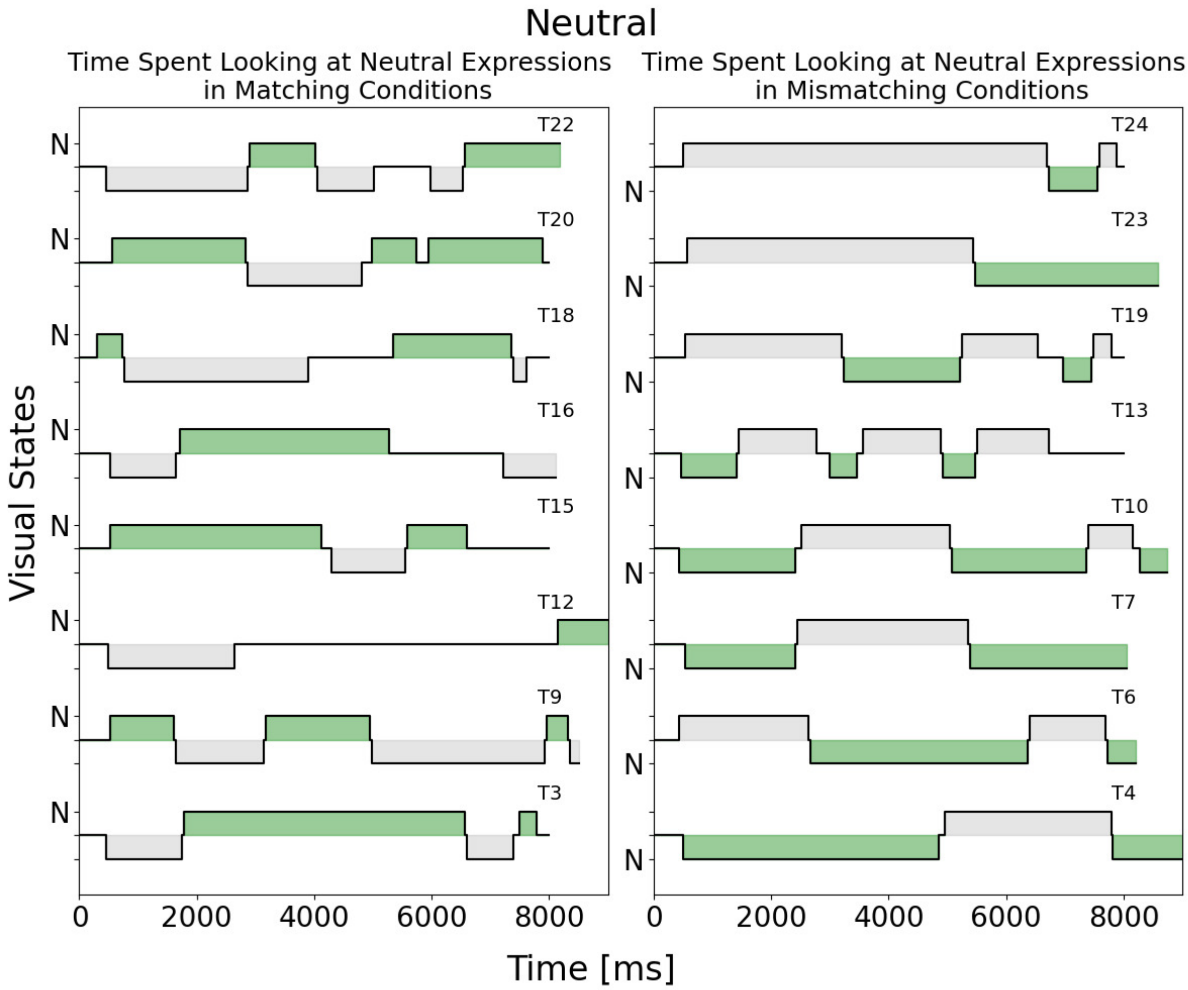


Figure 3.5 Trial-by-trial looking behavior of Participant 12 in the Neutral condition.

In addition to comparing infants' attention to congruent versus incongruent facial expressions, we also considered the potential influence of the specific facial pairings used in the task. Infants' matching performance may vary depending on the relative salience of the paired expressions. For instance, compared with neutral faces, happy or angry faces may be more attention-grabbing; consequently, in trials where the neutral expression served as the "matching" face, infants might allocate more attention to the paired emotional face, leading to weaker matching performance. To address this possibility, we calculated infants' trial-by-trial performance separately for each facial pairing condition (Happy–Angry, Happy–Neutral, Neutral–Angry). Figure 3.6 illustrates the performance of a representative participant (Subject 1) across the three pairing conditions. Trial-level data for all participants are provided in the Appendix.

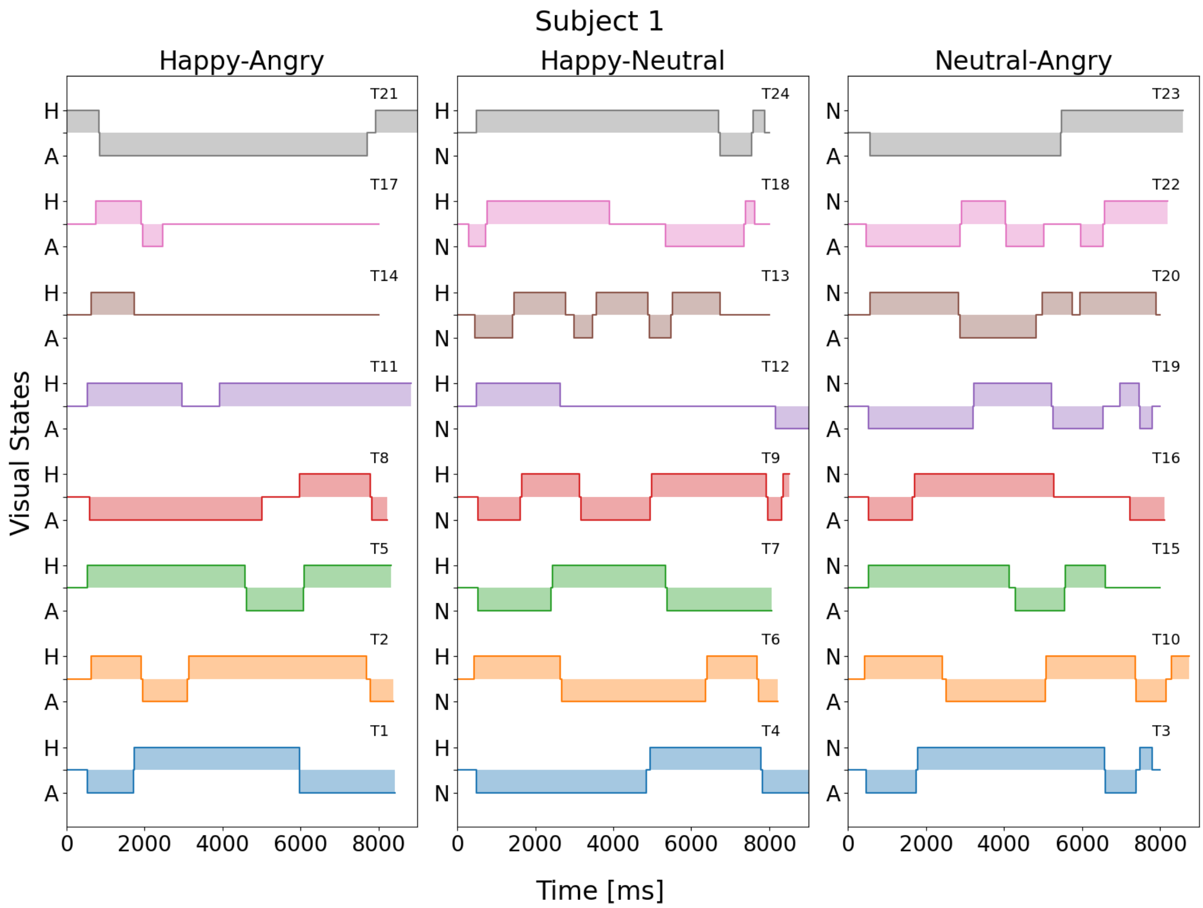


Figure 3.6 Trial-by-trial looking behavior of Participant 1 across the three facial pairing conditions: Happy–Angry (left), Happy–Neutral (middle), and Neutral–Angry (right). Each horizontal bar represents a single trial, with time on the x-axis and visual state (looking at one expression vs. the alternative) on the y-axis. Colored segments indicate fixations on the designated target face. Trial-level data for all participants are available in the Appendix.

3.3.2 Looking Time

To examine group-level performance across different facial pairing conditions, we compared infants' looking times toward the congruent facial expression, the incongruent facial expression, and off-screen (non-interest, N.I.) areas. As shown in Figure 3.7, in the Happy condition, a one-way ANOVA revealed a significant effect of region of interest, F (2, df) = 135.11, p < .001. Post-hoc tests showed that infants looked longer at the happy face than at the alternative face (angry/neutral) (p < .05), while both faces were looked at significantly more than the non-interest area (p < .001).

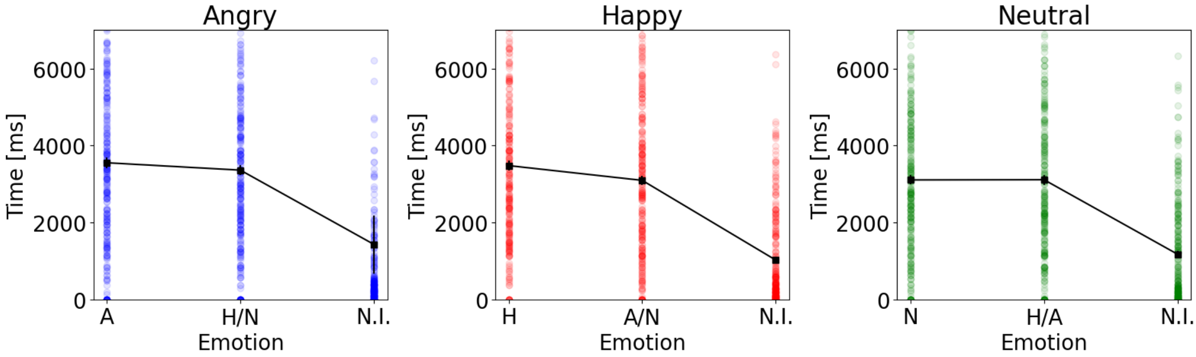


Figure 3.7 Mean looking times across regions of interest in the Angry (left), Happy (middle), and Neutral (right) target conditions. Each panel shows infants' looking times toward the target face (matching expression, left), the alternative face (non-matching expression, middle), and the non-interest area (N.I., right). Black squares with error bars represent the group mean ± SEM, and colored dots indicate individual participants' data.

To further examine the potential influence of facial pairings on infants' matching performance, we compared looking times across the Happy–Angry, Happy–Neutral, and Neutral–Angry conditions. Contrary to our prediction, no significant differences were found between pairing types. As illustrated in Figure 3.8, infants attended to both congruent and incongruent faces across all pairing conditions, with no systematic modulation by the specific facial combinations.

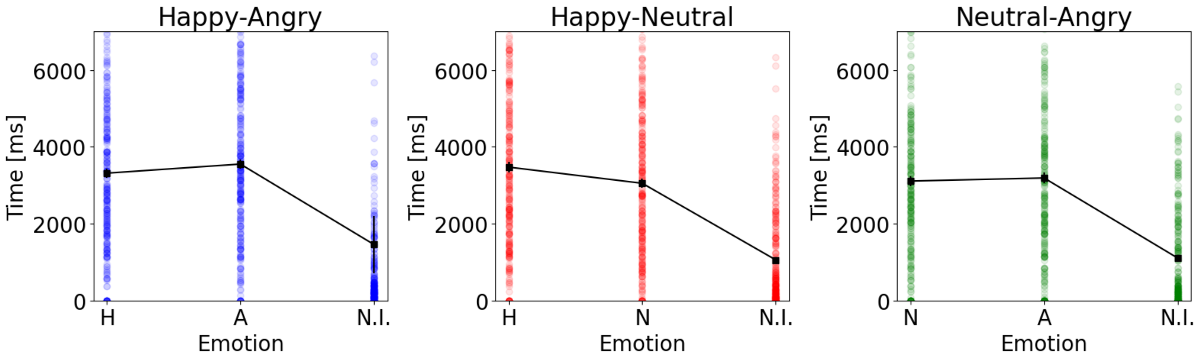


Figure 3.8 Mean looking times across facial pairing conditions. The three panels show infants' attention allocation in the Happy–Angry (left), Happy–Neutral (middle), and Neutral–Angry (right) conditions. Each panel presents looking times toward the matching face (left), the non-matching face (middle), and the non-interest area (N.I., right). Black squares with error bars represent group *M* ± *SEM*, and colored dots represent individual participants. Statistical analyses revealed no significant differences between pairing conditions.

3.3.3 Do Infants Show Evidence of Emotion Matching?

While raw looking times provide important information about infants' general attentional allocation, they do not directly capture whether infants exhibit a preferential tendency to match congruent emotional expressions. To address this, we computed EMS, defined as the difference between looking times to the congruent and incongruent faces. This relative measure controls for individual differences in overall attention and allows for a more direct assessment of infants' intermodal matching performance across conditions. Based on the previously described method for calculating, we computed EMS values for each infant under the three emotional conditions: Happy, Angry, and Neutral. The results are illustrated in Figure 3.10, which presents the individual EMS values across the various conditions. By definition, a positive EMS reflects a stronger visual preference for the emotional face that matched the accompanying vocal cue in the congruent condition.

As shown in Figure 3.9, the distribution of EMS values differed across emotional conditions. In the Happy condition, scores were broadly distributed toward positive values. This indicates that infants generally demonstrated higher EMS when responding to happy stimuli compared with other emotional contexts. The median EMS in the Happy condition was slightly above zero, suggesting that, on average, infants exhibited a stronger visual preference for congruent happy face–voice pairings. This pattern is consistent with the predicted congruency effect, whereby infants preferentially attend to facial expressions that are emotionally aligned with concurrent vocal cues.

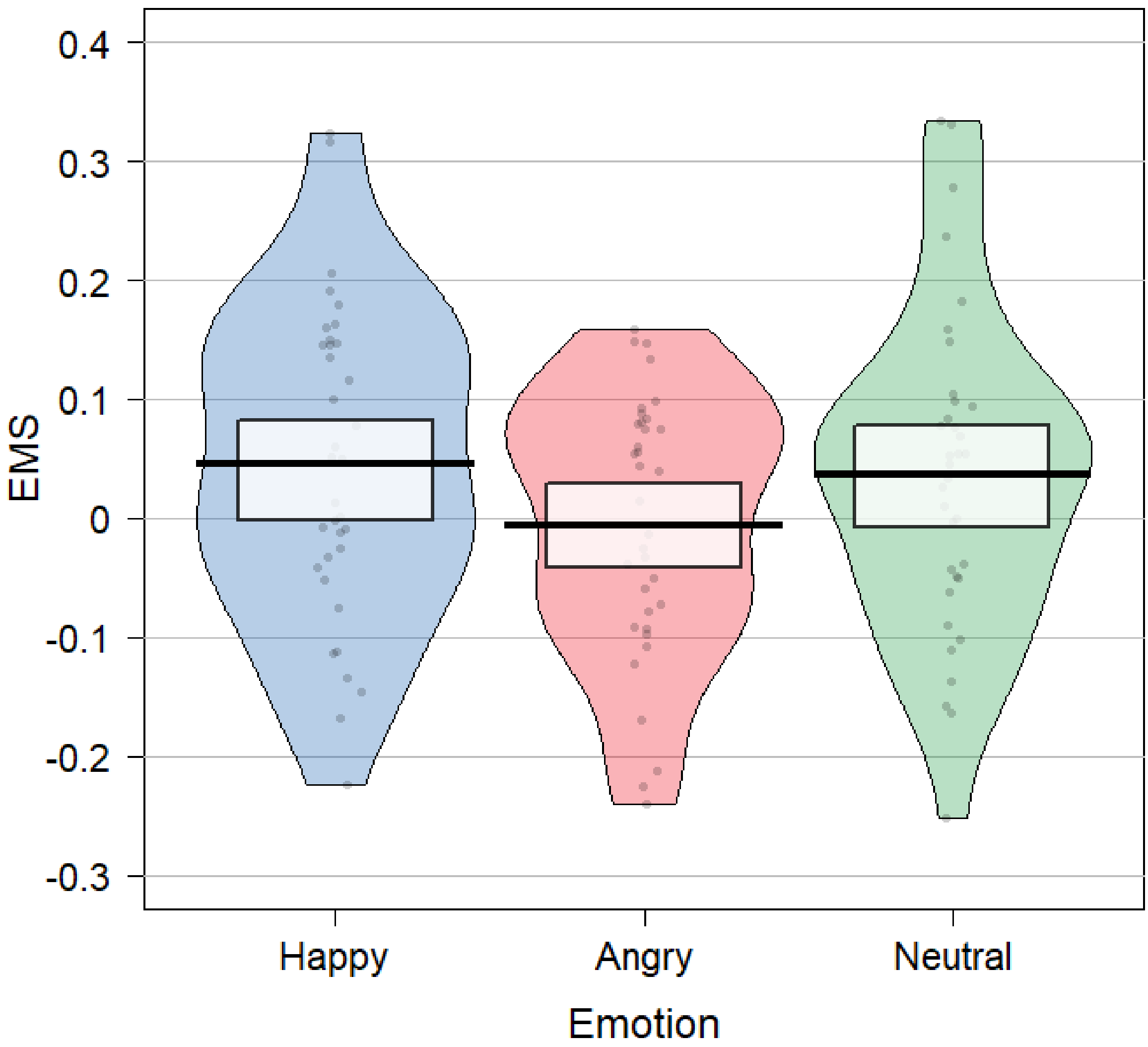


Figure 3.9 Distribution of Emotion Matching Scores across Happy, Angry, and Neutral conditions.

To examine whether EMS differed across emotional conditions statistically, a repeated-measures ANOVA was conducted with three levels: happy, angry, and neutral. The main effect of emotion was statistically significant, F (2, 68) = 3.89, p = .025, indicating that EMS varied by emotional condition. Pairwise comparisons with Bonferroni correction revealed that EMS was significantly higher in the happy condition compared to the angry condition (p = .046). There were no significant differences between the happy and neutral conditions (p = 1.000), nor between the angry and neutral conditions (p = .122).

To further assess whether EMS values significantly differed from chance, we performed one-sample t-tests against zero for each condition. In the Happy condition, EMS approached significance, t (34) = 2.02, p = .052 (two-tailed), with a mean score of 0.045 (95% CI [–0.0003, 0.0906]) and a small-to-moderate effect size (d = .34). In contrast, no evidence of congruency effects was found in the other conditions. For the Angry condition, EMS did not significantly differ from zero, t (34) = –0.21, p = .839, M = –0.004, 95% CI [–0.042, 0.035], d = –0.04. Similarly, in the Neutral condition, EMS was not significantly different from zero, t (34) = 1.60, p = .118, M = 0.037, 95% CI [–0.010, 0.084], d = 0.27. As shown in Table 3.3.

Table 3.3 Descriptive statistics of Emotion Matching Scores (EMS) across emotional conditions

Condition	***M***	***SD***	**95% CI Lower**	**95% CI Upper**
Happy	0.045	0.132	–0.0003	0.0906
Angry	–0.004	0.108	–0.042	0.035
Neutral	0.037	0.118	–0.010	0.084

These results suggest that infants exhibited relatively stronger emotion matching for happy expression, although the effect was modest, whereas no reliable congruency effects were observed for angry or neutral stimuli.

3.3.4 Does Age Modulate Emotion Matching?

To assess whether emotion matching varied as a function of age, infants were divided into three groups: 6 months, 7 months, and 8 months and older. Figure 3.10 illustrates individual EMS across age groups for each emotional condition (Happy, Angry, Neutral) and for the overall mean across conditions. Each dot represents one infant's EMS, with vertical lines connecting the score to the zero baseline. This visualization highlights both age-related trends and substantial inter-individual variability. In the Happy condition, the majority of 6-month-olds showed positive EMS values, indicating a consistent preference for congruent happy face–voice pairings. By contrast, EMS values among 7- and >8-month-olds were more widely dispersed around zero, suggesting a weakening or absence of this effect. In the Angry condition, EMS values were clustered symmetrically around zero across all age groups, reflecting little evidence of congruency effects. In the Neutral condition, 6-month-olds tended to show slightly higher EMS than older groups, although the pattern was less pronounced than in the Happy condition. The overall mean EMS panel revealed a general decline in congruency sensitivity with age, but the Happy condition primarily drove this effect. Relevant data are summarized in Table 3.

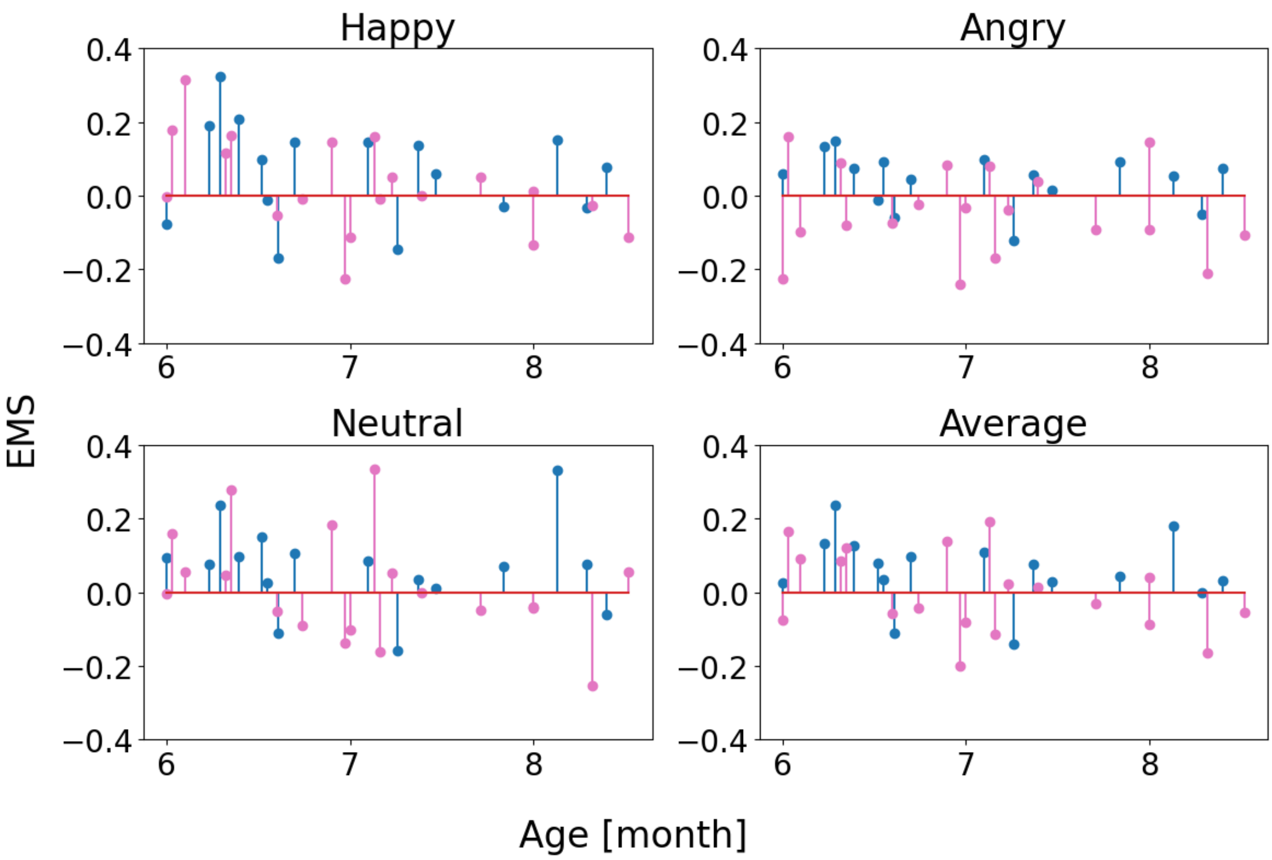


Figure 3.10 Emotion Matching Scores (EMS) by Age and Emotion Condition.

To further illustrate potential age-related effects, we plotted EMS distributions for each age group across the three emotional conditions (Figure 3.11). Descriptively, an age-related decline was most apparent in the Happy and Neutral conditions, with 6-month-olds showing the highest median EMS, 7-month-olds displaying moderately lower values, and infants older than 8 months showing the lowest. In contrast, EMS distributions largely overlapped across age groups in the Angry condition. In the Happy condition, a significant negative association was observed between infants' age and EMS, $\beta = -0.06$, $SE = 0.03$, $t\ (33) = -2.13$, $p = .041$. The model accounted for approximately 12% of the variance in EMS ($R^2 = .12$). Younger infants thus showed stronger matching tendencies for happy expressions, which decreased with age. In the Neutral condition, regression analysis revealed a negative but non-significant association between infants' age and EMS, $\beta = -0.04$, $SE = 0.03$, $t\ (33) = -1.41$, $p = .169$. The model accounted for only about 6% of the variance (R^2 = .06), indicating that age was not a reliable predictor of matching performance in this condition. In the Angry condition, regression analysis revealed no significant association between age and EMS, β = –0.02, SE = 0.02, t (33) = –0.88, p = .385. The model explained little variance in EMS (R^2 = .02), indicating that infants' matching performance in the Angry condition was not influenced by age.

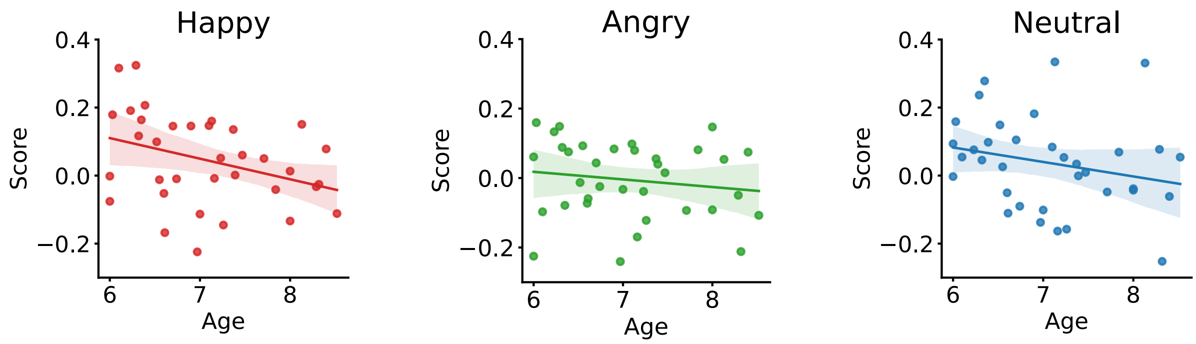


Figure 3.11 EMS as a function of age across the three emotional conditions.
Note. Each panel depicts individual infants' EMS values (colored dots) and the fitted regression line with 95% confidence intervals (shaded area). Positive scores indicate a preference for the congruent facial expression. A negative association with age was observed in the Happy and Neutral conditions, whereas no clear age-related trend emerged in the Angry condition.

As an exploratory step, we computed time-binned proportion-to-target curves for each nominal age group (<7 months, 7–<8 months, ≥8 months) and each emotion condition (Happy, Angry, Neutral), to further investigate the temporal dynamics of infants' attention to emotional stimuli across conditions (See Figure 3.12). For each group, we averaged across infants at each 1-s bin (8 bins per trial) to obtain group-level trajectories. We then compared age groups using one-way ANOVAs on the group-level time-binned means for each emotion, followed by Tukey's HSD for pairwise contrasts. Because bins are not independent observations and within-group averaging removes participant-level variance, these analyses are descriptive and are interpreted with caution; confirmatory inference was conducted with participant-level repeated-measures models.

In the Happy condition, 6-month-olds exhibited the highest and most sustained EMS across the 8-s trial window, with values rising early in the trial and remaining elevated throughout. By comparison, 7-month-olds displayed moderately lower but relatively stable trajectories, whereas infants older than 8 months showed a marked decline, with EMS values approaching zero toward the end of the trial. In contrast, the Angry condition yielded largely overlapping trajectories across age groups that hovered around zero, and the Neutral condition showed only a mild age-related decline.

Group-level one-way ANOVA confirmed these descriptive patterns. For Happy, robust age differences were observed, $F\,(2, 21) = 858.63$, $p < .001$. Tukey's HSD showed that 6–7-month-olds differed significantly from both 7-month-olds ($\Delta M = -0.0460$, 95% CI [−0.0511, −0.0409], $p < .001$) and ≥8-month-olds ($\Delta M = -0.0904$, 95% CI [−0.0956, −0.0853], $p < .001$), and that 7-month-olds also differed reliably from ≥8-month-olds ($\Delta M = -0.0444$, 95% CI [−0.0495, −0.0393], $p < .001$). For Angry, the ANOVA also indicated significant group differences, $F\,(2, 21) = 197.88$, $p < .001$. Post hoc tests revealed that 6–7-month-olds differed from 7-month-olds ($\Delta M = -0.0250$, 95% CI [−0.0306, −0.0194], $p < .001$) and from ≥8-month-olds ($\Delta M = -0.0478$, 95% CI [−0.0534, −0.0421], $p < .001$), and that 7-month-olds differed from ≥8-month-olds ($\Delta M = -0.0227$, 95% CI [−0.0284, −0.0171], $p < .001$). For Neutral, the analysis revealed significant differences among age groups, $F\,(2, 21) = 258.33$, $p < .001$. Tukey's HSD indicated stepwise contrasts: 6–7 vs. 7 months, $\Delta M = -0.0162$, 95% CI [−0.0211, −0.0113], $p < .001$; 6–7 vs. ≥8 months, $\Delta M = -0.0470$, 95% CI [−0.0519, −0.0420], $p < .001$; and 7 vs. ≥8 months, $\Delta M = -0.0308$, 95% CI [−0.0357, −0.0259], $p < .001$.

Taken together, these analyses indicate that group-level EMS values generally increased with age, with the most pronounced developmental separations observed for Happy stimuli. These between-group ANOVAs on group-averaged bin means are provided for descriptive purposes only. Because time bins are not independent samples and averaging across participants removes within-group variability, inferential claims are based on participant-level repeated-measures models with continuous age as a covariate.

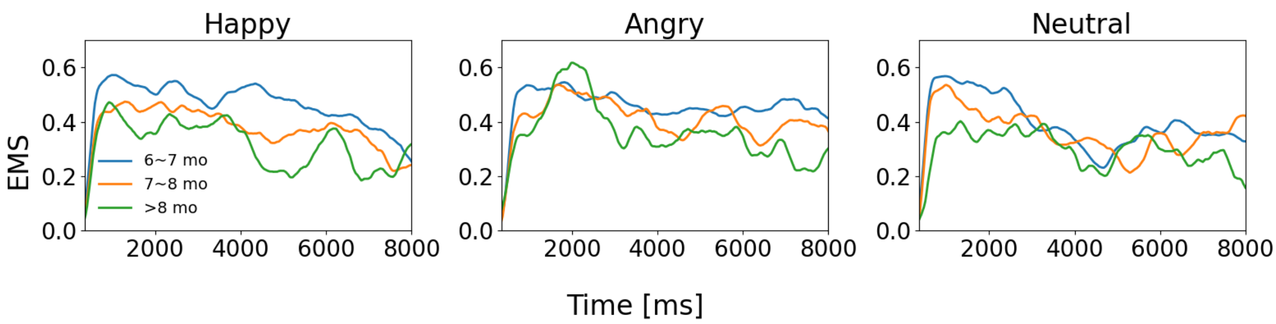


Figure 3.12 Time-Course of Proportion of Total Looking by Emotion and Age Group

Fraction-time curves of Emotion Matching Scores (EMS), expressed as proportion of target looking, plotted across the 8000 ms trial window for each age group (6–7, 7–8, and >8 months) and emotional condition (Happy, Angry, Neutral). In the Happy condition, 6-month-olds demonstrated the highest and most sustained EMS throughout the trial, followed by 7-month-olds, whereas >8-month-olds showed a marked decline with values approaching zero. In contrast, Angry condition curves overlapped across all age groups, while Neutral condition curves revealed only a mild decline with age.

3.3.5 Emotion Matching and Family Expressiveness

We then examined whether infants' emotion-matching performance was associated with the expressiveness of their family environment, as measured by FEQ. Both the total FEQ score and its subscales for negative expressiveness (NE) and positive expressiveness (PE) were considered.

Regression analyses revealed that total FEQ scores positively predicted infants' EMS in the Angry condition, F (1, 33) = 5.41, p = .026, accounting for approximately 14.1% of the variance (R^2 = .141). The standardized regression coefficient was β = 0.38, The unstandardized slope was B = .002, 95% CI [.000, .003]. Bootstrap estimates (1,000 samples) confirmed the robustness of this effect, yielding a similar confidence interval [.000, .003] and a significant two-tailed test (p = .025). A similar model for the Happy condition approached significance, F (1, 33) = 3.05, p = .090, R^2 =.085, The standardized regression coefficient was β = .29, and the unstandardized slope was B = .001, 95% CI [–.00001, .003], while the Neutral condition did not reach significance, F (1, 33) = 2.38, p = .132, R^2 = .067, the standardized regression coefficient was β = .26, with an unstandardized slope of B = .001, 95% CI [.000, .003]. Bootstrap estimates (1,000 samples) produced a similar confidence interval [.000, .003] and a non-significant two-tailed test (p = .113). As illustrated in Figure 3.13, regression slopes were positive across conditions, suggesting that higher overall family expressiveness was generally linked to greater EMS, although the association was most robust for angry stimuli.

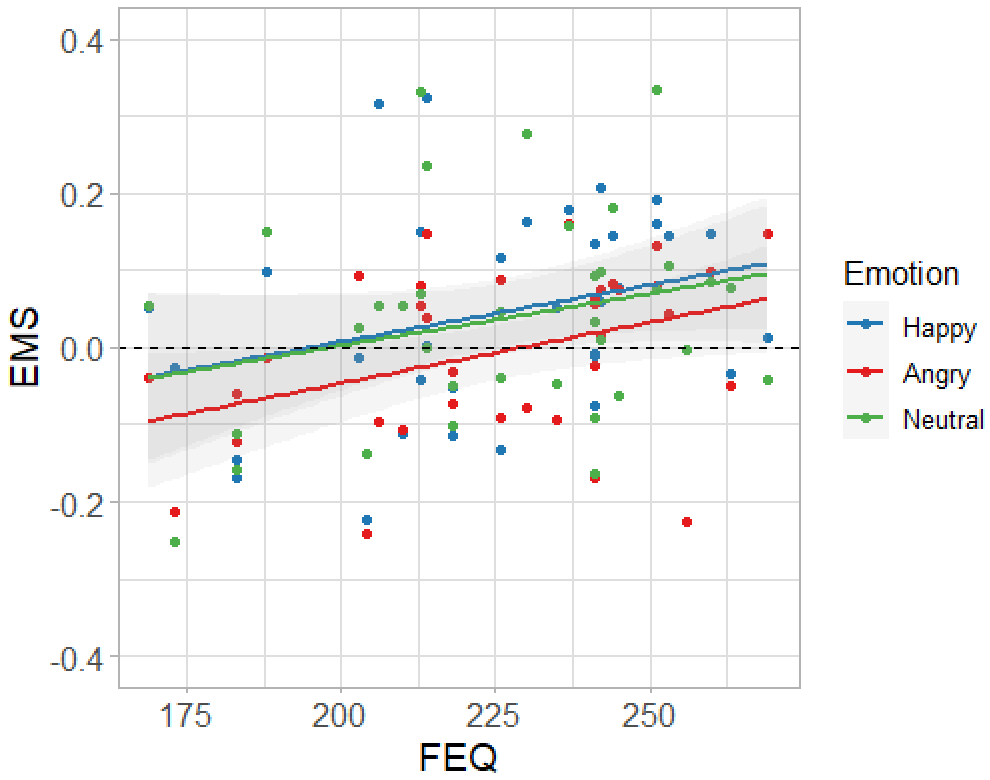


Figure 3.13 Scatterplots with regression lines showing the relation between FEQ total scores and infants' EMS across the three emotional conditions (Happy, Angry, Neutral).

Note. Each dot represents one infant, and shaded areas indicate 95% confidence intervals. Across conditions, higher family expressiveness was generally associated with higher EMS, with the association reaching significance in the Angry condition, approaching significance in the Happy condition, and not significant in the Neutral condition.

Correlation analysis further differentiated the contributions of NE and PE. For NE, significant positive associations emerged between family expressiveness and infants' EMS. At the global level ($N = 105$), NE scores were significantly correlated with overall EMS ($r = 0.33$, $p < .001$). At the condition level ($N = 35$ per condition), EMS was positively associated with NE in the Happy condition ($r = 0.33$, $p = .049$) and the Angry condition ($r = 0.36$, $p = .033$), with a marginal trend in the Neutral condition ($r = 0.32$, $p = .062$). See Figure 3.14. In contrast, PE was not significantly associated with EMS, either at the global level ($r = 0.10$, $p = .333$, $N = 105$) or within specific emotion conditions (Happy: $r = 0.07$, $p = .673$; Angry: $r = 0.20$, $p = .249$; Neutral: $r = 0.04$, $p = .833$; all $N = 35$). These findings indicate that infants' multimodal emotion matching is more strongly linked to negative family expressiveness than to positive expressiveness. The associations were particularly pronounced for angry stimuli, suggesting that frequent exposure to negative affect in the home may heighten infants' sensitivity to congruent affective cues, even for emotions that are less frequently expressed in typical caregiver–infant interactions.

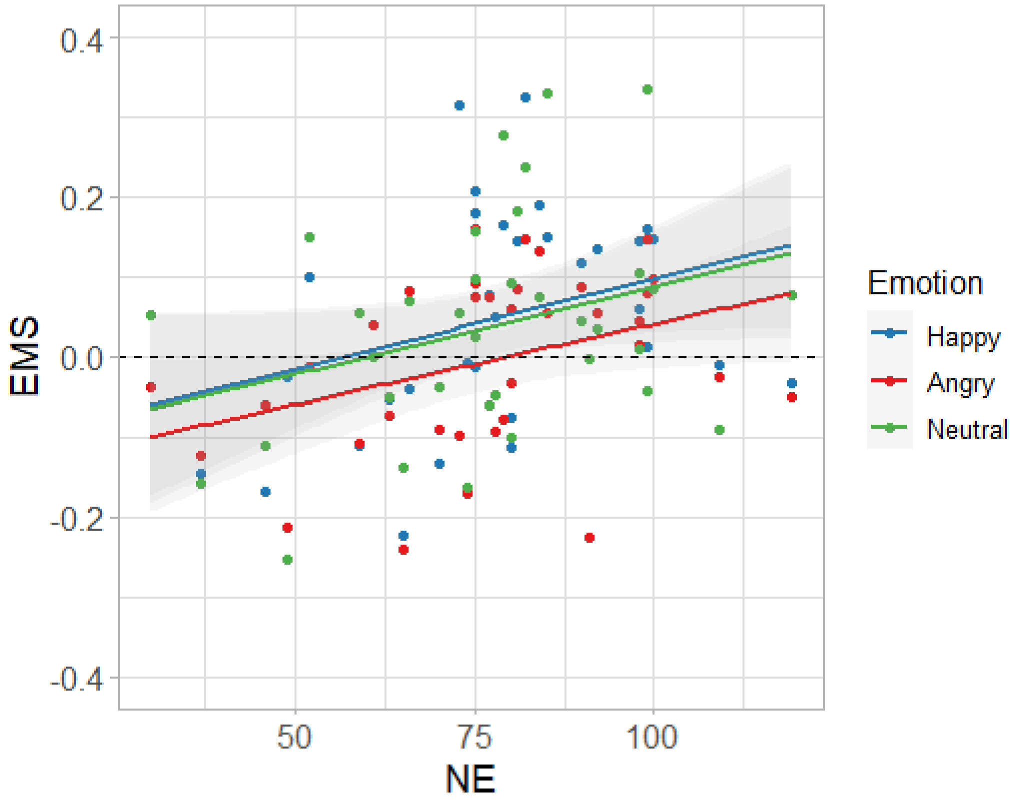


Figure 3.14 Scatterplots with regression lines showing the relation between FEQ negative subscale scores and infants' EMS across the three emotional conditions (Happy, Angry, Neutral).

Together, correlation analyses revealed positive associations between family expressiveness (total FE, PE, and NE) and infants' EMS, particularly in the Happy and Angry conditions (Figure 3.15). Infants from families with richer emotional expressiveness tended to show stronger matching performance, whereas associations with the Neutral condition were weaker.

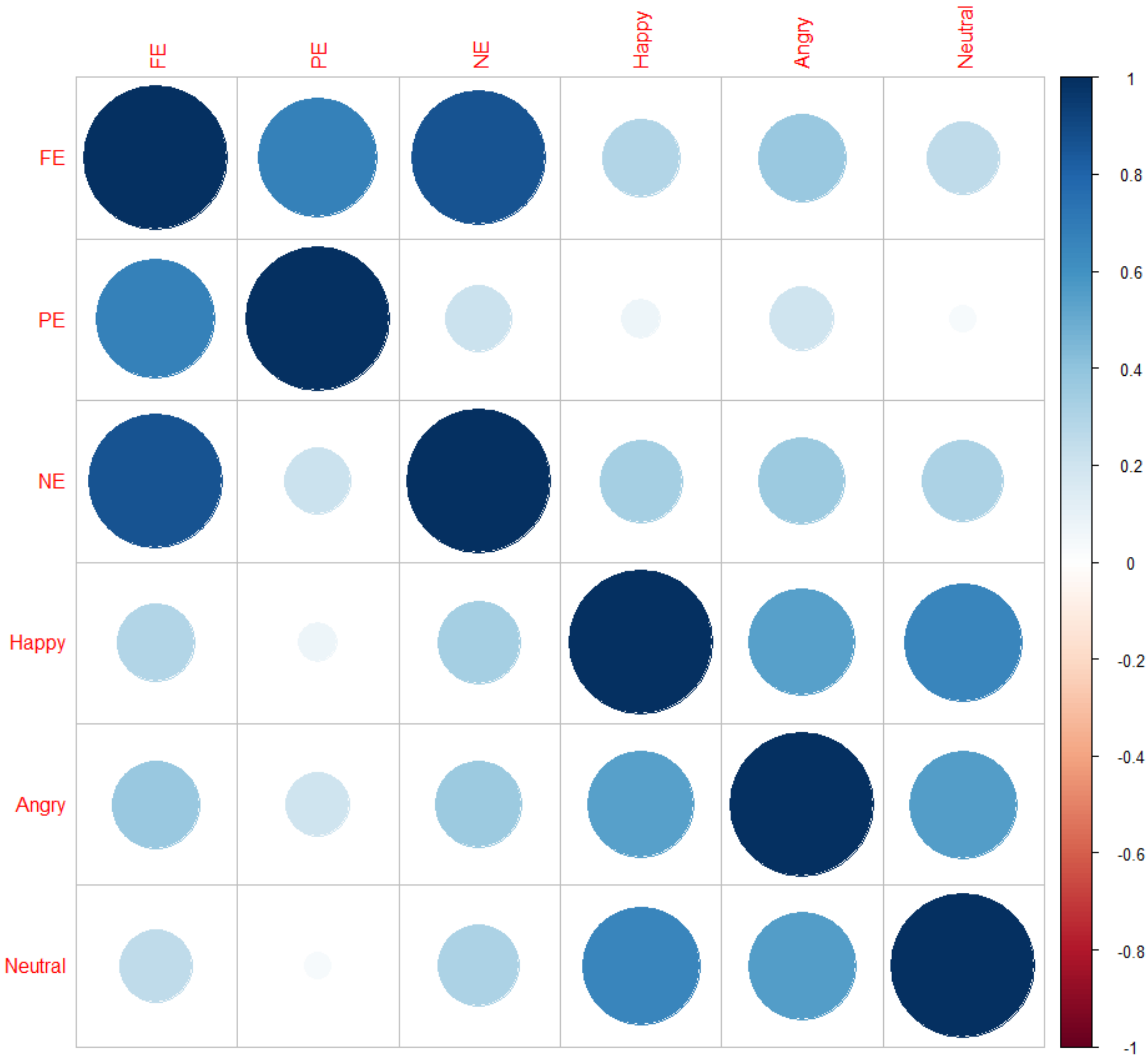


Figure 3.15 Correlation matrix between family expressiveness scores and infants' Emotion Matching Scores (EMS). FE = Family Expressiveness (total score), PE = Positive Expressiveness, NE = Negative Expressiveness. Circle size and color indicate the strength and direction of Pearson correlations (blue = positive, red = negative). Results show that family expressiveness (particularly FE, PE, and NE) was positively associated with infants' EMS in the Happy and Angry conditions, whereas associations with the Neutral condition were weaker.

3.3.6 Integrated Analyses Using Linear Mixed-effects Models

To examine the combined effects of multiple factors on infants' emotion matching while accounting for individual variability, we fitted a series of six nested linear mixed-effects models (LMMs), as detailed in the Methods section. Model selection was guided by likelihood ratio tests and model fit indices (AIC, BIC). The comparisons indicated that the model including both Emotion and Age as fixed effects (Model 2) provided the best fit to the data. This model revealed that both emotional condition and age significantly predicted emotion matching scores, whereas the inclusion of higher-order interactions or Gender did not significantly improve model fit. Thus, the selected model parsimoniously captured the main effects of Emotion and Age while controlling for individual-level variability in infants' performance.

The null model (m0), including only a random intercept for Subject, served as the baseline. Adding the Emotion condition (m1) significantly improved model fit compared to the null model, χ^2 (2) = 7.59, p = .023. The subsequent addition of Age (m2) provided a further, albeit marginal, improvement over m1, χ^2 (1) = 3.17, p = .075. Model 2 achieved the lowest AIC (–164.32) and a comparably favorable BIC (–148.39), indicating the best overall fit. More complex models, including Emotion × Age (m3) and additional terms for Gender (m4–m5), did not significantly improve model fit (all p > .07).

Analyses of fixed effects from Model 2 revealed that Emotion condition significantly predicted EMS, χ^2 (2) = 7.78, p = .020, while Age showed a marginal trend, χ^2 (1) = 3.13, p = .077. Relative to the Angry condition, infants showed significantly higher EMS in the Neutral condition (β = 0.043, SE = 0.020, p = .033, 95% CI [0.004, 0.081]) and in the Happy condition (β = 0.051, SE = 0.020, p = .011, 95% CI [0.013, 0.089]). By contrast, Age was negatively associated with EMS (β = –0.042, SE = 0.024, p = .086, 95% CI [–0.090, 0.006]), suggesting a marginal trend toward decreasing congruency sensitivity with increasing age.

Estimated marginal means (EMMs) further highlighted these effects. At mean-centered age, EMS was close to zero in the Angry condition (emmean = –0.006, 95% CI [–0.048, 0.036]) and in the Neutral condition (emmean = 0.037, 95% CI [–0.005, 0.079]), whereas it was significantly above zero in the Happy condition (emmean = 0.045, 95% CI [0.003, 0.087]). Pairwise comparisons with Bonferroni correction revealed that EMS in the Happy condition was significantly higher than in the Angry condition (estimate = 0.051, p = .034), while no other contrasts were significant. The overall slope for Age was negative (β = –0.042, SE = 0.024, 95% CI [–0.090, 0.006], p = .086), reinforcing the developmental decline in congruency sensitivity.

Model fit indices indicated that the fixed effects of emotion and age explained approximately 9% of the variance in EMS (marginal R^2 = 0.088), while the full model, including random effects for Subject, accounted for about 60% of the variance (conditional R^2 = 0.603). As illustrated in Figure 3.16, predicted EMS values derived from Model 2 were highest in the Happy condition, followed by Neutral, and lowest in Angry, with a modest decline as age increased.

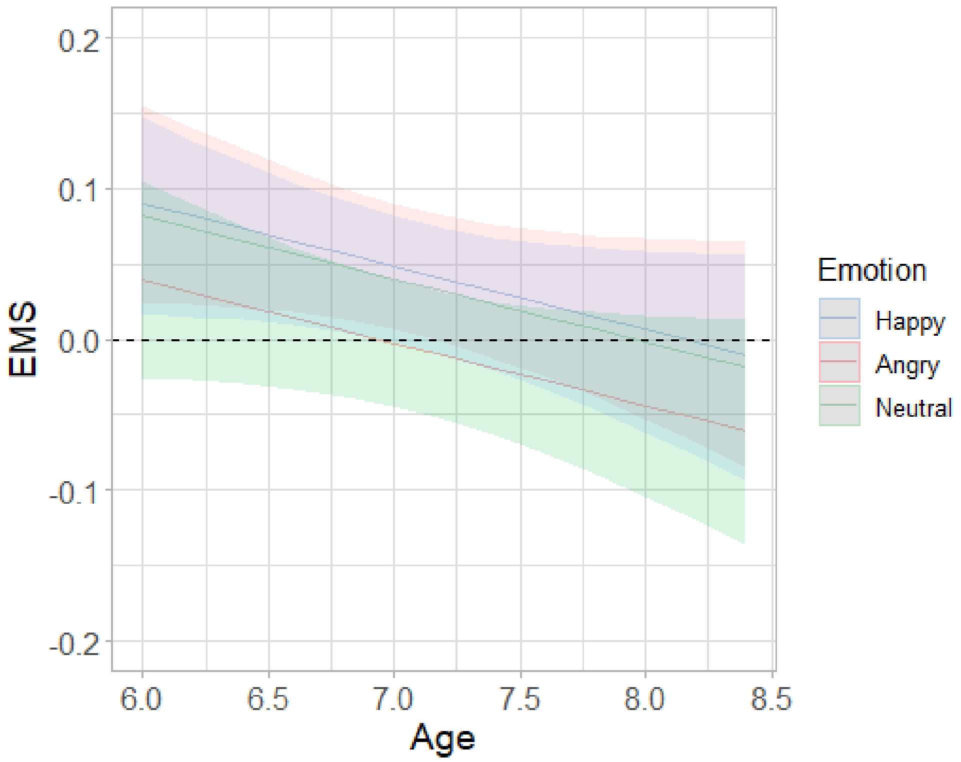


Figure 3.16 Predicted EMS from the best-fitting linear mixed-effects model (Model 2), which included fixed effects of emotion condition and age and a random intercept for Subject. Points represent estimated marginal means (EMMs) for each emotion condition (Happy, Neutral, Angry) at the mean-centered age, with error bars indicating 95% confidence intervals. Predicted values show that EMS was significantly higher in the Happy condition compared to the Angry condition, while Neutral did not differ significantly from either. A modest negative slope for Age was also observed, indicating a trend toward decreasing EMS with increasing age.

3.4 Discussion

The present study employed an intermodal matching paradigm using dynamic audiovisual stimuli to investigate 6- to 8-month-old infants' ability to recognize and match emotions, with particular attention to the potential contribution of age and family expressiveness. Although infants did not show statistically significant congruency sensitivity for angry or neutral expression conditions at the group level, they demonstrated a clear and significant tendency to match emotional expressions in the happy condition. Moreover, overall FEQ scores were positively associated with infants' EMS value, with this correlation being particularly pronounced in the angry condition. These findings provide an important insight into the development of emotion perception during the critical window of 6-8 months, highlighting how both developmental stage and family expressiveness contribute to infants' emerging socio-emotional competencies. More broadly, this study contributes to our understanding of how maturational factors and the family environment jointly influence one's early socio-emotional development. In the following sections, we discuss these findings in detail with the perspective on (1) infants' emotion-matching performance, (2) the relation between emotion matching and age, (3) the association between emotion matching and family expressiveness, and (4) the interplay of developmental and family environmental factors.

3.4.1 Emotion Matching

Although infants as a group did not show robust intermodal congruency sensitivity across all emotional conditions, they demonstrated a clear advantage for happy emotions. When facial expressions were happy, infants tended to allocate greater attention to these stimuli, with their EMS value generally exceeding zero. Moreover, task performance in the Happy condition was significantly superior to that in the Angry condition. This pattern suggests that, even in the absence of a consistent congruency effect across emotions, infants may still display an early preference toward positive affective information, with happiness occupying a privileged position in the developmental trajectory of emotion processing. This interpretation is consistent with the "positive preference" frequently reported in previous research (Farroni, Menon, et al., 2007a; Flom & Whiteley, 2014). Importantly, the intermodal redundancy hypothesis offers a complementary account: when visual and auditory channels synchronously convey happy cues, redundant information increases salience and facilitates infants' ability to match affective signals under the Happy condition (Bahrick et al., 2004; Bahrick & Lickliter, 2000).

Several theoretical models have been proposed to explain infants' early advantage in perceiving positive emotions. From an evolutionary perspective, happy expressions often signal safety, attachment, and the availability of care during caregiver–infant interactions in early infancy, enabling infants to adjust their behavior accordingly and adapt more effectively to the external environment (Farroni et al., 2007; Leppänen & Hietanen, 2004). From a socialization perspective, caregivers tend to display positive emotions more frequently in everyday interactions, providing infants with extensive exposure to the statistical regularities of happy facial and vocal cues. Such exposure may enhance infants' sensitivity to positive affective information (Eisenberg et al., 1998). At the perceptual level, happy expressions exhibit distinctive features, such as upturned mouth corners and narrowed eyes, that increase visual contrast and facilitate recognition (Jessen et al., 2016). In summary, these perspectives help explain why positive emotions, particularly happiness, tend to be processed earlier and more reliably than negative emotions in infancy, and why happiness is likely to represent the first emotional category that infants can consistently recognize and integrate across modalities.

In contrast, no congruency effect was observed for the angry condition, which diverges from previous findings, suggesting that infants around seven months of age begin to exhibit heightened sensitivity to threat-related emotions (Leppänen & Nelson, 2009; Soken & Pick, 1999). One possible explanation is that infants in this study did not integrate anger cues exclusively under congruent conditions but rather distributed their attention across both congruent and incongruent angry faces, resulting in EMS values near zero. Another possibility is that although infants may perceptually discriminate angry expressions, the ability to flexibly use congruency as a cue is still developing at this age. This interpretation highlights the importance of considering how different measurement approaches capture infants' attentional allocation strategies. Thus, the absence of a congruency effect in emotion-matching paradigms should not be taken as evidence of insufficient sensitivity to negative emotions.

Similarly, no congruency effect emerged in the neutral condition, which is not unexpected, as neutral faces lack distinctive affective features and may be perceived by infants as ambiguous or baseline stimuli. The relatively modest and non-significant positive EMS values observed here suggest that infants occasionally attended to neutral congruent pairings, but not systematically or reliably. This finding reinforces the view that early intermodal integration is emotion-specific, with happiness emerging first, followed by more gradual development of sensitivity to negative and ambiguous expressions.

It is also important to note that the congruency effect for happiness was only marginally statistically significant at the group level, and the variance explained by the emotion condition remains relatively limited. These limitations, together with the relatively small sample size, suggest that the present findings should be interpreted cautiously. Meanwhile, the consistent advantage for happiness, coupled with the absence of reliable effects for angry and neutral conditions, provides converging evidence that the development of intermodal emotion matching ability is not uniform but shaped by both the valence and intensity of emotional cues. Future research with larger samples and more sensitive measures (e.g., EEG, fNIRS) may help determine whether the trends observed here reflect genuine developmental constraints or methodological limitations. Furthermore, given that family emotional expressiveness has been shown to predict individual differences in infants' performance, incorporating environmental variables into future designs will be crucial for understanding how individual development and context jointly shape the emergence of multimodal emotion processing.

3.4.2 Emotion Matching and Age Group

The present study investigated infants' intermodal emotion-matching abilities and their relation to age during the critical window between 6 and 8 months. Our results suggested that younger infants, particularly those around 6 months, tended to show stronger and more sustained matching responses to happy faces, with EMS remaining positive across the trial. In contrast, 7-month-olds exhibited more variable trajectories, and infants older than 8 months appeared to show attenuated performance, with EMS values approaching zero toward the end of the trial. These patterns may suggest that sensitivity to congruent positive cues is relatively more pronounced in younger infants and diminishes with increasing age. This pattern was statistically confirmed by a significant negative association between age and EMS in the Happy condition, suggesting that sensitivity to positive emotional congruence is most robust early in this developmental window. By contrast, no reliable age effects emerged for Neutral or Angry expressions, with performance remaining relatively stable across age groups.

These results extend previous reports of a positivity bias in infancy, whereby positive emotions, especially happiness, are preferentially processed and integrated across modalities (Flom & Bahrick, 2007; Walker-Andrews, 1997). The age-related decline in matching for happy expressions may reflect a developmental reorganization of attentional strategies: while younger infants exhibit heightened sensitivity to salient positive cues, older infants may begin to allocate attention more broadly across multiple emotional and contextual signals, thereby reducing their selective bias for happiness.

One likely explanation for the lack of significant age effects for Neutral and Angry conditions is that neutral and negative cues may require additional social experience and contextual learning before they elicit systematic developmental differences. Also, the present sample size may limit statistical power, as suggested by the relatively low observed power values. Additionally, variability in infants' responses across trials may have obscured systematic age-related effects. Therefore, although the visual trend is broadly consistent with previous findings emphasizing the early salience of positive emotions in infancy, the current results should be interpreted with caution and regarded as preliminary. Future studies with larger samples and longitudinal approaches will be necessary to determine whether the observed decline in matching for happy expressions reflects a genuine developmental trajectory or sample-specific variability.

Several plausible explanations may account for this finding. In early infancy, discrimination and matching of multimodal emotions may rely strongly on bottom-up perceptual mechanisms and the availability of cross-modal redundancy. Together with infants' extensive exposure to happy expressions in daily caregiver interactions, these factors likely contribute to heightened sensitivity to happiness at 6 months (Eisenberg et al., 1998). By 7-8 months, infants' attentional strategies appear to undergo some shifts. Leppänen et al. (2009) suggested that infant brains are sensitive to positive emotions during the first six months, while after seven months they demonstrate stronger processing of threatening emotions such as anger and fear. Analysis of temporal trajectories from the present study revealed that older infants disengaged more quickly from congruent emotional cues, particularly happy expressions. This shift may reflect a transitional decline in performance as infants' cognitive and attentional systems become more flexible, reallocating resources toward broader exploratory behaviors and more complex socio-emotional processing. Alternatively, the reduced focus on congruent happy faces may indicate that infants at this age are beginning to distribute attention more evenly across multiple social signals, resulting in a decline in robust positive-bias effects.

Similarly, the absence of age-related changes in the angry and neutral conditions may suggest the emotion-specificity of these developmental patterns. Angry expressions, which often signal threat or social conflict, may require more prolonged exposure and neural maturation before infants can integrate them reliably across modalities (Leppänen & Nelson, 2009; Nelson & De

Haan, 1996). By contrast, neutral expressions, which lack salient affective features, may be processed as ambiguous or baseline cues, therefore reducing the likelihood of systematic age-related effects (Serrano et al., 1992). Taken together, these results are consistent with the view that the first half of the first year represents a critical period in which infants' perceptual advantage for positive affect consolidates, while their sensitivity to negative or ambiguous affective information remains relatively less developed (Grossmann, 2010). Moreover, the developmental trajectory observed here highlights the importance of examining temporal dynamics in emotion matching. Proportional looking time analyses revealed that the differences between age groups were most apparent in how attention unfolded across the trial. Six-month-olds not only showed higher overall EMS but also sustained their preference for congruent happy faces across time, whereas older infants' trajectories were more transient and less stable. Such findings emphasize that overall performance indices may obscure meaningful age-related differences in attentional allocation. Future research would benefit from combining longitudinal designs with fine-grained temporal analyses to capture the dynamic nature of emotion processing and to determine whether the observed decline represents a transient fluctuation or part of a broader developmental trajectory.

Based on the present findings, age can be considered a potential predictor of infants' cross-modal emotion matching, as it showed a statistical trend in predicting overall EMS. It may account for a meaningful proportion of variance, suggesting that chronological maturation may be one factor driving individual differences. Nevertheless, developmental outcomes are unlikely to be shaped by age alone. The quality and frequency of emotional input within the family environment are also likely to play a critical role in shaping infants' performance. We investigated this possibility below using FEQ-based measures.

3.4.3 Emotion Matching and Family Expressiveness

By analyzing infants' emotion matching performance alongside family emotional expressiveness, the present study demonstrates that infants raised in families with richer emotional expression are more likely to match facial and vocal cues. This association was particularly pronounced in the angry condition, with a weaker yet noticeable trend in the happy condition, whereas no reliable relationship was observed in the neutral condition. These findings suggest that the family emotional environment, as measured in this study by the Family Expressiveness Questionnaire, contributes to shaping infants' sensitivity to cross-modal emotional congruency, especially for emotions of greater intensity or social salience, such as happy or angry conditions.

More fine-grained analyses revealed that the observed effects were driven primarily by negative expressiveness (NE) rather than positive expressiveness (PE). Families characterized by frequent displays of anger, irritation, or other negative emotions may provide infants with a wider range of affective input, which in turn strengthens their ability to integrate negative emotional cues across modalities (Pollak et al., 2009). By contrast, positive expressiveness showed no significant association with infants' performance. This asymmetry likely reflects the fact that positive emotions such as happiness are pervasive in caregiver–infant interactions and may already be processed robustly due to biological predispositions, thereby reducing the incremental role of PE. Whereas negative emotions occur less frequently in infancy and may therefore rely more heavily on environmental input to consolidate infants' cross-modal matching skills.

The observed association from the present study can be interpreted in terms of emotion socialization processes. Families high in negative expressiveness may provide infants with richer and more frequent input about negative affective states, thereby enhancing their opportunities to learn the perceptual features and contingencies of such emotions. Over time, this repeated exposure may also serve adaptive functions by alerting infants to potential threats in their environment. The finding that NE, but not PE, predicted infants' performance is thus consistent with the notion that early socioemotional development reflects both developmental stage and experiential context. At the same time, although family emotional expressiveness explained a meaningful proportion of the variance in infants' performance, the effect sizes were modest. This suggests that other unmeasured factors, such as infants' individual temperament, broader cultural norms, or parent–infant interaction quality, also should be considered as factors contributing to variability in their emotion matching ability. Furthermore, the use of parent-reported questionnaires introduces potential biases in assessing family expressiveness. Future studies could therefore adopt longitudinal and multi-method approaches, incorporating naturalistic observation or video-based coding of parent–infant interaction to provide a more objective account of emotional input in the family expressiveness environment.

Overall, these findings suggest the important role of family expressiveness in influencing infants' emerging ability to integrate affective information across sensory modalities. Whereas positive emotion processing appears to be supported by both innate and ubiquitous social cues, the integration of negative emotional information is more malleable and strongly shaped by the expressiveness of the family environment. To comprehensively evaluate how developmental factors (age) and environmental factors (emotion types) jointly influence infants' intermodal emotion matching abilities, the following section turns to statistical modelling approaches that allow simultaneous examination of these influences.

3.4.4 Linear Mixed-Effects Model

As described in the Methods section, six nested linear mixed-effects models were compared. The results indicated that the model including the additive effects of emotional condition and age provided the best fit for infants' intermodal emotion matching performance, whereas the addition of higher-order interaction terms did not significantly improve model fit. This suggests that both emotion type and developmental stage independently influence infants' performance in multimodal emotion-matching tasks, with no strong evidence that age moderates the effect of specific emotional conditions.

From a methodological standpoint, the use of linear mixed-effects models offers distinct advantages for developmental research. By simultaneously accounting for within- and between-subject variability, these models provide a more robust framework for testing hypotheses about early emotion processing than traditional repeated-measures ANOVA. The present findings, therefore, highlight that emotion-matching performance in 6- to 8-month-old infants is shaped by contextual factors (emotional condition) and developmental factors (age), which appear to operate largely in parallel rather than interactively at this stage.

The predominance of additive rather than interactive effects supports two conclusions. First, happy expressions emerge earlier and with greater stability in intermodal emotion perception than other emotions. Second, developmental changes observed between 6 and 8 months appear progressive and generalizable rather than context dependent. Future research with larger samples and longitudinal designs may be better suited to detect interaction effects, particularly in examining how age and environmental factors, such as family expressiveness, jointly contribute to the development of emotion processing mechanisms. More broadly, the application of mixed-effects modelling represents a promising methodological approach for capturing the complexity of socioemotional trajectories in infancy.

The integrated analyses further confirmed that infants' emotion-matching performance was influenced by emotional condition and showed a marginal trend with age. Specifically, congruency sensitivity was significantly higher for Happy and Neutral expressions than for Angry expressions, consistent with the well-documented positivity bias in early infancy, whereby positive emotions such as happiness are preferentially attended to and more readily integrated by infants (Farroni et al., 2007; Grossmann, 2010). The marginal negative effect of age suggests that emotion matching performance may decline slightly with increasing age during the 6- to 8-month period. One possible explanation is that developmental changes in attentional strategies or emerging cognitive control may temporarily reduce sensitivity to congruent affective cues as infants transition toward more complex social-cognitive abilities (Lewkowicz & Hansen-Tift, 2012; Nelson & De Haan, 1996). However, these patterns should be interpreted cautiously. The age effect was only marginal, and much of the variability in performance reflected individual differences that were not explained by emotion condition or age. In addition, the modest sample size limits the ability to detect subtle developmental changes, raising the possibility that the apparent decline reflects short-term fluctuation rather than a robust developmental trajectory.

Future research with larger and more finely graded age cohorts would be necessary to determine whether the apparent decline in congruency sensitivity reflects a genuine developmental trajectory or a temporary shift within the second half of the first year. Incorporating longitudinal designs and alternative approaches, such as growth curve modelling, may provide a clearer picture of how age-related changes in multimodal emotion processing unfold across infancy.

3.5 Conclusion

In conclusion, the present study demonstrates that infants between 6 and 8 months of age show emerging sensitivity to intermodal emotional congruency, with a distinct preference for positive emotions. Six-month-olds showed stronger and more sustained matching responses to happy faces, highlighting the early recognition of positive affect and the progressive, bias-specific development of emotion-processing abilities. Moreover, family expressiveness was positively associated with infants' matching performance, especially in the angry condition, suggesting the formative role of socialization in shaping early socio-emotional abilities. Finally, model comparison analyses further indicated that emotional condition and age independently contributed to infants' emotion matching performance, whereas higher-order interactions did not provide additional explanatory power.

These findings make several contributions. Theoretically, they extend evidence for the early emergence of the positivity bias and reveal how developmental stage and family environment inform the developmental trajectory of emotion processing. Methodologically, they highlight the value of emotion matching scores and linear mixed-effects models as sensitive tools for capturing subtle individual differences in early socio-emotional development. Practically, the results suggest that variations in the emotional environment of the family may influence infants' capacity to process affective information, offering potential implications for early monitoring and intervention.

Several limitations should be acknowledged. The modest sample size reduced statistical power, and some effects only approached significance. In addition, family expressiveness was assessed through parental self-report, which may introduce reporting biases. Future research should adopt larger, longitudinal samples and incorporate observational measures of emotional communication in the home to clarify the causal pathways linking social environments and early emotion processing.

Taken together, this study advances our understanding of how infants begin to integrate affective information across sensory modalities, demonstrating that emotion processing in the first year is shaped not only by biological maturation but also by the quality of social experience. These findings highlight the importance of considering both developmental timing and environmental context in uncovering the origins of human socio-emotional cognition.

CHAPTER IV.

NEURAL SENSITIVITY TO MULTIMODAL EMOTIONAL CONGRUENCY IN INFANCY: AN fNIRS STUDY

Building on these behavioral findings, the next critical step is to examine the neural mechanisms underlying infants' sensitivity to multimodal emotional congruency. Our previous work demonstrated that by 6 to 8 months of age, infants not only show a tendency to detect congruency between facial and vocal expressions but also display a stronger preference for happy expressions. While these results highlight the emergence of emotion-matching abilities and a positivity bias at the behavioral level, they do not reveal how the infant brain differentiates between congruent and incongruent multimodal affective information.

To address this issue, the present study employs functional near-infrared spectroscopy (fNIRS) to investigate neural responses to multimodal emotional congruency in infants aged 6–7 months. Previous behavioral evidence has revealed a decline in intermodal emotion matching performance with increasing age, suggesting a developmental transition during this period. Therefore, by targeting infants within this specific age window, we aim to identify the neural correlations underlying the multimodal emotional congruency. Infants are presented with dynamic audiovisual stimuli in congruent (e.g., happy face + happy voice) and incongruent (e.g., happy face + angry voice) conditions while brain activation is recorded from bilateral temporal and prefrontal regions. Data is currently being collected, and analyses will focus on comparing oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) responses between conditions. This study aims to provide neurophysiological evidence that complements behavioral observations and advances our understanding of the early development of multimodal emotion processing.

4.1 Background and Rationale

Building on our behavioral findings, infants showed a tendency to exhibit intermodal congruency sensitivity for positive emotions, whereas such sensitivity did not generalize uniformly across all emotion conditions. Moreover, intermodal emotion matching performance appeared to decline with increasing age, suggesting a developmental shift that is not readily interpretable from behavior alone. Behavioral measures cannot determine whether this shift reflects altered neural sensitivity to multimodal congruency or, as we previously discussed, it reflects a reorganization of attentional strategies. Accordingly, a neural-level approach is needed to reveal the processing of multimodal emotional information that may remain covert at the behavioral level.

From a theoretical perspective, the cross-modal emotion matching theories proposed that by around 6–7 months of age, infants can integrate facial and vocal expressions based on shared emotional meaning rather than temporal alignment alone (Walker-Andrews, 1997). This ability reflects a development toward more abstract forms of multimodal integration, moving beyond reliance on temporal synchrony. By around 7 months of age, infants begin to form categorical representations of emotions (Soken & Pick, 1992). These emerging categories enable infants to detect emotional congruency sensitivity across modalities. Such multimodal emotion integration serves as an important social foundation: it allows infants to engage more effectively with their environment by rapidly and accurately interpreting others' emotional intentions, thereby guiding their own behavior (Campos et al., 1994; Klinnert et al., 1986). Our systematic review further indicates that a substantial proportion of infant multimodal emotion processing experiments (30/47) report sensitivity to multimodal congruency, supporting the view that, during the first year of life, infants develop the ability to detect common affect across bimodal, dynamic presentations of an emotional expression (Flom & Whiteley, 2014; A. R. Heck et al., 2016; Palama et al., 2018). This theoretical and practical framework provides a basis for formulating neural level predictions: if infants are sensitive to intermodal congruency, congruent facial-vocal expression pairings should elicit stronger hemodynamic responses than incongruent pairings. Complementary accounts also make discriminable predictions: prediction-error models would anticipate enhanced responses to unexpected (incongruent) pairings (Friston, 2005; Rothermund et al., 2011), whereas numerous empirical studies provide evidence that anticipate greater engagement for positively congruent cues (Flom & Whiteley, 2014; Kahana-Kalman & Walker-Andrews, 2001; Walker, 1982). Testing these competing patterns at the neural level provides a stringent evaluation of the underlying mechanisms.

Concurrently, the 6–7-month period marks rapid specialization within the infant social brain (Grossman, 2013; Johnson et al., 2009). Neuroimaging research has provided additional insights into the neural basis of early emotion perception. ERP research indicates that occipito-temporal N290/P400 components are sensitive to infants' processing of facial emotions, and they can be modulated by face–voice emotional congruency (De Haan et al., 2013; Grossmann et al., 2006). Infant fNIRS studies further revealed that temporal regions, including the superior temporal sulcus/gyrus (STS/STG), support audiovisual integration and voice-sensitive processing (Belin et al., 2002; Blasi et al., 2011; Grandjean et al., 2005), while ventromedial and orbitofrontal prefrontal areas (vmPFC/OFC) contribute to affective evaluation and social motivation (Di Lorenzo et al., 2019; Ravicz et al., 2015), the prefrontal cortex is recognized as a key node of the "social brain network" (Adolphs, 2009; Frith & Frith, 2007; Grossmann & Johnson, 2007). Moreover, neurophysiological studies suggest that synchronized audiovisual input enhances cortical specialization for social perception, with evidence of increased activation in temporal and prefrontal regions when infants are presented with redundant multimodal affective information (Grossmann, 2010; Hyde et al., 2011). If congruency sensitivity emerges during this period, we expect regional specificity, particularly stronger congruent-than-incongruent responses in temporal regions and congruency- or valence-modulated effects in vmPFC/OFC (Blasi et al., 2011; Xie et al., 2019).

Methodologically, fNIRS is well suited to infant research, the development of multi-channel fNIRS has enabled non-invasive measurement of localized hemodynamic responses in infants (Gervain et al., 2023) and enables the dissociation of prefrontal and temporal contributions under ecologically valid audiovisual stimulation. We use the alternating vs. non-alternating design, whereby non-alternating blocks only contain stimuli from one condition (only congruent or only incongruent), whereas alternating blocks contain stimuli from both conditions in strict alternation. The alternating vs. non-alternating paradigm has been successfully used in infant NIRS studies to prove fine-grained discrimination (Gervain et al., 2012; Sato et al., 2010). Our alternating versus non-alternating block design further separates effects of stable congruent or incongruent perception in the non-alternating blocks from switch-related responses in the alternating blocks, permitting a joint test of congruency advantages, habituation dynamics, and prediction-error signals. Critically, fNIRS may reveal neural differentiation even when behavioral measures appear insensitive. For instance, in our behavioral study, we did not observe clear differences between congruent and incongruent trials when the target facial expression was angry. This absence of behavioral effects may conceal underlying neural distinctions, reflecting different processing mechanisms, as has been found in some fNIRS studies (Martinez-Alvarez et al., 2023). By examining such cases at the neural level, fNIRS can help clarify whether different performance in the emotion matching task reflects a genuine reduction in integrative capacity or a developmental reorganization of attentional strategies. On this basis, we propose the following aims and hypotheses.

- Do infants' cortical responses differentiate between congruent and incongruent face–voice pairings? If so, are responses stronger for congruent stimuli?
- Does stable cross-modal redundancy enhance infants' neural sensitivity to emotional congruency? Specifically, is there more activation in the non-alternating than in the alternating blocks?
- Which regions of the infant social brain are engaged by multimodal emotional congruency? Are effects strongest in temporal areas and modulated by valence in vmPFC/OFC?
- How are neural congruency sensitivities linked to behavior and environment? Are they associated with infants' behavioral performance and family expressiveness?

Based on prior theoretical and empirical evidence, we mainly formulate the following hypotheses:

- We hypothesize greater responses (more HbO increase and HbR decrease) for congruent than incongruent stimuli in temporal and prefrontal ROIs. Alternatively, if incongruent stimuli elicit stronger responses, this would support prediction-error models.
- We predict more activation in the non-alternating than in the alternating blocks.
- We anticipate temporal regions involvement (STS/STG) and congruency- or valence-modulated responses in vmPFC/OFC.
- By comparing neural sensitivity with infants' behavior performance and family expressiveness experiment, we predict positive associations between neural congruency indices and behavior performance, moderated by family expressiveness.

4.2 Methods

4.2.1 Participants

A total of 45 healthy, full-term infants aged between 6 and 7 months will be recruited from the BabyLab at the University of Padua. All infants will be born at term (37–42 weeks of gestation) following uncomplicated deliveries, will have no history of neurological or developmental disorders, and will be from Italian-speaking families. Written informed consent was obtained from parents before participation.

4.2.2 Materials

The experimental stimuli consisted of dynamic audiovisual clips combining facial and vocal emotional expressions. The auditory stimuli were recorded by two native Italian speakers, who produced short two-second sentences in either happy or angry emotional tones. A set of semantically neutral sentences was selected to allow for positive, neutral, or negative valence to be carried exclusively by prosody while preserving natural prosodic features. Here are some examples: “Sei piccolo” (“You are little”), “Aspettami” (“Wait for me”), and “Chi sei” (“Who are you”). Speakers were instructed to articulate each sentence with prosody clearly expressing the target emotions. All audio files were edited using Ocenaudio and Praat to standardize amplitude, duration, intensity, and background noise.

The visual component was recorded by the same speakers, who displayed the target emotions through facial expressions while reading the same scripted short sentences. Recordings captured the upper body against a neutral, non-white background. The speakers maintained direct gaze toward the camera and were allowed only minimal natural head movements to ensure ecological validity while controlling extraneous motion. Audio and video were recorded separately and subsequently synchronized using the OpenShot video editor. The final stimuli presented combinations of emotional facial expressions and vocal tones. In the congruent condition, facial and vocal expressions conveyed the same emotion (e.g., a happy face paired with a happy voice), whereas in the incongruent condition, the expressions conveyed different emotions (e.g., a happy face paired with an angry voice). A stimulus example is shown below (Figure 4.1).

Figure 4.1 Example of visual stimuli. The figure shows one of the models displaying a happy facial expression.

The audiovisual stimuli were adapted from those validated and acoustically analyzed in Study 2. In addition, we administered the same Italian-translated version of the Family Expressiveness Questionnaire (FEQ) as in Study 2 to assess the family emotional environment.

4.2.3 Procedure

The experiment was programmed using E-Prime 2.0 software and conducted in a sound-attenuated, dimly lit room. Infants were seated on their caregiver's lap approximately 60 cm from the monitor. Caregivers were instructed to maintain a neutral demeanor and not to influence the infant's behavior. An experimenter monitored the session in real time to ensure the infant's attentiveness, and testing was terminated if the infant became excessively fussy or lost sustained attention.

Visual stimuli were presented on a 70 × 40 cm monitor (resolution: 1280 × 720 pixels). Identical loudspeakers were positioned 10 cm to the left and right of the monitor, delivering auditory stimuli at a comfortable intensity of approximately 70 dB, ensuring directional consistency and clear sound presentation.

The experiment followed an alternating versus non-alternating block design adapted for infant neuroimaging research. Each sequence began with a baseline phase of 22–25 seconds, during which a neutral face was presented silently. This was followed by a stimulation block lasting

16–18 seconds, composed of eight 2-second trials of multimodal emotional stimuli. In non-alternating blocks, all trials were congruent (happy face with happy voice) or all incongruent (happy face with angry voice). In alternating blocks, congruent and incongruent pairings occurred in sequence. The total task lasts approximately 12 minutes. An overview of the procedure is shown in Figure 4.2.

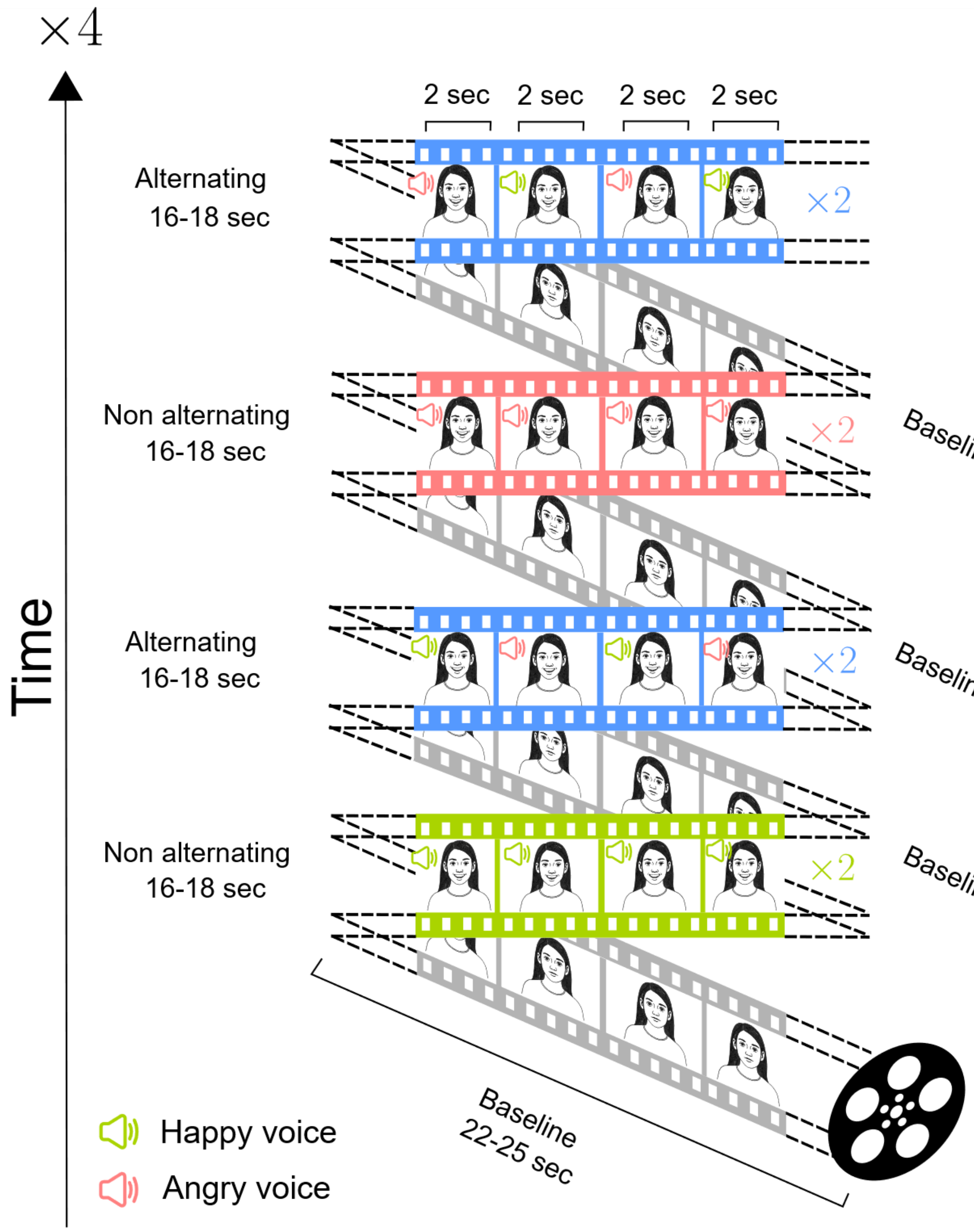


Figure 4.2 Schematic illustration of the experimental paradigm.

Note. Each sequence started with a baseline phase (22–25 s) presenting a silent neutral face, followed by a stimulation block (16–18 s) consisting of eight 2-s trials of multimodal emotional stimuli. In non-alternating blocks, all trials were either congruent (happy face with happy voice) or incongruent (happy face with angry voice). In alternating blocks, congruent and incongruent pairings alternated within the block.

4.2.4 Data Recording

4.2.4.1 Video Recording

Infants' visual fixations on the emotional stimuli were recorded with a video camera. While both stimulus presentation and infant behavior were simultaneously captured using OBS Studio (Open Broadcaster Software, Version 30.1.2). All video recordings were coded using Datavyu software. Infant visual fixation was analyzed on a frame-by-frame basis, with each coding unit corresponding to 30 frames. Two trained researchers independently coded infants' fixation time for each trial; coding discrepancies were jointly reviewed and resolved through discussion, with inconsistent segments re-coded when necessary.

4.2.4.2 fNIRS Recording

fNIRS data were acquired using the NIRSport2 system (NIRx Medical Technologies, Berlin, Germany) controlled by Aurora acquisition software. The system employs dual-wavelength light sources (760 and 850 nm), and data were sampled at 7.81 Hz. Infants wore an EasyCap, into which the optodes were inserted. A total of 12 sources (6 per hemisphere) and 14 detectors (7 per hemisphere) were mounted on the cap, yielding 38 measurement channels with an inter-optode distance of approximately 30 mm. The optode montage was designed to cover the bilateral temporal regions and the prefrontal cortex, including the STS, STG, vmPFC, and OFC. Figure 4.3 presents a montage. Signal quality was checked before each session to ensure optimal optode-scalp coupling.

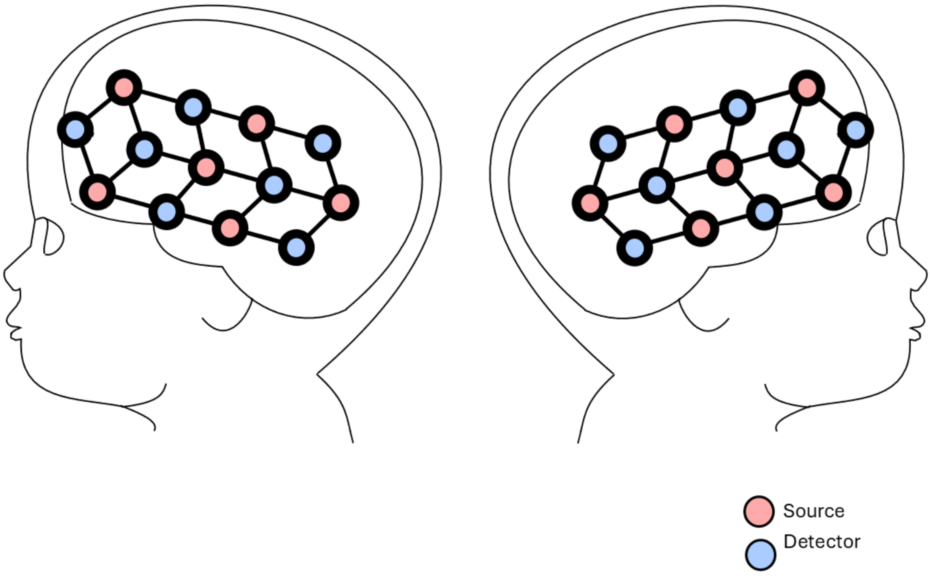


Figure 4.3 Schematic illustration of the fNIRS optode array. The montage consisted of 12 sources (red) and 14 detectors (blue), resulting in 38 source–detector channels with 30-mm separation. Optodes were positioned bilaterally over the temporal and prefrontal cortices following the 10–5 system.

4.2.5 Data Analysis

4.2.5.1 Preprocessing

fNIRS light intensity measures will first be converted to optical densities. Optical densities will then be converted to oxygenated (HbO) and deoxygenated (HbR) concentration changes, using the modified Beer–Lambert Law, using the following absorption coefficients (μa, mM−1 × mm−1): μa (HbO, 760 nm) = 0.1496, μa (HbO, 850 nm) = 0.2526, μa (HbR, 760 nm) = 0.3865 and μa (HbR, 850 nm) = 0.1798. The product of the optical pathlength and the differential pathlength factor will be set to 1, so that the resulting concentration changes will be expressed in mM x mm. Data will then be band-pass filtered using a digital fft (fast Fourier transform) filter, between 0.01-0.7 Hz. Finally, a block in a given channel will be rejected if the light intensity reaches the saturation value (1.2 V), if the block contains a motion artifact, or both. Motion artifacts will be defined as signal changes larger than 0.1 mM × mm over two samples. Finally, for the non-rejected blocks, a baseline will be linearly fit between the means of the 5 s preceding the onset of the block and the 5 sec starting 39 sec after the onset of the block (16 seconds of stimulation plus 23 sec of resting period). For each block, data will be averaged over a 16-second time window corresponding to the stimulation (Gemignani & Gervain, 2021).

4.2.5.2 Data Analysis

All analyses will be conducted within the general linear model (GLM) framework to assess infants' hemodynamic responses to multimodal emotional congruency. At the individual channel level, generating channel-wise *t* maps, and subsequently at the regional level using repeated-measures ANOVAs on predefined regions of interest (ROIs). The oxyHb and deoxyHb concentrations will be averaged across blocks for each condition. Two bilateral ROIs will be examined: the temporal cortices (STS/STG), associated with audiovisual integration and voice-sensitive processing (Belin et al., 2002; Blasi et al., 2011; Grandjean et al., 2005), and the prefrontal regions (mainly vmPFC/OFC), which have been implicated in affective evaluation and social motivation (Blasi et al., 2011; Minagawa-Kawai et al., 2009; Xie et al., 2019).

As a first analysis, channel-wise *t* maps will be constructed to compare concentration changes in oxyHb and deoxyHb between conditions in each channel. Specifically, we will contrast Congruent versus Incongruent non-alternating blocks to test for congruency effects, and Non-alternating versus Alternating blocks to test for alternation effects. Statistical significance will be assessed at $p < .05$ (uncorrected), consistent with prior work (Gervain et al., 2012).

In a second analysis, we will conduct repeated-measures ANOVAs with factors Condition (Congruent/Incongruent; Non-alternating/Alternating), Hemisphere (LH/RH), and ROI (temporal/frontal), run separately for oxyHb and deoxyHb. This analysis allows us to test the overall effects of the condition as well as interactions with hemisphere and ROI.

Finally, to explore whether neural sensitivity to emotional congruency was associated with their behavioral performance and family expressiveness environments, we will conduct the correlation analysis between hemodynamic congruency effects and infants' behavioral performance during the task, as well as parent-reported scores on the FEQ. For each infant, congruency indices (ΔHbO and ΔHbR = Congruent – Incongruent) will be calculated within each ROI. These indices will then be correlated with (a) infants' looking times during the task and (b) parent-reported scores on the FEQ, including both total scores and positive/negative expressiveness subscales. This analysis aims to test whether greater emotional expressiveness in the family context was associated with stronger cortical differentiation between congruent and incongruent multimodal cues.

4.3 Expected Results

4.3.1 Data Retention and Quality

At the time of writing, data collection for the present study is still ongoing. We aim to recruit approximately 45 infants aged 6–7 months. In line with practices in infant fNIRS research, the final size is expected to be reduced following the application of standard exclusion criteria, such as excessive fussiness, an insufficient number of valid trials, or poor signal quality resulting from motion artifacts. Based on prior studies using comparable paradigms, we anticipate that approximately 70–80% of the initially recruited infants will provide usable data for the final analyses.

4.3.2 Analytical Predictions

Congruency effect (Congruent vs. Incongruent).

At the channel level, *t* maps are expected to reveal significantly greater activation in response to the Congruent than Incongruent non-alternating blocks in a subset of channels within the temporal and frontal areas for oxyHb, with corresponding stronger decreases for deoxyHb ($p < .05$, uncorrected). We anticipate that most channels within the predefined ROIs will show positive β estimates relative to baseline for emotional stimulation, with particularly robust increases during congruent blocks. A stronger involvement of the right hemisphere (RH) is anticipated, with more discriminating channels in RH than LH.

At the ROI level, repeated-measures ANOVAs are expected to indicate a main effect of Condition (Congruent > Incongruent), in line with the channel-wise maps, and potentially a Condition × Hemisphere interaction reflecting greater RH involvement. We further expect that temporal ROIs (STS/STG) will show the most reliable Congruent > Incongruent effects, whereas prefrontal ROIs (vmPFC/OFC) may exhibit congruency- or valence-modulated responses (Blasi et al., 2011; Di Lorenzo et al., 2019). Such findings would be consistent with prior behavioral studies showing emerging sensitivity to intermodal congruency, especially for happy expressions, and extend them to the neural level by demonstrating enhanced cortical engagement for emotionally congruent cues.

Alternation effect (Non-alternating vs. Alternating).

At the channel level, *t* maps are expected to show significantly greater activation for non-alternating than for the Alternating blocks in both temporal and frontal areas for oxyHb, with stronger deoxyHb decreases for non-alternating. More discriminating channels are expected in temporal than frontal areas, suggesting broader temporal involvement in sustained processing of congruent or incongruent pairings.

At the ROI level, repeated-measures ANOVAs are expected to reveal the main effect of Condition (Non-alternating > Alternating). A significant main effect of ROI is also expected to be observed because of greater general activation in the temporal than in the frontal ROI. A Condition × ROI interaction may also emerge, with stronger non-alternating advantages in temporal compared with frontal regions, whereas hemispheric asymmetries are expected to be weaker or absent. Together with the congruency analyses, these effects would support the prediction of a redundancy boost, whereby congruent information in non-alternating blocks elicits stronger cortical responses than alternating sequences.

Across analyses, both congruent and incongruent conditions are expected to elicit stronger activation than the neutral baseline, with congruent responses particularly robust. OxyHb increases and deoxyHb decreases are anticipated to show convergent but oppositely signed effects. Although variability across infants is expected, the group-level pattern should reveal reliable congruency and alternation effects in bilateral temporal and prefrontal ROIs. Taken together, these expected findings would extend previous behavioral findings by providing neurophysiological support for the involvement of temporal and prefrontal cortices in early socioemotional development in 6–7-month-old infants.

Alternatively, if incongruent stimuli elicit stronger responses, this pattern would align with prediction-error models, suggesting that infants' brains are sensitive to violations of multimodal expectations (Friston, 2005; Rothermund et al., 2011). One likely explanation is that such incongruent emotional pairings are relatively rare in everyday interactions; the neural response may therefore reflect increased processing of the incongruent multimodal emotional expressions. Conversely, if no reliable differences are observed, this may indicate either developmental changes in integrative capacity or limited detectability of congruency sensitivity with infant fNIRS.

4.3.3 Expected Exploratory Correlations

For the Brain-Behavior-Environment Link, we expect positive correlations between neural response differences and infants' behavioral performance during the task (looking time), indicating that stronger cortical differentiation accompanies greater attentional allocation to congruent multimodal emotional information. We also anticipate positive associations between neural responses and FEQ score, but not particularly the positive or negative expressiveness subscale, as we have found the positive correlation between EMS and the negative subscale from the behavioral experiment, suggesting that families provide infants with a wider range of affective input, which strengthens their ability to integrate emotional cues across modalities. Align with the view that families characterized by more emotionally expressive interactions are predicted to have infants who show stronger neural differentiation between congruent and incongruent emotional stimuli. Moreover, Correlations are expected to be strongest in temporal ROIs ($r \approx .25$–$.35$). Family expressiveness may moderate the brain–behavior link: the correlation between neural responses with looking time is expected to be stronger at higher FEQ levels.

4.4 Discussion

4.4.1 Summary of Expected Findings

The present study employed fNIRS to investigate whether 6- to 7-month-old infants show neural sensitivity to multimodal emotional congruency. In particular, exploring how shared affective information conveyed through facial and vocal emotional cues shapes infants' brain responses to different multimodal expressions, with a specific focus on the temporal and prefrontal cortices, as well as the associations between neural congruency effects, infants' behavioral performance, and parental reports of family expressiveness. Although data collection is still ongoing, several expected patterns can be outlined based on our experimental design and existing literature.

First, we anticipate obtaining approximately 70-80% of valid data from the recruited sample, revealing individual-level neural responses to multimodal emotional stimuli, particularly in the bilateral temporal and frontal regions. Second, at the block level, we expect that congruent multimodal emotional information, operationalized in this study as happy facial and vocal expressions, will elicit greater neural responses than incongruent pairings. This prediction is consistent with our behavioral findings, which indicated that infants are particularly sensitive to the congruency of positive emotions. At the neural level, this sensitivity is expected to manifest as significantly greater activation for congruent than incongruent conditions within non-alternating blocks, as well as across the entire experimental blocks.

Furthermore, we predict that reliable congruency and alternation effects will be observed. Regarding regions of interest, based on prior neuroimaging research, we anticipate that congruency sensitivity will be most pronounced in the bilateral temporal cortices, whereas both sensitivity and emotional valence are expected to drive activation in the prefrontal regions.

Finally, we will explore the associations between neural congruency effects, infants' behavioral performance, and parent-reported scores on the FEQ. We expect that larger neural congruency effects will be positively correlated with infants' behavioral performance, and the FEQ scores will correlate with both neural and behavioral measures. FEQ is expected to moderate the relationship between neural sensitivity and behavioral performance, such that infants from more emotionally expressive families will show stronger brain–behavior coupling.

Together, these expected findings would provide neurophysiological evidence that extends our behavioral results, supporting the view that 6- to 7-month-old infants are sensitive to the congruency of multimodal emotional information.

4.4.2 Expected Findings and Interpretations

Stronger neural responses to congruent multimodal emotional expressions than incongruent pairings would support accounts emphasizing the important role of shared affective information across sensory modalities in early emotion perception (Flom & Bahrick, 2007; Walker-Andrews, 1997). It might be easier for infants to process congruent affective information, which enhances their ability to perceive and integrate emotional expressions. Previous fNIRS studies also demonstrate increased activation for socially congruent stimuli (Grossmann, 2010; Lloyd-Fox et al., 2018). Moreover, the greater activation in non-alternating than in alternating blocks reflects the importance of stable audiovisual redundancy, aligning with IRH theory, suggesting that redundant multimodal cues promote the detection of shared affective information (Bahrick & Lickliter, 2000; Vaillant-Molina et al., 2013). Yet there is a potential pattern that incongruent pairings might elicit stronger activation compared to congruent multimodal emotions, which would align with the prediction-error models (Friston, 2005), suggesting that infants' brains are already sensitive to violations of cross-modal expectations. In this theory, heightened activation reflects the additional neural resources required when predictions are violated. A lack of reliable differences would point to either developmental changes in integrative capacity or the limited detectability of congruency effects with infant fNIRS.

Regardless of congruency, multimodal emotional expressions are expected to elicit greater activation than the baseline (neutral facial expressions with no sound). Such findings would confirm that infant brains are engaged by emotional expressions per se, indicating the neural sensitivity to social cues (Grossmann & Johnson, 2007; Lloyd-Fox et al., 2018); this may be due to the development of the infants' social brain. Congruency effects are predicted to localize to temporal cortices, which support audiovisual integration (Belin et al., 2002; Grossmann et al., 2010), with supplementary responses in prefrontal regions, which contribute to affective evaluation and social motivation (Di Lorenzo et al., 2019). A greater activation pattern in temporal areas is plausible, consistent with evidence of hemispheric specialization for socio-emotional processing (Davidson & Fox, 1989; Grossmann, 2013). Observing these regional patterns would provide further evidence for the early specialization of the infant social brain.

Exploratory analyses are expected to show that stronger neural congruency effects are associated with longer looking times, suggesting alignment between attentional engagement and neural differentiation. In addition, greater family emotional expressiveness is anticipated to correlate with stronger cortical differentiation and to moderate the relationship between neural responses and behavior performance (Ogren et al., 2018). Such findings would highlight the role of environmental input in shaping individual differences in early socio-emotional development.

The expected findings are consistent with a growing body of behavioral research demonstrating that infants become sensitive to intermodal emotional congruency within the first year of life. As reviewed in prior work and confirmed in our behavioral study, infants aged 5-7 months often look longer at congruent than incongruent facial-vocal expression pairings (Flom & Whiteley, 2014; Phillips et al., 1990; Soken & Pick, 1999). Predicted stronger neutral responses to congruent happy expressions in the present study would further align with evidence of a positivity bias in infancy, whereby positive emotions are preferentially attended to and processed (Grossmann, 2010). Moreover, the anticipated association between neural congruency effects and family expressiveness would extend previous behavioral findings showing that infants' emotion-matching performance is shaped by emotionally expressive environments. Together, these findings would bridge behavioral sensitivity and neural mechanisms, highlighting the continuity between observable preference patterns and underlying cortical processes. More broadly, these results would support the view that early multimodal integration provides a foundation for the later emergence of socioemotional cognition.

4.4.3 Developmental and Applied Implications

These expected findings may offer new perspectives for theories of early emotion cognition in infancy. First, by extending behavioral evidence of sensitivity to emotional congruency to the neural level, the study provides neurophysiological support for the view that infants are able to extract cross-modal redundant affective information when processing multimodal emotions (Bahrick et al., 2004; Bahrick & Lickliter, 2000; Vaillant-Molina & Bahrick, 2012). Second, by contrasting intersensory integration accounts with prediction-error models (Friston, 2005), it highlights the possibility that congruent and incongruent cues are processed by distinct neural mechanisms, offering insights beyond behavioral performance alone. Third, the results provide further evidence for regional specificity within the infant social brain, emphasizing the early involvement of bilateral temporal and prefrontal regions (Grossmann, 2013; Minagawa-Kawai et al., 2009). Finally, the study highlights the role of emotional environmental factors, suggesting that neural sensitivity is shaped not only by intrinsic developmental processes but also by the broader emotional environment (Halberstadt & Eaton, 2002; Ogren et al., 2018).

From a practical perspective, the anticipated findings carry important implications for understanding both typical and atypical trajectories of socio-emotional development. Identifying sensitivity to multimodal emotional congruency at 6–7 months may provide a potential early indicator of risk in populations with socioemotional difficulties, such as infants at elevated likelihood for autism spectrum disorder (ASD) or other developmental conditions associated with atypical emotion processing. Moreover, the expected associations between family emotional expressiveness and infants' neural responses underscore the potential of early social environments to shape neurocognitive development. This suggests that interventions fostering emotionally rich caregiver–infant interactions may strengthen the foundations of socio-emotional cognition. More broadly, by integrating behavioral and neuroimaging approaches, the present study provides a methodological framework for future longitudinal research aimed at tracking how early neural sensitivity to emotional congruency supports the later emergence of social communication, language, and mentalizing abilities.

4.4.4 Limitations and Future Directions

Several limitations should be acknowledged. First, as is typical in infant neuroimaging research, data attrition is expected due to fussiness, insufficient valid trials, or poor signal quality, which may result in a relatively modest final sample size. Although the anticipated retention rate of 70–80% is consistent with previous fNIRS studies, larger samples would be needed in future work to increase statistical power and generalizability.

Second, experimental design and stimuli introduce certain constraints. The current paradigm primarily focuses on happy and angry expressions, thereby limiting conclusions about the processing of a broader range of emotions, such as sadness and fear. Moreover, the stimuli were derived from a small set of native Italian speakers, which may reduce ecological validity and cross-cultural generalizability. Future research should incorporate more diverse emotional categories, multiple speakers, and culturally varied materials.

Third, interpretational limitations might be considered. fNIRS provides good temporal resolution for cortical surface responses but cannot capture deeper subcortical structures, such as the amygdala, that are known to play a central role in emotion processing (Adolphs, 2009; Johnson et al., 2009). Moreover, the observed associations between family expressiveness and neural sensitivity to emotional congruency are correlational in nature and cannot establish causality. Longitudinal or intervention-based studies will be necessary to determine whether the family environment shapes the neural mechanisms of multimodal emotion perception.

Finally, several directions can be considered for future research. Longitudinal studies following infants into early childhood would clarify how early neural sensitivity to emotional congruency supports later outcomes such as social referencing, the development of secure attachment relationships, emerging empathy, and the regulation of emotional responses within social interactions. Cross-cultural comparisons would help determine whether the developmental trajectory of emotion processing is universal or shaped by cultural contexts. Studies in high-risk populations, such as infants with a familial likelihood of autism spectrum disorder, could assess whether atypical sensitivity to multimodal cues functions as an early biomarker of socioemotional difficulties. Methodologically, the integration of fNIRS with EEG or fMRI, as well as the application of machine learning approaches such as multivoxel pattern analysis (MVPA), may provide richer spatiotemporal characterizations of infant emotion processing.

4.5 Conclusion

In conclusion, the present study provides neural level evidence for emerging sensitivity to multimodal emotional congruency in 6- to 7-month-old infants. It highlights how the shared emotional information across modalities and brain regional specialization shapes early emotion perception. Moreover, by linking the neural responses, behavioral performance, and family emotional expressiveness, it underscores the interplay between biological mechanisms and environmental influences in socio-emotional development. Together, these expected findings would advance our understanding of how 6- 7-month-old infants integrate and interpret emotional cues, further promoting emotional development and regulation.

CHAPTER V.

ADVANCING KNOWLEDGE: FROM FINDINGS TO IMPACT

5.1 General Discussion

Emotions are central to human experience, functioning as primary regulators of behavior, communication, and social interaction. From birth, infants are embedded in a world rich with emotional exchanges, long before they acquire language. A core challenge of early development is therefore how preverbal infants recognize others' emotional states and use this information to guide their own behavior. Facial expressions and vocal prosody serve as the primary emotional cues through which infants interpret others' affect, enabling them to regulate their actions and engage in meaningful social interactions. Sensitivity to multimodal emotional signals not only enables rapid recognition of positive and threatening cues, supporting adaptive responses such as seeking protection and strengthening social bonds, but also lays the foundation for socioemotional development, including secure attachment, social referencing, empathy, and prosocial behavior. Studying the development of this ability therefore offers critical insights into both the adaptive functions of emotion processing and its role in shaping the broader trajectory of socio-emotional development.

To investigate the multimodal affective information perception in preverbal infants, three studies were conducted throughout my PhD. By integrating a systematic review, a behavioral experiment, and an fNIRS study, this dissertation provides converging evidence for the developmental, behavioral, and neural foundations of multimodal emotion processing in early infancy. Study 1 synthesized empirical findings on infants' emotion perception during the first year of life, establishing both the conceptual framework and the empirical basis for subsequent investigations. Building on this foundation, Study 2 offered behavioral evidence for intermodal emotion matching, highlighting developmental changes and individual differences in emotion processing among 6- to 8-month-old infants. Finally, Study 3 examined the neural mechanisms underlying infants' sensitivity to emotional congruency at 6–7 months, providing novel insights into the cortical bases of early socio-emotional processing. Collectively, these three studies contribute an integrated perspective on how infants begin to perceive, match, and neurally represent multimodal emotional signals during a critical period of socio-emotional development.

5.2 Conceptual and Empirical Foundations

The first study systematically reviewed research on how infants in the first year of life perceive and integrate affective information across sensory modalities. Most research reviewed focused on infants' discrimination of multimodal emotions (Caron et al., 1988; Flom & Bahrick, 2007; Soken & Pick, 1992), their ability to match emotional information across modalities (Grossmann et al., 2006; Kahana-Kalman & Walker-Andrews, 2001; Vaillant-Molina et al., 2013), and the developmental course of emotion preferences during the first year of life (Nawrot, 2003; Soken & Pick, 1992). The available evidence suggests that the developmental trajectory of multimodal emotion processing is gradual yet uneven. During the first few months after birth, infants' perceptual and social capacities are still developing, and their responses to emotional signals largely remain at the level of basic perception rather than meaningful integration (Flom & Bahrick, 2007; Walker-Andrews et al., 2011; Walker-Andrews & Lennon, 1991). Consequently, research in this early period primarily focused on whether infants can discriminate among different emotional expressions, rather than on more complex processes such as emotion matching. The visual limitations present during this stage further reinforce this emphasis on discrimination-based paradigms. In contrast, emotion matching has been most frequently examined in infants aged 5 to 9 months, a developmental window in which increasing social exchanges with caregivers and others provide a richer context for interpreting external emotional signals (Vogel et al., 2012; Xiao & Emberson, 2019; Yong & Ruffman, 2016). Sensitivity to cross-modal emotional congruency appears to stabilize in this window, particularly when temporally synchronous and dynamic audiovisual stimuli are presented (Bahrick & Lickliter, 2012; Vaillant-Molina & Bahrick, 2012). Remarkably, when the emotional expressions are produced by a familiar individual, evidence suggests that such sensitivity may emerge as early as 3.5 months (Kahana-Kalman & Walker-Andrews, 2001; Montague & Walker-Andrews, 2002; Walker-Andrews et al., 2011). Research on infants' emotion preferences spans the entire first year, yet findings indicate substantial variability: positive emotions, especially happy expressions, consistently attract infants' attention across tasks and modalities, whereas preferences for negative emotions are less frequent and tend to emerge only in the latter half of the first year (Nawrot, 2003; Otte et al., 2015). Although behavioral paradigms such as visual preference, habituation–dishabituation, and intermodal matching have dominated this literature, the collective findings nonetheless provide an essential framework both conceptually and methodologically for guiding subsequent empirical research.

These findings are broadly consistent with theoretical accounts emphasizing the role of intersensory redundancy. According to IRH, the presentation of redundant information across modalities enhances infants' sensitivity to multimodal signals (Bahrick et al., 2004; Bahrick & Lickliter, 2000). The reviewed evidence tends to align with this view, showing that congruency sensitivity is frequently modulated by stimulus synchrony, among other contextual factors. Importantly, however, some studies also suggest that infants can still recognize and prefer matching emotions even when multimodal stimuli are asynchronous (Soken & Pick, 1999). This finding implies that, beyond temporal redundancy, the semantic redundancy of affective information itself may be accessible to infants (Grossman, 2013; Grossmann, 2010). Congruency sensitivity appears to be especially pronounced between 6 and 8 months of age, a developmental window often regarded as critical for socio-emotional growth. In addition, infants' sustained attentional bias toward positive affect is consistent with the positivity bias hypothesis. At the same time, the literature reveals notable inconsistencies: congruency effects vary as a function of stimulus synchrony, dynamic versus static presentation, and task demands. These discrepancies suggest that while theoretical models capture broad developmental trends, the mechanisms underlying emotion-specific trajectories are more context-dependent and complex than previously assumed.

A further contribution of the review lies in clarifying how methodological approaches shape conclusions about early emotion processing. Habituation paradigms provide fine-grained evidence of discrimination abilities but are resource-intensive and susceptible to infant fatigue (Fantz, 1964). By contrast, intermodal matching tasks offer higher ecological validity but may underestimate subtle differences in processing. Comparisons between static and dynamic stimuli suggest that temporally coordinated audiovisual cues elicit stronger and more consistent responses, underscoring the importance of ecological validity in experimental design (A. R. Heck et al., 2016; Soken & Pick, 1999). Collectively, these methodological insights not only advance our understanding of early multimodal emotion processing but also provide both theoretical and empirical guidance for the design of our behavioral experiment. While at the same time highlighting the rationale for combining behavioral and neurophysiological approaches, such as fNIRS, to capture the complexity of early socio-emotional development.

At the same time, several limitations characterize existing literature. A striking imbalance concerns the types of emotions investigated: research has disproportionately emphasized positive emotions, particularly happiness. This overrepresentation is not necessarily the result of explicit theoretical bias but reflects the methodological tendency to include happiness as a prototypical category whenever multiple emotions are presented. Consequently, developmental accounts of positive emotions are relatively well documented, whereas evidence on negative emotions is fragmented and inconsistent, leaving their trajectories less clearly understood. Methodological constraints also persist. Studies of multimodal emotion perception have relied predominantly on behavioral measures, with only limited use of neuroimaging approaches to illuminate the underlying neural mechanisms. This provides a central rationale for our integration of behavioral paradigms with fNIRS. Finally, infant research is frequently constrained by small sample sizes and heterogeneous designs, which complicate cross-study comparisons and likely contribute to the inconsistencies observed across the literature. Taken together, these limitations highlight the need for more comprehensive and standardized approaches that extend beyond happiness and encompass the full spectrum of emotional expressions.

By synthesizing existing evidence and outlining conceptual and methodological needs, this review offers direct guidance for subsequent empirical work. Converging findings highlight 6–8 months as a transitional window in socio-emotional development, thereby justifying the focus of both behavioral and neuroimaging investigations on this age range. Evidence favoring dynamic multimodal stimuli further supports the use of naturalistic audiovisual recordings rather than static or artificially constructed materials. Notably, our experiments employed asynchronous stimulus presentation to ensure that infants' responses reflected sensitivity to the redundancy of emotional meaning rather than mere temporal synchrony. In addition, multiple studies have reported that familiarity with the expresser influences infants' emotion perception, underscoring the role of family and broader caregiving environments. To address this, both our behavioral and fNIRS studies incorporated family-level measures, such as the Family Expressiveness Questionnaire, to examine how variations in family emotional expressiveness shape infants' multimodal emotion processing. Taken together, the review not only clarifies the current state of knowledge but also provides a coherent theoretical and methodological rationale for the empirical investigations presented in Studies 2 and 3.

5.3 Behavioral Evidence from Emotion Matching Tasks

The second study investigated whether 6- to 8-month-old infants are able to match facial and vocal expressions of emotion in a visual-preference-based intermodal matching paradigm. Using video-coded measures of infants' looking behavior, we derived an Emotion Matching

Score (EMS) to quantify their preference for congruent versus incongruent face–voice pairings. The results revealed a consistent matching response in the happy condition, whereas performance in the angry and neutral conditions was less robust. The results align with the positivity bias hypothesis, showing that infants' ability to match happy expressions emerges earlier and with greater stability than for negative emotions. Interestingly, during testing, several infants appeared inclined to interact socially when presented with happy faces. This early advantage for positive emotions likely reflects both developmental factors, such as the maturation of socio-emotional capacities, and experiential influences, including the frequent exposure to positive affect in early caregiver–infant interactions. Moreover, age-related differences emerged: six-month-old infants showed stronger and more sustained matching responses to happy expressions, consistent with a positivity bias, whereas older infants displayed a more variable attentional distribution across conditions. Taken together, these results indicate that dynamic audiovisual stimuli elicit reliable cross-modal matching in early infancy, particularly for positive emotions. The stronger and more stable responses to happiness suggest that both developmental mechanisms and rich experiential input contribute to infants' early advantage for positive affect, whereas matching for negative and neutral emotions remains less robust. Moreover, the observed age differences highlight an ongoing reorganization of attentional strategies within this critical developmental window.

The behavioral findings align with theoretical accounts emphasizing the salience of redundant audiovisual affective information for early perceptual learning. Not only in line with IRH, but also future support accounts of cross-modal emotion recognition (Bahrick & Lickliter, 2000; Grossmann, 2010), even when visual and auditory channels asynchronously convey happy cues, redundancy enhances salience and facilitates infants' ability to match affective signals. The strong preference for congruent happy expressions further supports the notion of a positivity bias in infancy (Grossmann, 2010), whereby positive affective signals are prioritized in early social information processing. By contrast, the weaker and less consistent responses to angry and neutral expressions suggest that developmental trajectories may vary across emotion categories. In the angry condition, for instance, EMS values were close to zero, indicating comparable looking times toward congruent and incongruent angry faces. One interpretation is that 6–8 months represents a transitional window in emotional development during which infants' attentional strategies toward different emotions are being reorganized. Rather than integrating angry cues exclusively in the congruent condition, infants in our sample may have distributed attention across both congruent and incongruent angry faces. Alternatively, although infants may be able to perceptually discriminate angry expressions, the capacity to flexibly use congruency as a cue could still be developing at this age. Accordingly, the absence of a congruency effect in the anger condition should not be taken as evidence of insensitivity to negative emotions. Importantly, the age-related differences observed in this study resonate with previous accounts that highlight the 6–8 month period as a developmental transition, during which attentional strategies become increasingly flexible and adaptive. Finally, our findings further indicate that infants raised in families with higher levels of emotional expressiveness showed better performance in the emotion-matching task, with the association being strongest in the anger condition, weaker but still present in the happy condition, and absent for neutral expressions. These results suggest that the emotional atmosphere of the family may shape infants' sensitivity to cross-modal congruency, particularly for emotions with greater intensity or social salience (Ogren et al., 2018). Nonetheless, caution is warranted, as reliance on parent-report measures may introduce subjective bias.

Beyond the specific behavioral outcomes, this study also carries methodological implications for research on infant emotion processing. First, by employing dynamic audiovisual recordings rather than static photographs or artificially edited stimuli, the paradigm enhanced ecological validity and more closely approximated the multimodal emotional signals that infants encounter in everyday interactions. Second, the use of asynchronous audiovisual stimuli ensured that infants' matching responses reflected sensitivity to cross-modal emotional redundancy rather than simple temporal synchrony. Third, the derivation of the Emotion Matching Score provided a quantitative index of infants' relative preference for congruent versus incongruent pairings, thereby allowing for the detection of subtle individual and age-related differences. In sum, these methodological advances emphasized the value of naturalistic multimodal stimuli and fine-grained behavioral measures in elucidating the early developmental trajectory of emotion perception.

Several limitations of this study should be acknowledged. First, the sample size was modest, which may have reduced the statistical power to detect reliable age-related or condition-specific effects. Second, the stimuli were restricted to two female models and a limited set of emotions (happy, angry, neutral), constraining the generalizability of the findings to broader affective categories and more diverse social contexts. Third, although video-based coding provided a direct measure of infants' attentional behavior, it could not capture the underlying neural mechanisms driving intermodal matching, nor fully disentangle perceptual sensitivity from motivational influences on looking time. For example, while infants directed attention to angry expressions, their matching performance in the angry condition was unstable, suggesting that perceptual discrimination and integrative use of emotional cues may not always converge. Finally, because the paradigm was conducted in a controlled laboratory setting, it remains uncertain how infants' cross-modal matching operates in more naturalistic caregiver–infant interactions. Addressing these limitations will be essential for building a more comprehensive account of early emotion perception.

The behavioral findings from this study provide a compelling rationale for further investigating the neural bases of multimodal emotion processing in infancy. The consistent preference for congruent happy expressions suggests that neural systems may be particularly sensitive to positive affect during early development, a hypothesis that can be directly tested using fNIRS. In addition, the weaker and less consistent responses to angry and neutral conditions highlight the need to examine whether specific brain regions, such as the prefrontal and temporal cortices, differentially contribute to the processing of specific multimodal emotions. Moreover, the observed association between family expressiveness and infants' matching performance points to the importance of considering environmental influences alongside neural measures, thereby encouraging integrative approaches that link behavior, brain activity, and social context.

5.4 Neural Mechanisms of Multimodal Integration

The third study employed fNIRS to investigate the neural mechanisms underlying infants' processing of multimodal emotional information. Based on our hypotheses and experimental design, we predicted that all multimodal emotional stimuli would elicit greater neural activation compared to the baseline condition involving silent neutral faces. More specifically, congruent face–voice pairings were expected to evoke stronger hemodynamic responses in prefrontal and temporal regions relative to incongruent pairings, reflecting infants' sensitivity to cross-modal emotional congruence. Furthermore, this neural differentiation was anticipated to be more pronounced in the non-alternating group than in the alternating group. We also hypothesized that predefined ROIs would exhibit stronger activation patterns, given their established role as core components of the infant social brain (Belin et al., 2002; Davidson & Fox, 1989; Minagawa-Kawai et al., 2009). In addition, we predicted that individual differences in family emotional expressiveness, as measured by the FEQ, would positively correlate with neural sensitivity in these regions. These anticipated findings would extend the behavioral evidence reported in Study 2 by providing convergent support at the neural level and by linking brain activation to environmental input. While the current discussion is grounded in these hypothesized patterns, the analyses are still ongoing. The results section will be updated once analyses are complete, and revisions will be integrated to refine and strengthen the interpretations presented here.

The anticipated neural patterns are consistent with established theoretical accounts and prior empirical findings on early multimodal perception. Specifically, we predicted that congruent emotional cues would elicit stronger activation in prefrontal and temporal regions, in line with IRH and the framework of multimodal emotion recognition (Bahrick & Lickliter, 2000; Grossmann, 2010; Grossmann et al., 2006). These perspectives suggest that infants' brains are sensitive to redundant cross-modal signals, encompassing not only temporal synchrony but also affective meaning. Moreover, the predicted heightened response to positive affect converges with both the behavioral findings from Study 2 and prior evidence for a positivity bias in infancy, indicating that neural systems may already differentiate emotional valence before the emergence of language. Finally, the hypothesized association between family expressiveness and neural sensitivity accords with broader developmental theories emphasizing the role of environmental input in shaping socio-emotional processing. While these connections provide a coherent interpretive framework, it should be noted that they remain based on anticipated results; forthcoming analyses will be essential for testing and refining these theoretical implications.

Methodologically, this study highlights the value of using fNIRS to capture the neural correlations of infant emotion processing. By adopting an alternating versus non-alternating block design, the paradigm enabled direct comparisons of hemodynamic responses to congruent and incongruent audiovisual stimuli, thereby extending the scope of traditional behavioral approaches. The inclusion of both prefrontal and temporal cortices further provided complementary insights into higher-order socio-emotional processing and perceptual integration. Finally, by combining neural measures with environmental indices such as the FEQ, this study illustrates the promise of multi-level approaches that integrate behavior, brain function, and family context. Together, these methodological advances offer a framework for future research aimed at capturing the complexity of early socio-emotional development.

Several limitations of this study should be considered. One limitation concerns the relatively modest sample size, which restricts the generalizability of the findings and may reduce sensitivity to subtle developmental or individual differences. Methodological constraints also apply to the use of fNIRS: although well suited for infant testing, its limited spatial resolution precludes investigation of deeper limbic structures, such as the amygdala, that play a central role in emotion processing. A further consideration relates to the restricted set of emotional categories examined (happy, angry, neutral), which does not reflect the broader range of affective signals encountered in natural environments. Finally, despite the use of dynamic audiovisual stimuli to enhance ecological validity, the paradigm remained laboratory-based, leaving open questions about how these neural responses manifest in everyday caregiver–infant interactions. Addressing these limitations will be essential for developing a more comprehensive understanding of the neural foundations of early socio-emotional development.

Building on these anticipated findings, several avenues for future research emerge. A key priority will be to combine fNIRS with complementary techniques such as EEG, enabling a more fine-grained characterization of the temporal dynamics and attentional mechanisms involved in multimodal emotion processing. Longitudinal approaches are also needed to examine whether early neural sensitivity to emotional congruence predicts later socio-emotional outcomes, including attachment quality, empathy, and language development. In addition, future work should broaden the range of emotional categories under investigation, for example, by incorporating fear, sadness, or surprise, and explore potential cultural variations in infants' neural responses to affective cues. Finally, advancing beyond controlled laboratory settings toward naturalistic caregiver–infant interaction paradigms will be critical for enhancing ecological validity and clarifying how neural mechanisms of multimodal integration support everyday social communication. Collectively, these directions will deepen our understanding of how brain, behavior, and environment interact to shape the developmental foundations of emotion processing.

5.5 Overall Significance of the Three Studies

In conclusion, the three studies presented in this dissertation provide comprehensive evidence for the developmental, behavioral, and neural foundations of multimodal emotion processing in preverbal infants. Study 1 synthesized existing evidence from the first year of life to establish the conceptual and empirical landscape, identifying a critical developmental window and highlighting methodological challenges. Study 2 provided direct behavioral evidence that infants between 6 and 8 months are able to match happy face–voice pairings but display less consistent responses to angry and neutral emotions, thereby supporting the notion of a positivity bias and emerging age-related differences. Study 3 extended these insights to the neural level, using fNIRS to identify how prefrontal and temporal regions are engaged in processing multimodal emotional expressions, as well as to investigate how family expressiveness shapes individual variability in neural sensitivity. By integrating systematic review, behavioral observation, and neuroimaging, this work demonstrates that early emotion perception should be viewed both as a key developmental milestone and as the outcome of ongoing exchanges between brain development, perceptual experience, and the broader social environment.

Beyond the individual contributions of each study, their combined behavioral, neural, and contextual evidence allows this dissertation to advance a more integrated developmental framework for understanding early multimodal emotion processing. The findings indicate that infants' understanding of emotional information arises from the interplay of multiple factors. Different discrete emotional signals, particularly happiness, serve distinct functions in shaping early attentional biases and cross-modal integration. Infants' preference for happy expressions, together with their enhanced cortical responses to affectively congruent happy stimuli, suggests that early multimodal emotion processing follows emotion-specific developmental pathways. Such patterns imply that infants' emerging emotion understanding is structured around differentiated sensitivities rather than a generalized, undifferentiated affective system.

Importantly, the present work also advances a mechanistic account of why infants exhibit a positivity bias. Converging evidence across the studies points to at least three interacting mechanisms. First, at the perceptual level, happy expressions provide highly salient and cross-modally redundant audiovisual cues, such as distinctive facial dynamics and exaggerated prosodic contours, and high ecological frequency in caregiver–infant interactions, making them easier for infants to align across senses, facilitate infants' intermodal matching. Second, at the neural level, activation in prefrontal and temporal regions suggests that early cortical specialization may support the integration of positive affective cues, potentially reflecting infants' preferential processing of prosocial and affiliative affective signals. Third, at the contextual level, family expressiveness appears to modulate infants' behavioral response and neural sensitivity, indicating that positivity bias is not solely driven by innate predispositions but is also shaped by the emotional characteristics of the caregiving environment.

By bringing together perceptual, neural, and socio-emotional perspectives, this dissertation demonstrates that early multimodal emotion processing is best understood as emerging from dynamic interactions between biological preparedness, experiential input, and environmental context. This multi-level integration provides a more coherent and comprehensive account of how infants begin to perceive and interpret emotional signals across modalities during the first year of life.

From a broader disciplinary perspective, this dissertation also contributes at a meta-level by illustrating the value of integrating systematic review, controlled behavioral paradigms, and neuroimaging approaches within a unified developmental framework. Such an approach enables early emotion processing to be conceptualized as a coordinated cross-level system rather than as a set of isolated behavioral manifestations or neural responses. Whereas previous research often examined these levels independently, the present work demonstrates that cross-method and cross-level integration may yield more comprehensive insights into the developmental organization of infants' socio-emotional abilities. The aim of this dissertation is not to propose a fully elaborated theoretical model but to advance a more comprehensive and methodologically coherent research pathway for studying early emotion perception, one that highlights the importance of considering behavior, brain development, and environmental context as interrelated dimensions of a single developmental process.

Taken together, these findings enrich existing theoretical models in two key ways: first, by shifting the focus from broad valence-based accounts to emotion-specific developmental trajectories; and second, by moving from single-level analyses to a multi-level integrative approach. By synthesizing empirical evidence across methods and levels, this dissertation provides a foundation for future holistic investigations of how infants begin to organize, differentiate, and interpret emotional information during the first year of life.

5.6 Future Directions and Implications

The findings of this dissertation open several promising avenues for future research on early multimodal emotion processing. At the methodological level, future studies would benefit from employing stimuli that are more naturalistic, context-rich, and interaction-based in order to capture how infants process emotional cues within real-world social environments. Incorporating cross-cultural approaches will be equally essential. Emotional expressions, prosodic contours, infant-directed speech, and caregiver–infant interaction styles vary substantially across cultural contexts. Such variability is likely to shape both the perception of emotional cues and infants' opportunities for exposure. Cross-cultural and multi-site approaches would therefore enable researchers to examine whether early multimodal emotion processing reflects universal developmental mechanisms or culturally shaped learning trajectories. Methodological advances can also be achieved by integrating cross-cultural sampling with fine-grained behavioral measures such as eye-tracking, computational models of visual attention, and multimodal physiological indices, including gaze patterns, pupil dilation, facial expressions, EEG, and data from wearable biosensors. Together, these tools can enhance the sensitivity and ecological validity of assessments in infancy research. In terms of neuroimaging, combining fNIRS with EEG or applying multivariate and connectivity-based analytical approaches may further clarify the spatiotemporal dynamics underlying infants' integration of emotional information across modalities.

Future research should also broaden the range of affective categories investigated during infancy. Emotions such as fear, sadness, surprise, and disgust remain relatively understudied in multimodal paradigms, despite their distinct communicative functions, differential developmental trajectories, and relevance for early social adaptation. Expanding the scope of inquiry to include these emotions would offer a more complete picture of the early affective landscape and clarify whether infants' processing advantages, such as positivity bias or preference for congruent cues, extend to non-positive or less frequently encountered emotional signals. In addition, examining how infants respond to a wider array of emotions across cultural contexts, vocal registers, and varying levels of emotional intensity will be critical for understanding both the specificity and generalizability of early emotion processing. Such work could reveal whether developmental patterns observed for happiness and anger represent universal perceptual tendencies or instead reflect culture-specific exposure frequencies, caregiver interaction styles, or prosodic norms. A more comprehensive characterization of infants' sensitivity to diverse emotional cues will help establish a richer, more ecologically grounded framework for understanding the emergence of early social–emotional cognition.

Another important direction involves mapping developmental trajectories through longitudinal designs. Tracking infants across multiple time points would provide critical insights into how early positivity biases unfold over development, whether multimodal matching skills predict later socio-emotional competencies, and how early cortical responsiveness contributes to developmental cascades such as social referencing, joint attention, empathy, and emotion regulation. Longitudinal approaches may also help clarify the temporal ordering of these emerging skills, helping to determine whether early perceptual and neural sensitivities exert downstream influences on broader social–emotional outcomes. Moreover, longitudinal approaches offer a powerful way to identify early markers of resilience and vulnerability. By examining how individual differences in early attention, neural activation patterns, or family emotional environments predict later adaptive or maladaptive trajectories, future research may better characterize the developmental pathways that support or hinder socio-emotional well-being. Such knowledge holds the potential to inform early identification and intervention efforts aimed at promoting healthy socio-emotional development.

In addition to informing typical developmental pathways, the present findings have implications for understanding atypical development. Because multimodal emotion processing relies on coordinated interactions among perceptual, neural, and contextual systems, early deviations in any of these components may serve as sensitive indicators of emerging socio-emotional difficulties. Conditions such as autism spectrum disorder, developmental language disorder, and social communication impairments often involve atypical patterns of attention, cross-modal integration, or neural responsiveness to affective cues. Examining how such disruptions manifest during infancy could help identify developmental signatures that precede clinically observable symptoms.

Incorporating multimodal behavioral and neurophysiological markers, such as gaze-based matching indices, pupil-linked arousal measures, or fNIRS–EEG markers of affective congruency, into early assessment frameworks may increase the precision and predictive validity of early screening practices. These multimodal markers may also reveal distinct developmental pathways that differentiate transient delays from more persistent socio-emotional vulnerabilities. Moreover, the observed influence of family expressiveness on infants' neural sensitivity highlights the potential value of early, family-centered interventions aimed at enhancing emotional availability, parent–infant synchrony, and positive affective exchanges. Strengthening these relational contexts may help buffer children who show early risk markers, thereby supporting more adaptive socio-emotional trajectories over time.

Taken together, the future directions outlined above provide a coherent roadmap for advancing the theoretical and empirical contributions of this dissertation. By integrating methodological innovation, cross-cultural perspectives, longitudinal approaches, mechanistic analyses, and contextual considerations, future research can generate a more comprehensive and developmentally grounded account of how infants learn to interpret and navigate the emotional landscapes of their early social worlds. Such an integrative research program would refine current models of multimodal emotion processing, as well as strengthen efforts to identify early markers of atypical developmental trajectories. In turn, this knowledge can inform the design of targeted, developmentally sensitive interventions aimed at promoting adaptive socio-emotional functioning from the earliest stages of life.

Acknowledgements

As I arrive at the final chapter of my doctoral journey, I write these words with a heart full of profound gratitude. These three years have been long and richly woven, and every step forward has been shaped by the guidance, companionship, and steadfast support of my supervisors, colleagues, friends, and family who have walked beside me.

First and foremost, I would like to express my heartfelt thanks to my supervisor, Professor Teresa Farroni. Over the past three years, you have encouraged me with patience and understanding and guided me with unwavering professionalism. Both academically and personally, you have always offered timely help, providing invaluable learning resources, fostering my growth, and gently reminding me to care for my health and daily rhythm. Your constant support for my development has been a source of comfort and warmth throughout my years of studying abroad. Being your student has been a true privilege.

I am equally grateful to my co-supervisor, Professor Judit Gervain. Throughout these years, you have consistently understood my circumstances and offered help whenever I needed it. Thank you for guiding me through new knowledge, encouraging me, and caring about my well-being. I feel truly lucky to have had the opportunity to work with you.

My sincere appreciation extends to the University of Padua, DPSS, and the Babylab, for providing an inspiring research environment. My deepest thanks go to all the faculty members and staff for their guidance and assistance.

My heartfelt thanks go to my colleagues Letizia, Laura, and Giulia, as well as all my classmates. Thank you for filling my days with warmth, encouragement, and understanding. Thank you for all the small yet beautiful moments we lived together.

My deepest gratitude goes to my family, whose unwavering love and faith carried me through every challenge, no matter the distance between us.

I am also grateful to the China Scholarship Council for its generous support, which allowed me to devote myself fully to research.

Finally, to everyone I met along the way, thank you for bringing pieces of the world into my life, and for leaving me with beauty that has made my world deeper, wider, and brighter.

Thank you all for helping me become who I am today.

References

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. *Annual Review of Psychology*, *60*(1), 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Ainsworth, M. S. (1979). Infant–mother attachment. *American Psychologist*, *34*(10), 932.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*(2), 190–201. https://doi.org/10.1037/0012-1649.36.2.190

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. *Multisensory Development*, 183–206.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory Redundancy Guides the Development of Selective Attention, Perception, and Cognition in Infancy. *Current Directions in Psychological Science*, *13*(3), 99–102. https://doi.org/10.1111/j.0963-7214.2004.00283.x

Barrera, M. E., & Maurer, D. (1981). The perception of facial expressions by the three-month-old. *Child Development*, 203–206.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G. J., Renvall, V., Deoni, S., Gasston, D., Williams, S. C. R., Johnson, M. H., Simmons, A., & Murphy, D. G. M. (2011). Early Specialization for Voice and Emotion Processing in the Infant Brain. *Current Biology*, *21*(14), 1220–1224. https://doi.org/10.1016/j.cub.2011.06.009

Bornstein, M. H. (1985). *Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis.* https://psycnet.apa.org/record/1985-97644-011

Bornstein, M. H., & Arterberry, M. E. (2003). Recognition, discrimination, and categorization of smiling by 5-month-old infants. *Developmental Science*, *6*(5), 585–599. https://doi.org/10.1111/1467-7687.00314

Bowlby, J. (1979). The bowlby-ainsworth attachment theory. *Behavioral and Brain Sciences*, *2*(4), 637–638.

Brazelton, T. B., Koslowski, B., & Main, M. (1974). *The origins of reciprocity: The early mother-infant interaction.* https://psycnet.apa.org/record/1974-22730-007

Bridgett, D. J., Gartstein, M. A., Putnam, S. P., McKay, T., Iddins, E., Robertson, C., Ramsay, K., & Rittmueller, A. (2009). Maternal and contextual influences and the effect of temperament development during infancy on parenting in toddlerhood. *Infant Behavior and Development*, *32*(1), 103–116.

Campos, J. J. (1982). Human emotions: Their new importance and their role in social referencing. 乳幼児発達臨床センター年報, *4*, 1–7.

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel Broadens the Mind. *Infancy*, *1*(2), 149–219.

https://doi.org/10.1207/S15327078IN0102_1

Campos, J. J., Campos, R. G., & Barrett, K. C. (1989). Emergent themes in the study of emotional development and emotion regulation. *Developmental Psychology*, *25*(3), 394–402. https://doi.org/10.1037/0012-1649.25.3.394

Campos, J. J., Mumme, D., Kermoian, R., & Campos, R. G. (1994). A functionalist perspective on the nature of emotion. *Japanese Journal of Research on Emotions*, *2*(1), 1–20.

Camras, L. A., Bakeman, R., Chen, Y., Norris, K., & Cain, T. R. (2006). Culture, ethnicity, and children's facial expressions: A study of European American, Mainland Chinese, Chinese American, and adopted Chinese girls. *Emotion*, *6*(1), 103.

Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant Discrimination of Naturalistic Emotional Expressions: The Role of Face and Voice. *Child Development*, *59*(3), 604–616. https://doi.org/10.2307/1130560

Cheung, C. S., Pomerantz, E. M., Wang, M., & Qu, Y. (2016). Controlling and Autonomy-Supportive Parenting in the United States and China: Beyond Children's Reports. *Child Development*, *87*(6), 1992–2007. https://doi.org/10.1111/cdev.12567

Cole, P. M., LeDonne, E. N., & Tan, P. Z. (2013). A Longitudinal Examination of Maternal Emotions in Relation to Young Children's Developing Self-Regulation. *Parenting*, *13*(2), 113–132. https://doi.org/10.1080/15295192.2012.709152

Cole, P. M., & Tan, P. Z. (2007). *Emotion socialization from a cultural perspective.* https://psycnet.apa.org/record/2006-23344-020

Darwin, C., & Darwin, S. F. (1872). *The expression of the emotions in man and animals* (Vol. 3). John Murray London. https://psycnet.apa.org/fulltext/2004-16316-000-FRM.pdf

Davidson, R. J., & Fox, N. A. (1989). Frontal brain asymmetry predicts infants' response to maternal separation. *Journal of Abnormal Psychology*, *98*(2), 127.

De Haan, M., Johnson, M. H., & Halit, H. (2013). Development of face-sensitive event-related potentials during infancy. In *Infant EEG and event-related potentials* (pp. 77–99). Psychology Press. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203759660-4/development-face-sensitive-event-related-potentials-infancy-michelle-de-haan-mark-johnson-hanife-halit

De Haan, M., & Nelson, C. A. (1999). Brain activity differentiates face and object processing in 6-month-old infants. *Developmental Psychology*, *35*(4), 1113.

DeCasper, A. J., & Fifer, W. P. (1980). Of Human Bonding: Newborns Prefer Their Mothers' Voices. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional Neuroimaging of Speech Perception in Infants. *Science*, *298*(5600), 2013–2015. https://doi.org/10.1126/science.1077066

Della Longa, L., Carnevali, L., & Farroni, T. (2023). The role of affective touch in modulating emotion processing among preschool children. *Journal of Experimental Child Psychology*, *235*, 105726. https://doi.org/10.1016/j.jecp.2023.105726

Denham, S. A. (1993). Maternal Emotional Responsiveness and Toddlers' Social-Emotional Competence. *Journal of Child Psychology and Psychiatry*, *34*(5), 715–728. https://doi.org/10.1111/j.1469-7610.1993.tb01066.x

Di Lorenzo, R., Blasi, A., Junge, C., van den Boomen, C., van Rooijen, R., & Kemner, C. (2019). Brain Responses to Faces and Facial Expressions in 5-Month-Olds: An fNIRS Study. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01240

Draghi-Lorenz, R., Reddy, V., & Costall, A. (2001). Rethinking the development of "nonbasic" emotions: A critical review of existing theories. *Developmental Review*, *21*(3), 263–304.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5%253C178::AID-EVAN5%253E3.0.CO;2-8

Eisenberg, N., Cumberland, A., & Spinrad, T. L. (1998). Parental Socialization of Emotion. *Psychological Inquiry*, *9*(4), 241–273. https://doi.org/10.1207/s15327965pli0904_1

Ekman, P. (1992a). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (1992b). *An argument for basic emotions? Cognition and Emotiont, 6 (3/4), 169-200*. Lawrence Erlbaum Associates Limited. Disponível em: https://www. paulekman ….

Ekman, P. (1994). *Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique.* https://psycnet.apa.org/record/1994-24317-001

Ekman, P., Rolls, E. T., Perrett, D. I., & Ellis, H. D. (1997). Facial expressions of emotion: An old controversy and new findings. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 63–69. https://doi.org/10.1098/rstb.1992.0008

Fantz, R. L. (1961). The origin of form perception. *Scientific American*, *204*(5), 66–73.

Fantz, R. L. (1964). Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones. *Science*, *146*(3644), 668–670. https://doi.org/10.1126/science.146.3644.668

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. https://doi.org/10.1073/pnas.152159999

Farroni, T., Massaccesi, S., Menon, E., & Johnson, M. H. (2007). Direct gaze modulates face recognition in young infants. *Cognition*, *102*(3), 396–404.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze Following in Newborns. *Infancy*, *5*(1), 39–60. https://doi.org/10.1207/s15327078in0501_2

Farroni, T., Menon ,Enrica, Rigato ,Silvia, & and Johnson, M. H. (2007). The perception of facial expressions in newborns. *European Journal of Developmental Psychology*, *4*(1), 2–13. https://doi.org/10.1080/17405620601046832

Feldman, R. (2007a). Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *Journal of Child Psychology and Psychiatry*, *48*(3–4), 329–354. https://doi.org/10.1111/j.1469-7610.2006.01701.x

Feldman, R. (2007b). Parent–Infant Synchrony: Biological Foundations and Developmental Outcomes. *Current Directions in Psychological Science*, *16*(6), 340–345. https://doi.org/10.1111/j.1467-8721.2007.00532.x

Feldman, R. (2015). The adaptive human parental brain: Implications for children's social development. *Trends in Neurosciences*, *38*(6), 387–399.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293.

Field, T. M., Woodson, R., Cohen, D., Greenberg, R., Garcia, R., & Collins, K. (1983). Discrimination and imitation of facial expressions by term and preterm neonates. *Infant Behavior and Development*, *6*(4), 485–489.

Flom, R., & and Whiteley, M. O. (2014). The dynamics of intermodal matching: Seven- and 12-month-olds' intermodal matching of affect. *European Journal of Developmental*

Psychology, *11*(1), 111–119. https://doi.org/10.1080/17405629.2013.821059

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*(1), 238–252. Scopus. https://doi.org/10.1037/0012-1649.43.1.238

Flom, R., Bahrick, L. E., & Pick, A. D. (2018). Infants Discriminate the Affective Expressions of their Peers: The Roles of Age and Familiarization Time. *Infancy*, *23*(5), 692–707. https://doi.org/10.1111/infa.12246

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications. https://books.google.com/books?hl=en&lr=&id=SfNrDwAAQBAJ&oi=fnd&pg=PP13&dq=An+R+Companion+to+Applied+Regression&ots=EqLn89anXz&sig=VIe6FTGRHUtIeOlcAEKo71rVxtM

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16), R724–R732.

Gemignani, J., & Gervain, J. (2021). Comparing different pre-processing routines for infant fNIRS data. *Developmental Cognitive Neuroscience*, *48*, 100943.

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience*, *24*(3), 564–574.

Gervain, J., Minagawa, Y., Emberson, L., & Lloyd-Fox, S. (2023). Using functional near-infrared spectroscopy to study the early developing brain: Future directions and new challenges. *Neurophotonics*, *10*(2), 023519–023519.

Gibson, E. J. (1991). Perceptual development and the reduction of uncertainty. *An Odyssey in Learning and Perception*, 353–363.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* https://psycnet.apa.org/record/1966-35026-000

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, *56*(4), 544–549.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146.

Grossman, T. (2013). The Early Development of Processing Emotions in Face and Voice. In P. Belin, S. Campanella, & T. Ethofer (Eds), *Integrating Face and Voice in Person Perception* (pp. 95–116). Springer New York. https://doi.org/10.1007/978-1-4614-3585-3_5

Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, *28*(2), 219–236. https://doi.org/10.3233/RNN-2010-0499

Grossmann, T. (2013). The role of medial prefrontal cortex in early social cognition. *Frontiers in Human Neuroscience*, *7*, 340.

Grossmann, T., & Johnson, M. H. (2007). The development of the social brain in human infancy. *European Journal of Neuroscience*, *25*(4), 909–919. https://doi.org/10.1111/j.1460-9568.2007.05379.x

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, *65*(6), 852–858.

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science*, *9*(3), 309–315. https://doi.org/10.1111/j.1467-7687.2006.00494.x

Grossmann, T., Striano, T., & Friederici, A. D. (2007). Developmental changes in infants' processing of happy and angry facial expressions: A neurobehavioral study. *Brain and Cognition*, *64*(1), 30–41. https://doi.org/10.1016/j.bandc.2006.10.002

Halberstadt, A. G. (1986). Family socialization of emotional expression and nonverbal communication styles and skills. *Journal of Personality and Social Psychology*, *51*(4), 827.

Halberstadt, A. G., Cassidy, J., Stifter, C. A., Parke, R. D., & Fox, N. A. (1995). Self-expressiveness within the family context: Psychometric support for a new measure. *Psychological Assessment*, *7*(1), 93.

Halberstadt, A. G., Crisp, V. W., & Eaton, K. L. (1999). *Family expressiveness: A retrospective and new directions for research.* https://psycnet.apa.org/record/1999-04108-004

Halberstadt, A. G., & Eaton, K. L. (2002). A Meta-Analysis of Family Expressiveness and Children's Emotion Expressiveness and Understanding. *Marriage & Family Review*, *34*(1–2), 35–62. https://doi.org/10.1300/J002v34n01_03

Haley, D. W., & Stansbury, K. (2003). Infant Stress and Parent Responsiveness: Regulation of Physiology and Behavior During Still-Face and Reunion. *Child Development*, *74*(5), 1534–1546. https://doi.org/10.1111/1467-8624.00621

Heck, A., Hock, A., White, H., Jubran, R., & Bhatt, R. S. (2016). The development of attention to dynamic facial emotions. *Journal of Experimental Child Psychology*, *147*, 100–110. https://doi.org/10.1016/j.jecp.2016.03.005

Heck, A. R., Panneton, R. K., & Mills-Smith, L. (2016). Bimodal dynamics of faces and voices influence older infants' perception of emotion. *Infancy*, *21*(4), 522–531. APA PsycInfo. https://doi.org/10.1111/infa.12121

Hiatt, S. W., Campos, J. J., & Emde, R. N. (1979). Facial patterning and infant emotional expression: Happiness, surprise, and fear. *Child Development*, 1020–1035.

Hoehl, S., & Striano, T. (2010). The development of emotional face and eye gaze processing. *Developmental Science*, *13*(6), 813–825. https://doi.org/10.1111/j.1467-7687.2009.00944.x

Hooper, D., Coughlan, J., & Mullen, M. (2008). Evaluating model fit: A synthesis of the structural equation modelling literature. *7th European Conference on Research Methodology for Business and Management Studies*, *2008*(2), 195–200. https://books.google.com/books?hl=en&lr=&id=ZZoHBAAAQBAJ&oi=fnd&pg=PA195&dq=Hooper+et+al.,+2008&ots=gX-T3yYw7a&sig=TUSE8l_yQ-T1nZ_slbXFEodXEFU

Hyde, D. C., Jones, B. L., Flom, R., & Porter, C. L. (2011). Neural signatures of face–voice synchrony in 5-month-old human infants. *Developmental Psychobiology*, *53*(4), 359–370. https://doi.org/10.1002/dev.20525

Isomura, T., & Nakano, T. (2016). Automatic facial mimicry in response to dynamic emotional stimuli in five-month-old infants. *Proceedings. Biological Sciences*, *283*(1844), 20161948. https://doi.org/10.1098/rspb.2016.1948

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes.

Psychological Review, *100*(1), 68.

Izard, C. E. (1994). *Innate and universal facial expressions: Evidence from developmental and cross-cultural research.* https://psycnet.apa.org/fulltext/1994-24321-001.html

Izard, C. E., Haynes, O. M., Chisholm, G., & Baak, K. (1991). Emotional Determinants of Infant-Mother Attachment. *Child Development*, *62*(5), 906. https://doi.org/10.2307/1131142

Jessen, S., Altvater-Mackensen, N., & Grossmann, T. (2016). Pupillary responses reveal infants' discrimination of facial emotions independent of conscious perception. *Cognition*, *150*, 163–169. https://doi.org/10.1016/j.cognition.2016.02.010

Jessen, S., & Grossmann, T. (2014). Unconscious discrimination of social cues from eye whites in infants. *Proceedings of the National Academy of Sciences*, *111*(45), 16208–16213. https://doi.org/10.1073/pnas.1411333111

Jessen, S., & Grossmann, T. (2015). Neural signatures of conscious and unconscious emotional face processing in human infants. *Cortex*, *64*, 260–270. https://doi.org/10.1016/j.cortex.2014.11.007

Jessen, S., & Grossmann, T. (2016). The developmental emergence of unconscious fear processing from eyes during infancy. *Journal of Experimental Child Psychology*, *142*, 334–343. https://doi.org/10.1016/j.jecp.2015.09.009

Johnson, M. H. (2011). Face perception: A developmental perspective. *The Oxford Handbook of Face Perception*, 3–14.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Johnson, M. H., Grossmann, T., & Kadosh, K. C. (2009). Mapping functional brain development: Building a social brain through interactive specialization. *Developmental Psychology*, *45*(1), 151.

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The Role of Person Familiarity in Young Infants' Perception of Emotional Expressions. *Child Development*, *72*(2), 352–369. https://doi.org/10.1111/1467-8624.00283

Kaneshige, T., & Haryu, E. (2015). Categorization and understanding of facial expressions in 4-month-old infants. *Japanese Psychological Research*, *57*(2), 135–142. https://doi.org/10.1111/jpr.12075

Kirby, G. (2024). Early experiences matter: Fostering positive relationships and secure attachments with babies and toddlers. *Early Years Educator*, *24*(6), 16–17. https://doi.org/10.12968/eyed.2024.24.6.16

Klinnert, M. D., Emde, R. N., Butterfield, P., & Campos, J. J. (1986). Social referencing: The infant's use of emotional signals from a friendly adult with mother present. *Developmental Psychology*, *22*(4), 427.

Kotsoni, E., de Haan, M., & Johnson, M. H. (2001). Categorical Perception of Facial Expressions by 7-Month-Old Infants. *Perception*, *30*(9), 1115–1125. https://doi.org/10.1068/p3155

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *Neuroimage*, *37*(4), 1445–1456.

Lamb, M. E., Thompson, R. A., Gardner, W., & Charnov, E. L. (2013). *Infant-mother attachment: The origins and developmental significance of individual differences in* *strange situation behavior*. Routledge. https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780203781791&type=googlepdf

LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. Simon and Schuster. https://books.google.com/books?hl=en&lr=&id=7EJN5I8sk2wC&oi=fnd&pg=PA9&dq=The+Emotional+Brain.+New+York:+Simon+%26+Schuster.&ots=09aiazDo-J&sig=DNUTB2kekzF7Cxnf53RqzwD1ZgU

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research*, *69*(1), 22–29.

Leppänen, J. M., Kauppinen, P., Peltola, M. J., & Hietanen, J. K. (2007). Differential electrocortical responses to increasing intensities of fearful and happy emotional expressions. *Brain Research*, *1166*, 103–109.

Leppänen, J. M., Moulson, M. C., Vogel-Farley, V. K., & Nelson, C. A. (2007). An ERP Study of Emotional Face Processing in the Adult and Infant Brain. *Child Development*, *78*(1), 232–245. https://doi.org/10.1111/j.1467-8624.2007.00994.x

Leppänen, J. M., & Nelson, C. A. (2009). Tuning the developing brain to social signals of emotions. *Nature Reviews Neuroscience*, *10*(1), 37–47.

Lewis, M. (2000). The emergence of human emotions. *Handbook of Emotions*, *2*, 265–280.

Lewis, M. (2007). Self-conscious emotional development. *The Self-Conscious Emotions: Theory and Research*, 134–149.

Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (2010). *Handbook of emotions*. Guilford Press.

https://books.google.com/books?hl=en&lr=&id=uIQQskejGwUC&oi=fnd&pg=PR1&dq=+Lewis,+Haviland-Jones,+%26+Barrett,+2010&ots=3RWpTOlcw6&sig=3RiZSmofndEguqoimgD-2icoxOg

Lewis, M., Sullivan, M. W., Stanger, C., & Weiss, M. (1989). Self development and self-conscious emotions. *Child Development*, 146–156.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience & Biobehavioral Reviews*, *34*(3), 269–284.

Lloyd-Fox, S., Blasi, A., Pasco, G., Gliga, T., Jones, E. J. H., Murphy, D. G. M., Elwell, C. E., Charman, T., Johnson, M. H., & the BASIS Team. (2018). Cortical responses before 6 months of life associate with later autism. *European Journal of Neuroscience*, *47*(6), 736–749. https://doi.org/10.1111/ejn.13757

Malatesta, C. Z., & Haviland, J. M. (1982). Learning display rules: The socialization of emotion expression in infancy. *Child Development*, 991–1003.

Marks, L. E. (1978). Multimodal perception. In *Perceptual coding* (pp. 321–339). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780121619084500152

Martinez-Alvarez, A., Gervain, J., Koulaguina, E., Pons, F., & De Diego-Balaguer, R. (2023). Prosodic cues enhance infants' sensitivity to nonadjacent regularities. *Science Advances*, *9*(15), eade4083. https://doi.org/10.1126/sciadv.ade4083

Melanie, S., Reimchen ,Melissa, Sauter ,Disa, & and Morgan, J. L. (2017). Do infants discriminate non-linguistic vocal expressions of positive emotions? *Cognition and Emotion*, *31*(2), 298–311. https://doi.org/10.1080/02699931.2015.1108904

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 702–709.

Mesman, J., van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2009). The many faces of the Still-Face Paradigm: A review and meta-analysis. *Developmental Review*, *29*(2), 120–162.

Minagawa-Kawai, Y., Matsuoka, S., Dan, I., Naoi, N., Nakamura, K., & Kojima, S. (2009). Prefrontal activation associated with social attachment: Facial-emotion recognition in mothers and infants. *Cerebral Cortex*, *19*(2), 284–292.

Missana, M., Grigutsch, M., & Grossmann, T. (2014). Developmental and individual differences in the neural processing of dynamic expressions of pain and anger. *PloS One*, *9*(4), e93728.

Missana, M., Rajhans, P., Atkinson, A. P., & Grossmann, T. (2014). Discrimination of fearful and happy body postures in 8-month-old infants: An event-related potential study. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00531

Montague, D. P. F., & Walker-Andrews, A. S. (2002). Mothers, fathers, and infants: The role of person familiarity and parental involvement in infants' perception of emotion expressions. *Child Development*, *73*(5), 1339–1352. APA PsycInfo. https://doi.org/10.1111/1467-8624.00475

Nakato, E., Otsuka, Y., Kanazawa, S., Yamaguchi, M. K., Watanabe, S., & Kakigi, R. (2009). When do infants differentiate profile face from frontal face? A near-infrared spectroscopic study. *Human Brain Mapping*, *30*(2), 462–472. https://doi.org/10.1002/hbm.20516

Nawrot, E. S. (2003). The Perception of Emotional Expression in Music: Evidence from Infants, Children and Adults. *Psychology of Music*, *31*(1), 75–92. https://doi.org/10.1177/0305735603031001325

Nelson, C. A. (1987). The Recognition of Facial Expressions in the First Two Years of Life: Mechanisms of Development. *Child Development*, *58*(4), 889–909. https://doi.org/10.2307/1130530

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, *10*(1–2), 3–18. https://doi.org/10.1002/icd.239

Nelson, C. A., & De Haan, M. (1996). Neural correlates of infants' visual responsiveness to facial expressions of emotion. *Developmental Psychobiology*, *29*(7), 577–595. https://doi.org/10.1002/(SICI)1098-2302(199611)29:7%253C577::AID-DEV3%253E3.0.CO;2-R

Nelson, C. A., & McCleery, J. P. (2008). Use of event-related potentials in the study of typical and atypical development. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(11), 1252–1261.

Ogren, M., & and Johnson, S. P. (2020). Intermodal emotion matching at 15 months, but not 9 or 21 months, predicts early childhood emotion understanding: A longitudinal investigation. *Cognition and Emotion*, *34*(7), 1343–1356. https://doi.org/10.1080/02699931.2020.1743236

Ogren, M., Burling, J. M., & Johnson, S. P. (2018). Family expressiveness relates to happy emotion matching among 9-month-old infants. *Journal of Experimental Child Psychology*, *174*, 29–40. https://doi.org/10.1016/j.jecp.2018.05.003

Otte, R. A., Donkers, F. C. L., Braeken, M. a. K. A., & Van den Bergh, B. R. H. (2015). Multimodal processing of emotional information in 9-month-old infants II: Prenatal exposure to maternal anxiety. *Brain and Cognition*, *95*, 107–117. https://doi.org/10.1016/j.bandc.2014.12.001

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Bmj*, *372*. https://www.bmj.com/content/372/bmj.n71.short

Palama, A., Malsert, J., & Gentaz, E. (2018). Are 6-month-old human infants able to transfer emotional information (happy or angry) from voices to faces? An eye-tracking study. *PLoS ONE*, *13*(4). APA PsycInfo. https://doi.org/10.1371/journal.pone.0194579

Palama, A., Malsert, J., Grandjean, D., Sander, D., & Gentaz, E. (2022). The cross-modal transfer of emotional information from voices to faces in 5-, 8- and 10-year-old children and adults: An eye-tracking study. *Emotion (Washington, D.C.)*, *22*(4), 725–739. https://doi.org/10.1037/emo0000758

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford university press. https://books.google.com/books?hl=en&lr=&id=qqcRGagyEuAC&oi=fnd&pg=PR13&dq=Affective+neuroscience:+The+foundations+of+human+and+animal+emotions.+1998&ots=-QL41-ziqA&sig=ucIRVDPfAZui7e35kAsMyK2Kqjk

Parker, S. W., Nelson, C. A., & The Bucharest Early Intervention Project Core Group. (2005). The Impact of Early Institutional Rearing on the Ability to Discriminate Facial Expressions of Emotion: An Event-Related Potential Study. *Child Development*, *76*(1), 54–72. https://doi.org/10.1111/j.1467-8624.2005.00829.x

Pascalis, O., De Haan, M., & Nelson, C. A. (2002). Is Face Processing Species-Specific During the First Year of Life? *Science*, *296*(5571), 1321–1323. https://doi.org/10.1126/science.1070223

Pelaez, M., Virues-Ortega, J., & Gewirtz, J. L. (2012). ACQUISITION OF SOCIAL REFERENCING VIA DISCRIMINATION TRAINING IN INFANTS. *Journal of Applied Behavior Analysis*, *45*(1), 23–36. https://doi.org/10.1901/jaba.2012.45-23

Peltola, M. J., Hietanen, J. K., Forssman, L., & Leppänen, J. M. (2013). The Emergence and Stability of the Attentional Bias to Fearful Faces in Infancy. *Infancy*, *18*(6), 905–926. https://doi.org/10.1111/infa.12013

Peltola, M. J., Leppänen, J. M., Mäki, S., & Hietanen, J. K. (2009). Emergence of enhanced attention to fearful faces between 5 and 7 months of age. *Social Cognitive and Affective Neuroscience*, *4*(2), 134–142. https://doi.org/10.1093/scan/nsn046

Peltola, M. J., Leppänen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful faces modulate looking duration and attention disengagement in 7-month-old infants. *Developmental Science*, *11*(1), 60–68. https://doi.org/10.1111/j.1467-7687.2007.00659.x

Pérez-Edgar, K., Vallorani, A., Buss, K. A., & LoBue, V. (2020). Individual differences in infancy research: Letting the baby stand out from the crowd. *Infancy*, *25*(4), 438–457. https://doi.org/10.1111/infa.12338

Phillips, R. D., Wagner, S. H., Fells, C. A., & Lynch, M. (1990). Do infants recognize emotion in facial expressions? Categorical and 'metaphorical' evidence. *Infant Behavior & Development*, *13*(1), 71–84. APA PsycInfo. https://doi.org/10.1016/0163-6383(90)90006-T

Pollak, S. D., Messner, M., Kistler, D. J., & Cohn, J. F. (2009). Development of perceptual expertise in emotion recognition. *Cognition*, *110*(2), 242–247.

Poncet, F., Leleu, A., Rekow, D., Damon, F., Dzhelyova, M. P., Schaal, B., Durand, K., Faivre, L., Rossion, B., & Baudouin, J.-Y. (2022). A neural marker of rapid discrimination of facial expression in 3.5- and 7-month-old infants. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.901013

Ravicz, M. M., Perdue, K. L., Westerlund, A., Vanderwert, R. E., & Nelson, C. A. (2015). Infants' neural responses to facial emotion in the prefrontal cortex are correlated with temperament: A functional near-infrared spectroscopy study. *Frontiers in Psychology*, *6*, 922.

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*(5), 704.

Rothermund, K., Gast, A., & Wentura, D. (2011). Incongruency effects in affective processing: Automatic motivational counter-regulation or mismatch-induced salience? *Cognition and Emotion*, *25*(3), 413–425. https://doi.org/10.1080/02699931.2010.537075

Ruba, A. L., & Repacholi, B. M. (2020). Do Preverbal Infants Understand Discrete Facial Expressions of Emotion? *Emotion Review*, *12*(4), 235–250. https://doi.org/10.1177/1754073919871098

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102.

Russell, L., Henrik, S., Jonathon, L., Paul, B., & Maxime, H. (2018). Estimated marginal means, aka least-squares means. *The American Statistician*, *34*, 216–221.

Saarni, C. (1999). *A Skill-Based Model of Emotional Competence: A Developmental Perspective.* https://eric.ed.gov/?id=ED430678

Saarni, C. (2000). Emotional competence: A developmental perspective. In *The handbook of emotional intelligence: Theory, development, assessment, and application at home, school, and in the workplace* (pp. 68–91). Jossey-Bass/Wiley.

Saarni, C. (2008). The interface of emotional development with social context. *Handbook of Emotions*, *3*, 332–347.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Development of hemispheric specialization for lexical pitch–accent in Japanese infants. *Journal of Cognitive Neuroscience*, *22*(11), 2503–2513.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Schwartz, G. M., Izard, C. E., & Ansul, S. E. (1985). The 5-month-old's ability to discriminate facial expressions of emotion. *Infant Behavior and Development*, *8*(1), 65–77.

Serrano, J. M., Iglesias, J., & Loeches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear, and surprise in 4- to 6-month-old infants. *Developmental Psychobiology*, *25*(6), 411–425. https://doi.org/10.1002/dev.420250603

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' Listening Preferences: Baby Talk or Happy Talk? *Infancy*, *3*(3), 365–394. https://doi.org/10.1207/S15327078IN0303_5

Soken, N. H., & Pick, A. D. (1992). Intermodal Perception of Happy and Angry Expressive Behaviors by Seven-Month-Old Infants. *Child Development*, *63*(4), 787–795. https://doi.org/10.1111/j.1467-8624.1992.tb01661.x

Soken, N. H., & Pick, A. D. (1999). Infants' Perception of Dynamic Affective Expressions: Do Infants Distinguish Specific Expressions? *Child Development*, *70*(6), 1275–1282. https://doi.org/10.1111/1467-8624.00093

Song, L., Spier, E. T., & Tamis-Lemonda, C. S. (2014). Reciprocal influences between maternal language and children's language and cognitive development in low-income families. *Journal of Child Language*, *41*(2), 305–326.

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

Spelke, E. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, *8*(4), 553–560.

Spinrad, T. L., & Eisenberg, N. (2019). Socialization of moral emotions and behavior. *The Oxford Handbook of Parenting and Moral Development*, 57–72.

Sroufe, L. A. (1982). The organization of emotional development. *Psychoanalytic Inquiry*, *1*(4), 575–599. https://doi.org/10.1080/07351698209533421

Sroufe, L. A. (1997). *Emotional development: The organization of emotional life in the early years*. Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=w9EDGhyXYGoC&oi=fnd&pg=PR11&dq=Emotional+development:+The+organization+of+emotional+life+in+the+early+years&ots=mLIvF9UFKW&sig=AaDumT2Qs1CQ6dMxS_1lr_PcxTQ

Tang, Y., Triesch, J., & Deák, G. O. (2023). Variability in infant social responsiveness: Age and situational differences in attention-following. *Developmental Cognitive Neuroscience*, *63*, 101283.

Taylor-Colls, S., & Pasco Fearon, R. M. (2015). The Effects of Parental Behavior on Infants' Neural Processing of Emotion Expressions. *Child Development*, *86*(3), 877–888. https://doi.org/10.1111/cdev.12348

Tomasello, M. (2003). The key is social cognition. *Language in Mind: Advances in the Study of Language and Thought*, 47–57.

Tomasello, M. (2014). Joint attention as social cognition. In *Joint attention* (pp. 103–130). Psychology Press. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315806617-6/joint-attention-social-cognition-michael-tomasello

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry*, *17*(1), 1–13.

Tronick, E. D., Als, H., & Brazelton, T. B. (1977). Mutuality in mother–infant interaction. *Journal of Communication*. https://psycnet.apa.org/record/1979-08081-001

Tronick, E. Z., & Cohn, J. F. (1989). Infant-mother face-to-face interaction: Age and gender differences in coordination and the occurrence of miscoordination. *Child Development*, 85–92.

Tsui, A. S. M., Ma, Y. K., Ho, A., Chow, H. M., & Tseng, C. (2016). Bimodal emotion congruency is critical to preverbal infants' abstract rule learning. *Developmental Science*, *19*(3), 382–393. https://doi.org/10.1111/desc.12319

Urakawa, S., Takamoto, K., Ishikawa, A., Ono, T., & Nishijo, H. (2015). Selective medial prefrontal cortex responses during live mutual gaze interactions in human infants: An fNIRS study. *Brain Topography*, *28*(5), 691–701. APA PsycInfo. https://doi.org/10.1007/s10548-014-0414-2

Vaillant-Molina, M., & Bahrick, L. E. (2012). The role of intersensory redundancy in the emergence of social referencing in 5½-month-old infants. *Developmental Psychology*, *48*(1), 1–9. https://doi.org/10.1037/a0025263

Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young Infants Match Facial and Vocal Emotional Expressions of Other Infants. *Infancy*, *18*(s1), E97–E111. https://doi.org/10.1111/infa.12017

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383.

Vogel, M., Monesson, A., & Scott, L. S. (2012). Building biases in infancy: The influence of race on face and voice emotion matching. *Developmental Science*, *15*(3), 359–372. https://doi.org/10.1111/j.1467-7687.2012.01138.x

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Walden, T. A., & Ogan, T. A. (1988). The development of social referencing. *Child Development*, 1230–1240.

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, *33*(3), 514–535. https://doi.org/10.1016/0022-0965(82)90063-7

Walker-Andrews, A. S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? *Developmental Psychology*, *22*(3), 373–377. https://doi.org/10.1037/0012-1649.22.3.373

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*(3), 437.

Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development*, *6*(4), 491–498.

Walker-Andrews, A. S., Krogh-Jespersen, S., Mayhew, E. M. Y., & Coffield, C. N. (2011). Young infants' generalization of emotional expressions: Effects of familiarity. *Emotion*, *11*(4), 842–851. https://doi.org/10.1037/a0024435

Walker-Andrews, A. S., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior & Development*, *14*(2), 131–142. APA PsycInfo. https://doi.org/10.1016/0163-6383(91)90001-9

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Weinberg, M. K., & Tronick, E. Z. (1996). Infant Affective Reactions to the Resumption of Maternal Interaction after the Still-Face. *Child Development*, *67*(3), 905. https://doi.org/10.2307/1131869

White, H., Chroust, A., Heck, A., Jubran, R., Galati, A., & Bhatt, R. S. (2019). Categorical Perception of Facial Emotions in Infancy. *Infancy*, *24*(2), 139–161. https://doi.org/10.1111/infa.12275

White, M. (1995). Preattentive analysis of facial expressions of emotion. *Cognition & Emotion*, *9*(5), 439–460. https://doi.org/10.1080/02699939508408975

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

Xiao, N. G., & Emberson, L. L. (2019). Infants use knowledge of emotions to augment face perception: Evidence of top-down modulation of perception early in life. *Cognition*, *193*. APA PsycInfo. https://doi.org/10.1016/j.cognition.2019.104019

Xie, W., McCormick, S. A., Westerlund, A., Bowman, L. C., & Nelson, C. A. (2019). Neural correlates of facial emotion processing in infancy. *Developmental Science*, *22*(3), e12758. https://doi.org/10.1111/desc.12758

Yong, M. H., & Ruffman, T. (2016). Domestic dogs and human infants look more at happy and angry faces than sad faces. *Multisensory Research*, *29*(8), 749–771. APA PsycInfo. https://doi.org/10.1163/22134808-00002535

Zadra, J. R., & Clore, G. L. (2011). Emotion and perception: The role of affective information. *WIREs Cognitive Science*, *2*(6), 676–685. https://doi.org/10.1002/wcs.147

Zhang, D., Liu, Y., Hou, X., Sun, G., Cheng, Y., & Luo, Y. (2014). Discrimination of fearful and angry emotional voices in sleeping human neonates: A study of the mismatch brain responses. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00422

Zhang, D., Zhou, Y., Hou, X., Cui, Y., & Zhou, C. (2017). Discrimination of emotional prosodies in human neonates: A pilot fNIRS study. *Neuroscience Letters*, *658*, 62–66. https://doi.org/10.1016/j.neulet.2017.08.047

Appendix

a. Scripts used in Study 2 and Study 3

	Group	Content
1	[24, 45, 41]	Quand'è? \ Che sorpresa! \ Sei piccolo!
2	[32, 16, 7]	Perché? \ Aiutami ora \ Chiudi la porta
3	[1, 26, 18]	Vieni qui \ Perche fai questo? \ Arrivederci!
4	[2, 25, 54]	Siediti \ Quale di questo vuoi? \ Attento!
5	[3, 19, 49]	Mangialo \ Che cos'é questo? \ Non credo
6	[4, 22, 38]	Guardami \ Sei sicuro? \ Che coincidenza!
7	[5, 23, 36]	Andiamo \ Chi sei? \ Che giornata!
8	[9, 29, 52]	Aspettami \ Che ore sono? \ Che grande!
9	[11, 17, 42, 48]	Guarda \ Sono qui \ Non dimenticarmi! \ Grazie!
10	[14, 50, 37]	Buongiorno! \ Magari domani \ Quanto é piccolo!
11	[20, 31, 46]	Dove siete? \ Sei assonnato? \ Che sorriso!
12	[27, 34, 55]	Ti piace questo? \ Dove é mamma? \ Presta attenzione!

b. Example data structure for one representative participant

LookingLeft. onset	LookingLeft. offset	Duration	LookingLeft (Face1	Match	Looking Happy	Looking Angry	Looking Neutral	LookingRigh t.ordinal	LookingRigh t.onset	LookingRigh t.offset	Duration	LookingRight (Face2	Match	Looking Happy	Looking Angry	Looking Neutral
17787	18810	1023	T1 CH3A3 (A	N	1023			0	16269	17688	1419	T1 CH3A3 (A	Y		1419	
19998	21252	1254	T1 CH3A3 (A	N	1254			1	18843	19899	1056	T1 CH3A3 (A	Y		1056	
22638	23397	759	T1 CH3A3 (A	N	759			2	21285	22143	858	T1 CH3A3 (A	Y		858	
27984	29700	1716	T2 EA4H4 (A	Y		1716		3	22374	22605	231	T1 CH3A3 (A	Y		231	
32439	34485	2046	T2 EA4H4 (A	Y		2046		4	23430	23727	297	T1 CH3A3 (A	Y		297	
37158	38280	1122	T3 CN1A1 (N	Y			1122	5	27027	27918	891	T2 EA4H4 (A	N	891		
42141	45540	3399	T3 CN1A1 (N	Y			3399	6	29733	30954	1221	T2 EA4H4 (A	N	1221		
48048	49467	1419	T4 EH2N2 (H	Y	1419			7	31647	32406	759	T2 EA4H4 (A	N	759		
52635	53460	825	T4 EH2N2 (H	Y	825			8	38313	42042	3729	T3 CN1A1 (N	N		3729	
59400	60687	1287	T5 CA5H5 (A	Y		1287		9	49500	52602	3102	T4 EH2N2 (H	N			3102
62172	63228	1056	T5 CA5H5 (A	Y		1056		10	55011	55704	693	T4 EH2N2 (H	N			693
64317	65802	1485	T5 CA5H5 (A	Y		1485		11	60720	62073	1353	T5 CA5H5 (A	N	1353		
69498	71709	2211	T6 EN6H6 (H	N			2211	12	63327	64284	957	T5 CA5H5 (A	N	957		
73161	78243	5082	T6 EN6H6 (H	N			5082	13	65967	67023	1056	T5 CA5H5 (A	N	1056		
80718	83061	2343	T7 CH7N7 (H	Y	2343			14	71742	73062	1320	T6 EN6H6 (H	Y	1320		
84579	84909	330	T7 CH7N7 (H	Y	330			15	83094	84447	1353	T7 CH7N7 (H	N			1353
85338	87186	1848	T7 CH7N7 (H	Y	1848			16	84942	85305	363	T7 CH7N7 (H	N			363
91212	91773	561	T8 EH8A8 (A	N	561			17	87219	88704	1485	T7 CH7N7 (H	N			1485
92928	94083	1155	T8 EH8A8 (A	N	1155			18	91872	92829	957	T8 EH8A8 (A	Y		957	
94710	97680	2970	T8 EH8A8 (A	N	2970			19	94116	94677	561	T8 EH8A8 (A	Y		561	
98637	99198	561	T8 EH8A8 (A	N	561			20	97713	98604	891	T8 EH8A8 (A	Y		891	
99561	99594	33	T8 EH8A8 (A	N	33			21	99231	99528	297	T8 EH8A8 (A	Y		297	
102003	104577	2574	T9 CH9N9 (N	N	2574			22	104610	105105	495	T9 CH9N9 (N	Y			495
105138	105501	363	T9 CH9N9 (N	N	363			23	107118	108702	1584	T9 CH9N9 (N	Y			1584
106326	107019	693	T9 CH9N9 (N	N	693			24	115368	116391	1023	T10 EA10N10 (A	N			1023
114246	115137	891	T10 EA10N10 (A	Y		891		25	117150	118437	1287	T10 EA10N10 (A	N			1287
116424	117117	693	T10 EA10N10 (A	Y		693		26	128040	128601	561	T11 CA11H11 (H	Y	561		
118602	118767	165	T10 EA10N10 (A	Y		165		27	136851	137709	858	T12 EN12H12 (N	N	858		
125367	128007	2640	T11 CA11H11 (H	N		2640		28	147279	149556	2277	T13 EH1N1 (H	N			2277
128634	129525	891	T11 CA11H11 (H	N		891		29	171501	172392	891	T15 EA3N3 (N	Y			891
136587	136752	165	T12 EN12H12 (N	Y			165	30	176418	177870	1452	T15 EA3N3 (N	Y			1452

Trial	Looking Match	Looking Happy	Looking Angry	Looking Neutral	Looking Mismatch	Looking Happy	Looking Angry	Looking Neutral
T1CH3A3 (A	3861		3861		3036	3036		
T2EA4H4 (A	3762		3762		2871	2871		
T3CN1A1 (N	4521			4521	3729		3729	
T4EH2N2 (H	2244	2244			3795			3795
T5CA5H5 (A	3828		3828		3366	3366		
T6EN6H6 (H	1320	1320			7293			7293
T7CH7N7 (H	4521	4521			3201			3201
T8EH8A8 (A	2706		2706		5280	5280		
T9CH9N9 (N	2079			2079	3630	3630		
T10EA10N10 (A	1749		1749		2310			2310
T11CA11H11 (H	561	561			3531		3531	
T12EN12H12 (N	1650			1650	858	858		
T13EH1N1 (H	726	726			2277			2277
T14CA2H2 (A	3168		3168		0			
T15EA3N3 (N	2343			2343	6072		6072	
T16CN4A4 (N	3267			3267	825		825	
T17EH5A5 (H	1881	1881			4752		4752	
T18CN6H6 (N	4620			4620	2013	2013		
T19EN7A7 (A	0				0			
T20CA8N8 (N	693			693	1980		1980	
T21EH9A9 (H	759	759			825		825	
T22CN10A10 (N	4917			4917	2211		2211	
T23EN11A11 (A	0				0			
T24CN12H12 (H	2376	2376			3795			3795
Sum	57552	14388	19074	24090	67650	21054	23925	22671
Ave	2398							
	Total looking time							
Angry	42999							
Happy	35442							
Neutral	46761							

c. How we calculated the EMS value

The EMS was computed as the difference in the average proportion of time infants spent looking at the target image in matching versus non-matching conditions, according to the following formula:

Emotion Matching Score (EMS) = *Average fraction of time looking at the target image when it matches the vocal emotion* − *Average fraction of time looking at the same image when it does not match the vocal emotion*

For each trial, i where the target image matched the vocal emotion, the total duration of fixation time was defined as: $t_{total}(i) = t_{matching}(i) + t_{mismatching}(i)$. Where: $t_{matching}(i)$ means Time spent looking at the target image during the matching condition, $t_{mismatching}(i)$ means Time spent looking at the target image during the mismatching condition, the fraction of time spent looking at the target image in the matching condition was calculated as: $fr_{match}(i) = t_{matching}(i) / t_{total}(i)$

For trials, j where the target image did not match the vocal emotion, the fraction of time spent looking at the target image was defined as: $fr_{mismatch}(j) = t_{mismatching}(j) / t_{total}(j)$. The EMS was then calculated as follows:

$$EMS = \frac{1}{N_1} \sum_{i=1}^{N_1} fr_{match}\,(i) - \frac{1}{N_2} \sum_{j=1}^{N_2} fr_{mismatch}\,(j),$$

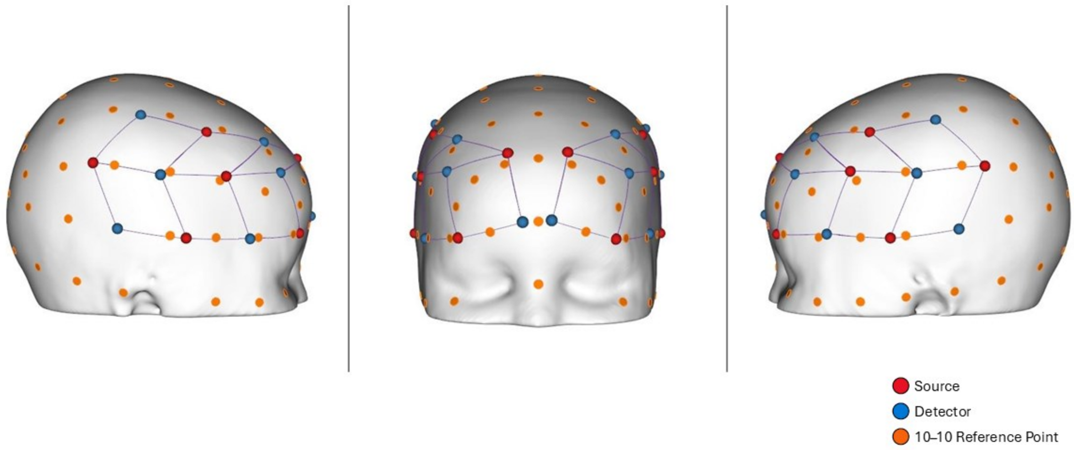
Where: N_1 = Number of trials under the matching condition; N_2 = Number of trials under the mismatching condition

For instance, if an infant viewed a pair of facial expressions, including a 'Happy' face and another emotion, and the vocal stimulus conveyed a 'Happy' emotion, the infant looked at the 'Happy' face for 60% of the time. When the vocal stimulus conveyed a different emotion, the infant looked at the 'Happy' face for 20% of the time. The FDS for 'Happy' would be calculated as:

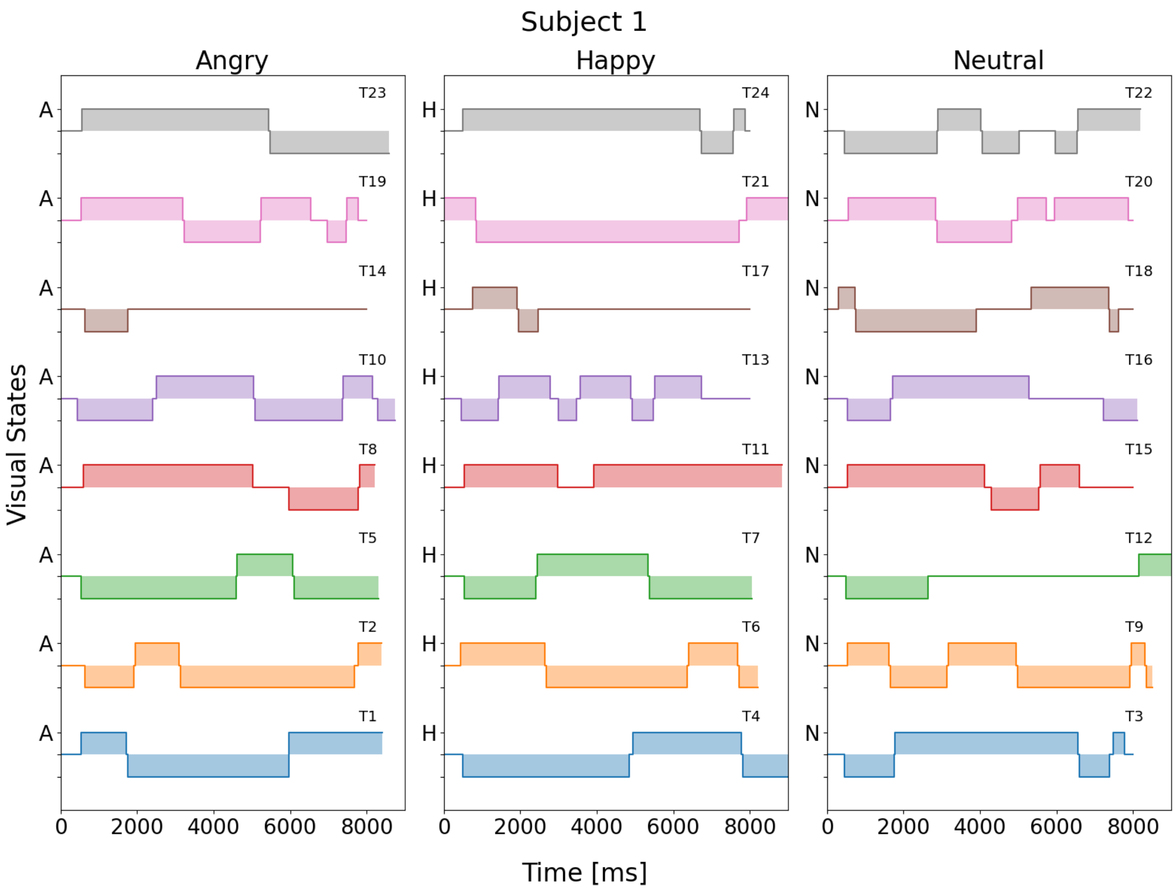
EMS = 0.6 - 0.2 = 0.4

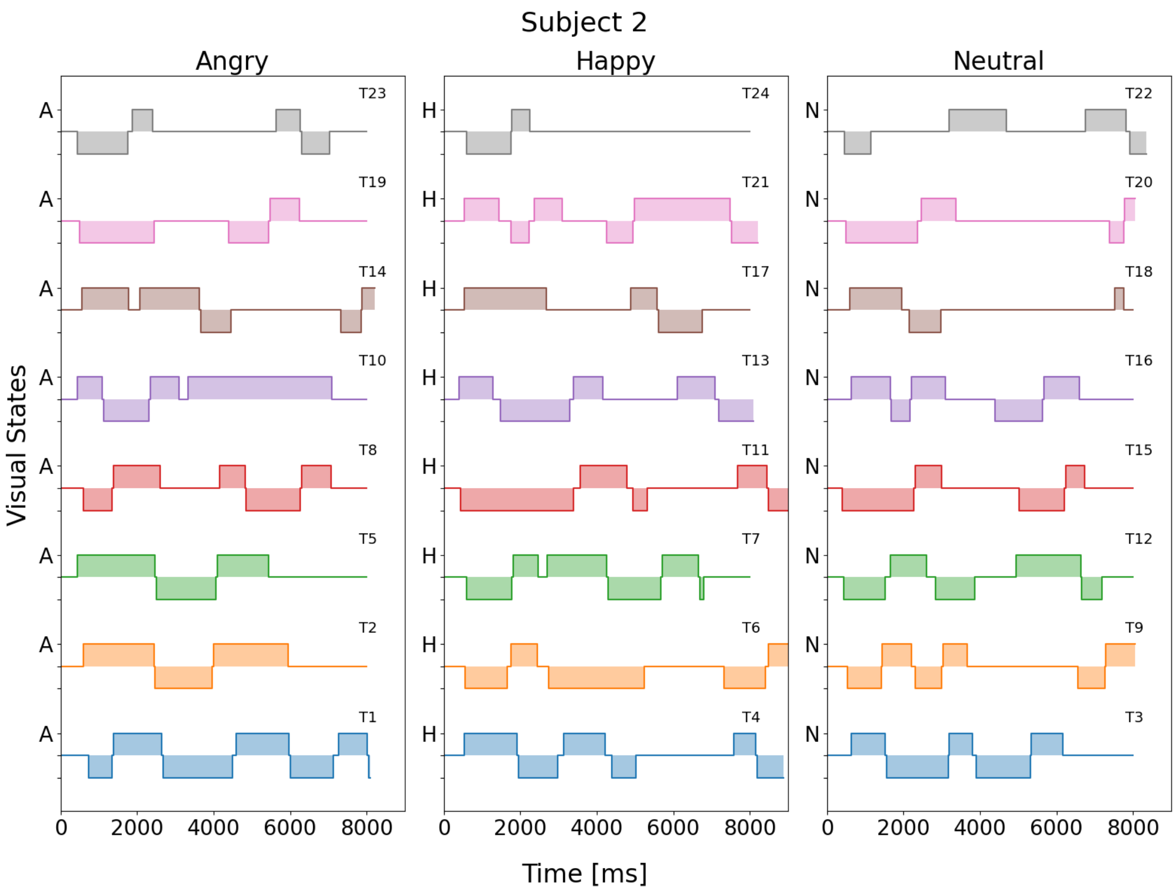
A positive EMS value (e.g., +0.4) indicates that the infant spent more time attending to the facial expression that matched the vocal emotion compared to when it did not match, thereby reflecting sensitivity to emotional congruency across modalities.

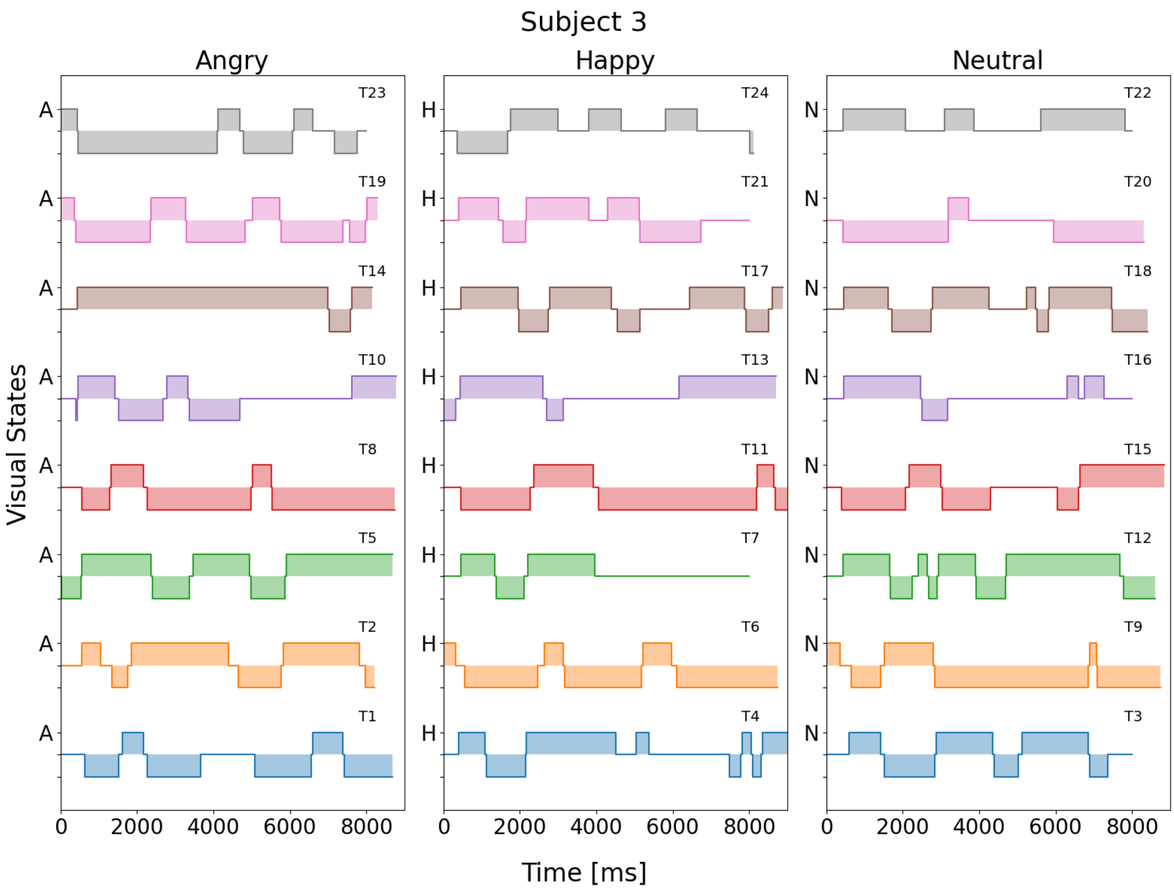
d. Montage used for Study 3

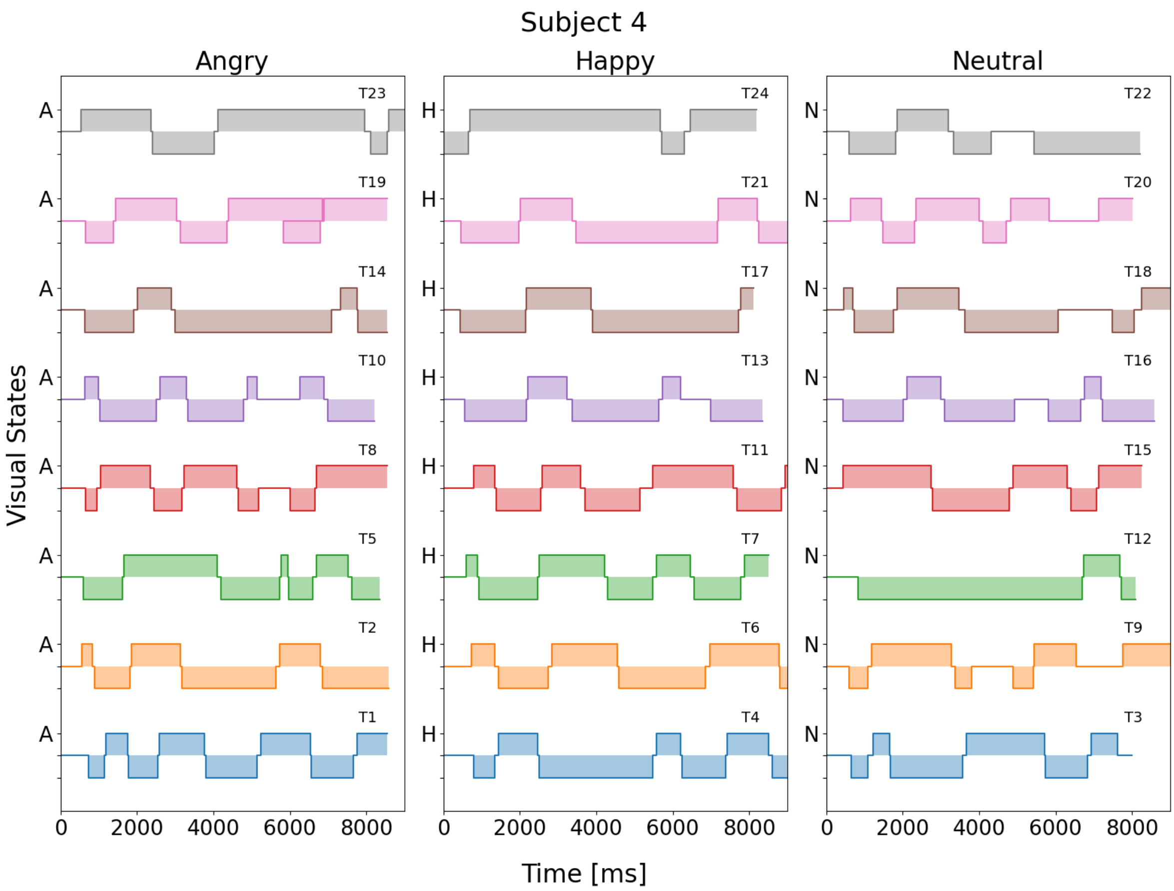


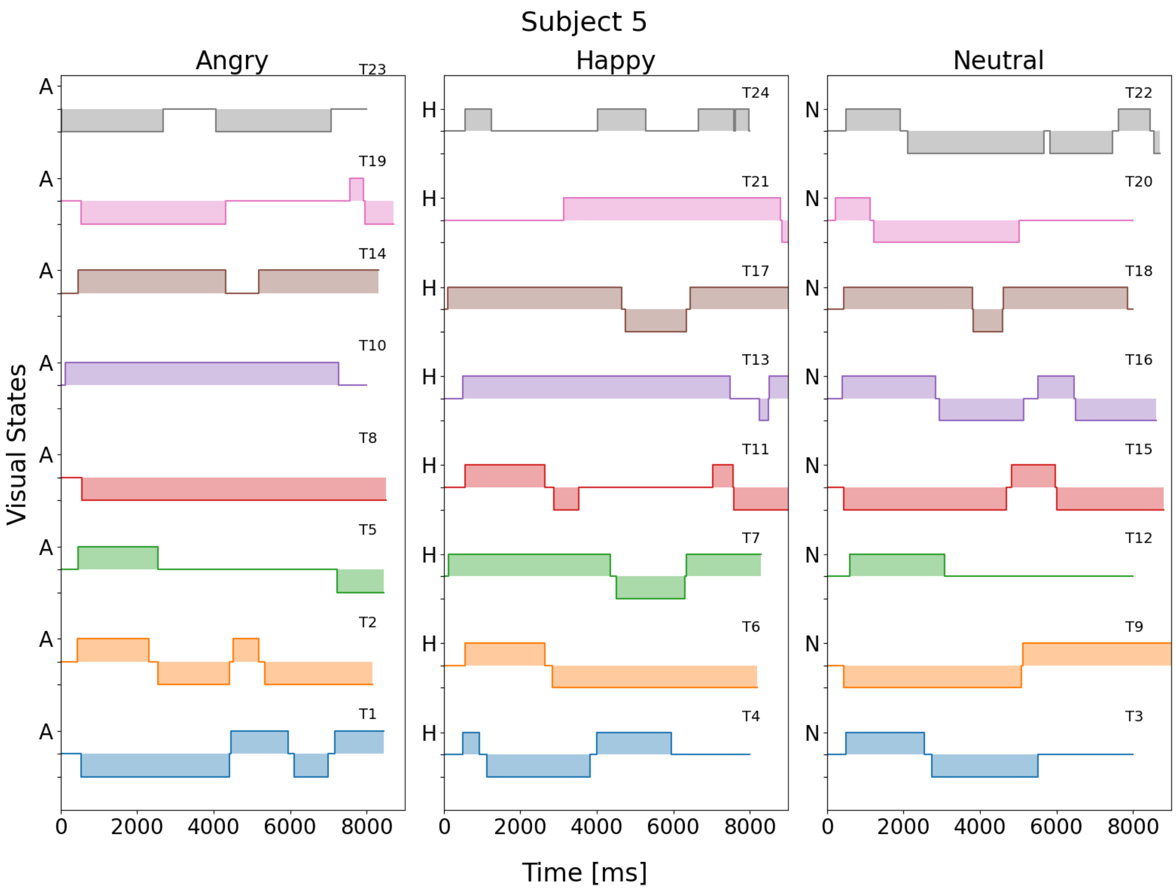
e. Individual plots of the emotion matching task

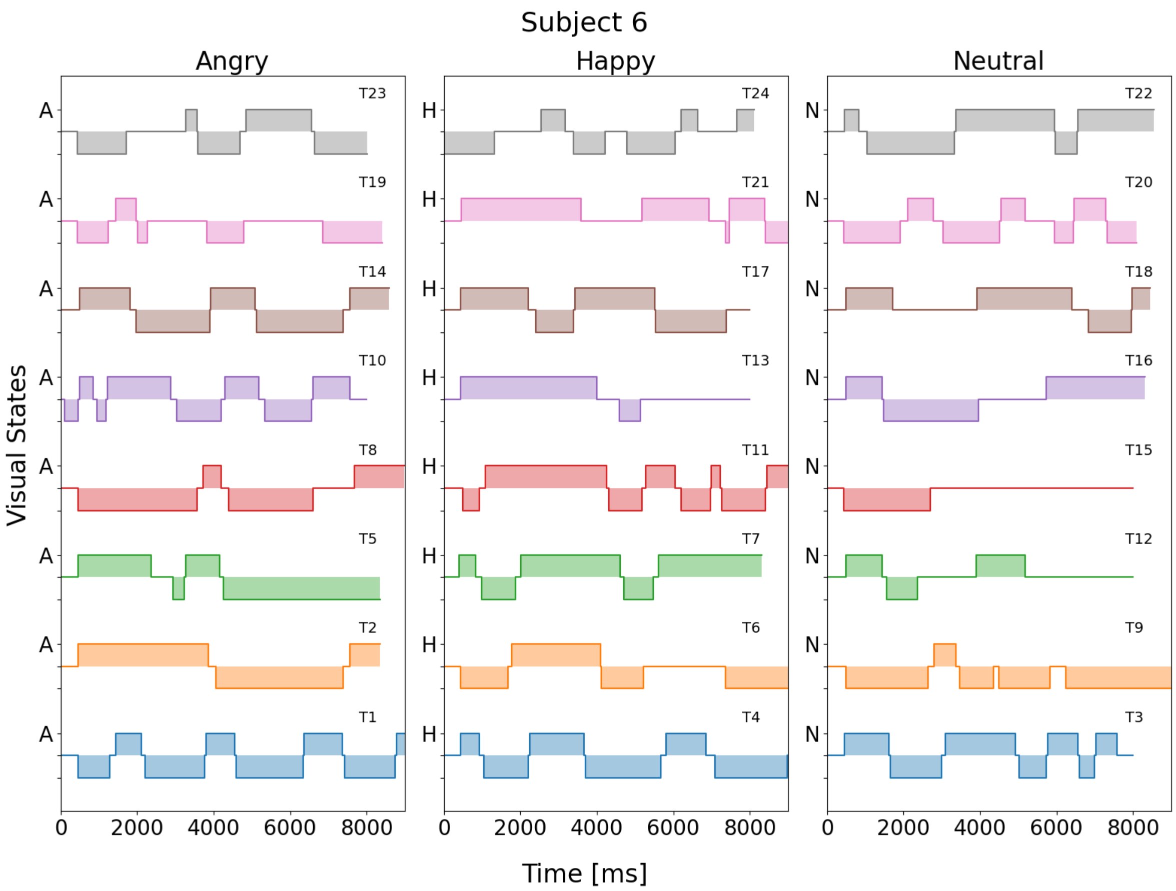

Subject 1
Angry
Happy
Neutral
Visual States
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]

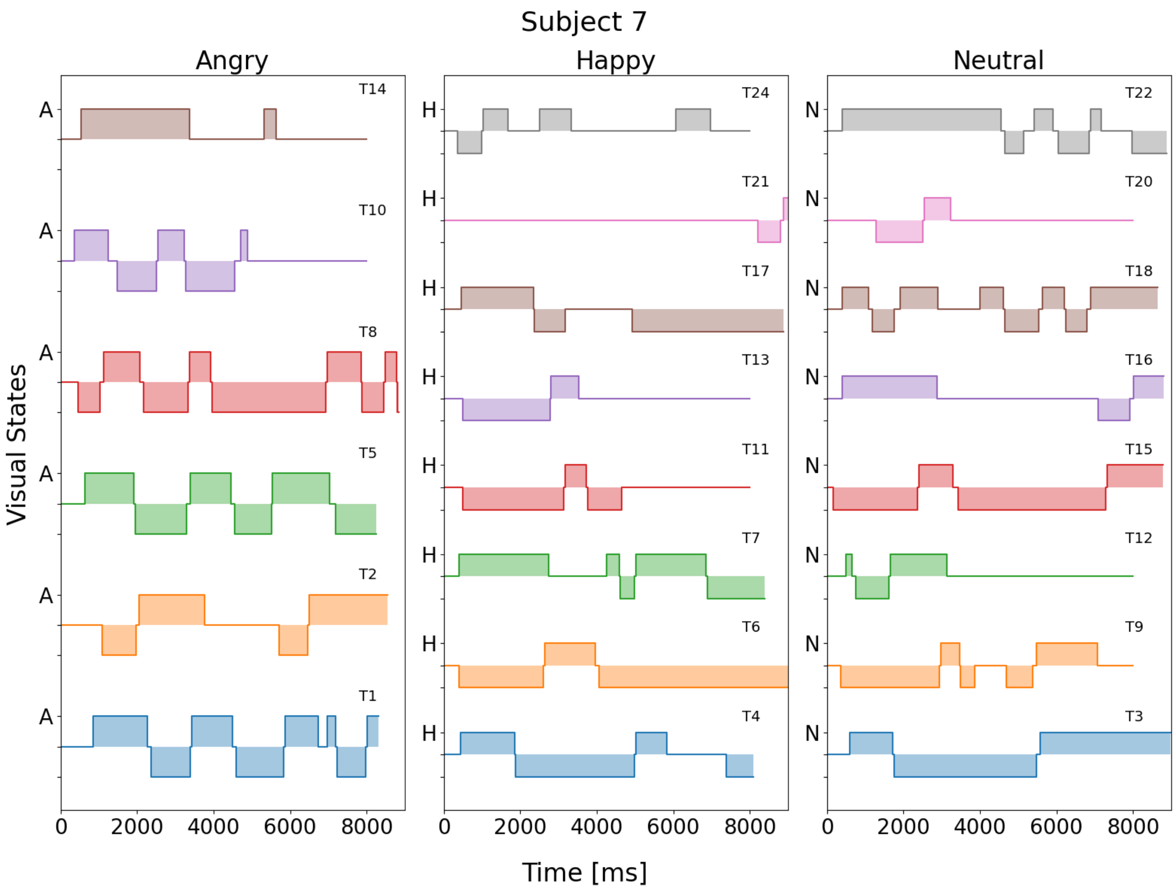

Subject 2
Angry
Happy
Neutral
Visual States
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]


Subject 3
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]

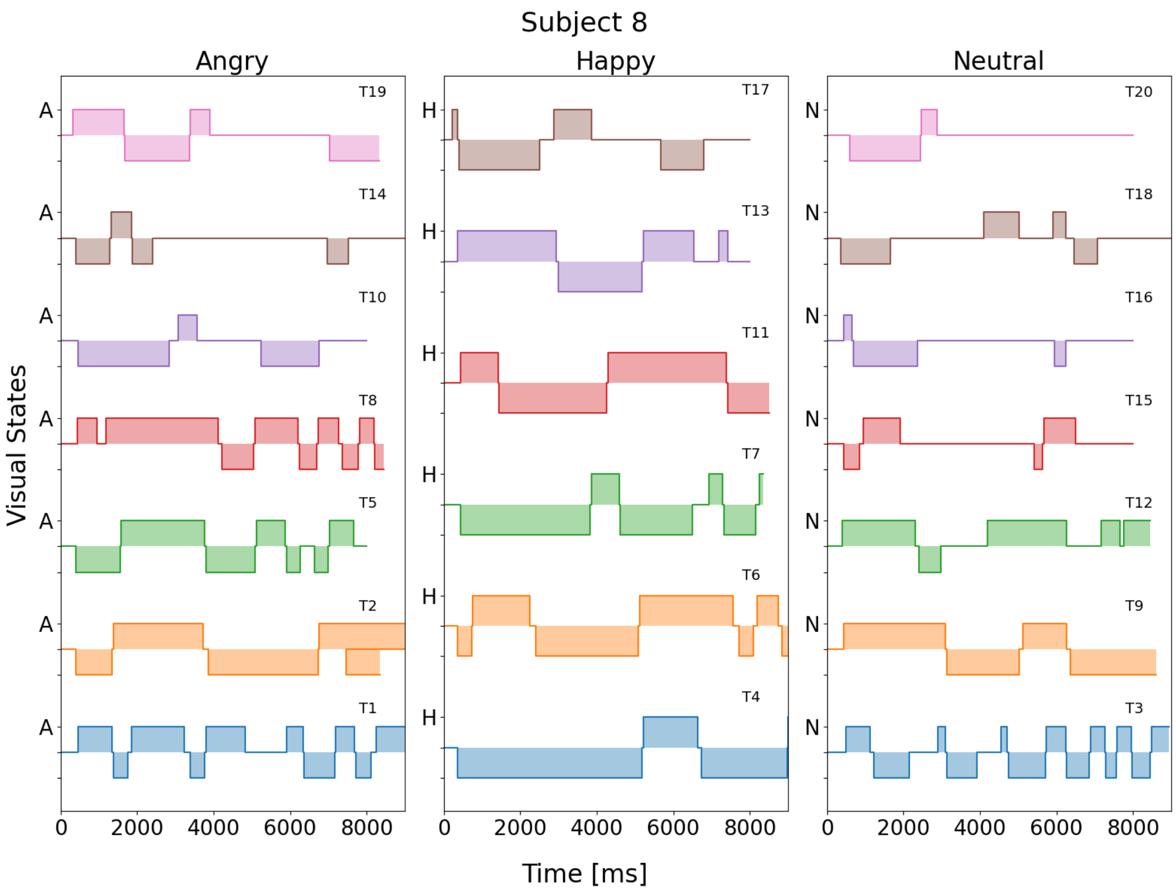

Subject 4
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]


Subject 5
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]

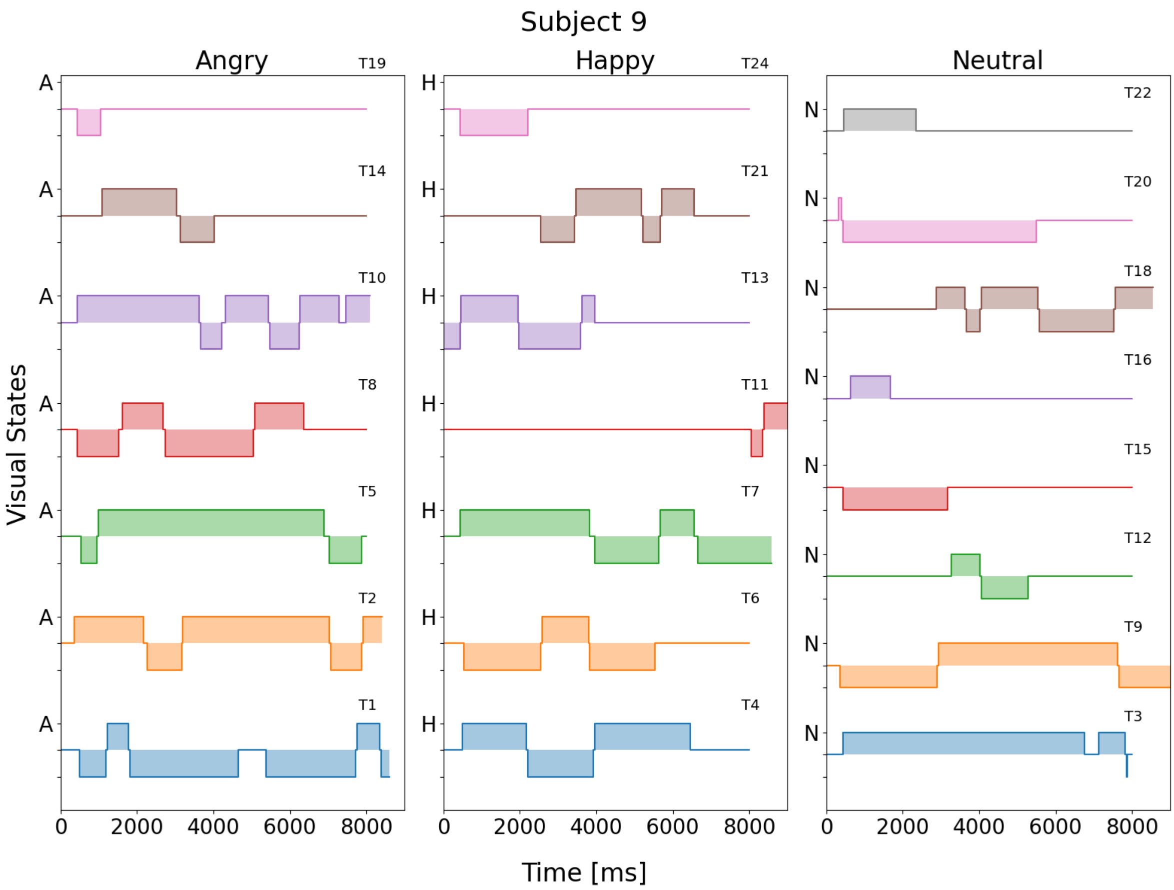

Subject 6
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]

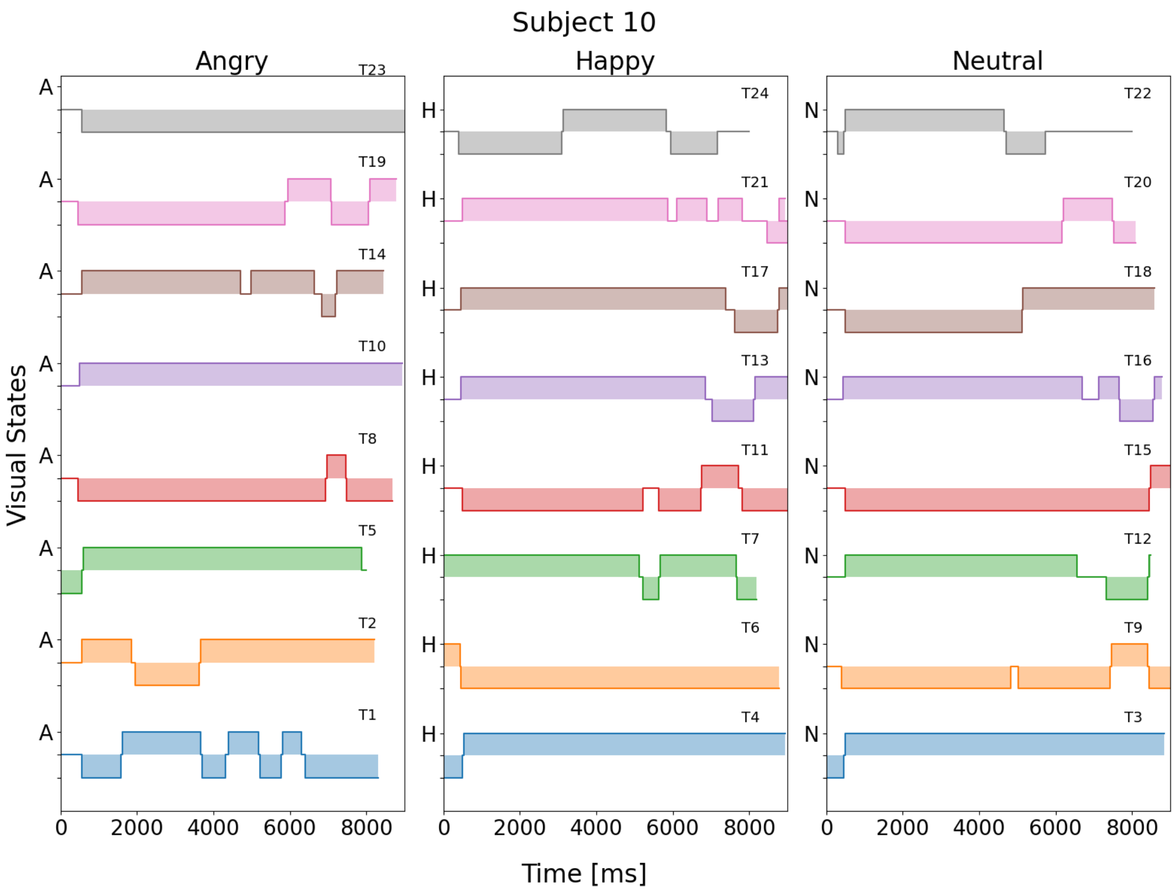


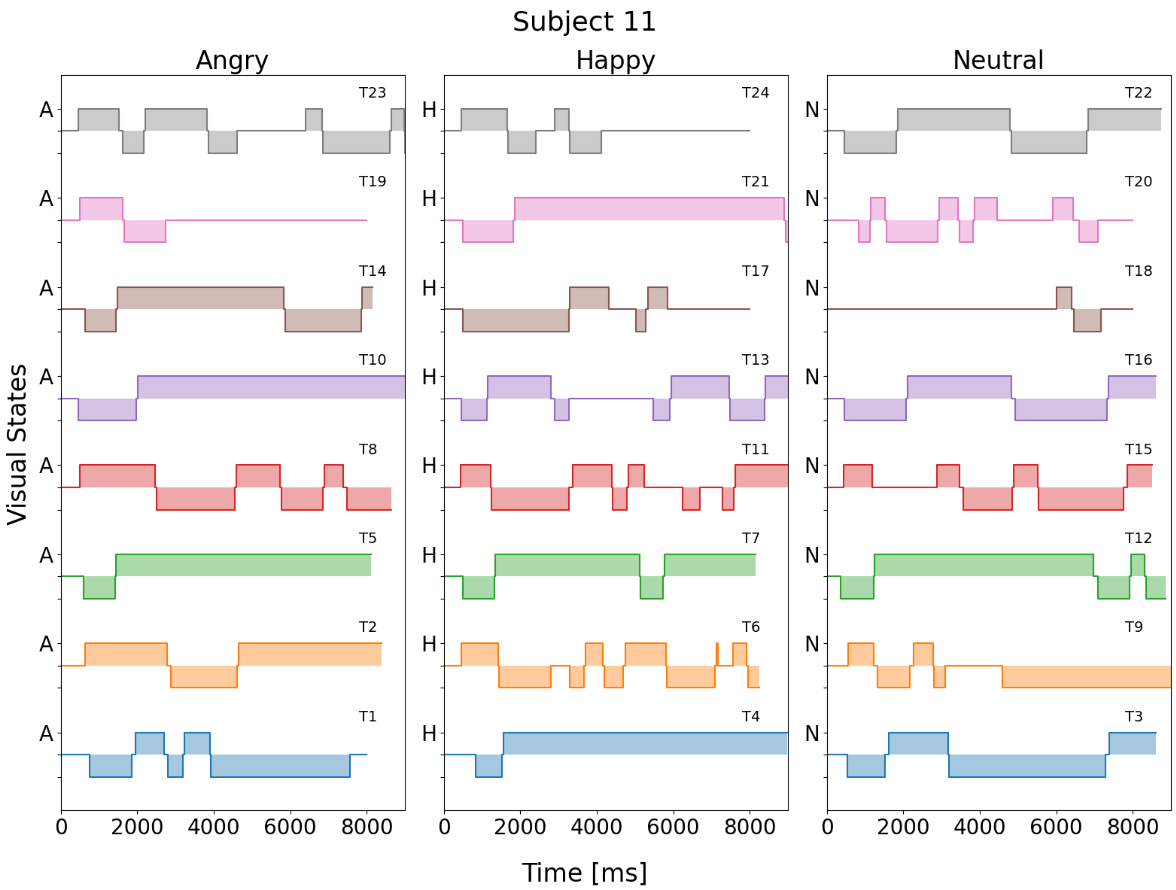
Subject 7
Angry
Happy
Neutral
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
Visual States
0
2000
4000
6000
8000
Time [ms]

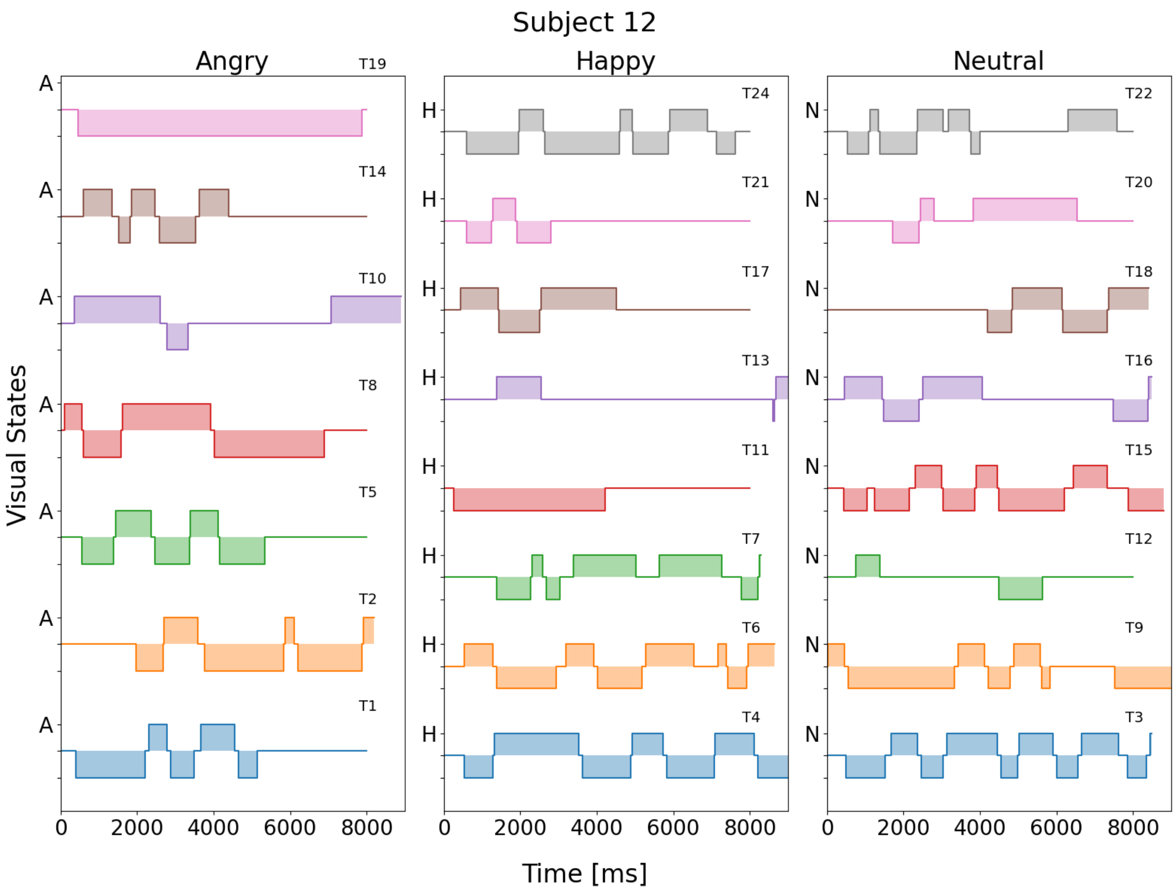


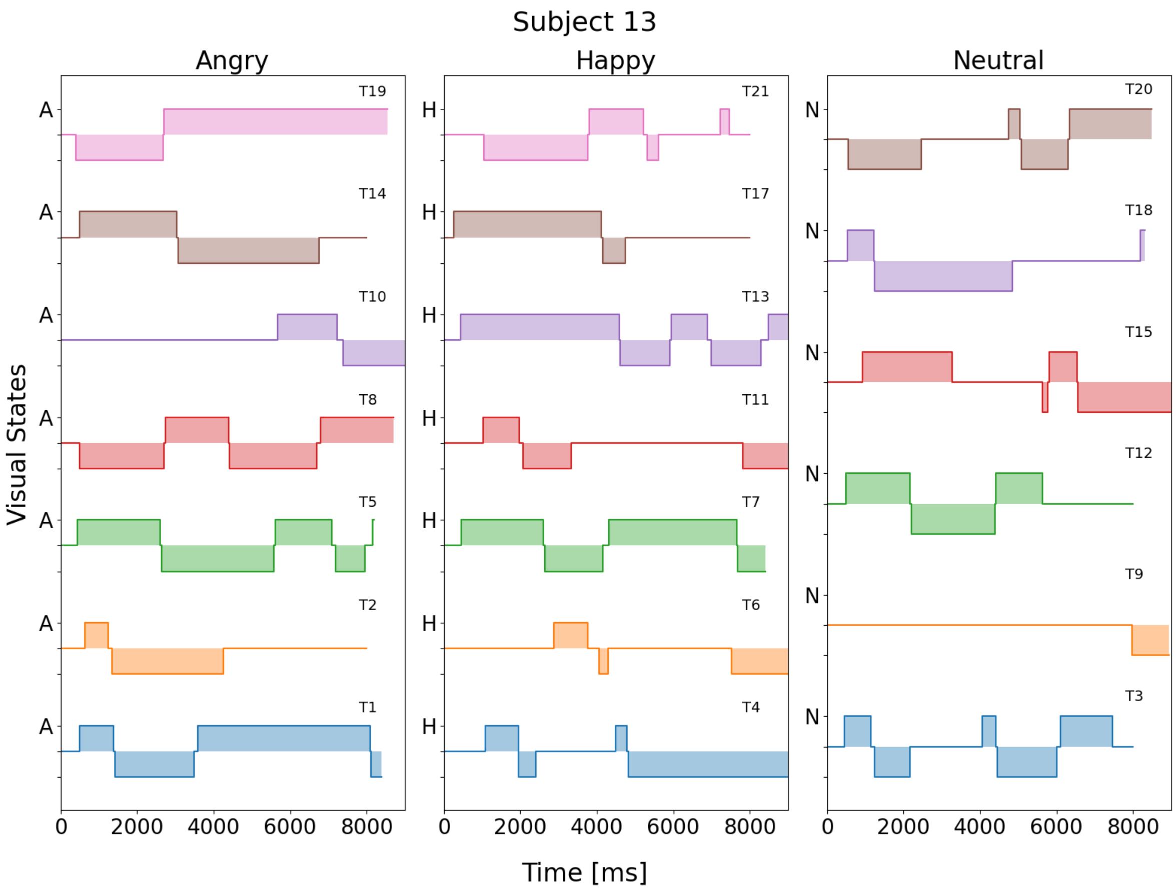
Subject 8
Angry
Happy
Neutral
T19
T14
T10
T8
T5
T2
T1
T17
T13
T11
T7
T6
T4
T20
T18
T16
T15
T12
T9
T3
A
H
N
Visual States
0
2000
4000
6000
8000
Time [ms]

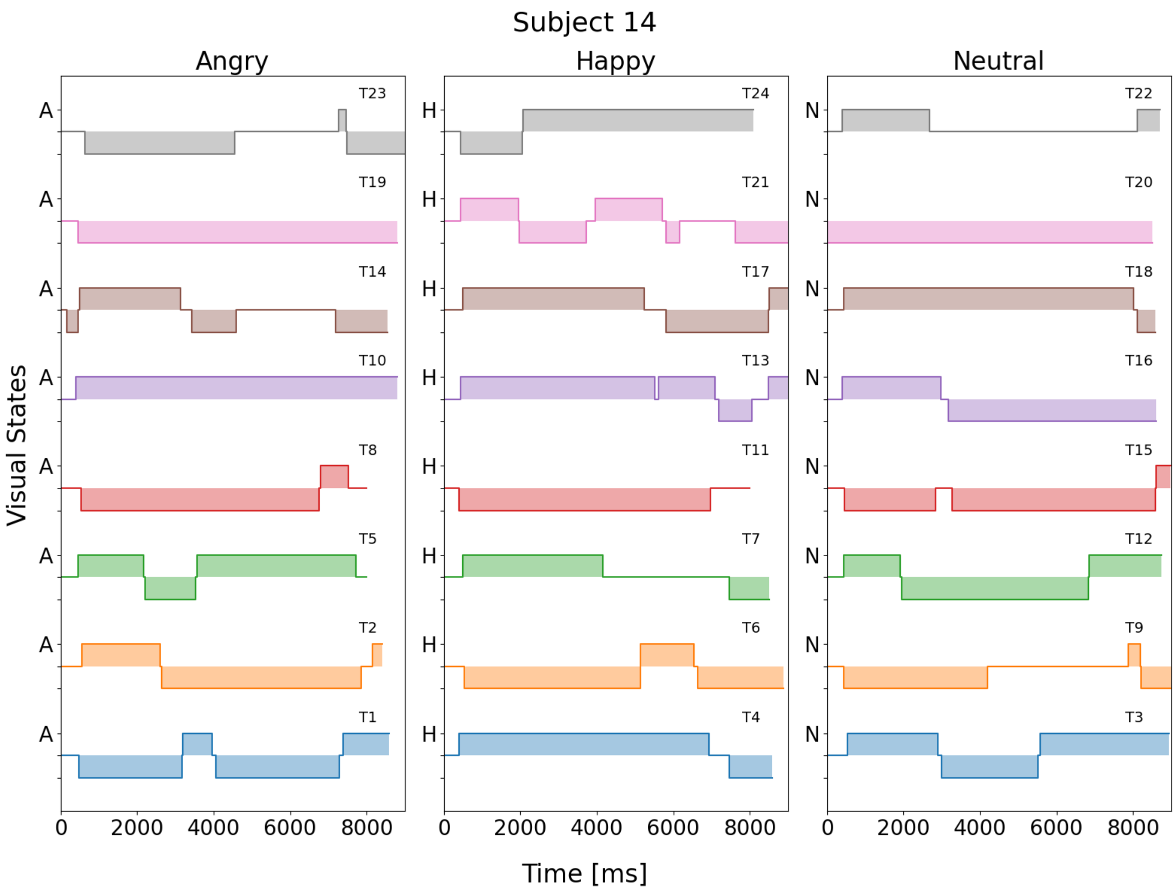

Subject 9
Angry
Happy
Neutral
T19
T14
T10
T8
T5
T2
T1
T24
T21
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
Visual States
0
2000
4000
6000
8000
Time [ms]

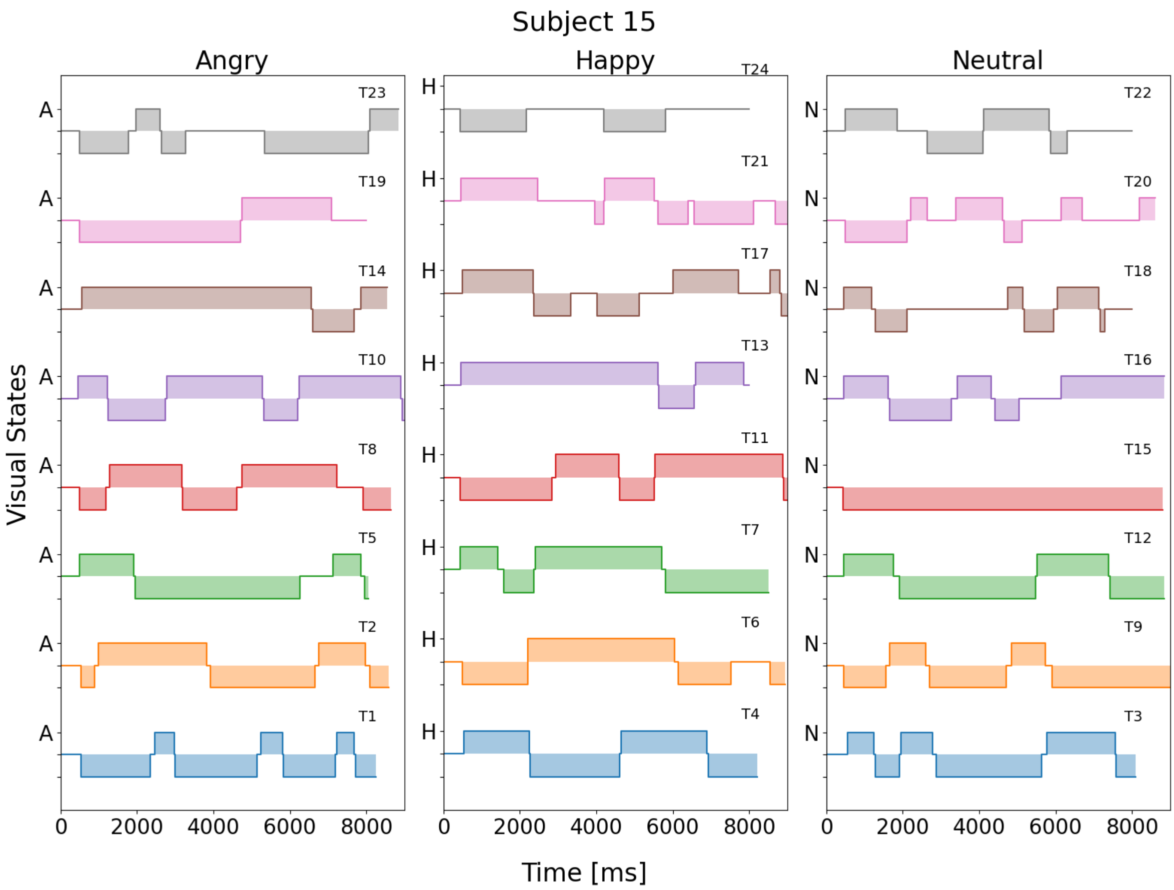

Subject 10
Angry
Happy
Neutral
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
Visual States
0
2000
4000
6000
8000
Time [ms]

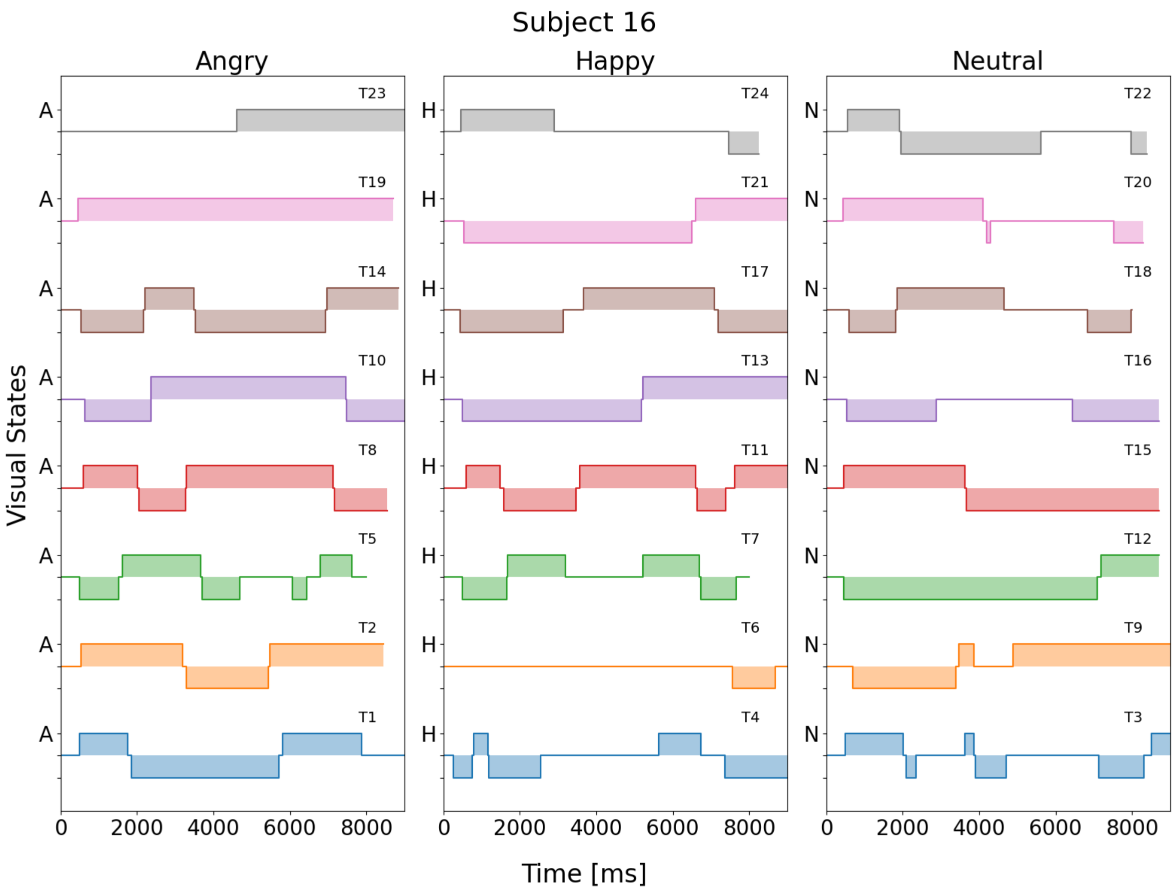

Subject 11
Angry
Happy
Neutral
Visual States
A
H
N
T23 T19 T14 T10 T8 T5 T2 T1
T24 T21 T17 T13 T11 T7 T6 T4
T22 T20 T18 T16 T15 T12 T9 T3
0 2000 4000 6000 8000
Time [ms]

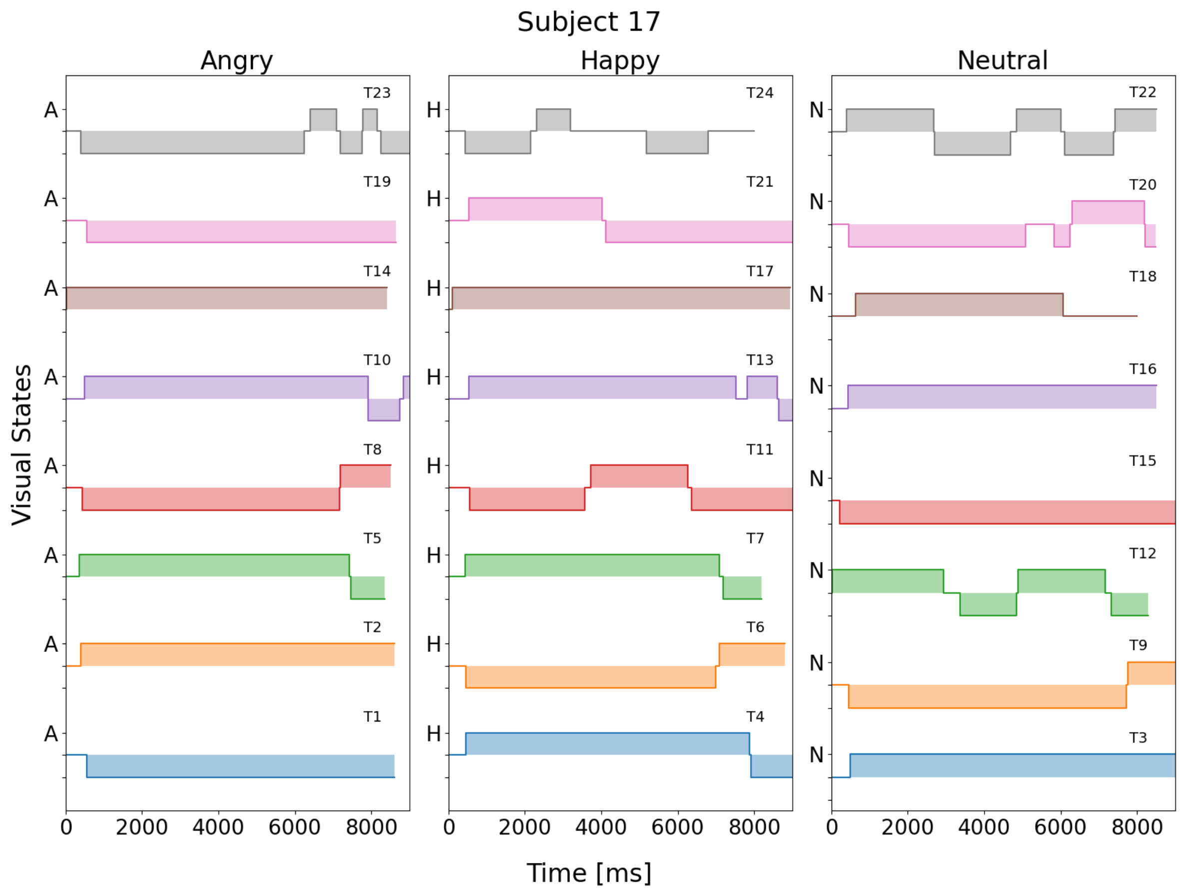

Subject 12
Angry
Happy
Neutral
Visual States
A
H
N
T19 T14 T10 T8 T5 T2 T1
T24 T21 T17 T13 T11 T7 T6 T4
T22 T20 T18 T16 T15 T12 T9 T3
0 2000 4000 6000 8000
Time [ms]

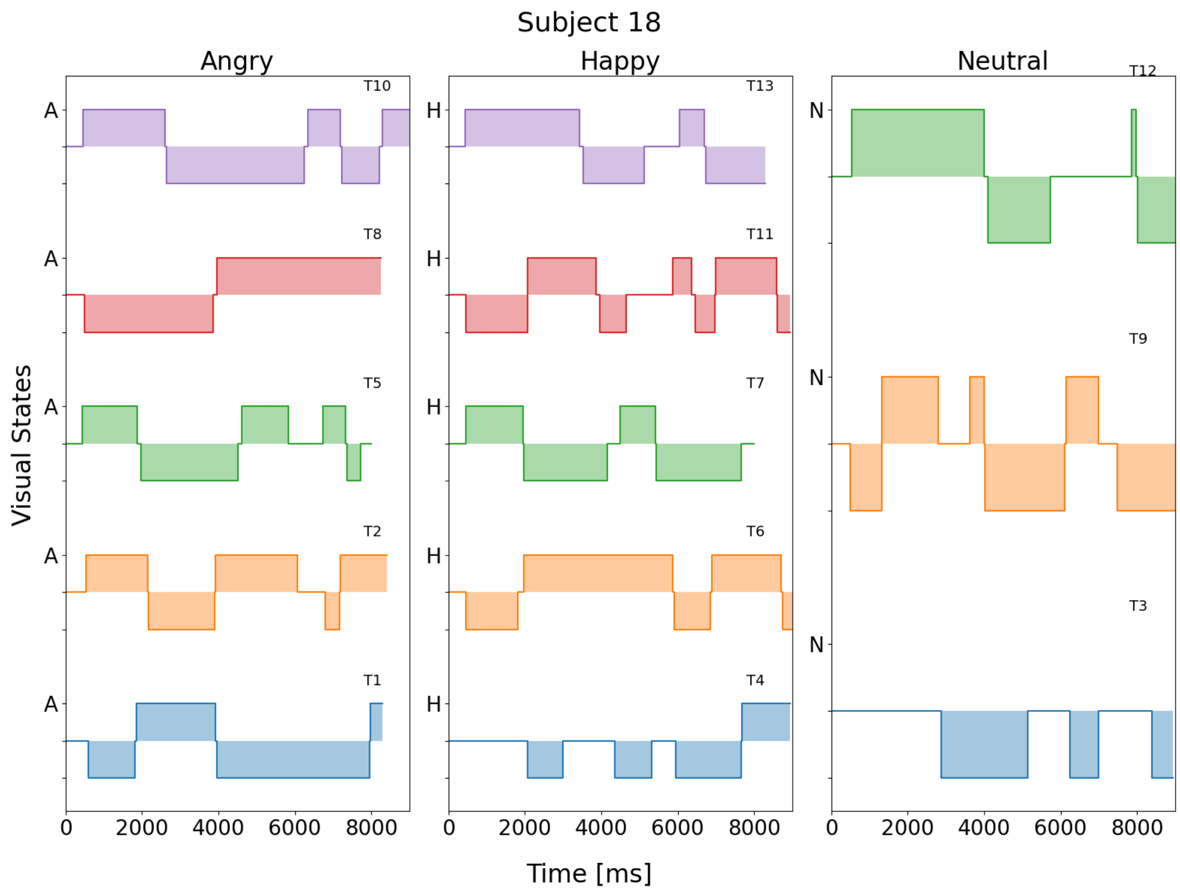

Subject 13
Angry
Happy
Neutral
Visual States
T19
T14
T10
T8
T5
T2
T1
T21
T17
T13
T11
T7
T6
T4
T20
T18
T15
T12
T9
T3
A
H
N
0
2000
4000
6000
8000
Time [ms]

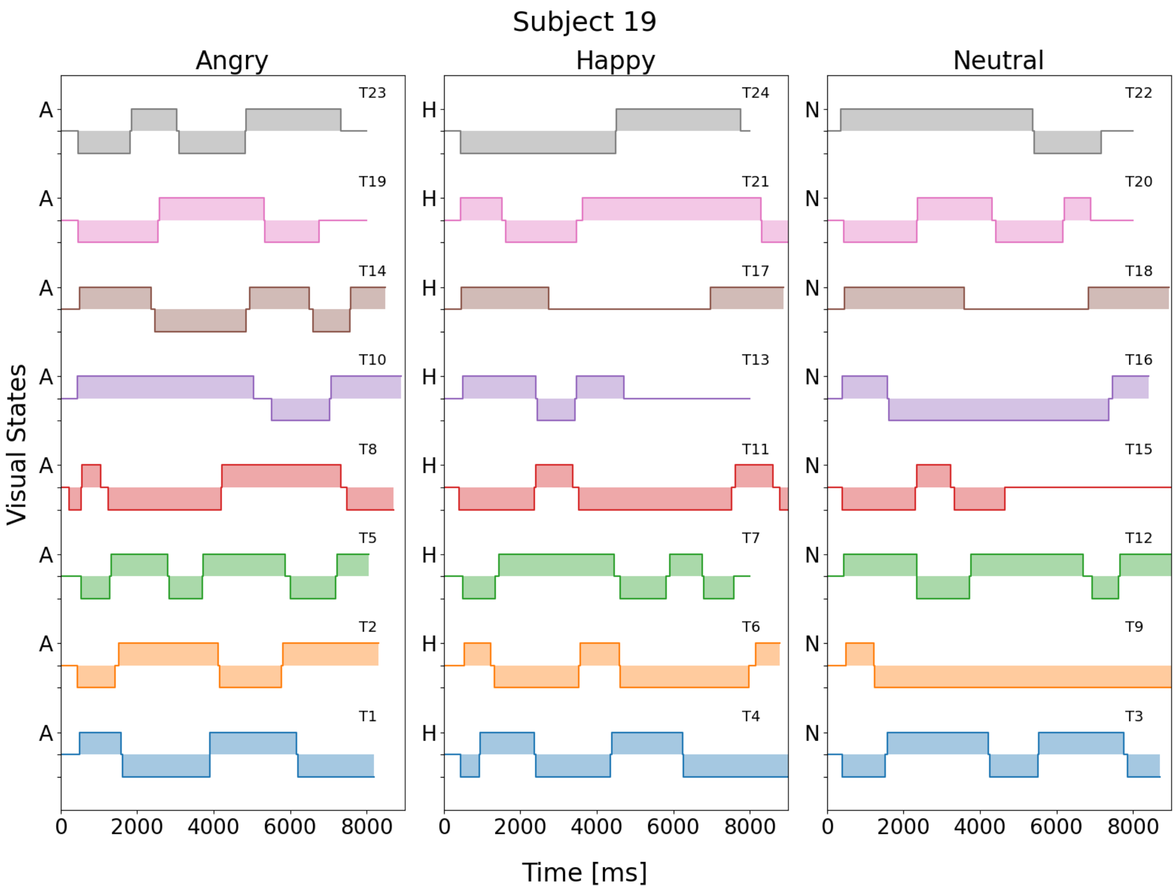

Subject 14
Angry
Happy
Neutral
Visual States
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
0
2000
4000
6000
8000
Time [ms]

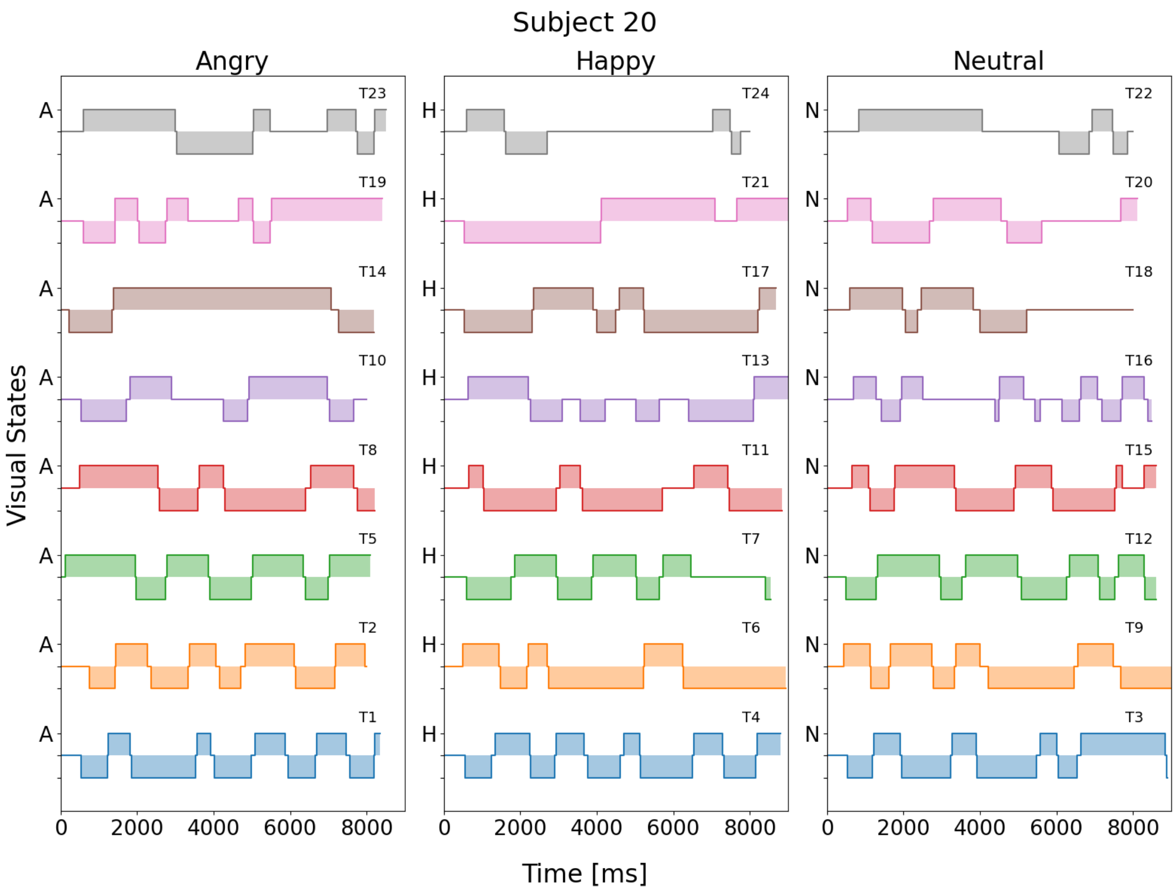

Subject 15
Angry
Happy
Neutral
Visual States
Time [ms]
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
0
2000
4000
6000
8000

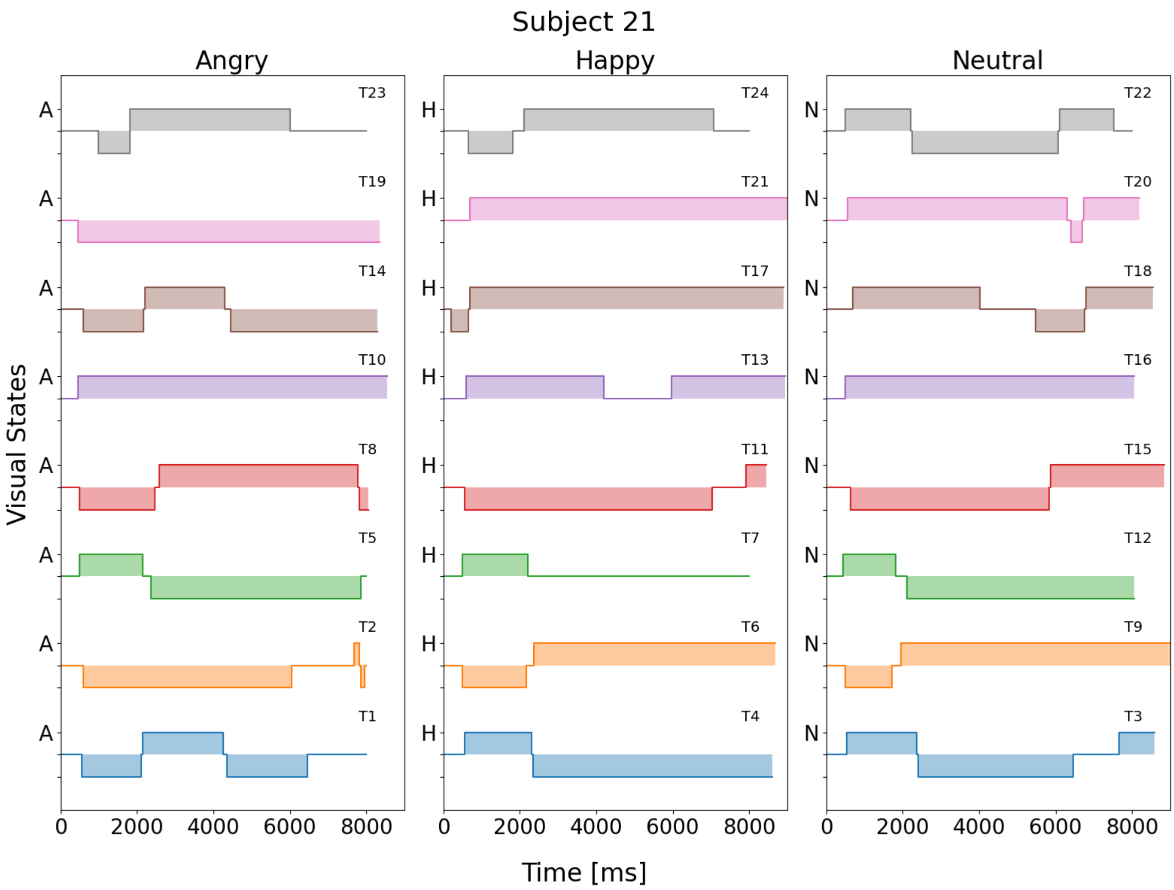

Subject 16
Angry
Happy
Neutral
Visual States
Time [ms]
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
0
2000
4000
6000
8000


Subject 17
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]

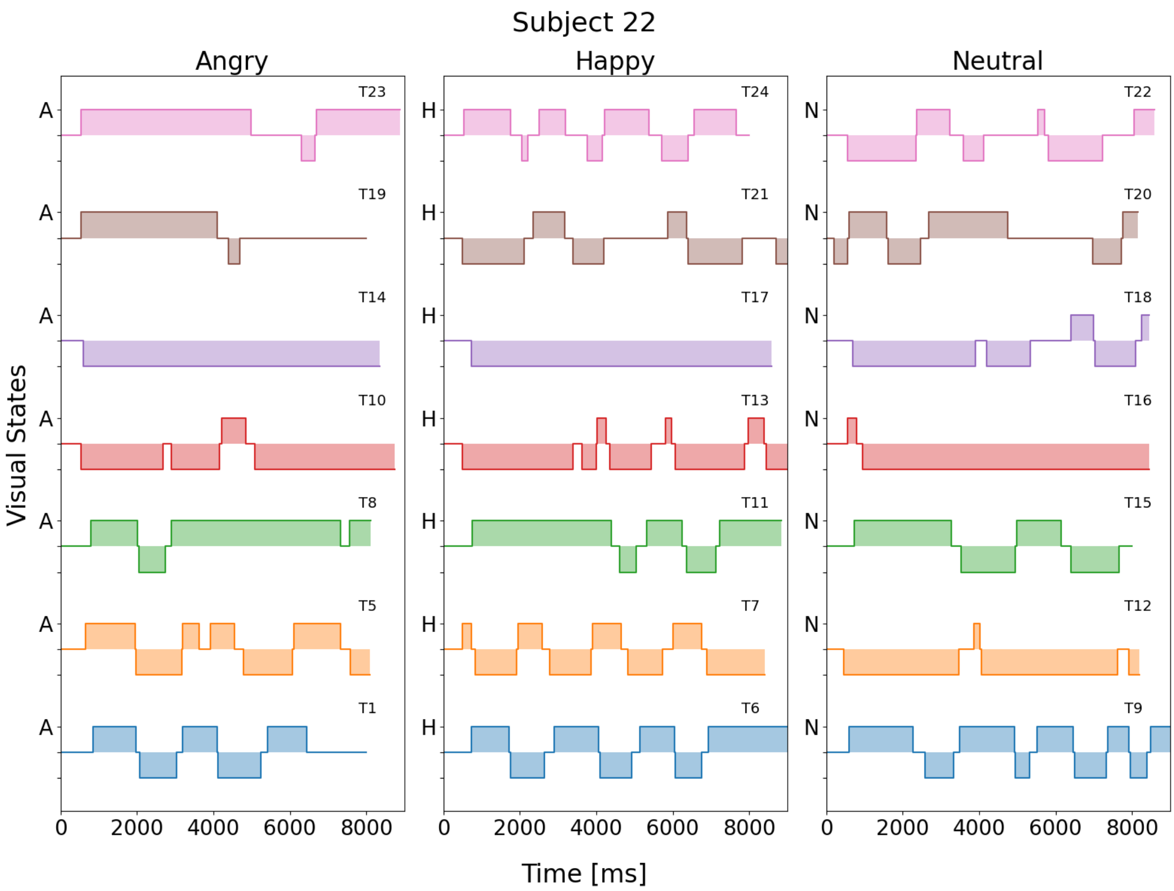

Subject 18
Angry
Happy
Neutral
Visual States
A
H
N
T10
T8
T5
T2
T1
T13
T11
T7
T6
T4
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]


Subject 19
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]

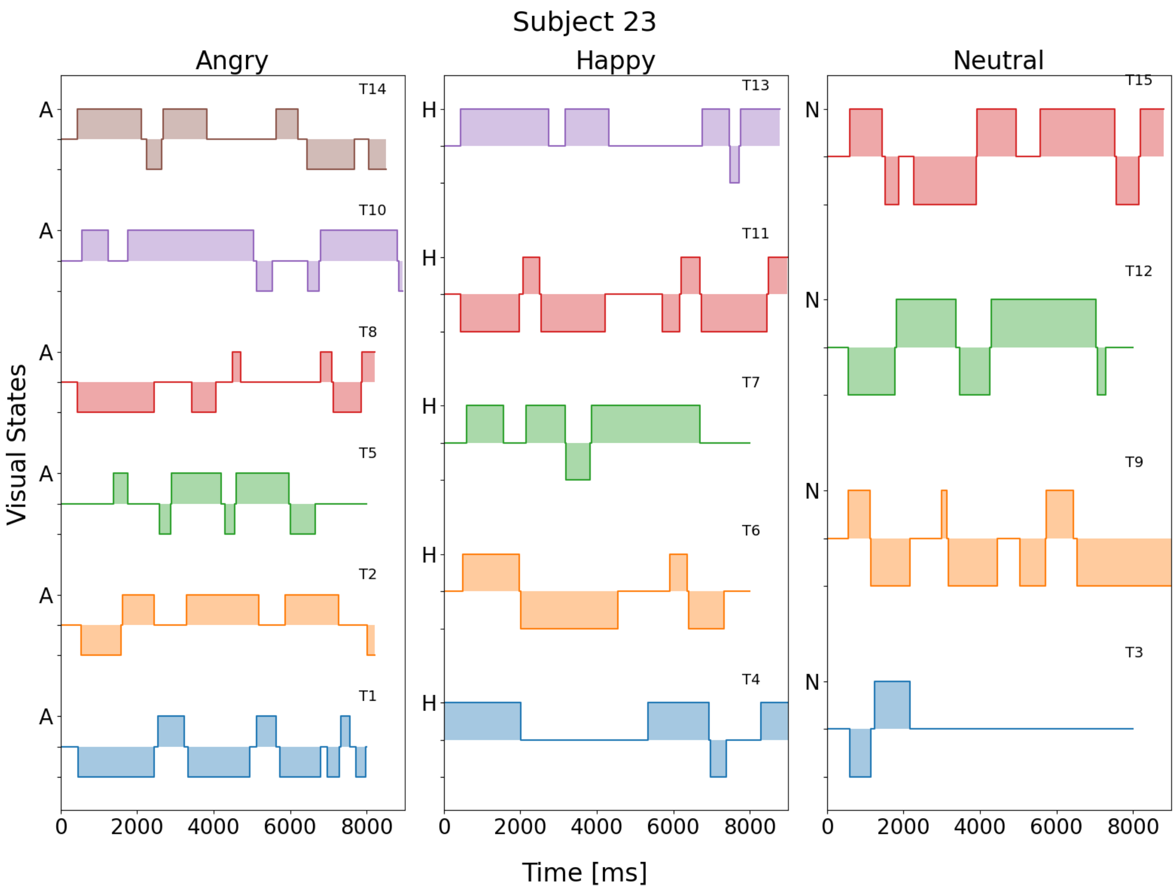

Subject 20
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]



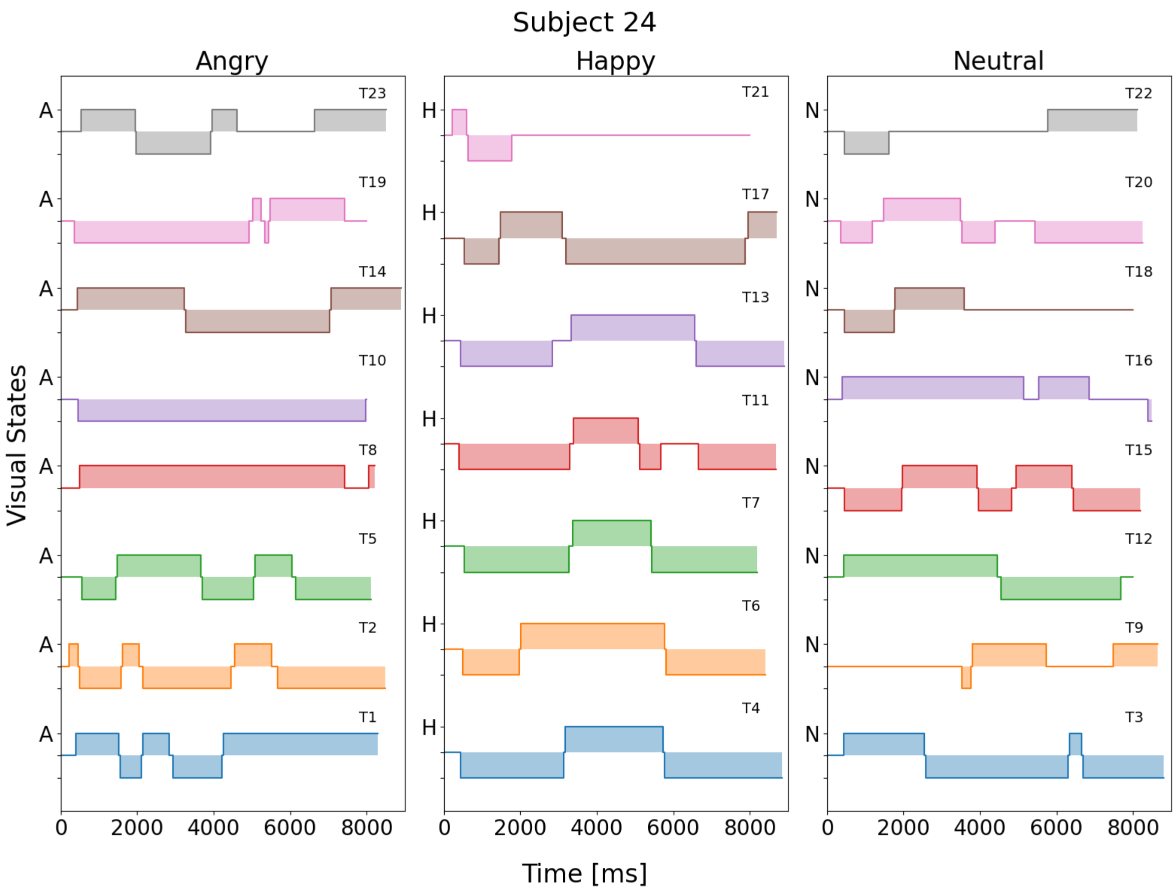
Subject 21
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]



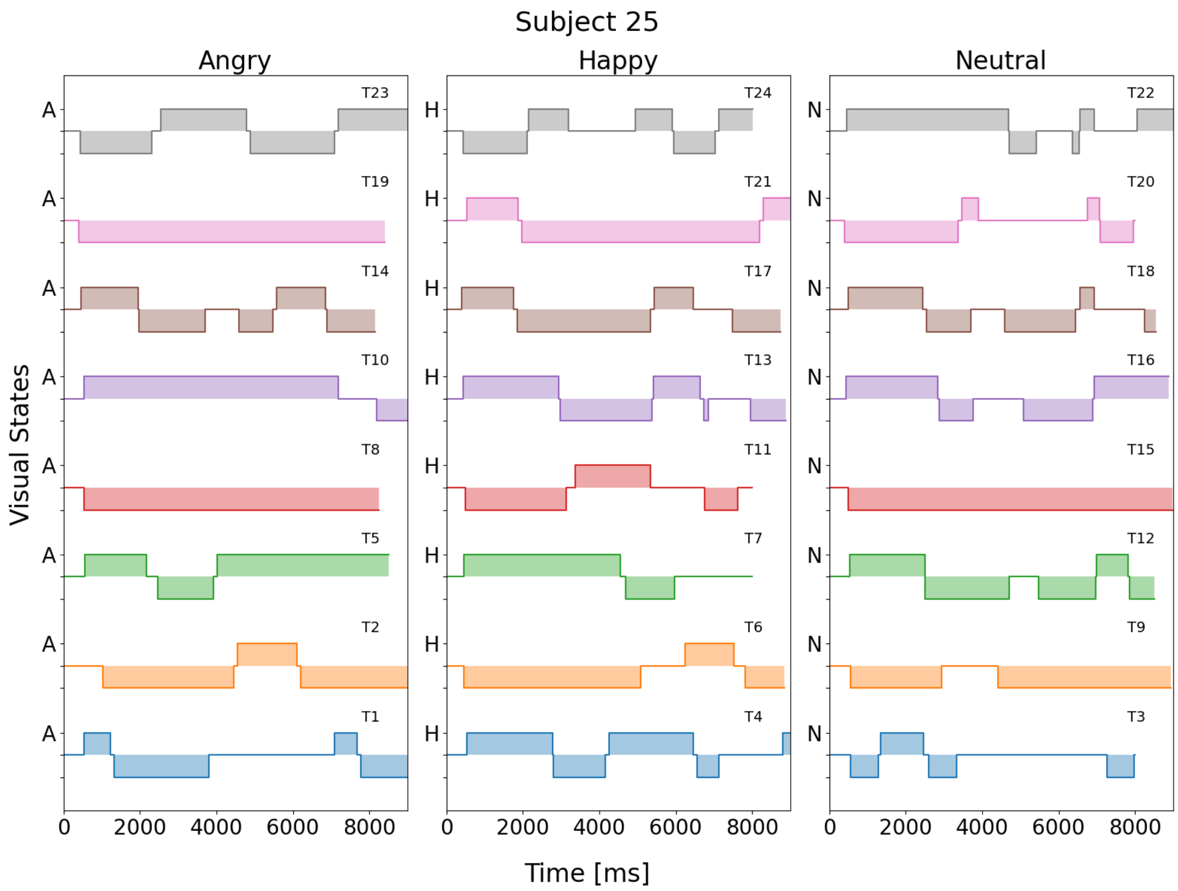
Subject 22
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T1
T24
T21
T17
T13
T11
T7
T6
T22
T20
T18
T16
T15
T12
T9
0
2000
4000
6000
8000
Time [ms]

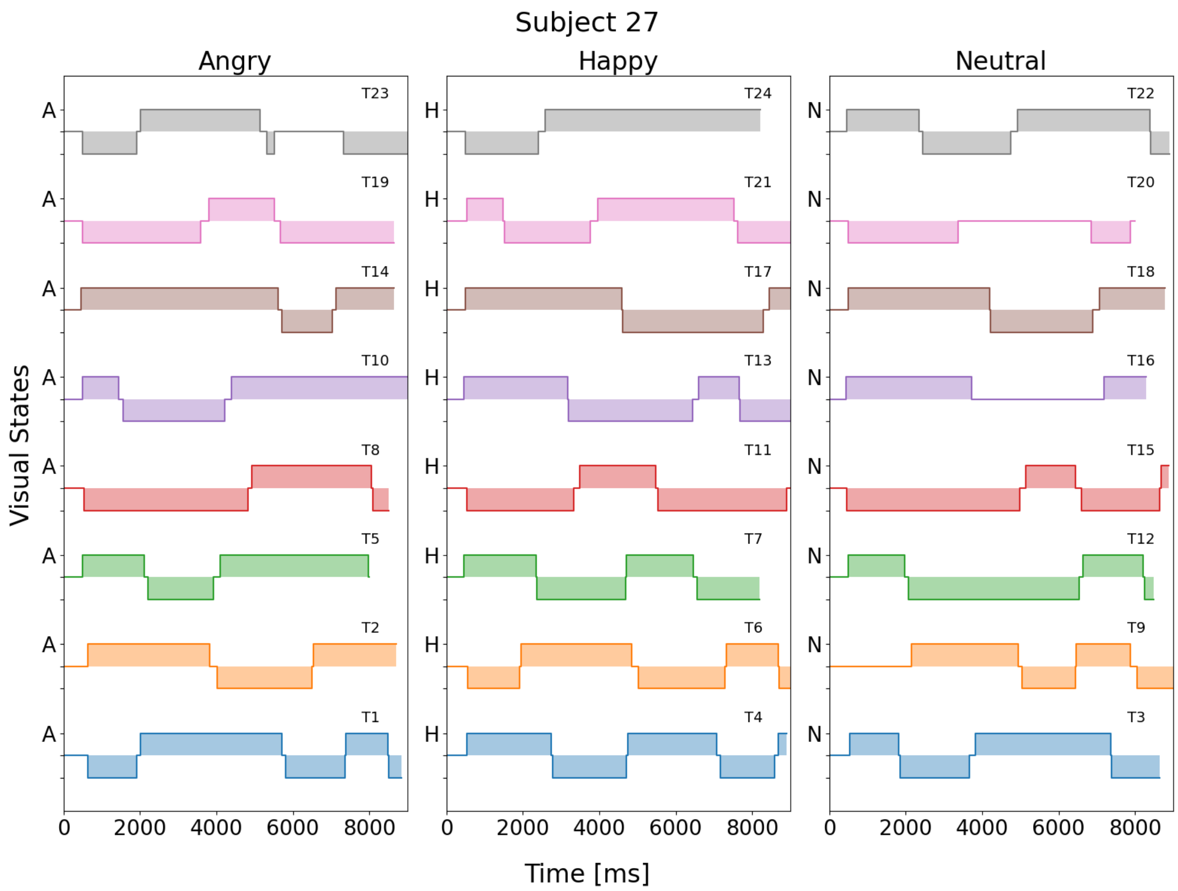


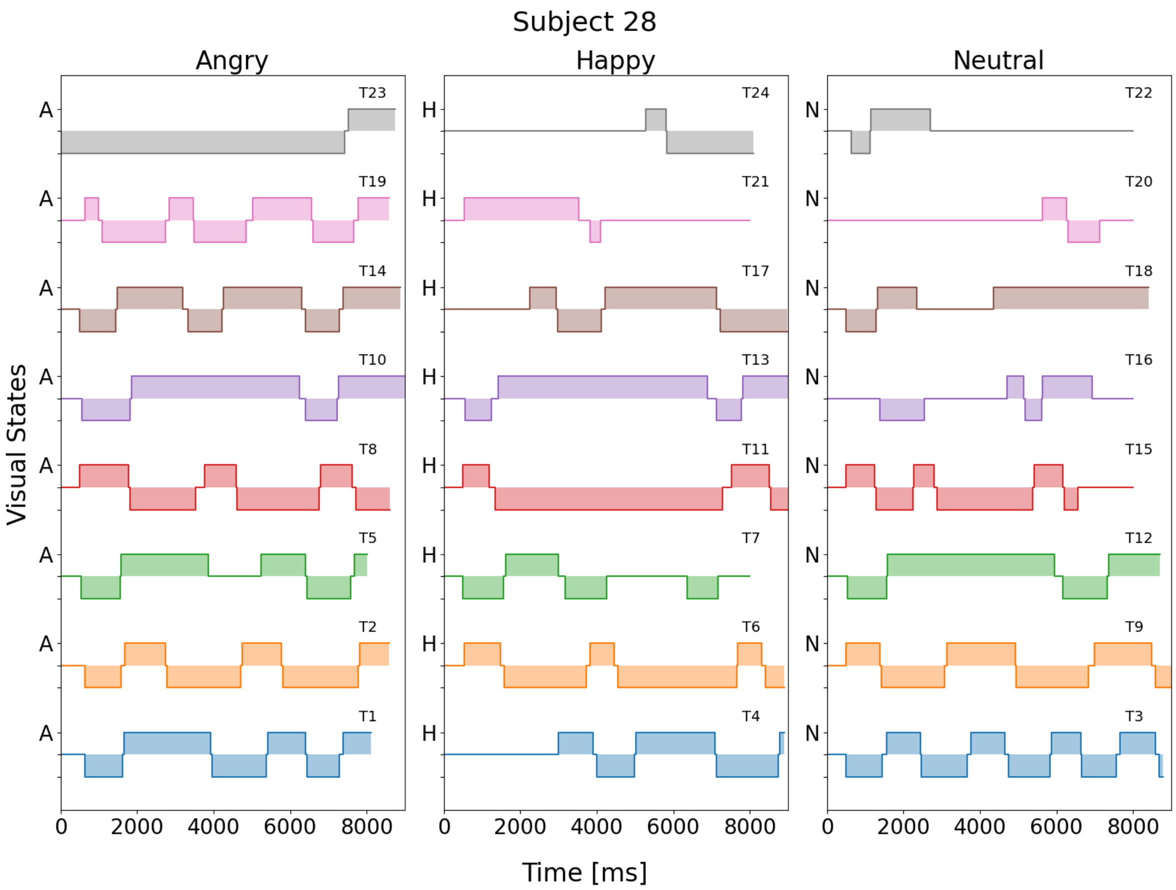
Subject 23
Angry
Happy
Neutral
Visual States
A
H
N
T14
T10
T8
T5
T2
T1
T13
T11
T7
T6
T4
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]

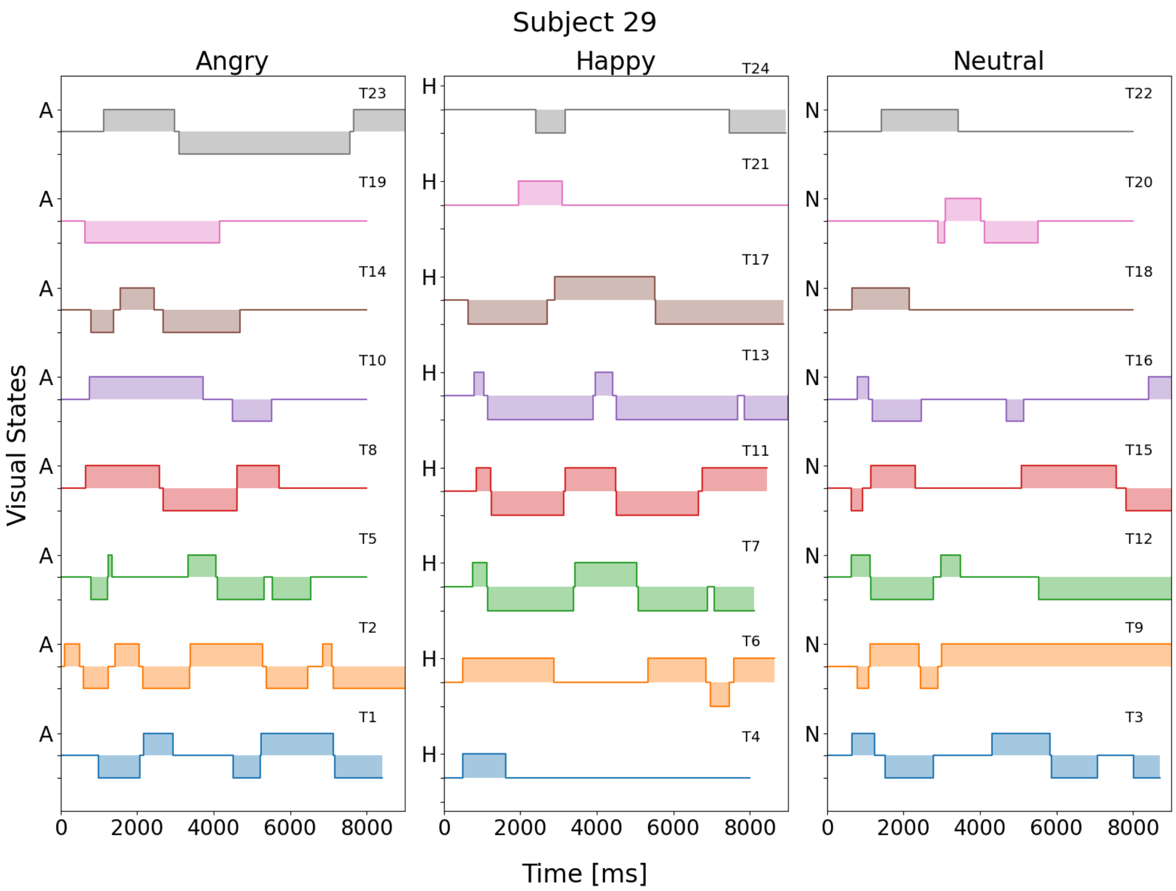


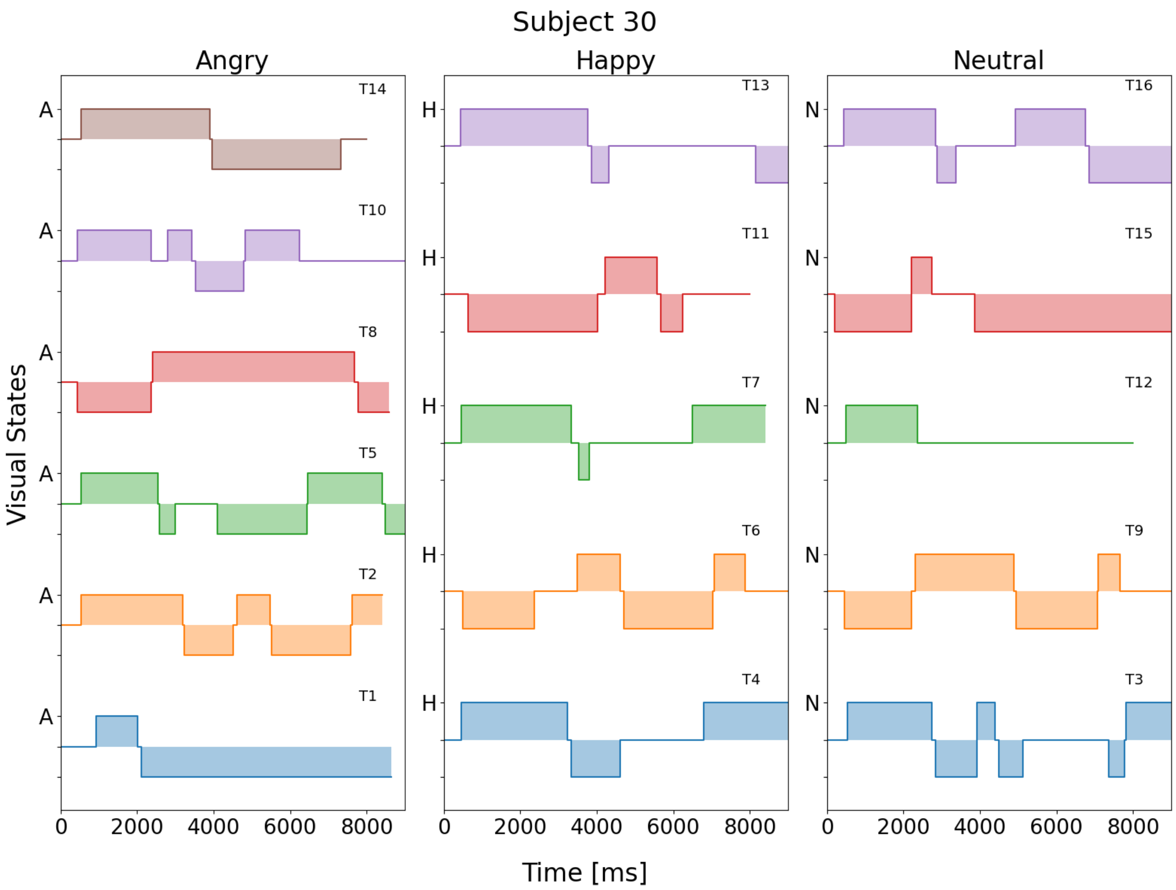
Subject 24
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]


Subject 25
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]

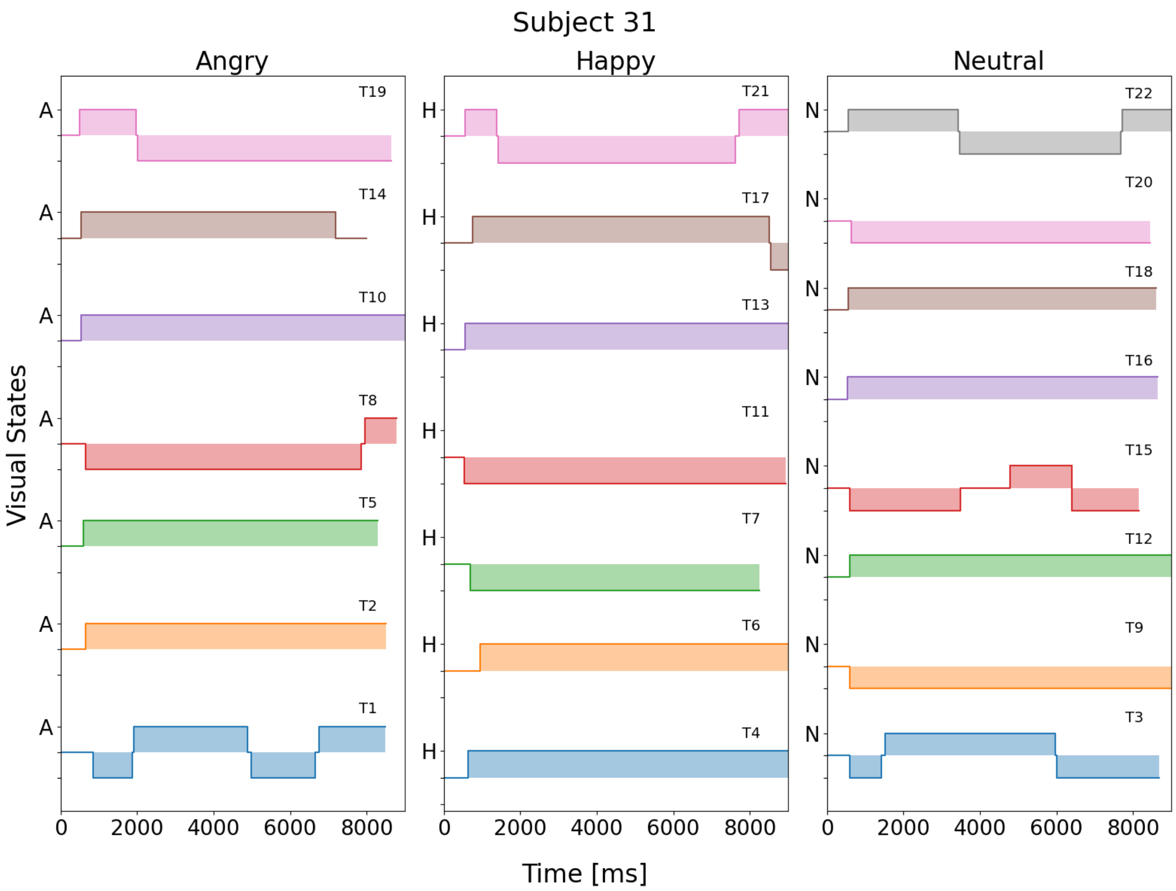

Subject 27
Angry
Happy
Neutral
Visual States
A
H
N
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
0
2000
4000
6000
8000
Time [ms]


Subject 28
Angry
Happy
Neutral
Visual States
Time [ms]
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
0
2000
4000
6000
8000

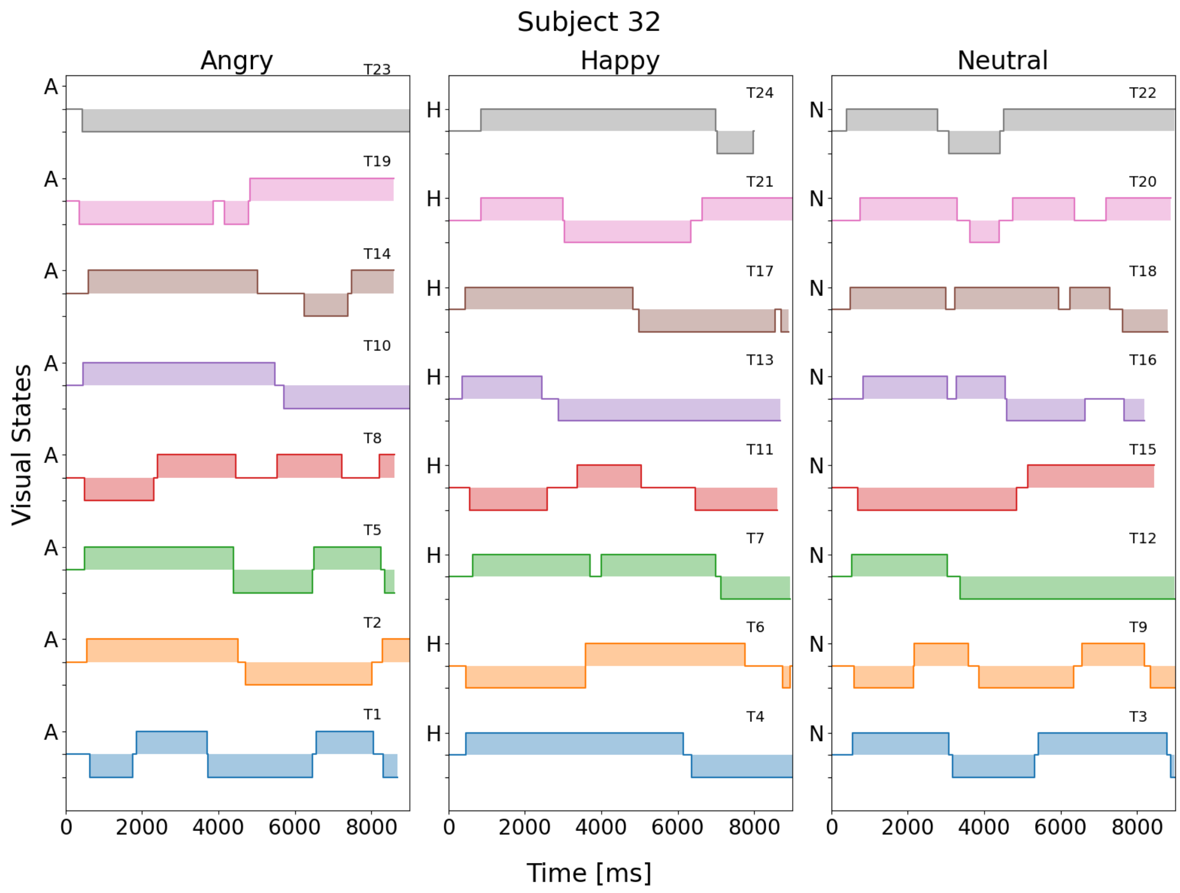

Subject 29
Angry
Happy
Neutral
Visual States
Time [ms]
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
0
2000
4000
6000
8000

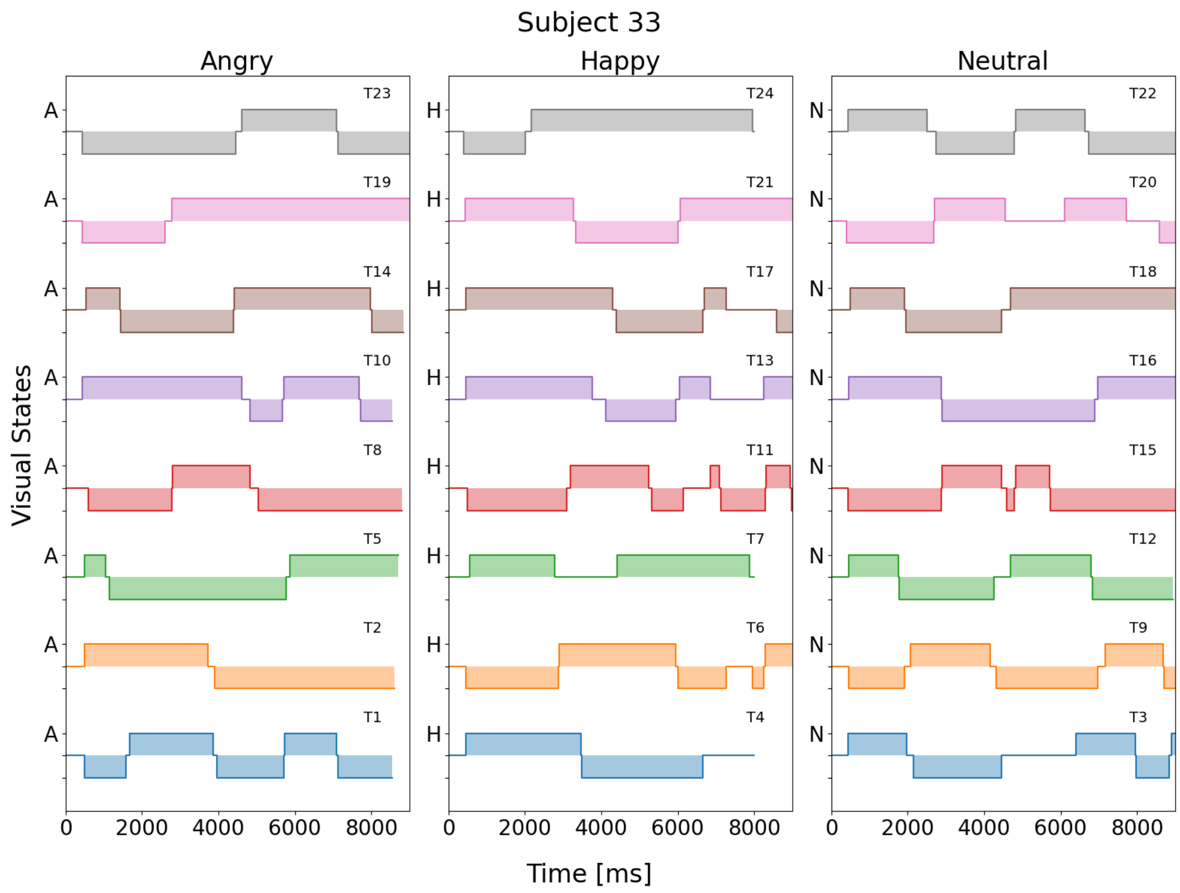


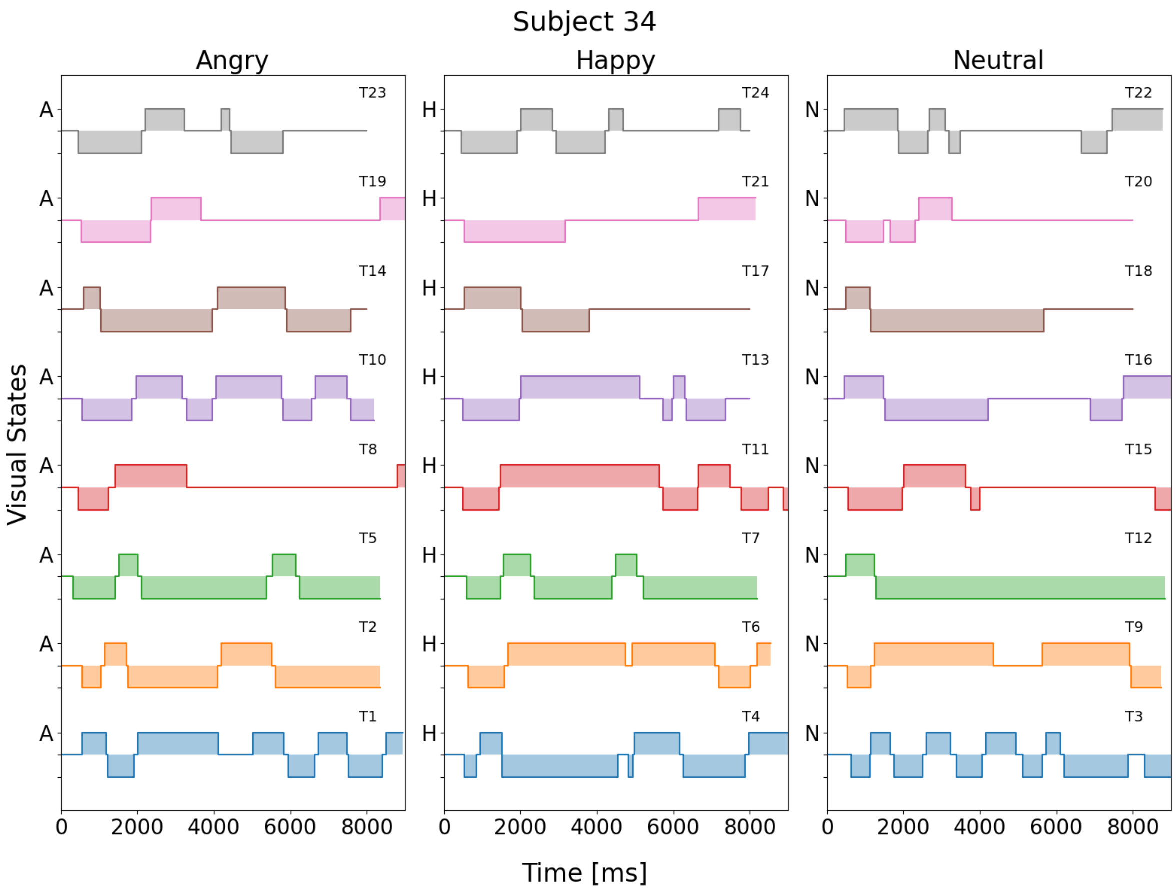
Subject 30
Angry
Happy
Neutral
T14
T10
T8
T5
T2
T1
T13
T11
T7
T6
T4
T16
T15
T12
T9
T3
A
H
N
Visual States
0
2000
4000
6000
8000
Time [ms]



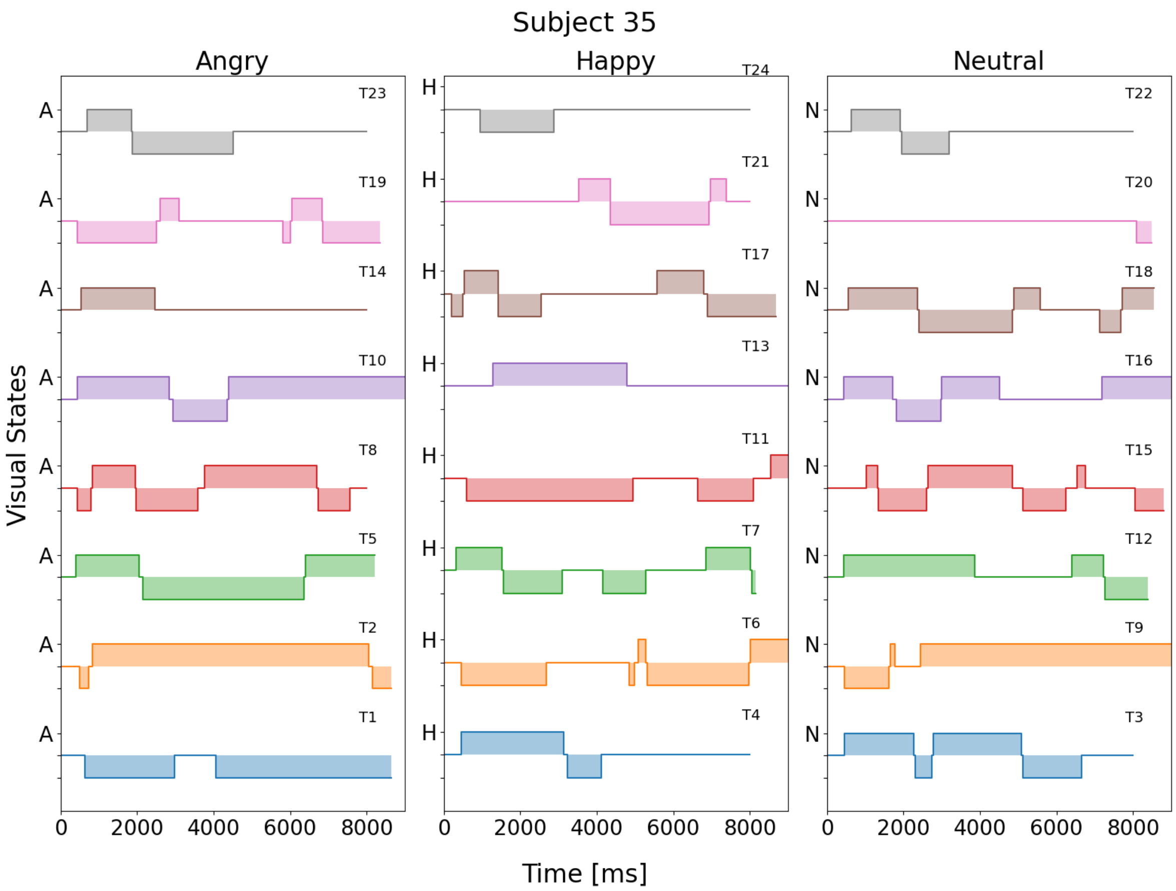
Subject 31
Angry
Happy
Neutral
T19
T14
T10
T8
T5
T2
T1
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
Visual States
0
2000
4000
6000
8000
Time [ms]


Subject 32
Angry
Happy
Neutral
Visual States
T23 T19 T14 T10 T8 T5 T2 T1
T24 T21 T17 T13 T11 T7 T6 T4
T22 T20 T18 T16 T15 T12 T9 T3
A
H
N
0 2000 4000 6000 8000
Time [ms]

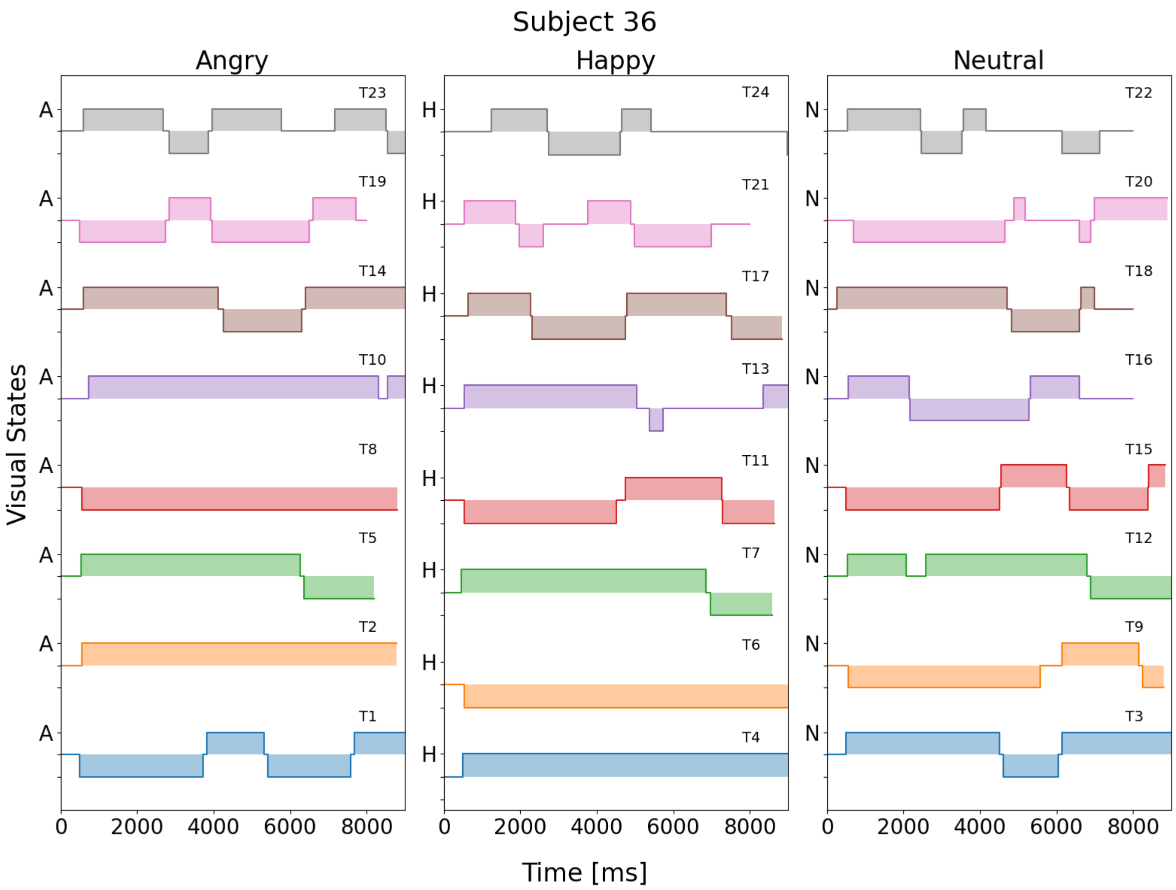

Subject 33
Angry
Happy
Neutral
Visual States
T23 T19 T14 T10 T8 T5 T2 T1
T24 T21 T17 T13 T11 T7 T6 T4
T22 T20 T18 T16 T15 T12 T9 T3
A
H
N
0 2000 4000 6000 8000
Time [ms]

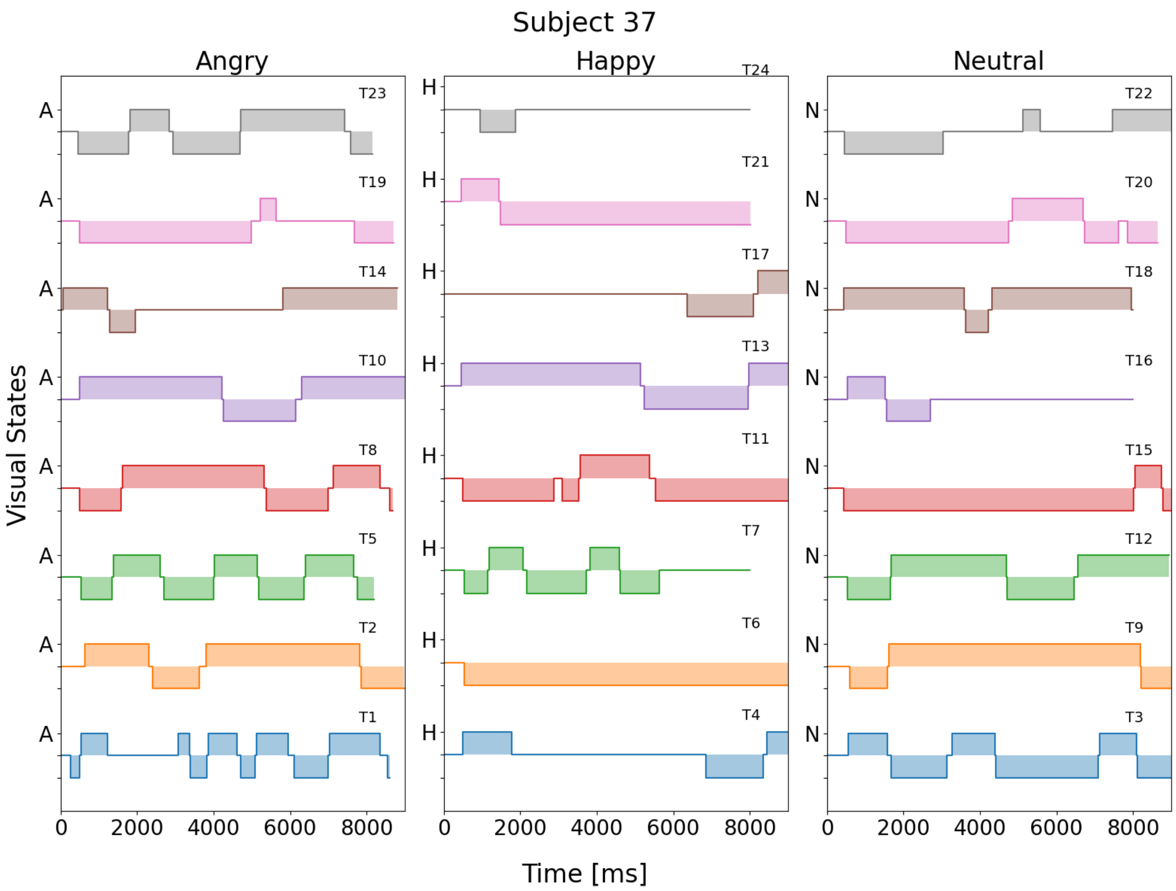


Subject 34
Angry
Happy
Neutral
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
Visual States
0
2000
4000
6000
8000
Time [ms]



Subject 35
Angry
Happy
Neutral
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
Visual States
0
2000
4000
6000
8000
Time [ms]


Subject 36
Angry
Happy
Neutral
Visual States
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
0
2000
4000
6000
8000
Time [ms]


Subject 37
Angry
Happy
Neutral
Visual States
T23
T19
T14
T10
T8
T5
T2
T1
T24
T21
T17
T13
T11
T7
T6
T4
T22
T20
T18
T16
T15
T12
T9
T3
A
H
N
0
2000
4000
6000
8000
Time [ms]

f. Additional figures for Study 2

Figure 3.17

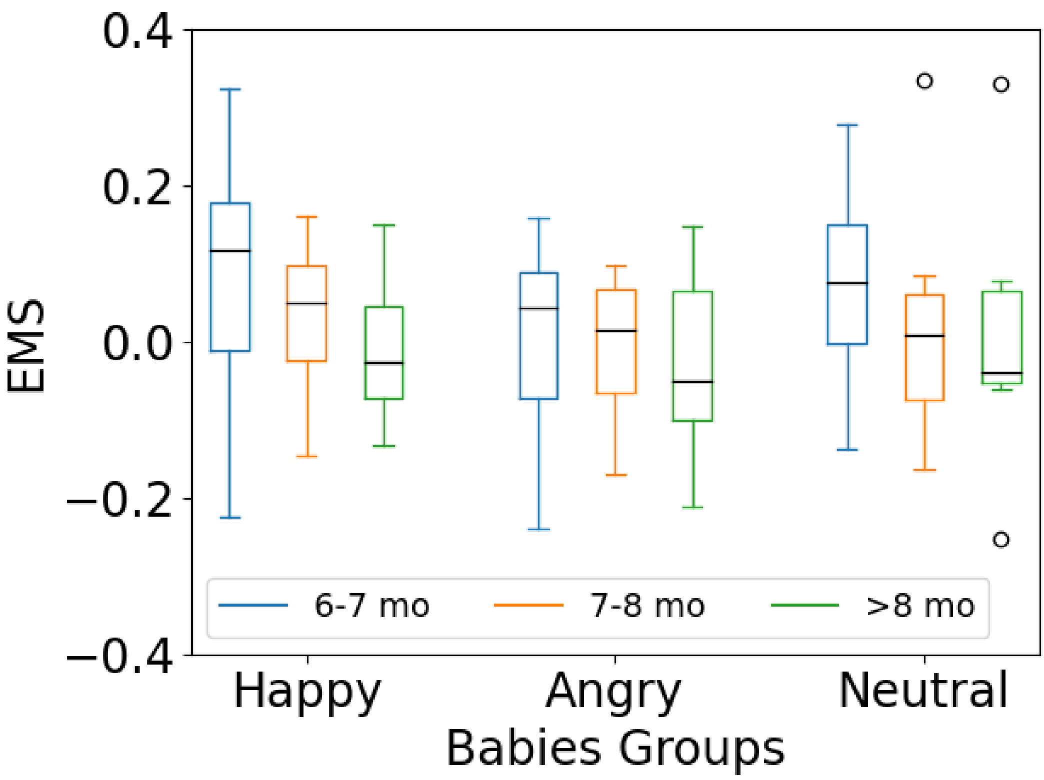


Figure 3.17 Boxplots of Emotion Matching Scores for each age group (6–7, 7–8, and >8 months) across the three emotional conditions (Happy, Angry, Neutral). The age-related decline was most pronounced in the Happy condition, where 6-month-olds exhibited the highest median EMS, followed by 7-month-olds, and >8-month-olds or older had the lowest. No reliable age differences were evident in the Angry condition, while Neutral scores showed only a slight decrease with age.

Figure 3.18

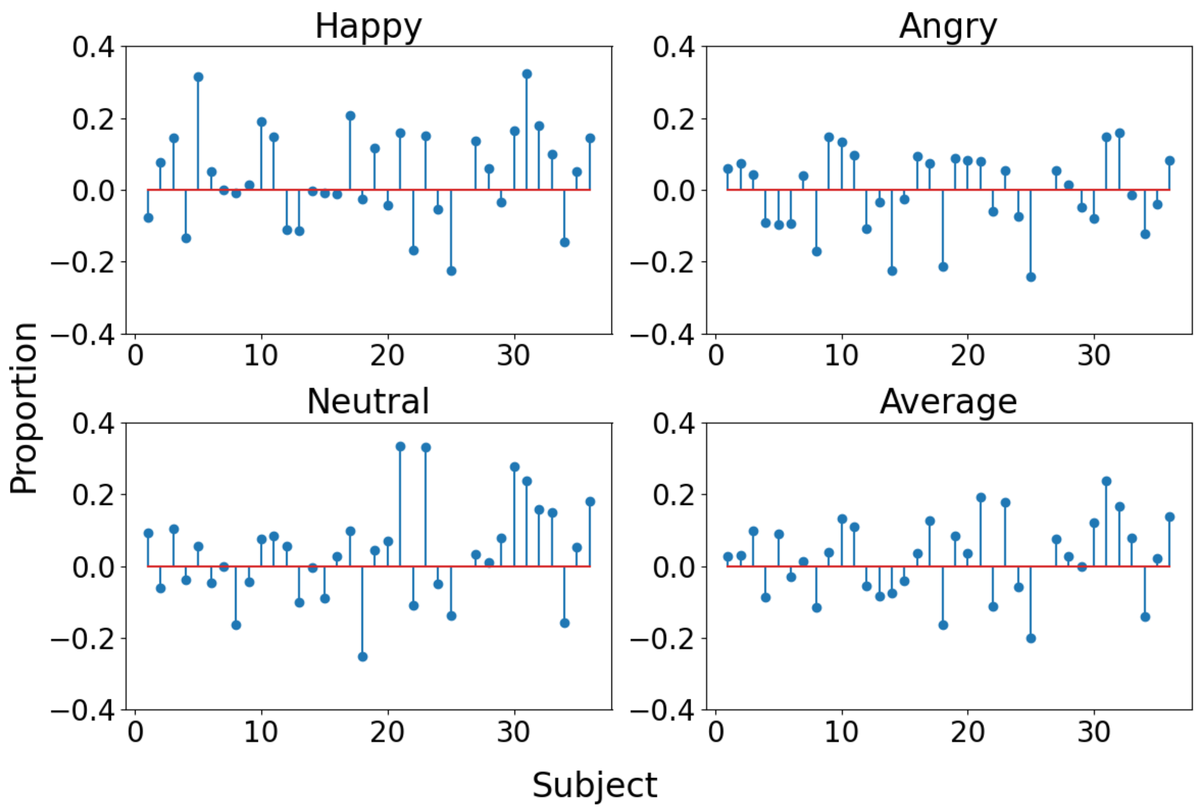


Figure 3.18 Individual EMS across participants for each emotion condition. The x-axis represents individual infants, and the y-axis represents their EMS values (proportion of looking time difference between congruent and incongruent trials). Separate panels display results for happy, angry, and neutral conditions, with the bottom-right panel showing the across-condition average. The red horizontal line indicates the chance level (EMS = 0).

To assess infants' attentional engagement across the experiment, we analyzed the number of eye movements directed toward the images over trials. This measure allows us to evaluate whether attention declined systematically during the task, which is important for verifying data quality and ensuring that subsequent analyses are not confounded by fatigue or habituation effects.

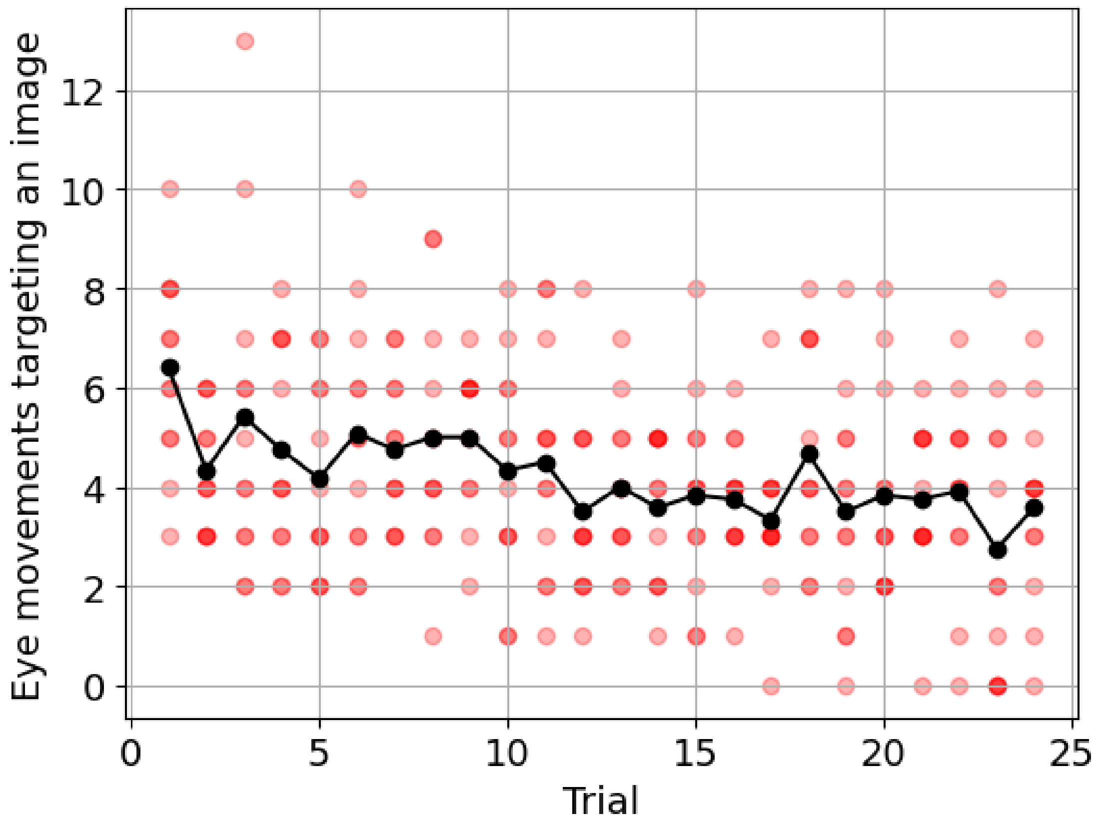


Figure 3.19 Eye movements targeting an image across trials for all infants. The x-axis represents trial number, and the y-axis indicates the number of eye movements directed toward the image within each trial. Red dots show individual data points across subjects, while the black line represents the across-subject average.

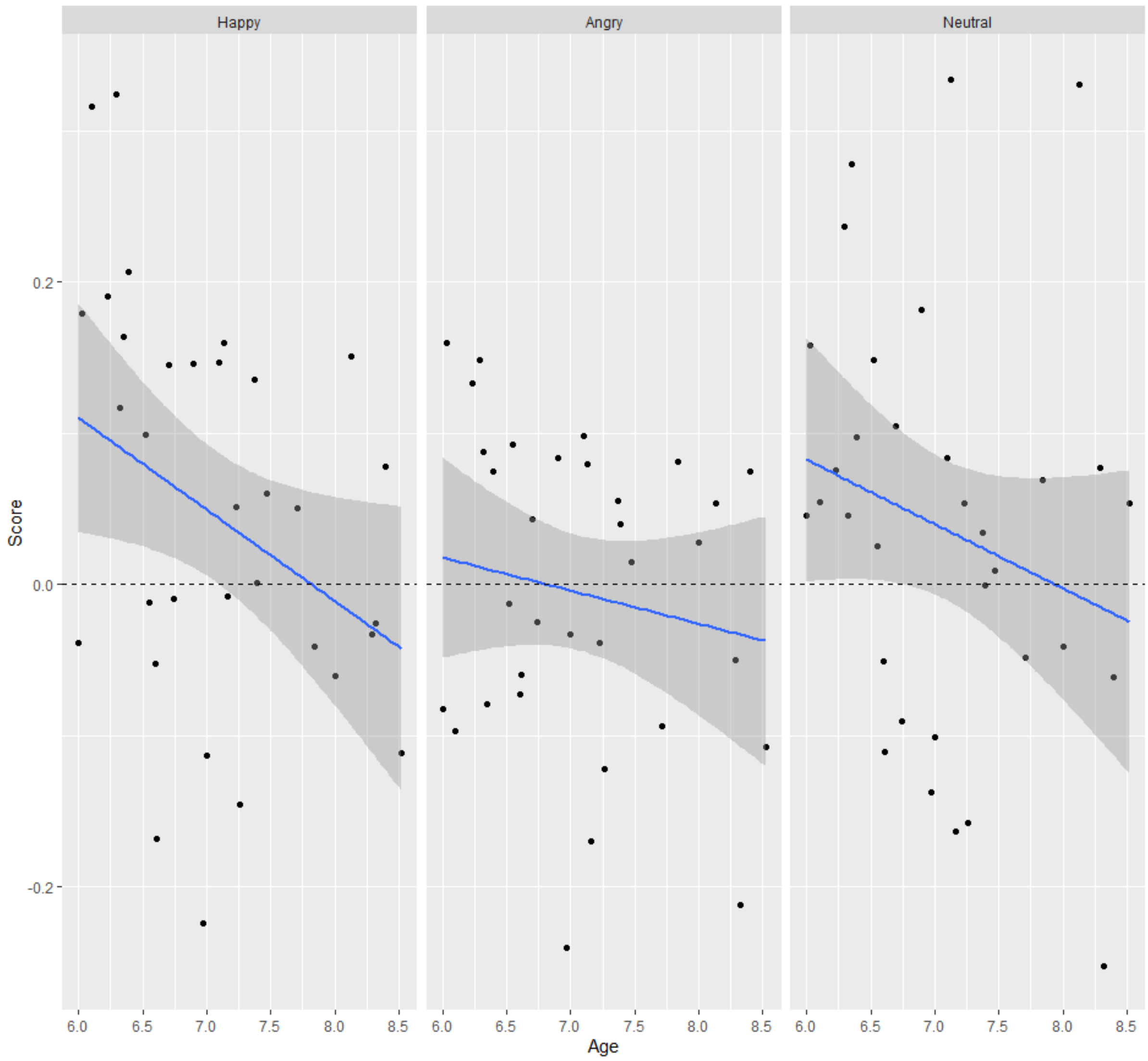


Figure 3.20 EMS as a function of age across the three emotional conditions.

Note. Each panel depicts individual infants' EMS values (black dots) and the fitted regression line with 95% confidence intervals (blue line and shaded area). Positive scores indicate a preference for the congruent facial expression. A negative association with age was observed in the Happy and Neutral conditions, whereas no clear age-related trend emerged in the Angry condition.

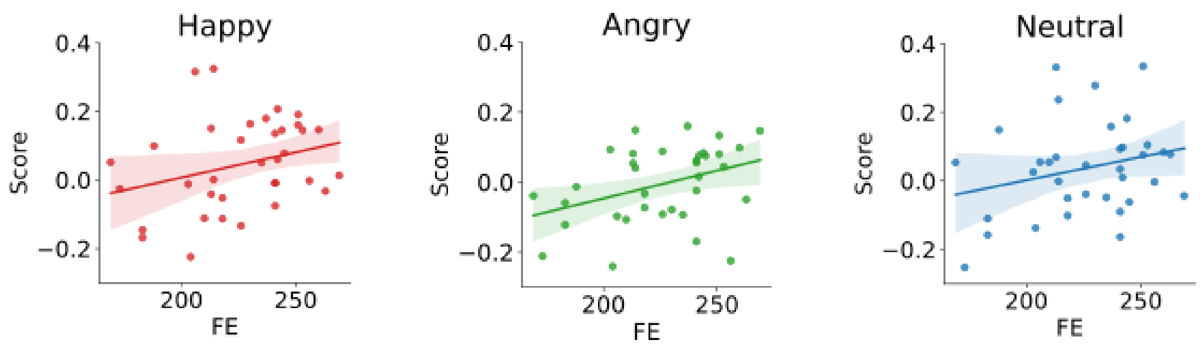


Figure 3.21 EMS as a function of family expressiveness (FE) across the Happy, Angry, and Neutral conditions.

Note. Each panel shows individual infants' EMS values (colored dots) together with the fitted regression line and 95% confidence intervals (shaded area). Positive scores reflect a preference for the congruent facial expression. Across all three emotion conditions, higher family expressiveness scores were associated with higher EMS values.

To explore how the family emotional environment influences infants' multimodal emotion processing, we conducted correlation analyses between family expressiveness (FE, PE, NE) and infants' EMS for Happy, Angry, and Neutral conditions.

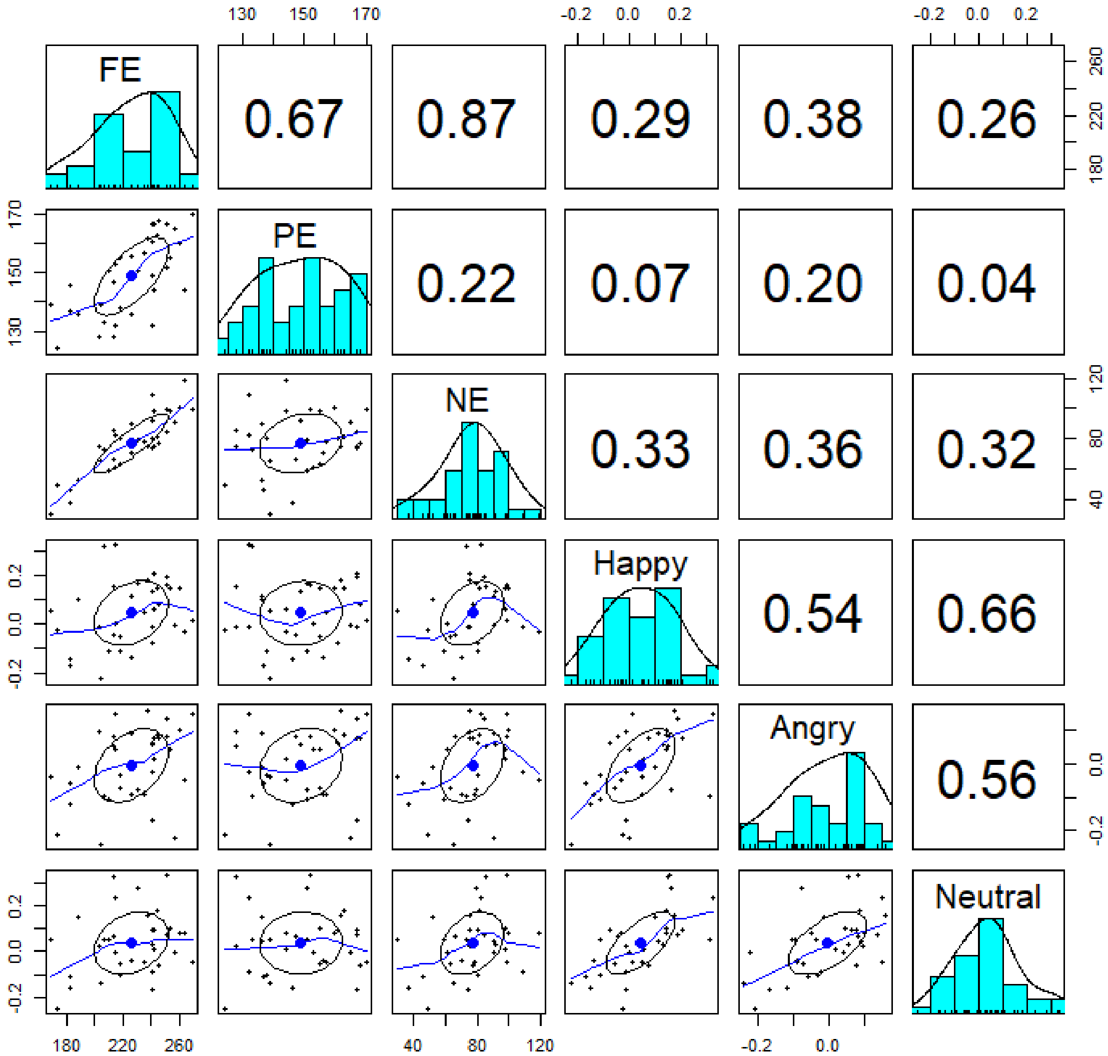


Figure 3.22 Correlation matrix showing the relationships between family expressiveness (FE, PE, NE) and infants' EMS across Happy, Angry, and Neutral conditions. The diagonal displays distributions of each variable, the lower panels show scatterplots with regression lines and confidence ellipses, and the upper panels present Pearson correlation coefficients. The results indicate positive associations between FE and infants' EMS, particularly in the Happy and Neutral conditions. This analysis addresses whether the emotional environment in the family influences infants' multimodal emotion processing.