THEORETICAL/REVIEW

How do human newborns come to understand the multimodal environment?

Arlette Streri[1] · Maria Dolores de Hevia[1]

Abstract

For a long time, newborns were considered as human beings devoid of perceptual abilities who had to learn with effort everything about their physical and social environment. Extensive empirical evidence gathered in the last decades has systematically invalidated this notion. Despite the relatively immature state of their sensory modalities, newborns have perceptions that are acquired, and are triggered by, their contact with the environment. More recently, the study of the fetal origins of the sensory modes has revealed that in utero all the senses prepare to operate, except for the vision mode, which is only functional starting from the first minutes after birth. This discrepancy between the maturation of the different senses leads to the question of how human newborns come to understand our multimodal and complex environment. More precisely, how the visual mode interacts with the tactile and auditory modes from birth. After having defined the tools that newborns use to interact with other sensory modalities, we review studies across different fields of research such as the intermodal transfer between touch and vision, auditory-visual speech perception, and the existence of links between the dimensions of space, time, and number. Overall, evidence from these studies supports the idea that human newborns are spontaneously driven, and cognitively equipped, to link information collected by the different sensory modes in order to create a representation of a stable world.

Keywords Newborn · Multimodal · Integration · Cognitive development

Introduction

Looking for your apartment keys in a handbag filled with various objects, turning your head towards an unusual sound, eating your meal, driving a car, talking while walking, are all behaviors that mobilize two or even several senses. While all of our activities rely on the simultaneous participation of our different sensory modalities, we pay little attention to the interactions between them. Likewise, most objects and events in our environment require the participation of several of our senses in order to be perceived. Consequently, the exchanges between the individual and his/her environment are *multimodal*. The difficulties lie in the fact that each system picks up information that is unique to itself. We know that receptors of vision are sensitive only to electromagnetic

waves (light), a first step to seeing the world in color; receptors of the skin only react to pressure, mechanical changes, and temperature; those of hearing only detect sound waves; and taste and smell process only chemical molecules. Each sensory system has access to a universe of properties that is specific to itself. Despite the specificities of receptors of each modality, adult humans see and act in a world that seems stable, coherent, organized, and meaningful; how is this process achieved? In order to accomplish this, the human organism has to resolve the paradox of evolving in a stable and unified world while extracting information from the external world through sensory modalities whose structure and functioning appear to differ profoundly.

In this review, we aim at complementing previous major revisions of the literature regarding intersensory perception in infants, children, and non-human animals (e.g., Bremner et al., 2012; Lewkowicz & Lickliter, 1994) by highlighting new elements that help us understand how multimodal perception is coped with at birth. One is the contribution of fetal studies in shedding light on neonates' understanding of the environment; this aspect has rarely, if ever, been mentioned before. In fact, taking into account evidence from neonatal

✉ Maria Dolores de Hevia
   dehevialola@gmail.com

   Arlette Streri
   arlette.streri@gmail.com

[1] Université Paris Cité, CNRS, Integrative Neuroscience and Cognition Center, F-75006 Paris, France

studies allows us to highlight an essential transnatal continuity that would constitute the basis for the links between sensory modalities at birth. Studies are reviewed that suggest that the fetal period sets the stage and prepares the neonate to navigate the postnatal period. Another main aspect of this review is the possibilities that new, more naturalistic stimuli (e.g., using videos of speaking faces), as well as the use of bi-modal stimulation, open for understanding the presence of critical newborns' abilities and the constraints operating on them that were previously unknown (e.g., the ability to represent and discriminate between quantities). We see how recent research on neonatal cognition has proven to be effective in better defining the tools the newborn relies on in order to perceive and understand the postnatal multimodal environment.

An old question in philosophy and psychology

Aristotle and other philosophers examined the question of how humans perceive a stable world using signals coming from different sensory modes from a theoretical point of view. For a very long time, it was primarily the particularities of the sense modalities that held the attention of philosophers and researchers (Révész, 1934). In order to unify the individual's environment, some have postulated the existence of mediators (von Helmholtz, 1885) such as language (Ettlinger, 1967), mental images, schemes, or codes through which the transfer of sensory information from one channel to another takes place (Bryant, 1974). This transfer would enable the creation of links between the different sensory modalities. The existence of suprasensory or "intermodal" dimensions postulated by the Gestaltists (Koffka, 1935; Köhler, 1964) represents the opposite hypothesis. Perceptions coming from different sensory systems are thought to have elements in common, such that one and the same perpetual quality can be obtained at once. The Gestaltists postulate that more than one correspondence exists between the modalities, and this is the result of the process of synesthesia. The stimulation of one modality influences the functioning of another. James Gibson' s theory (1950, 1967), which is very close to that of the Gestaltists, insists more specifically on the role of sensory input without, however, neglecting the role of action. According to this theory, the proximal stimulus (the retinal projection) carries a great amount of information about the external world, and the process of perception consists of discovering invariants in the stimulation (see Streri, 1993, for a brief historic review).

What has been argued with respect to the human infant, and in particular for the human newborn? It has long been thought that the abilities necessary in order to perceive and act in a stable world are nonexistent at birth. For a long time, philosophers, scientists, and medical doctors, among others, attested that the human newborn brain was like a "Tabula rasa." This epistemo-philosophical concept implies that the human mind is born virginal, without built-in content, and is marked, shaped, or "impressed" only by experience. Under this view, the mind is mainly characterized by its passivity in the face of sensorial experience. An often-cited example is the well-known Molyneux question regarding the integration between senses: "Will a blind person who recovers sight as an adult immediately be able to visually distinguish between a cube and a sphere?" (Bruno & Mandelbaum, 2010; Gallagher, 2005). For a long time, the answer was negative. Diderot, the first philosopher to compare a blind person to a neonate, claimed that "vision must be very imperfect in an infant that opens his/her eyes for the first time, or in a blind person just after his/her operation" (Diderot, 1749). However, the analogy between the blind person and the neonate is hardly justified, as the former has had an extensive experience with other perceptual systems and has plausibly developed alternative strategies to link senses between them.

How does the human newborn start to interact with the multidimensional world at the beginning of postnatal life? In this review, we first briefly introduce and discuss studies on the development of sensory modalities in utero, including the capacities and limits, which shed some light on the origins of newborns' perceptual abilities. Next, studies on human newborns' perceptual abilities are presented. Because traditionally these studies have been centered on isolated abilities such as face recognition or language, using experimental conditions that exposed newborns to single sensory modalities, this part refers to newborns' capacities when encountering these isolated sources of environmental stimulation. We later turn to studies using richer stimulation conditions, and therefore mobilizing two or more sensory modalities, which more faithfully reflect the real environment the newborn faces outside the womb. Some of the tools that might guide newborns when linking information coming from different senses are discussed as well. These include the detection of invariants (i.e., establishing stability despite the variation in the stimulation), the existence of amodal properties of objects (i.e., an object's feature can be expressed and detected by multiple senses), and the spatiotemporal synchrony (i.e., two concurrent stimulations that take place at the same time and/or place tend to be perceived as belonging to a unique event). We end by discussing research across three different domains: (1) intermodal transfer between vision and touch; (2) recognition of speaking faces; and (3) relationships between number, space, and time. Overall, these studies suggest that a human newborn must have the abilities necessary to perceive a stable world at least, using the capacity to meaningfully compare information from different senses and create expectations of congruency between them right from birth.

The fetal period: The origins of our senses; richness and limits

Empirical studies on the development of sensory modes during the fetal period were conducted relatively recently. Using ultrasonography it has been shown that the anatomy and functioning of different sensory modalities emerge during gestation, not at birth. What are the origins of our different senses? Will these origins allow newborns to perceive and understand the complexity and the richness of the environment?

Origins and development of our senses

The description and detailed analyses of the anatomy and function of our multiple sensory systems during gestation revealed that the sensory modes emerge and function in a successive manner. At 4 weeks/7 weeks of gestation, touch is the first to appear and to function, followed at 7 weeks/12–13 weeks by the chemical senses (olfactory and taste), the vestibular system (10 weeks/16 weeks), the auditory mode (9 weeks/24 weeks), with the visual system the latest to appear at about 22 weeks/28 weeks of gestation (see Bremner et al., 2012, for an adapted version from Gottlieb, 1971[1]). This analysis suggests, first, that despite the temporal succession in the anatomy of different senses, there is an overlap in their functioning. Also, because sensory receptors allow the fetus to receive and process information from the external world in a multimodal way, the beginning of integration could already take place at this prenatal stage, with the exception of vision, the only sense that might not be fully functional until birth. Moreover, all sensory modalities continue their maturation process beyond birth, plausibly at different speeds.

For certain modalities such as hearing and olfaction there is a continuity between the fetus and the newborn, suggesting that the newborn has already been prepared in the womb to perceive the world. Evidence comes from the memory trace formed during the fetal period that can be measured at birth. In the case of hearing, transnatal continuity has been evaluated concerning not only language-related stimuli, but also musical stimuli, and even meaningless sounds such as background noise. Studies on the fetal sound environment, as well as the physiological development of the auditory system, have shown that the auditory system of the fetus has a good resolution during the last weeks of pregnancy, being functional at birth, even if infants' hearing discrimination abilities continue to develop during the first two years of life

(see Streri et al., 2013, for a review). In utero, the transmission of external and maternal sounds has been evaluated in pregnant women, showing that frequent and repeated exposure to the mother's voice produces calming, orienting, and preference reactions in the newborn (DeCasper & Spence, 1986). Moreover, newborns are able to remember music, rhythms, or voices heard multiple times in utero: for instance, newborns stop crying and pay attention to familiar melodies as soon as they are heard. It has been shown that a prenatal memory of a short melody can last up to 6 weeks after birth, with a decrease in newborns' cardiac rhythm when exposed to it (Granier-Deferre et al., 2011). With regard to meaningless sounds, studies have shown that the fetus seems to be able to develop a specific and lasting adaptation mechanism to high-intensity noises such as the takeoff or landing of planes (Ando & Hattori, 1977). Newborns familiar with these noises during the fetal period do not wake up to the sound of an airplane emitted at 90 dB, whereas they wake up and cry to a control stimulus of lower intensity such as a musical sequence (see Granier-Deferre et al., 2004, for a review).

Continuity between fetal and the newborn's auditory memory skills seems to be even more remarkable for language. This finding supports the idea that humans are biologically endowed with the faculty of language, presenting innate predispositions for perceiving and learning their species' speech (Dehaene-Lambertz et al., 2002; Doupe & Kuhl, 1999). It has been recently suggested that neonates' speech perception abilities are already shaped by prenatal experience with speech, which for the fetus consists of a low-pass signal mainly preserving prosody (Nallet & Gervain, 2021). Prosody perceived during the very first experiences with language plays a fundamental role in speech perception and language development. The ambient language to which fetuses are exposed in the womb also starts to affect their perception of their native language at a phonetic level. Newborns discriminate between familiar versus unfamiliar vowels (Moon et al., 2013), prefer the mother's voice versus an unfamiliar one (DeCasper & Spence, 1991), and their native language versus a foreign one (Moon et al., 1993). Using the fNIRS technique, Benavides-Varela et al. (2011a, b) showed that newborns are able to remember words after interference or a brief silent period, suggesting that fetal stimulation with language seems to boost the newborn's memory for the native language. Overall, studies on newborns' language perceptual abilities underscore the idea that a multisystem view better characterizes language acquisition abilities, with the participation of, at least, the genetic endowment, the prenatal contextual stimulation, as well as the postnatal cultural environment (Lewontin, 2000). Moreover, the fact that during the first months of life humans are able to discriminate phonemes belonging to any language (Jusczyk, 1997; Kuhl, 2004), together with the fact that right from birth newborns

---

[1] The first date indicates the anatomy and appearance of the receptors of the modality. The second date corresponds to the functioning of that same modality.

are able to recognize, and prefer to attend to, their native language (Moon et al., 1993), suggests that there is a rather important degree of flexibility that eventually allows humans to adapt to contrasting cultural environments.

When considering the olfactory modality, it is known that newborns possess a memory trace of the scent and odor components exposed to during fetal life. Using the preferential choice technique, newborns were presented with the smell of amniotic fluid they had bathed in, the smell of unfamiliar amniotic fluid, and the smell of a neutral substance. They showed a preference for the familiar scent of amniotic fluid over both the unfamiliar and the neutral scent, while they preferred the unfamiliar smell of amniotic fluid over the neutral substance (Klaey-Tassone et al., 2021; Schaal et al., 1998). Very quickly, during the first two or three postnatal days, newborns develop a preference for the odor of colostrum compared to the familiar amniotic fluid; later, this preference gives way to that of breast milk. Newborns therefore display a succession of adaptive behaviors depending on the diversity of substances available to feed (Schaal, 2005).

Furthermore, it is very likely that fetuses simultaneously receive different stimulations from external events, selectively reacting to them, as well as stimuli such as speech, movements, and tactile pressures directly from the mother. For example, during the third trimester of gestation, fetuses show evidence of communicative engagement with the mother: they display more self-touch responses in reaction to the mother's touch, and open their mouth for a longer time when mothers talk (Marx & Nagy, 2015; Nagy et al., 2021). However, despite extensive research on fetal stimulation (see Granier-Deferre et al., 2004, for a review), the integration between senses in fetuses is difficult to evaluate.

The limits

Studies reviewed above point towards the richness of fetal life; however, because the visual mode is not functional before birth, we cannot know whether these different and multimodal stimulations are integrated in the same object or person from birth. For instance, hearing the mother's voice does not allow the fetus to link the voice to the mother's face. In the external environment, seeing an object involves that the latter reflects the light it receives on the retinal receptors of the observer. This process cannot take place inside the womb, where the fetus lives in a closed and dark environment.

While the external environment already imprints fetuses' perceptual abilities, birth is the first moment newborns truly share our environment. Inside the womb, sounds are distorted and attenuated because the properties of the uterus render many phonemic differences imperceptible (Querleu et al., 1988), the chemical senses are mixed, and the fetus is fed via the umbilical cord. The haptic mode is less used because the fetus does not have any opportunity to grasp an object except for the umbilical cord, and the tactile sense is limited to touching the uterine wall, face, or body. Given that vision is the sense that provides us with the richest information about the world and is the most frequently used to perceive the environment, and that other senses such as auditory and tactile often interact with it, the question remains whether newborns are able to link these different sources of information in order to perceive and understand the complexity of the environment.

The beginning of research on neonatal perception across different sensory modalities

When newborns emerge from their close environment they are confronted with a richer, more complex one where a variety of stimuli are immediately available, sometimes in the form of only auditory, visual, or tactile information, and more often in a multimodal form such as audiovisual as well as visuo-tactile. In this new context, dynamic objects and moving people speaking or silent regularly occupy the vast spatial surrounding. Given the complexity of this new environment, and because the sensory modalities of newborns are not fully developed, it has historically been a difficult question to address whether humans at birth are capable of creating links between the senses.

The visual mode: Face perception

The first studies on the perception of newborns were mainly devoted to showing if and how the newborn would react to the presentation of stimuli in a single sensory modality, surprisingly the least efficient at birth: vision. For example, Fantz (1963) revealed a preference for the outline of the face over other targets. Using the preferential gaze time procedure, the results revealed that newborns were able to discriminate between different targets and showed greater attention to faces. A decade later, other studies confirmed the attraction of newborns to the human face when tested just a few minutes after birth (Goren et al., 1975). Subsequently, intrigued by this preference for the human face, several studies designed experiments to understand which parts of the face were especially attractive to the newborn (Morton & Johnson, 1991; Turati et al., 2002). To test for this visual bias, researchers primarily presented two-dimensional targets and simplified facial schema structures as a three-square configuration, with two squares for the eyes and one for the mouth (Macchi Cassia et al., 2004; Valenza et al., 1996), or abstract faces (Simion et al., 2003), thus revealing that face structure is systematically preferred at birth when presented in a canonical orientation, but not when presented upside

down. Finally, using photographs of human faces, Turati and colleagues presented facial stimuli along different orientations and found that newborns show a preference for the face perceived from the front (Turati et al., 2008). However, in a face there are two fundamental elements that trigger the newborn's attention: the eyes and the mouth. In addition, the mouth produces sounds and especially language that are important for newborns. How do we know if the newborn binds the face to the language?

The auditory mode: Language perception

Initial studies provided some evidence that the newborn has an ability to process language. For example, newborns are able to memorize nonsense syllables heard for 1 h on the day of birth that are presented 24 h later for a recognition test (Benavides-Varela, Gómez, Macagno, et al., 2011a; Swain et al., 1993), prefer a story heard during the last weeks of pregnancy to a new one (DeCasper & Spence, 1986), and his/her mother tongue to a foreign language (Moon et al., 1993). Phonetic discrimination has also been extensively investigated in neonates (Eimas et al., 1971; Mehler et al., 1988). Researchers have shown that between birth and 4 months of age humans can discriminate between all the phonetic contrasts used in the world's languages (Jusczyk, 1997; Kuhl, 2004). This early ability, therefore, will necessarily develop under the influence of the stimulation of the language in which the child is immersed, most often the mother tongue, in order to adapt to adult speakers. Mehler and colleagues proposed that speech flow segmentation procedures should be based on an early sensitivity to prosody (the music of language), and more specifically to speech rhythm (Mehler et al., 1996), which would not be still at the syllable level. Additional support for the neonates' brain specialization in language comes from brain imaging studies showing that speech and music are treated differently in the human brain, and that vocal and non-vocal sounds are processed by different areas of the brain (Benavides-Varela, Gómez, & Mehler, 2011b; Gervain et al., 2008).

The haptic mode and the chemical senses

The haptic modality has also been studied in humans early in postnatal life. Studies have shown that beyond the grasp and avoidance reflexes, the newborn is able to manipulate a small object (e.g., a cylinder or a prism) and retrieve certain information about it. It has been shown that if the same object is presented several times in the newborn's hand, a decrease in holding time is observed (i.e., habituation process) and the presentation of a novel object triggers an increase in holding time (i.e., reaction to novelty). Using this method, it was inferred that newborns are able to discriminate between

two objects in terms of their shape when presented in the tactile modality, and that this ability is observed with both the left and the right hands (Streri et al., 2000).

The functional properties of both chemical senses, taste and smell, have also been studied in newborns. We know that the taste of a sweet solution triggers newborns' sucking behavior, tongue protrusion, and a facial expression that resembles a smile, and has been interpreted as indicating pleasure (Steiner, 1979). On the other hand, a bitter or acidic solution immediately triggers in newborns crying, grimacing, head-turning reactions, as well as overt rejection behaviors. All these reactions observed in the gustatory modality have, however, been less frequently described in the olfactory modality, although newborns can show some hedonic reactions when stimulated with certain odors (Soussignan et al., 1997) by changing the facial expression, and in particular showing nasofacial and respiratory changes. Compared to organic odors (breast milk, amniotic fluid), newborns find a pure odor like vanilla to be pleasant, while butyric acid from rancid butter and Parmesan cheese is considered unpleasant (Soussignan et al., 1999; see also Tristão et al., 2021, for a full review of these sensory modalities).

Regardless of the sensory modality investigated, these studies have revealed some of the perceptual and cognitive abilities of newborns in their first days after birth by using methodologies widely used in studying the perceptual abilities of older infants. The findings from studies using behavioral (e.g., habituation, visual preference) or brain imaging (ERP, MEG) methods support the conclusion that newborns are capable of discriminating stimuli with all of their senses and to detect invariants in order to perceive information in a stable way.

It may seem paradoxical that researchers were only interested in the functioning of a single sensory modality despite the fact that virtually all behaviors involve the integration of various sources of information coming from our different senses. Only in this way is a significant interaction with the environment produced. The reason for this choice undoubtedly stems from the complexity of this integration, since it is carried out in very varied situations and cannot be reduced to an additive or subtractive phenomenon of the different sources of stimulation (Stein & Meredith, 1993). Furthermore, the modular or "isolationist" approach has the advantage of emphasizing the diversity and richness of the perceptive function by emphasizing the specificities of each sense, of its organ, of its own object. But it seems that many sensory inputs are common to the different senses, such as movement, magnitude, number, duration, and intensity, among others. This question of correspondence or communication between the senses was famously raised with great modernity by Aristotle in his treatise "On the soul."

The newborn in a multimodal context

A theoretical issue

Why is it important to study the human newborn from 1 to 3 days of postnatal life? Historically, the behaviors of newborn babies have been at the center of a number of controversies in both philosophy and psychology, and especially when it comes to the question of the integration of the senses. The origin of the behaviors present in the newborn was attributed to heredity and the origin of those acquired later was attributed to environmental influences. Despite evidence of transnatal continuity between the fetal sensory and motor periods and the performance of newborns at birth, and because vision is not functional before birth, the question of multimodal sensory integration remains at the center of the controversy between Piaget's and Elenaor Gibson's theories. On the one hand, Piaget's theory (Piaget, 1937) assumes that the newborn begins life with separate sensory worlds, while on the other hand Eleanor Gibson (1969) argues that the human being is capable of creating a unified perceptual world at birth.

According to Piaget, humans gain knowledge of the environment through active motor behavior; in fact, since newborns' motor behavior is weak, Piaget was not interested in infants' perceptions but rather in the spaces in which each system could act. For example, he considered the problem of grasping behavior in terms of schema coordination (Piaget, 1936), but ignored the question of cross-modal perception. Under this view, the newborn receives information from the world through reflexes, which are transformed into specialized and independent action schemes. When a baby manipulates or holds an object (tactile space) he/she does not try to see it; if the baby sees an object (visual space) he/she does not try to catch it. When at around 4/5 months of age babies can both see and manipulate the object, the merge of separate spaces takes place. Indirect support for Piaget's theory comes from studies on the influence of external spatial representation on young infants' orienting responses to tactile stimuli (Begum Ali et al., 2015). When 4- and 6-month-old infants' feet are stimulated in crossed and uncrossed legs conditions, only 6-month-olds show better localization in the uncrossed than in the crossed posture, whereas 4-month-olds perform equally in both conditions. The authors interpreted these findings as providing evidence that the youngest remain in a "state of tactile solipsism," perceiving touches only in relation to anatomical coordinates without taking into account visuo-spatial information. Related studies examining auditory, tactile, and audio-tactile localization abilities in infants provide evidence that visually impaired and sighted infants show comparable auditory localization abilities, but only infants deprived of visual input show difficulties in both tactile and audio-tactile stimulations, suggesting a crucial role of vision in the integration of senses in the course of development (Gori et al., 2021).

However, because visual perception of objects is functional in newborns, albeit still immature (Braddick & Atkinson, 2011), the visual modality can influence other sensory modalities from birth, supporting James Gibson's and Eleanor Gibson's theory of perception. Studies on (pre-) reaching actions at birth have shown that when newborns are stimulated with the presence of a moving object, they reach out their hand and open their arm towards it more frequently than when the object is not present (von Hofsten, 1982), thus supporting this view. Furthermore, newborns are able to link some visual information taken from an object with tactile and auditory information. Recently, neonatal manual specialization depending on acoustic stimulation, language, or music, has been observed (Morange-Majoux & Devouche, 2022). In these studies, when newborns are listening to music in the presence of an object they activate their left hand, while when they are listening to speech, they activate their right hand in the presence of the same object. Moreover, sucking movements were performed more often in the language than in the music condition. In the same vein, it has been shown that newborns' motor activity is enhanced when listening to their native language, as opposed to a foreign language (Hym et al., 2022), and they are able to use visual information regarding the texture of an object and adapt accordingly the actions made on the object when manipulating it (Molina & Jouen, 2001). These studies suggest that newborns are able to compare information across different modalities, such as touch, vision, and audition, and that a modification in their environment in one modality sensibilizes and modifies behavior measured in another modality. These findings could be interpreted, however, as a form of synesthesia as suggested by Gestalt's theory, without implying that newborns create a unified object by integrating information across different senses.

What tools are available to the newborn to integrate the information collected by the different senses?

The diversity of interactions between senses depends on the properties of objects or events as well as on the information the senses are able to extract from these properties. We propose three tools that would allow newborns to obtain a coherent and stable world: (1) detection of invariants, (2) amodal properties of objects, and (3) spatio-temporal synchrony.

The ability to detect invariants in the environment ensures a stable perception of the world and is one of the most critical features of humans' perceptuo-cognitive abilities. To

detect an invariant, either a moving stimulus is needed or the organism encounters the same stimulus several times to be recognized. Newborns' detection of invariants can be measured with the habituation/reaction to novelty method, a technique that can be used with all sensory modalities. When presenting the same object to a newborn several times, a decrease in attention (measured in seconds) during the course of consecutive trials (i.e., habituation) is often observed. Afterwards, when the newborn is presented with an object containing a different property from the first object, he/she recovers attention (i.e., reaction to novelty). This measurable increase of attention reveals that the newborn has detected a modification in the stimulation. With this technique one can evaluate both learning across trials (during habituation) as well as the detection of an invariant (during test trials or novelty presentation). Another technique frequently used is the preferential choice technique, in which newborns are presented with two objects simultaneously and the preference (e.g., longer looking time) is measured to infer a preference for one of the two. This technique does not allow the newborn to detect an invariant, but to distinguish between two stimuli.

The two other tools newborns can use depend on the properties of the environment. Indeed, some objects' and events' properties are shared across several modalities. This is the case for the shape and the texture properties of an object, as these are two sources of information that can be captured through both vision and touch. These properties are qualified as *amodal*, as information is not specific to a particular sensory system. More abstract properties have been hypothesized to be amodal in nature, such as the dimensions of space and time (Gibson, 1969). In cases where amodal properties are available, the interaction/integration between senses might be facilitated. Other properties are *specific* to one modality; for example, color for vision, weight for touch, and sounds for audition. In these conditions, the interactions between senses could be more difficult to observe. Finally, *spatiotemporal synchrony* can help the integration between senses. For example, a rattle that is shaken and makes noise or music can be considered as coming from the same object even if the shape of the object is captured by vision and the sound is captured by audition (two specific properties captured by two different modalities). This synchrony in space and time has been investigated in detail by Bahrick (1987, 1988), who established the foundations of the intersensory redundancy theory (Bahrick & Lickliter, 2000, 2012; Bremner et al., 2012). This theory has highlighted the important contribution of intersensory redundancy in the development of perceptual, cognitive, and social learning in young infants. Some links between audition and vision already exist in newborns when the conditions of temporal and spatial synchrony are met. Morrongiello et al. (1998) have provided evidence that newborns can associate objects and sounds based on the combined cues of collocation and synchrony. They are even able to learn arbitrary auditory-visual associations (e.g., between an oriented colored line and a syllable), but only in the condition where the visual and auditory information are presented synchronously (Slater et al., 1999).

However, newborns are not always stimulated by synchronous events or objects. For example, they can occasionally touch or grasp an object in their bed; will they be able to recognize it afterwards when visually presented with that same object? Sometimes they hear a sound or speech without seeing the source object or the face; will they recognize that object or the silent face if visually presented afterwards? In this context, recognition involves identification when stimuli are presented sequentially, not simultaneously. Are newborns capable of intersensory integration in the absence of spatial and/or temporal synchrony? Studies on varied domains that attempt to address this question are reviewed next. First, studies on the relation between touch and vision in cross-modal transfer tasks; second, studies on the relation between audition and vision during the presentation of speaking faces; and third, by examining the Eleanor Gibson (1969)'s proposition that spatial, temporal and magnitude dimensions are amodal in nature, i.e., they are available to all sensory modalities right from birth.

Intermodal transfer between touch and vision in human newborns

The links between the haptic and visual modalities are not well established at birth and will strengthen until the age of 15 years. The period of coordination between vision and prehension, by 4–5 months of age, is critical to link these two senses as infants are able to grasp an object they see and to look at an object that is put in their hand. Infants can therefore simultaneously explore objects both manually and visually in order to obtain a complete representation of them. Because newborns cannot engage in bimodal visual-haptic exploration of an object, a cross-modal paradigm can be used at birth to uncover the nature of the links between these two modalities.

The cross-modal transfer task was used to study Molyneux's question, i.e., the transfer of information from touch to vision, in 1-year-old infants (Rose et al., 1981). The task involves two successive phases: a habituation phase in one modality, followed by a recognition phase in a second modality. The mouth and the hands are the best organs for perceiving and knowing the properties of the environment using the tactile modality. Cross-modal transfer of texture (smooth vs. granular) and substance (hard vs. soft) from the oral modality to vision has been shown in 1-month-old infants (Gibson & Walker, 1984; Meltzoff & Borton, 1979),

supporting the idea that detection of shape invariants across touch and vision is functional early in the first month of life, and therefore independent of extensive experience in simultaneous tactual–visual exploration.

Streri and Gentaz (2003) provided the first evidence of cross-modal transfer from hand to eyes in newborns aged from a few hours to up to 3 days. Before engaging in the cross-modal task, the authors made sure that newborns were able to discriminate the objects separately in each modality. In the "tactile-to-visual modality" task, an object is put in the newborn's hand (cylinder or prism), and in the second visual phase, the familiar and a novel object are visually presented in alternation across four trials. A transfer is considered successful if the newborns look at the novel object longer than at the familiar object, compared to a control group where newborns look at the two objects without tactile familiarization. In the "visual-to-tactile modality" task, newborns are first visually habituated to an object, and then, in the second phase, the familiar and the novel object are tactually presented to their right hand in alternation across four trials. The results are also compared against a control group where newborns receive in their hand the objects in alternation without visual familiarization (Sann & Streri, 2007; Streri & Gentaz, 2003). The results have revealed that the transfer is unidirectional, with visual recognition observed following haptic habituation, but not the reverse, no haptic recognition was found following visual habituation. Moreover, cross-modal transfer seems to be modulated by handedness in newborns, with no evidence of cross-modal recognition when the left hand is stimulated (Streri & Gentaz, 2004). This asymmetry is observed as well in 2-month-old infants: infants can visually recognize an object they have previously held, but do not manifest tactile recognition of an object previously seen (Streri, 1987). This asymmetry is surprising because the properties of the objects should give rise to the same representation via the detection of invariants in the two modalities.

A number of explanations can account for this pattern of performance. First, the two modalities differently process the object shape, with vision processing shapes in a global manner, and touch processing information sequentially. Second, newborns and 2-month-olds do not use efficient tactile exploratory procedures, "*Contour Following,*" to create a stable representation of shapes (Lederman & Klatzky, 1987). Third, the format of representation created by each modality is not similar enough to ensure the exchange of information between sensory modalities. Fourth, a plausible hypothesis could be that before the coordination between vision and prehension is well in place, the intermodal transfer between touch and vision would be hard to establish.

Several studies have also revealed that, over the course of development, the links between the haptic and the visual modes are fragile, are often not bi-directional, and the representation of the objects is never complete. These findings have been shown not only in 5- to 6-month-old infants (Rose & Orlian, 1991; Streri & Pêcheux, 1986), but also in children (Gori et al., 2008) and adults (Kawashima et al., 2002). Indeed, cross-modal transfer of information is rarely reversible, and even when it is bi-directional it generally displays an asymmetrical pattern (Hatwell et al., 2003). The links between sensory modalities regarding object shape appear therefore to be flexible rather than immutable across development.

These findings are puzzling given that from 5 months of age infants have the possibility to simultaneously compare objects' visual and haptic information through a bimodal exposure to them. In other words, it appears that intersensory redundancy (see Bahrick & Lickliter, 2000) enabled by this learning is not sufficient to contribute to a bi-directional exchange of information between these two modalities in intermodal transfer tasks. Intermodal transfers between these modalities require a change of format and seem to be more difficult for newborns because of the higher level of abstraction involved. Indeed, intermanual transfer of shape is bidirectional in 2-day-old newborns despite the immaturity of the corpus callosum (Sann & Streri, 2008). When a newborn holds an object in one hand, left or right, its shape is recognized by the other hand. Newborns are able to intramodally and intermodally differentiate a curvilinear from a rectilinear shape.

Shape and texture are two amodal properties that are processed through both vision and touch. Studying these two properties could therefore allow us to test the hypothesis of amodal perception in newborns anew, as well as to shed light on the processes involved in gathering information by both sensory modalities. However, shape is best processed by the visual modality, whereas texture is thought to be best detected by touch (see Bushnell & Boudreau, 1998; Klatzky et al., 1987). According to Guest and Spence (2003), texture is "more ecologically suited" to touch than to vision. In many studies concerning shape (a macrogeometric property), transfer from haptics to vision has been found to be easier than transfer from vision to haptics in both children and adults (Connolly & Jones, 1970; Jones & Connolly, 1970; Juurmaa & Lehtinen-Railo, 1994; Newnham & McKenzie, 1993; cf. Hatwell, 1994). In contrast, when the transfer concerns texture (a microgeometric property), for which touch is as efficient as vision, if not better, this asymmetry does not appear.

Sann and Streri (2007) studied the ability to detect the texture of an object using bidirectional intermodal transfer tasks, and compared this ability to newborns' ability to detect the shape of an object intermodally. These authors attempted to understand how visual and tactile modalities collect and process information and therefore to shed light on the perceptual mechanisms of newborns. If the perceptual mechanisms involved in collecting information about object properties are equivalent in both modalities at birth, a reverse

cross-modal transfer would be observed. Conversely, if the perceptual mechanisms differ in the two modalities, an irreversible transfer should be found. The material was a smooth cylinder versus a granular cylinder (a cylinder with beads on its surface). The results revealed the presence of cross-modal texture recognition between modalities in both directions.

In conclusion, despite the various discrepancies between haptic and visual modalities (i.e., asynchrony in the maturation and development of these senses, distal vs. proximal inputs, and parallel (for vision) vs. sequential (for haptic) processing of information), both systems can detect both regularities and irregularities in an event or object. Numerous studies have shown that this ability is functional from birth. In addition, despite the lack of spatial and temporal synchrony due to the characteristics of the transfer task paradigm, these studies show that 2- to 3-day-old newborns are able to establish links between what they see and what they feel with their hands.

Intersensory perception between audition and vision: the case of auditory-visual speech

In our natural environment, auditory-visual (AV) events are very frequent and diversified: some examples are the sound of a car starting, the movements of an orchestra conductor in tune with the music, the sound of the impact of a falling object to the ground, or the movements of the lips when a person is speaking. In all these cases, the integration of AV information is mainly based on its temporal synchrony (Bahrick & Lickliter, 2012).

Several studies have revealed that the sense of hearing is excellent at birth, mainly because it is already functional in the womb. In fact, starting at 35 weeks of gestation, various reception mechanisms are mature enough to allow fetuses to extract auditory information from outside the womb. In utero, it is mainly speech sounds that are heard well (see Streri et al., 2013, for a review of fetal auditory maturation and competencies). However, although near-term fetuses are already involved in a "pre-social" environment (Marx & Nagy, 2015; Nagy et al., 2021), it is only at birth that babies will share our social environment with the visual perception of faces.

Several investigations have provided evidence that newborns already possess a series of abilities that allow them to detect and respond to the social environment. The preference for the human face over any other stimulus has been clearly demonstrated (e.g., Fantz, 1963; Goren et al., 1975). Other studies have revealed that, from the earliest moments of postnatal life, human infants show sensitivity to the perceptual attributes of speech, both its acoustic and visual properties, preferring to listen to speech over complex nonspeech stimuli (Vouloumanos & Werker, 2007). Newborns

are also sensitive to the gestures of adults. In fact, neonatal facial imitation of tongue protrusion or mouth opening/closing (Meltzoff & Moore, 1977), and neonatal facial imitation of emotions (Field et al., 1982) are all probably done in response to social gestures made by an adult model. Furthermore, these responses could be considered as non-verbal exchanges. The question is how newborns link auditory speech perception with visually perceived faces and whether they consider speech-producing faces as a unit.

Kuhl and Meltzoff (1982) were the first to propose an experimental paradigm designed to reveal how 4- to 5-month-old infants are capable of "lip-reading." The infants viewed a movie with two side-by-side faces articulating two different vowels emitted from a central speaker. The auditory and visual stimuli were aligned in such a way that the temporal synchronization was perfect. The babies looked longer at the face that matched the sound. Babies recognize the sound as /i/ produced by the face with retracted lips, while /a/ is recognized by the face with open lips. However, the question of whether this link was acquired or was already present at birth remained open.

Some researchers supported the idea of an innate origin of the relationships between the motor patterns of the mouth and its auditory outputs (Aldridge et al., 1999; Walton & Bower, 1993). Using an operant sucking procedure, Aldridge and colleagues presented infants several hours old with matched and mismatched audio-visual presentations of one face articulating vowels (native: /a/, /u/, /i/, and non-native: /y/). Results revealed that infants significantly preferred matched to mismatched presentations. Assuming that these findings were not the result of learning, particularly with regard to non-native vowels, the authors concluded that knowledge about audiovisual speech correspondences is based on an innate process.

Other studies have shown that newborns are capable of relating auditory information to motor actions. Using the neonatal facial imitation paradigm, Coulon et al. (2013) asked do newborns associate the vowels /a/ with an "opening mouth" and /i/ with "spreading lips"? The authors compared the imitative responses of infants who were presented with only visual information, congruent audiovisual information, or incongruent audiovisual information. The results revealed that the newborns imitated the model's mouth movements significantly faster in the condition with audiovisual congruence. Furthermore, when presented with an incongruous audiovisual model, the infants did not produce any imitative behavior. Another study has shown that newborns who hear the sound /a/ make more instances of mouth-opening movements, and when they hear the sound /m/ they make more instances of mouth-clenching movements (Chen et al., 2004). All of these findings highlight the influence of speech perception on imitative responses in newborns and thus suggest that the neural architecture that enables the production-perception link already exists at birth.

Guellaï et al. (2016) conducted a study inspired by a previous one with 8-month-old infants (Kitamura et al., 2014) to assess the ability of newborns to detect congruence between auditory speech and faces. Stimuli consisted of sentences, more complex than vowels, and the faces were represented only by displays of dots and lines. Using an intermodal matching procedure, French newborns were confronted with two silent point-line displays representing the same face uttering different sentences in English while listening to a vocal-only expression that matched one of the two stimuli. Nearly all of the newborns looked longer at the matching point-line face than at the mismatching one. This result suggests the existence of intrinsic links between the visual movements of the face (head and eyebrow, articulation during production) and the sounds or rhythm of speech. This finding reveals the abstraction skills that newborns need to make this bond, and provides evidence that newborns can map auditory speech (even from an unknown language) to abstract representations of speech motion.[2] All of these studies performed at birth confirm the early involvement of the motor domain in the match between visual articulation and auditory sounds before the onset of the first babbling (see Cox et al., 2022, for a recent meta-analysis).

Newborns, a few minutes after birth, already encounter many different faces, either talking or silently moving. How do they process and recognize them, and which are the cues that are important for early face recognition? A recognition task, irrespective of the stimuli (faces or events), involves different cognitive processes such as encoding, storing and retrieval of information, and therefore appears to be more difficult than matching A-V (audiovisual) stimuli. Numerous studies on face perception and recognition have mainly used two-dimensional (2D) supports such as faces presented as photographs (Gava et al., 2008; Pascalis & de Schonen, 1994; Turati et al., 2006, 2008). In contrast, recognition of talking faces has rarely been tested in newborns. In an early study (Sai, 2005), one group of mothers was encouraged to talk to their babies immediately after birth, while another group was asked not to interact with them verbally. In the test session, an average of 7 h later, both the mother's face and a stranger's face were presented side by side, and the newborns looked longer, and oriented more towards, their mother's face, only if her mother had previously spoken with them (Sai, 2005). Because fetuses hear and prefer their mother's voice at birth (DeCasper & Spence, 1986), newborns who received verbal interaction were reinforced soon after birth and helped to encode and memorize their mother's face.

This conjecture has been tested in more recent studies using familiarization and testing phases, and using videos of dynamic unknown faces (Coulon et al., 2011). The newborns were presented with two different conditions: During the familiarization phase, the newborns saw a video of a woman's face speaking to them or moving her lips but not producing speech sounds. In the testing phase, the newborns viewed photographs of both the familiar face and the new face. Most of the newborns exhibited a visual preference for the familiar face over the new one only when the face had been speaking to them during the familiarization phase. This result is important because it reveals that the presence of speech increases the attention of the newborn, as has been shown previously (Coulon et al., 2013), and that spatio-temporal synchrony plays a fundamental role in face learning and recognition, in line with the intersensory redundancy hypothesis (Bahrick & Lickliter, 2000). Furthermore, recent work following studies on neonatal imitation have analyzed the mouth movements of newborns as indicators of social interest, revealing that a talking face elicits more motor feedback from the newborns than a silent one (Guellaï & Streri, 2022).

Direct gaze has been shown to be an effective cue for directing attention in adults and infants, and several studies have highlighted the importance of gaze as a social cue for understanding the human mind and behavior (Baron-Cohen, 1995; Batki et al., 2000; Senju & Csibra, 2008). In this context, several studies have tested the importance of gaze in infants with schematic faces or photographs, and have provided evidence that newborns prefer direct gaze to averted gaze (Farroni et al., 2002, 2004). However, the question of which of the two cues in the human face is more important in recognizing a person, spatio-temporally synchronized speech or gaze, remains open.

Guellaï and colleagues (Guellaï et al., 2020; Guellaï & Streri, 2011) tested the role of eye gaze orientation while a person is talking to newborns. In the familiarization phase, newborns saw a woman who talked to them with a direct gaze, with an averted gaze (Guellaï & Streri, 2011) or a faraway gaze (Guellaï et al., 2020). In the test phase, the photos of a familiar face and a new face were presented to newborns. The results revealed that newborns recognized and preferred to look at the previously seen face in photograph only when this person talked to them with direct gaze, but not when the person talked to them with an averted gaze or with a faraway gaze. This finding was replicated with a woman who talked to them in a non-native language, i.e., in English (Guellaï et al., 2015), suggesting that irrespective of the language (native or non-native) used by the talking face during the familiarization phase, newborns can encode and later show a visual preference for someone who talked to them with direct gaze. Therefore, the spatio-temporal synchrony of speech alone is not a sufficient cue for newborns to recognize a face, while direct eye gaze enhances attention and allows a better encoding and memorization of the event.

---

[2] French adults have been exposed to the same experiment and have presented some difficulties to match audio-visual stimuli (unpublished results).

The understanding of magnitude at birth

The ability to estimate numerical quantities, without using language or counting, is universal in humans (Dehaene, 1992), and is supported by a cognitive system that is evolutionary ancient (Piantadosi & Cantlon, 2017), preverbal (Xu & Spelke, 2000), and functional at birth (Izard et al., 2009). This ability allows us to approximately estimate the numerosity of a set of items, as well as to compare, and operate on, such numerical representations (Zorzi et al., 2005). The precision with which we are able to compare two sets of items is governed by the ratio between them, such that the larger the ratio difference the more accurate the comparison.

The ability to create an abstract representation of numbers by newborns has been revealed through the use of intermodal stimulation, whereby newborns are simultaneously presented with numerical information through visual and auditory modalities. These studies were able to reveal the ability of newborns to match number quantities across different sensory modalities, who look longer at a visual number display that matches an auditory number stream (Coubart et al., 2014; Izard et al., 2009). Importantly, newborns display this ability in the absence of perfect audiovisual synchrony between the two modality events. Furthermore, this bimodal stimulation has recently been adapted to investigate newborns' understanding of an even more abstract and dimensionless concept of magnitude/quantity, testing their ability to link the dimensions of number, space, and time (de Hevia et al., 2014). In particular, the ability of newborns to react to congruent changes in magnitude was assessed across two quantitative dimensions presented through two different modalities: when auditory numerosity and/or duration changes, they must form an expectation of change in the length of the line in the same (congruent) direction (i.e., both increasing or both decreasing).

In these studies, infants are first familiarized with an auditory sequence of syllables of varying number, either 6 or 18, and simultaneously shown a visual object of varying size, either small or large. After 60 s of exposure to this bimodal (i.e., auditory and visual) information, they experience a change in auditory number, either from 6 to 18 or from 18 to 6 in two trials, one where the visual object is new with respect to the one observed during the familiarization, while on another test trial the visual object is the same as in the familiarization. The results show that newborns systematically look longer at the object that corresponds to the new numerosity heard (that is, a small object if the new numerosity is 6 and a large object if the new numerosity is 18). Critically, this performance cannot be explained in terms of preference for visual novelty on the test, but rather by magnitude matching across dimensions: when infants are not given the opportunity to experience consistent changes

in magnitude from familiarization to test (e.g., when they are familiarized with a small auditory number and a large spatial extension, then tested with a large auditory number paired once with a small spatial extension, novel, and with the familiar spatial extension), they show no preference in the test. Thus, newborns appear to create links between the dimensions of number, space, and time, possibly using a common dimensionless representation, before extensive postnatal experience with these dimensions. The same performance pattern was observed in different experimental conditions that presented the dimensions of number and time confounded, as well as when each of these dimensions was presented in isolation.

Despite evidence showing that human newborns expect magnitude changes to occur in the same direction across different dimensions, these expectations appear to apply to number, space, and time, and do not seem to generalize to any quantitative dimension. In a follow-up study, the ability of newborns to create the same type of link with the brightness dimension was tested (Bonn et al., 2019). In this case, the auditory numerosities, either 6 or 18, were presented simultaneously with a visual object of varying brightness level, dark or light, against a black background. Under these conditions, the newborns seemed to accept any variation in quantity and were flexible enough to create a link between congruent and incongruent variations in magnitude through brightness and number. A control study, which used a change in visual shape and thus no quantitative dimension involved, confirmed that newborns were sensitive to quantitative variation in the dimensions of number and brightness. Importantly, their performance differed from their mappings in number, space, and time in that, for number-brightness pairings, newborns do not seem to exhibit a preferred directionality for matching. These findings suggest that a privileged mapping among the dimensions of number, space and time might exist at birth.

However, newborns do not treat these three dimensions in exactly the same way, but their performance suggests the existence of a degree of specificity in the representation of number, space, and time at birth (de Hevia et al., 2017). Using the same bimodal and cross-dimensional methodology, infants were presented with congruent changes in magnitude across dimensions (auditory changes in magnitude were always accompanied by congruent, expected changes in visual size); the critical test was whether the newborn preferred to see a decrease in auditory magnitude along with a congruent decrease in visual size on the left and an increase in auditory magnitude along with a congruent increase in visual size on the right, relative to the reverse association. Half of the participants were familiarized with a small number and the other half with a large number, paired with a centered geometric figure. Newborns were then tested with

the new numerosity paired with the new (corresponding) geometric figure in two trials, one on the left and one on the right side of the screen. Thus, as in the previous studies, infants familiarized with the small number (i.e., six syllables) were tested with the large number (i.e., 18 syllables), while infants familiarized with the large number were tested with the small number. Newborns were thus stimulated with both quantities (small and large) across the experimental session from familiarization to test, which we assumed allowed them to infer relative quantities. The results showed that newborns do, in fact, have a left-few/right-many bias (see also Di Giorgio et al., 2019 using unimodal, visual stimulation). More critically, this association was absent when, rather than number, infants were stimulated only with spatial extension or only with temporal duration (de Hevia et al., 2017). These findings suggest that the lateralized mapping of numbers at birth is not a general bias that applies to other magnitude dimensions, such as duration and spatial length, even though newborns can discriminate variations within those dimensions. Therefore, it is difficult to imagine that humans at birth possess an indistinguishable representation of magnitude that includes representations of number, space, and time; instead, their performance suggests that they can create a correspondence of magnitude changes across those dimensions at a certain level of abstraction using, however, differentiated representations for each of the dimensions.

Why do humans at birth associate relatively small numbers to the left and relatively large numbers to the right side of space? An evolutionary theory of this spatial representation of number has been recently proposed (de Hevia, 2016, 2021; de Hevia et al., 2012) according to which this bias might originate from a biological tendency to initially fixate on the left side of space. Some authors have previously proposed the existence of an innate visuo-spatial processing advantage for the left side of space (Bowers & Heilman, 1980; Kinsbourne, 1970), due to the right hemisphere of the brain preferentially processing the left visuo-spatial hemifield. While there is some evidence that this bias exists early in life (Deruelle & de Schonen, 1998; Nava et al., 2022), we tested the possibility that humans might have evolved this asymmetry in visuo-spatial processing by investigating whether a left visual advantage might be functional at birth (McCrink et al., 2020). We know that newborns attend significantly more to a visual stimulus showing the same numerical quantity they are hearing than to a visual stimulus showing a different, discriminable number (Izard et al., 2009). If newborns preferentially attend towards the left side of space, then they would show higher looking times towards a visual stimulus whose left side contains the same numerical magnitude as an auditory numerical sequence, as opposed to when the same stimulus is inversely oriented, therefore presenting the corresponding visual number on its right side. Infants were presented with visual chimeric

stimuli, which consisted of visual arrays of 24 items where individual elements were spatially arranged in such a way that the left and right sides of the array contained six and 18 elements. Newborns showed a general preference for fixating the visual stimulus whose left side numerically matched the auditory numerical sequence. This finding constitutes preliminary evidence for human newborns spontaneously privileging the left side of visual stimulation when presented with numerical information.

Recent work has provided evidence that newborns' abstraction abilities regarding the domain of quantity do not only apply to large numerosities, but they are functional when computing small sets of objects as well, for instance when discriminating two versus three objects (Martin et al., 2022). In particular, newborns are able to match AB versus ABB patterns across visual and auditory modalities, using either abstract representations of repetition (Bulf et al., 2011; Gervain et al., 2008), of numerosity (Coubart et al., 2014), or even possibly using both to detect the higher-level correspondence between the sets across modalities.

Newborns at birth therefore possess a rich concept of magnitude, which privileges its numerical, spatial, and temporal forms, and allows them to map corresponding magnitude changes across those dimensions. Humans also appear to possess at birth an inborn bias to associate small numbers to the left and large numbers to the right side of space; this signature, however, is specific to number, therefore suggesting the existence of separate representations for each of these dimensions (number, spatial extent, and temporal duration) at birth. Newborns possess an ability of abstraction in the magnitude domain that applies to both large and small numerical quantities. The human mind is therefore equipped with a sufficiently abstract notion of magnitude that allows cross-dimensional talk for the dimensions of space, number, and time, paving the way for future, mature correspondences between these core dimensions along which most of human experience is organized.

Conclusions

One of the central questions in perceptual and cognitive development is to understand whether and how humans (from birth) perceive and combine the multiple, related environmental signals coming from different sensory modalities. Here we reviewed a series of studies suggesting that, despite the asynchronies in the maturation of the senses and the later functioning of vision, humans possess basic integration abilities that are functional from birth and that allow them to relate sensory signals across modalities. Three mechanisms, at least, could be at play when newborns create a meaningful representation of the multimodal world: the detection of invariants present in sensory stimulation, which newborns

can detect irrespective of the sensory stimulation through which information is presented; the existence of amodal information that is shared across senses, and spatio-temporal synchrony that governs perception of specific information. Three different fields of research on newborn cognition were discussed. Studies on cross-modal transfer between vision and touch show that a link between these senses is functional at birth, although it is characterized by an asymmetrical exchange of information, as newborns can tactually recognize the shape of a previously seen object but they cannot recognize a visual shape that has been previously handled in their hand. In this case, the sensorial specificities of each modality are important. In the perception of auditory-visual speech, spatial synchrony is not always essential, since newborns are able to link the movement of lips with the vowels or the sentences heard. However, in the case of recognizing speaking faces, spatio-temporal synchrony is critical but sometimes not sufficient: newborns recognize a face speaking to them but not a silent one, and this ability is modulated by the presence of direct gaze. On the amodal domain of magnitude, newborns are able to link increases and decreases of magnitude across the dimensions of space, time, and number, even in the absence of a perfect spatio-temporal synchrony of the events. Moreover, newborns seem to truly possess and link specific, separate representations of number, space, and time, as they spatially structure these domains differently. Newborns' ability for abstraction in the quantity domain extends to small sets, detecting correspondences for two versus three across vision and audition.

The research reviewed here points out, on the one hand, the possibility of presenting moving stimuli that are essential to attract newborns attention while taking into account the visual characteristics of the newborn; on the other hand, it reveals that presenting newborns with conditions that are closer to those that he/she encounters in the environment facilitates their understanding of complex, multimodal situations (as we have seen for speaking faces). This research invites the use of varied and systematic bi-modal, or even tri-modal, stimulation. Studies reviewed here reveal that newborns, despite their immature state, possess the necessary skills that allow them to perceive our complex and multimodal world. Human beings at birth come equipped with a number of mechanisms to identify related sensory signals across modalities in order to obtain coherence. Although researchers have long simplified stimulation in experimental settings to study the perceptual and representational abilities of newborns, the notion that has emerged in recent years is that a complex, multimodal setting can facilitate newborns' performance and highlight their drive to understand the environment.

The immaturity of the sensory systems of the newborn does not allow a precise apprehension of the properties of objects and events. Nonetheless, research on the characteristics of newborns' perception and memory in the domains of language, face recognition, and object perception, among others, has provided the scientific community with a valuable understanding of the building blocks of human cognition. Yet, in the search for a stable and coherent environment, the human newborn is driven to actively integrate different sources of stimulation; therefore, it is in a multimodal context that the tools that enable the unity of perception are most easily revealed.

References

Aldridge, M. A., Braga, E. S., Walton, G. E., & Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Developmental Science, 2*(1), 42–46.

Ando, Y., & Hattori, H. (1977). Effects of noise on sleep of babies. *The Journal of the Acoustical Society of America, 62*(1), 199–204.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior & Development, 10*(4), 387–416.

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development, 59*(1), 197–209.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*(2), 190.

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In *Multisensory development* (pp. 183–206). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

Baron-Cohen, S. (1995). The eye direction detector (EDD) and the shared attention mechanism (SAM): Two cases for evolutionary psychology. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 41–59). Lawrence Erlbaum Associates Inc.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior & Development, 23*(2), 223–229.

Begum Ali, J., Spence, C., & Bremner, A. J. (2015). Human infants' ability to perceive touch in external space develops postnatally. *Current Biology, 25*(20), R978–R979.

Benavides-Varela, S., Gómez, D. M., Macagno, F., Bion, R. A. H., Peretz, I., & Mehler, J. (2011a). Memory in the neonate brain. *PLoS ONE, 6*(11), e27497.

Benavides-Varela, S., Gómez, D. M., & Mehler, J. (2011b). Studying neonates' language and memory capacities with functional near-infrared spectroscopy. *Frontiers in Psychology, 2*, 64.

Bonn, C. D., Netskou, M.-E., Streri, A., & de Hevia, M. D. (2019). The association of brightness with number/duration in human newborns. *PLOS ONE, 14*(10), e0223192.

Bowers, D., & Heilman, K. M. (1980). Pseudoneglect: Effects of hemispace on a tactile line bisection task. *Neuropsychologia, 18*(4–5), 491–498.

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research, 51*(13), 1588–1609.

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2012). *Multisensory development*. Oxford University Press.

Bruno, M., & Mandelbaum, E. (2010). Locke's answer to molyneux's thought experiment. *History of Philosophy Quarterly, 27*(2), 165–180 JSTOR.

Bryant, P. (1974). *Perception and Understanding in Young Children: An Experimental Approach* (1st ed.). Routledge. https://doi.org/10.4324/9781315534251

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition, 121*(1), 127–132.

Bushnell, E. W., & Boudreau, J. P. (1998). Exploring and exploiting objects with the hands during infancy. In K. J. Connolly (Ed.), *The psychobiology of the hand* (pp. 144–161). Mac Keith Press.

Chen, X., Striano, T., & Rakoczy, H. (2004). Auditory-oral matching behavior in newborns. *Developmental Science, 7*, 42–47.

Connolly, K., & Jones, B. (1970). A developmental study of afferent-reafferent integration. *British Journal of Psychology, 61*(2), 259–266.

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science, 17*(1), 11–22.

Coulon, M., Guellai, B., & Streri, A. (2011). Recognition of unfamiliar talking faces at birth. *International Journal of Behavioral Development, 35*(3), 282–287.

Coulon, M., Hemimou, C., & Streri, A. (2013). Effects of seeing and hearing vowels on neonatal facial imitation. *Infancy, 18*(5), 782–796.

Cox, C. M. M., Keren-Portnoy, T., Roepstorff, A., & Fusaroli, R. (2022). A Bayesian meta-analysis of infants' ability to perceive audio–visual congruence for speech. *Infancy, 27*(1), 67–96.

de Hevia, M. D. (2016). Link between numbers and spatial extent from birth to adulthood. In A. Henik (Ed.), *Continuous issues in numerical cognition: How many or how much* (pp. 37–58). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-801637-4.00002-0

de Hevia, M. D. (2021). How the human mind grounds numerical quantities on space. *Child Development Perspectives, 15*(1), 44–50.

de Hevia, M. D., Girelli, L., & Macchi Cassia, V. (2012). Minds without language represent number through space: Origins of the mental number line. *Frontiers in Psychology, 3*, 466.

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences, 111*(13), 4809–4813.

de Hevia, M. D., Veggiotti, L., Streri, A., & Bonn, C. D. (2017). At birth, humans associate "Few" with Left and "Many" with right. *Current Biology, 27*(24), 3879–3884.e2.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development, 9*(2), 133–150.

DeCasper, A. J., & Spence, M. J. (1991). Auditorily mediated behavior during the perinatal period: A cognitive view. In M. J. S. Weiss & P. R. Zelazo (Eds.), *Newborn attention: Biological constraints and the influence of experience* (pp. 142–176). Ablex Publishing.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*(1–2), 1–42.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional Neuroimaging of Speech Perception in Infants. *Science, 298*(5600), 2013–2015.

Deruelle, C., & de Schonen, S. (1998). Do the right and left hemispheres attend to the same visuospatial information within a face in infancy? *Developmental Neuropsychology, 14*(4), 535–554.

Di Giorgio, E., Lunghi, M., Rugani, R., Regolin, L., Dalla Barba, B., Vallortigara, G., & Simion, F. (2019). A mental number line in human newborns. *Developmental Science, 22*(6), e12801.

Diderot, D. (1749). *La lettre sur les aveugles à l'usage de ceux qui voient.* Garnier-Flammarion.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*(1), 567–631.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303–306.

Ettlinger, G. (1967). Analysis of cross-modal effects and their relationship to language. In F. L. Darley & C. H. Millikan (Eds.), *Brain mechanisms underlying speech and language* (pp. 53–60). New York: Grune & Stratton.

Fantz, R. L. (1963). Pattern vision in newborn infants. *Science, 140*(3564), 296–297.

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.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy, 5*(1), 39–60.

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expression by neonates. *Science, 218*(4568), 179–181.

Gallagher, S. (2005). *How the Body Shapes the Mind.* Oxford University Press. https://doi.org/10.1093/0199271941.001.0001

Gava, L., Valenza, E., Turati, C., & de Schonen, S. (2008). Effect of partial occlusion on newborns' face preference and recognition. *Developmental Science, 11*(4), 563–574.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105*(37), 14222–14227.

Gibson, J. J. (1950). *The perception of the visual world.* Boston: Houghton Mifflin.

Gibson, J. J. (1967). On the proper meaning of the term "stimulus". *Psychological Review, 74*(6), 533–534.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* Appleton-Century-Crofts.

Gibson, E. J., & Walker, A. S. (1984). Development of knowledge of visual-tactual affordances of substance. *Child Development, 55*(2), 453.

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.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology, 18*(9), 694–698.

Gori, M., Campus, C., Signorini, S., Rivara, E., & Bremner, A. J. (2021). Multisensory spatial perception in visually impaired infants. *Current Biology, 31*(22), 5093–5101.e5.

Gottlieb, G. (1971). Ontogenesis of sensory function in birds and mammals. In E. Tobach, L. R. Aronson, & E. Shaw (Eds.), *The biopsychology of development* (pp. 67–128). New York: Academic Press.

Granier-Deferre, C., Schaal, B., & DeCasper, A. J. (2004). Les prémices fœtales de la cognition. In R. Lecuyer (Ed.), *Le développement du Nourrisson* (pp. 59–94). Dunod.

Granier-Deferre, C., Bassereau, S., Ribeiro, A., Jacquet, A.-Y., & DeCasper, A. J. (2011). A melodic contour repeatedly experienced by human near-term fetuses elicits a profound cardiac reaction one month after birth. *PLoS ONE, 6*(2), e17304.

Guellaï, B., & Streri, A. (2011). Cues for early social skills: Direct gaze modulates newborns' recognition of talking faces. *PLoS ONE, 6*(4), e18610.

Guellaï, B., & Streri, A. (2022). Mouth movements as possible cues of social interest at birth: New evidences for early communicative behaviors. *Frontiers in Psychology, 13*, 831733.

Guellaï, B., Mersad, K., & Streri, A. (2015). Suprasegmental information affects processing of talking faces at birth. *Infant Behavior and Development, 38*, 11–19.

Guellaï, B., Streri, A., Chopin, A., Rider, D., & Kitamura, C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: Evidence from point-line displays of talking faces. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1275–1281.

Guellaï, B., Hausberger, M., Chopin, A., & Streri, A. (2020). Premises of social cognition: Newborns are sensitive to a direct versus a faraway gaze. *Scientific Reports, 10*(1), 9796.

Guest, S., & Spence, C. (2003). Tactile dominance in speeded discrimination of textures. *Experimental Brain Research, 150*(2), 201–207.

Hatwell, Y. (1994). Transferts intermodaux et intégration intermodale (crossmodal transfers and intersensory integration).

In M. Richelle, J. Requin, & M. Robert (Eds.), *Traité de psychologie expérimentale* (Vol. 1, pp. 543–584). PUF.

Hatwell, Y., Streri, A., & Gentaz, E. (2003). *Touching for knowing: Cognitive psychology of haptic manual perception* (Vol. 53). John Benjamins Publishing.

Hym, C., Dumuids, M., Anderson, D. I., Forma, V., Provasi, J., Brière-Dollat, C., Granjon, L., Gervain, J., Nazzi, T., & Barbu-Roth, M. (2022). Newborns modulate their crawling in response to their native language but not another language. *Developmental Science*. https://doi.org/10.1111/desc.13248

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences, 106*(25), 10382–10385.

Jones, B., & Connolly, K. (1970). Memory effects in cross-modal matching. *British Journal of Psychology, 61*(2), 267–270.

Jusczyk, P. W. (1997). *The discovery of spoken language.* (pp. xii, 314). The MIT Press.

Juurmaa, J., & Lehtinen-Railo, S. (1994). Visual experience and access to spatial knowledge. *Journal of Visual Impairment & Blindness, 88*(2), 157–170.

Kawashima, R., Watanabe, J., Kato, T., Nakamura, A., Hatano, K., Schormann, T., Sato, K., Fukuda, H., Ito, K., & Zilles, K. (2002). Direction of cross-modal information transfer affects human brain activation: A PET study. *European Journal of Neuroscience, 16*(1), 137–144.

Kinsbourne, M. (1970). The cerebral basis of lateral asymmetries in attention. *Acta Psychologica, 33*, 193–201.

Kitamura, C., Guellaï, B., & Kim, J. (2014). Motherese by eye and ear: Infants perceive visual prosody in point-line displays of talking heads. *PLoS ONE, 9*(10), e111467.

Klaey-Tassone, M., Durand, K., Damon, F., Heyers, K., Mezrai, N., Patris, B., Sagot, P., Soussignan, R., Schaal, B., & the MILKODOR Consortium. (2021). Human neonates prefer colostrum to mature milk: Evidence for an olfactory bias toward the "initial milk"? *American Journal of Human Biology, 33*(5), e23521.

Klatzky, R. L., Lederman, S. J., & Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General, 116*(4), 356–369.

Koffka, K. (1935). *Principles of Gestalt Psychology*. Lund Humphries.

Köhler, W. (1964). Unsolved problems in the field of figural aftereffects. *The Psychological Record, 15*(1), 63–83.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–841.

Kuhl, P. K., & Meltzoff, A. N. (1982). The Bimodal Perception of Speech In Infancy. *Science, 218*(4577), 1138–1141.

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology, 19*(3), 342–368.

Lewkowicz, D. J., & Lickliter, R. (Eds.). (1994). *The development of intersensory perception: Comparative perspectives*. Lawrence Erlbaum Associates, Inc.

Lewontin, R. C. (2000). *The Triple Helix: Gene, Organism, and Environment*. Harvard University Press.

Macchi Cassia, V., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15*(6), 379–383.

Martin, L., Marie, J., Brun, M., de Hevia, M. D., Streri, A., & Izard, V. (2022). Abstract representations of small sets in newborns. *Cognition, 226*, 105184.

Marx, V., & Nagy, E. (2015). Fetal behavioural responses to maternal voice and touch. *PloS One, 10*(6), e0129118.

McCrink, K., Veggiotti, L., & de Hevia, M. D. (2020). A left visual advantage for quantity processing in neonates. *Annals of the New York Academy of Sciences, 1477*(1), 71–78.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143–178.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101–116). Lawrence Erlbaum Associates Inc.

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature, 282*(5737), 403–404.

Meltzoff, A., & Moore, M. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*(4312), 74–78.

Molina, M., & Jouen, F. (2001). Modulation of manual activity by vision in human newborns. *Developmental Psychobiology, 38*(2), 123–132.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior & Development, 16*(4), 495–500.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced *in utero* affects vowel perception after birth: A two-country study. *Acta Paediatrica, 102*(2), 156–160.

Morange-Majoux, F., & Devouche, E. (2022). Neonatal manual specialization in language and music conditions: Consistency with the hemispheric specialization adult model. *Early Human Development, 168*, 105575.

Morrongiello, B. A., Fenwick, K. D., & Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior & Development, 21*(4), 543–553.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review, 98*(2), 164–181.

Nagy, E., Thompson, P., Mayor, L., & Doughty, H. (2021). Do foetuses communicate? Foetal responses to interactive versus non-interactive maternal voice and touch: An exploratory analysis. *Infant Behavior & Development, 63*, 101562.

Nallet, C., & Gervain, J. (2021). Neurodevelopmental preparedness for language in the neonatal brain. *Annual Review of Developmental Psychology, 3*(1), 41–58.

Nava, E., de Hevia, M. D., Bulf, H., & Cassia, V. M. (2022). Signatures of functional visuospatial asymmetries in early infancy. *Journal of Experimental Child Psychology, 215*, 105326.

Newnham, C., & McKenzie, B. E. (1993). Cross-modal transfer of sequential visual and haptic shape information by clumsy children. *Perception, 22*(9), 1061–1073.

Pascalis, O., & de Schonen, S. (1994). Recognition memory in 3– to 4-day-old human neonates. *NeuroReport, 5*(14), 1721–1724.

Piaget, J. (1936). *Origins of intelligence in the child*. Routledge & Kegan Paul.

Piaget, J. (1937). *La naissance de l'intelligence chez l'enfant*. Delachaux & Niestle.

Piantadosi, S. T., & Cantlon, J. F. (2017). True numerical cognition in the wild. *Psychological Science, 28*(4), 462–469.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 28*(3), 191–212.

Révész, G. (1934). System der optischen und haptischen Raumtäuschungen [A system of optic and haptic space illusions]. *Zeitschrift Für Psychologie Und Physiologie Der Sinnesorgane. Abt. 1. Zeitschrift Für Psychologie, 131*, 296–375.

Rose, S. A., & Orlian, E. K. (1991). Asymmetries in infant cross-modal transfer. *Child Development, 62*(4), 706–718.

Rose, S. A., Gottfried, A. W., & Bridger, W. H. (1981). Cross-modal transfer and information processing by the sense of touch in infancy. *Developmental Psychology, 17*(1), 90–98.

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development, 14*(1), 29–50.

Sann, C., & Streri, A. (2007). Perception of object shape and texture in human newborns: Evidence from cross-modal transfer tasks. *Developmental Science, 10*(3), 399–410.

Sann, C., & Streri, A. (2008). The limits of newborn's grasping to detect texture in a cross-modal transfer task. *Infant Behavior and Development, 31*(3), 523–531.

Schaal, B. (2005). From amnion to colostrum to milk: Odor bridging in early developmental transitions. In B. Hopkins & S. P. Johnson (Eds.), *Prenatal development of postnatal functions* (pp. 51–102). Praeger Publishers/Greenwood Publishing Group.

Schaal, B., Marlier, L., & Soussignan, R. (1998). Olfactory function in the human fetus: Evidence from selective neonatal responsiveness to the odor of amniotic fluid. *Behavioral Neuroscience, 112*(6), 1438–1449.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology, 18*(9), 668–671.

Simion, F., Macchi Cassia, V. M., Turati, C., & Valenza, E. (2003). Non-specific perceptual biases at the origins of face processing. In O. Pascalis & A. Slater (Eds.), *The development of face processing in infancy and early childhood: Current perspectives* (pp. 13–25). Nova Science Publishers.

Slater, A., Quinn, P. C., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory−visual pairings. *Developmental Science, 2*(3), 333–338.

Soussignan, R., Schaal, B., Marlier, L., & Jiang, T. (1997). Facial and autonomic responses to biological and artificial olfactory stimuli in human neonates. *Physiology & Behavior, 62*(4), 745–758.

Soussignan, R., Schaal, B., & Marlier, L. (1999). Olfactory alliesthesia in human neonates: Prandial state and stimulus familiarity modulate facial and autonomic responses to milk odors. *Developmental Psychobiology, 35*(1), 3–14.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT press.

Steiner, J. E. (1979). Human facial expressions in response to taste and smell stimulation. *Advances in Child Development and Behavior, 13*, 257–295.

Streri, A. (1987). Tactile discrimination of shape and intermodal transfer in 2-to 3-month-old infants. *British Journal of Developmental Psychology, 5*(3), 213–220.

Streri, A. (1993). *Seeing, reaching, touching: The relations between vision and touch in infancy* (pp. xvi, 224). The MIT Press.

Streri, A., & Gentaz, E. (2003). Cross-modal recognition of shape from hand to eyes in human newborns. *Somatosensory & Motor Research, 20*(1), 13–18.

Streri, A., & Gentaz, E. (2004). Cross-modal recognition of shape from hand to eyes and handedness in human newborns. *Neuropsychologia, 42*(10), 1365–1369.

Streri, A., & Pêcheux, M. (1986). Vision-to-touch and touch-to-vision transfer of form in 5-month-old infants. *British Journal of Developmental Psychology, 4*(2), 161–167.

Streri, A., Lhote, M., & Dutilleul, S. (2000). Haptic perception in newborns. *Developmental Science, 3*(3), 319–327.

Streri, A., de Hevia, M. D., Izard, V., & Coubart, A. (2013). What do we know about neonatal cognition? *Behavioral Sciences (Basel, Switzerland), 3*(1), 154–169.

Swain, I. U., Zelazo, P. R., & Clifton, R. K. (1993). Newborn infants' memory for speech sounds retained over 24 hours. *Developmental Psychology, 29*(2), 312–323.

Tristão, R. M., Lauand, L., Costa, K. S. F., Brant, L. A., Fernandes, G. M., Costa, K. N., Spilski, J., & Lachmann, T. (2021). Olfactory sensory and perceptual evaluation in newborn infants: A systematic review. *Developmental Psychobiology, 63*(7), e22201.

Turati, C., Simion, F., Milani, I., & Umiltà, C. (2002). Newborns' preference for faces: What is crucial? *Developmental Psychology, 38*(6), 875–882.

Turati, C., Macchi Cassia, V., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child Development, 77*(2), 297–311.

Turati, C., Bulf, H., & Simion, F. (2008). Newborns' face recognition over changes in viewpoint. *Cognition, 106*(3), 1300–1321.

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 892–903.

von Helmholtz, H. (1885). *On the sensations of tone as a physiological basis for the theory of music*. In A. Ellis (Ed.). New York: Dover Publications.

von Hofsten, C. (1982). Eye-hand coordination in the newborn. *Developmental Psychology, 18*(3), 450–461.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science, 10*(2), 159–164.

Walton, G. E., & Bower, T. G. R. (1993). Newborns form "Prototypes" in less than 1 minute. *Psychological Science, 4*(3), 203–205.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11.

Zorzi, M., Stoianov, I., & Umiltà, C. (2005). Computational modeling of numerical cognition. In J. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 67–84). New York: Psychology Press.