Newborns Are Sensitive To The Correspondence Between Auditory Pitch And Visuospatial Elevation

Peter Walker[1,2], J. Gavin Bremner[1], Marco Lunghi[3], Sarah Dolscheid[4], Beatrice Dalla Barba[5], & Francesca Simion[3]

1 Department of Psychology, Lancaster University, UK

2 Department of Psychology, Sunway University, Malaysia

3 Department of Developmental and Social Psychology, University of Padova, Italy

4 Department of Rehabilitation and Special Education, University of Cologne, Germany

5 Paediatric Department, University of Padova, Italy

Correspondence concerning this article should be addressed to:

Peter Walker, Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK

e-mail: p.walker@lancaster.ac.uk

tel: +44 (0) 1524 593163

fax: +44 (0) 1524 593744

Abstract

Amodal (*redundant*) and *arbitrary* cross-sensory feature associations involve the context-insensitive mapping of absolute feature values across sensory domains. Cross-sensory associations of a different kind, known as *correspondences*, involve the context-sensitive mapping of relative feature values. Are such correspondences in place at birth (like amodal associations), or are they learned from subsequently experiencing relevant feature co-occurrences in the world (like arbitrary associations)? To decide between these two possibilities, human newborns (median age = 44 hrs) watched animations in which two balls alternately rose and fell together in space. The pitch of an accompanying sound rose and fell either congruently with this visual change (pitch rising and falling as the balls moved up and down), or incongruently (pitch rising and falling as the balls moved down and up). Newborns' looking behaviour was sensitive to this congruence, providing the strongest indication to date that cross-sensory correspondences can be in place at birth.

Keywords: cross-sensory correspondences, amodal and arbitrary mappings, neonatal perception, pitch-elevation mapping

Newborns Are Sensitive To The Correspondence Between Auditory Pitch and Visuospatial Elevation

Most objects and events are encoded over multiple sensory feature channels, with the feature values identified in different channels varying in the extent of their co-occurrence, that is, in the extent to which they predict each other.[1]

At one extreme, the co-occurrence of cross-sensory feature values is invariant, as when vision and touch identify the same, metrically specified spatial location for an object. In effect, it is absolute, narrowly-tuned feature values that are here being mapped onto each other across sensory domains. Because of the redundancy in the information the two domains provide regarding, in this case, spatial location, together the domains specify an object's location in a modality-independent representation of space. Similar absolute mappings support other modality-independent representations, including: the temporal synchrony between a seen object collision and the impact sound it creates; the fineness of the surface texture of an object as seen and touched; elements of shape that are encoded with some precision both visually and haptically (e.g., the thickness and orientation of a rod); and, the temporal synchrony between seen movements of the mouth and the sounds of the vocalisations they create (e.g., variations in loudness). These cross-sensory mappings are thought of as being *amodal,* or *redundant* in nature. In light of the fact that they are available to provide necessary support for the learning of other

[1] A sensory feature channel refers to those elements of our sensory systems designed to encode a specific aspect of an item or event in the world, such as an object's spatial location, colour, surface brightness, direction of movement, and weight, or the loudness and acoustic frequency (pitch) of a simple sound. The particular level of surface brightness of an object, and the pitch of a simple sound, would be the feature values encoded in their respective channels.

types of cross-sensory association in newborns (i.e., *arbitrary* associations, see below), they appear to be in place at birth (Bahrick, Hernandez-Reif & Flom, 2005; Slater, Quinn, Brown & Hayes, 1999). Furthermore, because it is invariant absolute values of features that are being mapped onto each other, amodal cross-sensory mappings are context insensitive, that is, they are able to influence behaviour without needing other objects or events for comparison (e.g., convergence of the visual and haptic identification of the spatial location of an object can be done without reference to other objects).

At the other extreme, feature values identified across different sensory domains show no evidence of co-occurring (i.e., correlating), with the effect that they are unable to predict each other. The individual mapping of features is *arbitrary* and each has to be learned independently, as when we learn the mappings between: the shape and colour of a flower; the ringtone and colour of a friend's mobile phone; a person's facial features and the acoustic features identifying their spoken regional accent; and, for most words (in English at least), the sound of a word and the nature of the word's referent. Arbitrary mappings are learned by encountering them directly, or by hearing about them verbally (e.g., "All post boxes are red"). Finally, as with amodal mappings, arbitary mappings are able to influence behaviour without needing other objects, possessing different values for the same features, available for comparison (e.g., having an expectation that an object with the shape of a post box will be red does not require seeing it alongside other objects with a different shape and/or colour).

Cross-sensory mappings of a third type, known as *correspondences*, are intermediate in the strength of the perceived co-occurrences between feature values encoded in different sensory domains (i.e., in the degree to which they seem able to predict each other). A feature value in one domain can at best predict only the approximate feature value being encoded in a different domain, and will fail to do even this on many occasions. Examples of cross-sensory correspondences include: bigger objects and animals make sounds (impact sounds and vocalisations) that are lower in pitch than those made by smaller objects; darker objects look heavier than brighter objects; and, higher pitch sounds feel brighter than lower pitch sounds. These are regarded as 'tendencies' towards co-occurrence because they are contradicted on many occasions (e.g., many smaller people have lower pitch voices than bigger people, and many darker objects are lighter in weight than brighter objects). When Bahrick and her colleagues talk about this type of association they allow for such contradictions by referring to 'typical' cross-sensory associations (e.g., "Children typically have smaller, rounder heads and voices of a higher pitch than adults", see Bahrick, Netto & Hernandez-Reif, 1998, *p*. 1263).

Cross-sensory correspondences are especially evident in relation to auditory frequency (pitch). People judge higher-pitch sounds to be brighter, lighter in weight, pointier, smaller, and thinner than lower-pitch sounds (Boltz, 2011; Collier & Hubbard, 2004; Eitan & Timmers, 2010; Kussner & Leech-Wilkinson, 2013; Marks, 1978; Perrott, Musicant, & Schwethelm, 1980; Tarte, 1982; P. Walker & Smith, 1984). Similar cross-sensory mappings involving pitch appear in the sound-induced visual imagery experienced by visual-hearing synaesthetes (Chiou, Stelter & Rich, 2013; Ward, Huckstep & Tsakanikos, 2006), and in the congruity effects observed during the speeded classification of elementary stimulus features (e.g., Evans & Treisman, 2010). Of course, correspondences are not confined to auditory-visual mappings, but instead extend across all sensory domains, such as when seeing a relatively dark object induces expectations that it will be relatively heavy (P. Walker, 2012b; P. Walker, Francis, & L. Walker, 2010b; L. Walker, P. Walker & Francis, 2012), again despite there being many exceptions to this generalisation.

Parise (2015) points out that this third type of cross-sensory association involves the mapping of *relative* feature values across domains, rather than the mapping of absolute feature values, which explains the prominence of comparative adjectives in descriptions of the associations (*heavier, brighter, smaller, sharper, higher*). To illustrate, it is known that the same absolute pitch for a sound will map on to a higher or lower visuospatial location depending whether the alternative sound with which it appears is higher or lower in pitch than itself (Chiou & Rich, 2012). In other words, it is the status of a sound's features relative to the features of other sounds appearing in the same context that has functional significance (see also Gallace & Spence, 2006; Marks, 1987). It follows from this that, unlike amodal and arbitrary mappings, relative mappings are context sensitive, that is, they require multiple values for the salient features to be present in the same situation (Chiou & Rich, 2012; Gallace & Spence, 2006; L. Walker & P. Walker, 2015).

These three types of cross-sensory mapping can be thought of as lying on a continuum defined by the strength of the feature correlations they reveal. Notwithstanding this continuity, however, it is more appropriate to regard them as qualitatively different types of mapping, not least because of their distinctiveness in terms of: the relative or absolute nature of the feature values being mapped onto each other; the nature of the feature representations being mapped onto each other (i.e., whether they are sensory-perceptual or conceptual, see the Discussion below); how the mappings come to influence behaviour (e.g., whether their impact is context sensitive or context insensitive); and, whether or not they are learned. Regarding the latter difference, though it is clear that arbitrary mappings have to be learned, and amodal mappings are likely to be available at birth (e.g., because they are available to support the learning of arbitrary associations in newborns, see Slater, Quinn, Brown & Hayes, 1999), the situation regarding cross-sensory correspondences remains to be determined.

On the one hand, since cross-sensory correspondences can exist as natural co-occurrences among elementary stimulus features (in natural scene statistics), it could be through direct exposure to these that people learn the correspondences (see Peters, Balzer, & Shams, 2015, for an example of this). Prime candidates are the co-occurrences between size and pitch, visual thinness and pitch, and spatial elevation and pitch, the latter because sounds emanating from lower spatial locations can have weakened higher frequency components through their selective damping by the ground (Parise, Knorre, & Ernst, 2014).

On the other hand, there is evidence that cross-sensory correspondences are present in very young infants, pointing to the possibility that correspondences are available early in life and are perhaps present at birth. In particular, 3- to 4-month-olds are sensitive to the correspondence between auditory pitch and each of visuospatial elevation (Dolscheid, Hunnius, Casasanto & Majid, 2014; P. Walker et al., 2010a), visual pointiness (P. Walker et al., 2010a), visual thinness (Dolscheid et al., 2014) and visual size (Pena, Mehler, & Nespor, 2012), with the latter sensitivity allowing 3-month-olds to appreciate that lower pitch voices tend to belong to bigger animals (see Pietraszewski, Wertz, Bryant, & Wynn, 2017). Though this is compelling evidence that correspondences are present early in life, the first 3 months in a child's life is likely to be sufficient for correspondences to be learned through direct exposure to relevant natural co-occurrences (and it is known that neonates are able to learn statistical contingencies in their environment, see Bulf, Johnson & Valenza, 2011). It is therefore unclear whether cross-modal correspondences are learned from experiencing relevant feature co-occurrences in the world (as for arbitrary mappings) or whether they are in place at birth (as with amodal mappings).

To decide between these two possibilities, the focus of the present study was on a cohort whose exposure to relevant feature co-occurrences in the world is at a minimum: Newborns. For the first time, we examine if newborns are sensitive to a cross-sensory correspondence, specifically, the correspondence between auditory pitch and visuospatial elevation.

In two previous studies of this correspondence, 3- to 4-month-olds watched animations in which a ball moved up and down a screen. At the same time, a sound changed its pitch in time with the movement of the ball, sometimes in a manner that was congruent with the visual movement (i.e., rising and falling pitch as the ball moved up and down), and sometimes in a manner that was incongruent (i.e., falling and rising pitch as the ball moved up and down). The infants were observed to prefer looking at the ball in the congruent animation rather than at the ball in the incongruent animation, which was reflected both in longer individual fixations and longer total looking times across a full trial (with no difference in the number of fixations) (Dolscheid et al., 2014; P. Walker et al., 2010a). Though newborns might be expected to respond to congruence in the same way as older infants, this is not inevitable. It is already known, for example, that they show a relatively reduced tendency to shift their gaze away from any object they are currently fixating (i.e., they display 'sticky fixations'). Because of this tendency, the newborns were presented with two balls appearing side-by-side and moving up and down the screen together (see Figure 1). Though this modification might induce a different reaction to congruence, the prediction remained that the looking behaviour of the newborns would confirm their sensitivity to the manipulation of pitch-elevation congruence.

Method

Participants

It was planned to have data from 12 newborns to analyse, which is in the range of previous successful studies of the pitch-elevation correspondence in 3- to 4-month-olds (P. Walker et al., 2010a; where *N* = 16), and in the range of previous successful studies focusing on preferential looking in newborns (e.g., Bidet-Ildei et al., 2013; Di Giorgio, Leo, Pascalis, & Simion, 2012; Di Giorgio, Lunghi, Simion & Vallortigara, 2017; Di Giorgio et al., 2016; Leo & Simion, 2009; Macchi Cassia, Valenza, Simion, & Leo, 2008; Mascalzoni, Regolin, Vallortigara, & Simion, 2013; Simion, Regolin & Bulf, 2008; Simion, Valenza, Cassia, Turati, & Umilta, 2002; Turati, Simion, Milani & Umilta, 2002; where *N* ranges from 12 to 16, and where 13 of the 26 experiments reported have *N* = 12).

All 12 newborns completing the present study were healthy Caucasians born into middle-class families, having had normal, full-term deliveries (mean birth weight = 3.37 kg, range: 2.93 to 4.02 kg), with 5-min Apgar scores of 9 or 10. The 5 male and 7 female newborns had a mean age at testing of 49 hours (range: 6 to 122 hours). Data from a further 5 newborns could not be used. Three became fussy or sleepy during testing and stopped looking at the animations before the end of a trial (see below). Two showed a strong bias by fixating one of the two objects (see below) for more than 80% of the time.

The experimental procedures were licensed by the Ethics Committee of the Paediatric Clinic of the University of Padova, and all parents provided informed consent.

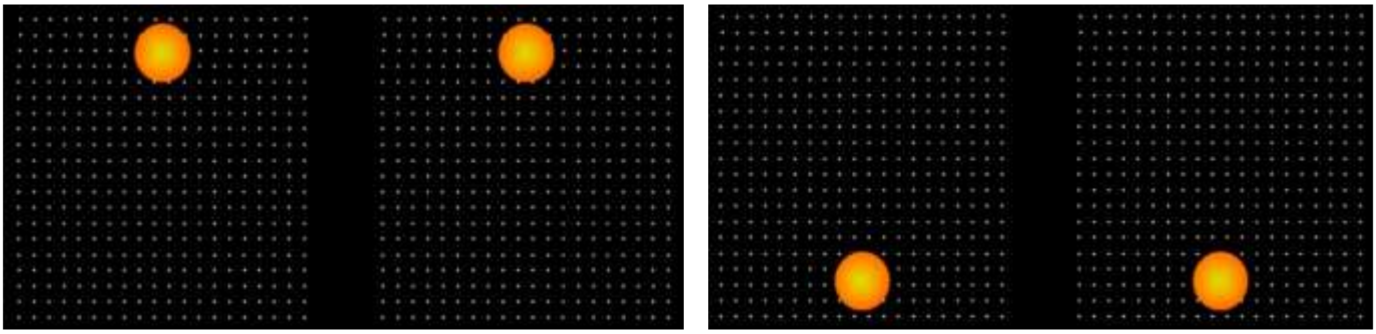

Figure 1. Illustrating the visual animations, with the moving balls shown at the extremes of their vertical trajectory.

Materials and procedure

The visual stimuli were QuickTime animations appearing on an Apple LED 2560 x 1600 pixel cinema display (Flat Panel 30”) with a refresh rate of 60 Hz. The screen was integrated into one side of a cubicle created with thick, white curtains, with ambient lighting set at 1.60 cd/m^2. Each newborn was held in a sitting position on an experimenter’s lap and, from a distance of 35 cm, viewed pairs of the same animation located side-by-side on the screen. This experimenter was naive to the hypothesis being tested and to the stimuli being presented to the baby, and was instructed to fix their gaze throughout the experiment on a camera located on the ceiling. A separate video camera placed above the monitor was utilized to monitor on-line if the babies were looking at the screen and to record their looking behaviour.

The newborns completed two test trials, each beginning with the presentation of a central red disc (to attract their attention) that grew and shrank back continuously (between 1.8 and 2.5 cm diameter) against the black background. Once the newborn’s gaze was aligned with the central disk, the 2 min animation for that trial was started. If the newborn shifted their gaze away from the animation for more than 10 s the trial was terminated and they were removed from the study (see above).

The animation comprised two 2.5 cm diameter (4.1 deg) orange balls (average luminance = 35 cd/m^2) moving together up and down a 13 cm (21.31 deg) vertical trajectory at a constant speed of 12.8 cm/sec, but pausing briefly (approximately 50 ms) at each endpoint of the trajectory (i.e., each phase of the animation lasted 2.5 sec) (Figure 1). The balls appeared at the left and right hand side of the screen, with a centre-to-centre separation of 15 cm, in front of a 20 x 20 grid of white dots, 13 x 13 cm, on an otherwise dark field. (see Figure 1). The upward and downward movement of the balls was accompanied by the sound of a sliding tone, whose fundamental frequency changed, at a constant rate, between 300 and 1700 Hz, over a period of time coinciding with a single phase of the balls' trajectory (i.e., the full upward or downward movement of the balls, which takes 2.5 s). The loudness of the sound increased and then decreased between 68 and 74 dBA (against an ambient laboratory noise level of 46 dBA) within each phase of the animation, peaking when the fundamental frequency of the sound was midrange between the two extreme values, but falling silent briefly to coincide with the ball ceasing to move at the extremes of its elevation.[2]

[2] Without knowing the function relating perceived loudness to auditory frequency (i.e., the equal loudness curves) for newborns listening to these sounds, it is impossible to arrange to hold loudness constant as pitch either rises or falls. With this in mind, loudness was removed from the equation in our study (and previous of our studies) by arranging for the sound to increase and then decrease very noticeably in loudness during each phase of the animation (i.e., during the upward or downward movement of the object and during the rise or fall in pitch). In this way, loudness was very effectively precluded from being confounded with pitch or visuospatial height, and thereby from inducing an element of congruence to the animations. In this way, only the correspondence between pitch and height was available to induce a congruence effect. In addition, with regard to the possibility that changes in loudness were congruent/incongruent with pitch and/or height during part of each phase of the

One of the two trials involved a congruent combination of changing pitch and changing visual-spatial elevation, with the pitch of the sound rising and falling as the two balls rose and fell. The other trial involved an incongruent combination, with the pitch of the sound falling and rising as the balls rose and fell. In addition, one of the trials started with the balls at the lowest location in their trajectory, and one started with the balls at the highest location in their trajectory. An equal number of newborns (three) successfully completed the two trials in each of the four possible orderings derived from combining these two factors, with newborns being randomly assigned to one of the orders.

Two independent coders, who were unaware of the nature of the animations being presented to the newborns, coded their visual behaviour off-line.

The following measures of infants' looking behaviour were extracted from the video recordings: i) The total time spent fixating either of the two objects in the congruent animation, expressed as a percentage of the total time spent looking at any object (whether congruent or incongruent) across both the congruent and incongruent animations (a preferred measure often used in studies of newborns' looking behaviour), for comparison against a value of 50%; ii) The total time spent looking directly at either of the two objects in the congruent animation, for comparison against the total time spent looking directly at either of the two objects in the incongruent animation (a measure used in previous studies of correspondences in older infants); iii) The average duration of a single fixation on either of the two objects in the congruent animation, for comparison against the average duration of a

animation, sufficient to induce an effect on overall looking times, this is ruled out by arranging for the balls to first appear equally often at the top or the bottom of the screen, and for the starting pitch to be equally often high or low (for discussion, see Lewkowicz & Minar, 2014; Walker et al., 2014).

single fixation on either of the two objects in the incongruent animation; and iv) The number of fixations towards an object in the congruent animation, for comparison against the number of fixations towards an object in the incongruent animation.

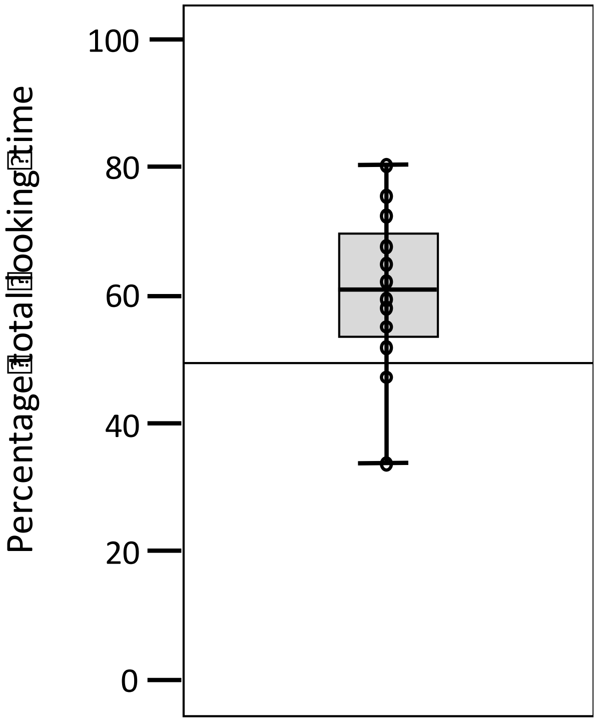

Figure 2. Box and whisker plot of the percentage of total looking time spent looking at objects in congruent animations, rather than at objects in incongruent animations, for each of the 12 newborns (where 50% represents no preference contingent on congruence).

Results

A high level of agreement was confirmed between the two coders in their coding of the newborns' visual fixations (mean Pearson *r* = .82, based on total looking times on the balls).

Looking times. A single sample t-test, with the comparison value set at 50%, confirmed that, relative to the total time a neonate spent looking directly at any object across both congruent and incongruent animations, a greater percentage of time was spent looking directly at objects in congruent animations than at objects in incongruent animations, 61%, CI [53, 69], t (11) = 2.90, p = .014. Cohen's d confirmed this as a large effect, d = .84, which was further confirmed by the fact that ten of the 12 newborns manifest a difference in this direction, p = .038 on a binomial test (see Figure 2). Analyses of variance (ANOVA) revealed that this preference for objects in congruent animations did not differ significantly according to whether it was the congruent or incongruent animation that was encountered first, 56%, CI [45, 68] and 65%, CI [54, 77], respectively, F (1,10) = 1.55, p = .24, $\eta_p^2 = .14$. Neither did it differ whether the balls were first seen moving upwards or downwards, 67%, CI [56, 77] and 55%, CI [44, 65], respectively, F (1,10) = 3.14, p = .11, $\eta_p^2 = .24$. In addition, when the 120 s duration of each trial was partitioned into four 30 s intervals, ANOVA failed to reveal a significant effect of interval on the percentage time spent looking directly at objects in congruent animations than at objects in incongruent animations, $F < 1$. It would appear, therefore, that the effect of congruence on looking behaviour remained steady during each trial.

A Wilcoxon matched-pairs signed-ranks test also confirmed that, on individual trials, significantly more time was spent looking directly at either object in a congruent animation than at either object in an incongruent animation, 63.6 s, CI [51.8, 75.3] and 42.3 s, CI [30.6, 54.0], respectively, Z = 2.27, p = .023, reflecting a medium to large effect, r = .46, (see Figure 3). A Wilcoxon test also confirmed that the mean duration of a single fixation on an object was significantly longer for congruent animations than for incongruent animations, 3.9 s, CI [3.5, 4.4] and 2.7 s, CI [2.3, 3.1], respectively, *Z* = 2.98, *p* = .003, reflecting a large effect, *r* = .61.

For all three of these indices of looking time, 10 of the 12 newborns showed a congruity difference in the same direction as the significant effect, and it was consistently the same 10 newborns. This consistency provides additional reassurance regarding the adequacy of the number of participants contributing data to the study.

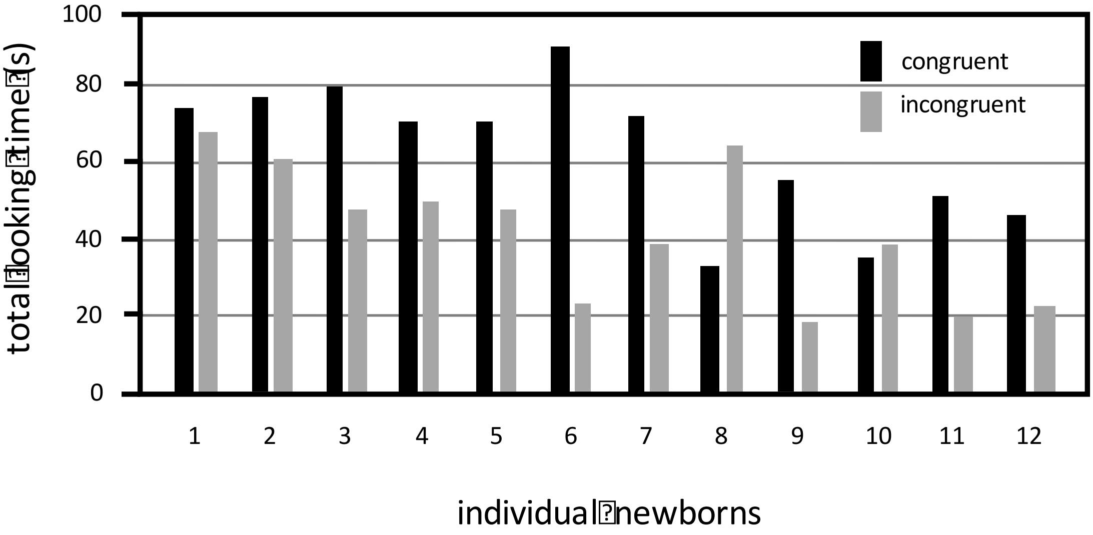

Figure 3. For each of the 12 newborns, the total time on each trial spent looking at objects in congruent and incongruent animations (max = 120 s)

Number of fixations. A Wilcoxon test failed to reveal a significant difference in the number of fixations directed towards objects in congruent animations than towards objects in incongruent animations, 16.5, CI [13.1, 19.9] and 15.1, CI [12.5, 17.6], respectively, *Z* = .826, *p* = .409. These data were then partitioned according to whether the immediately preceding fixation was directed at the other object in the animation or at some (any) point on the background (labelled *type of fixation shift*). The number of such fixations were submitted to an ANOVA with CONGRUENCE (congruent *versus* incongruent animation) and TYPE OF FIXATION SHIFT (direct shift to an object from the other object *versus* indirect shift to an object from any part of the background) as factors. Neither factor, nor their interaction, had a statistically significant effect, all $F < 1$, all $p > .25$.

Discussion

Within the first 2 or 3 days of life, newborns reveal their sensitivity to the cross-sensory correspondence between auditory pitch and visuospatial elevation. When the rise and fall in pitch of a simple tone is congruent, rather than incongruent, with the rise and fall of two balls, newborns fixate the moving balls for longer periods of time (but do not fixate them more frequently). This effect of the auditory-visual congruence on newborns' looking behaviour is the same as that observed previously when 3- to 4-month-old infants viewed a single moving ball (i.e., congruence induced longer individual fixations on the objects, but did not alter the number of fixations) (Dolscheid et al., 2014; P. Walker et al., 2010a).

In light of the limited opportunity these newborns had to learn the correspondence on the basis of encountering relevant feature co-occurrences in the external world, their sensitivity to it appears to have been in place at birth. This accords with claims that an appreciation of the cross-sensory correspondence between auditory pitch and visual size, that is exploited by animals in their vocalizations, also is innately determined (Ohala, 1994).[3] Assuming that being able to appreciate

[3] Many species of animals appreciate the correspondence between pitch and size: When they compete for food, or a mate, they lower the pitch of their vocalizations at the same time as making themselves appear bigger visually (e.g., through piloerection), with both manoeuvres designed to give the impression they are stronger than they would otherwise appear to be (Fitch, & Hauser, 2002; Morton, 1994; Ohala, 1994).

correspondences affords significant advantage early in life (e.g., by supporting knowledge that deeper vocalisations are likely to belong to bigger animals, to adults rather than children, and to men rather than women, see Pietraszewski et al., 2017), it would make sense to have the appropriate structural adaptations in place at birth (much as Parise et al., 2014, regard the shape of the outer ear as an adaptation to the co-occurrence of the spatial elevation of a sound source and the sound's acoustic frequency profile).

By seeming to be in place at birth, the pitch-elevation correspondence, and by implication other correspondences, appear more aligned with amodal cross-sensory mappings than with arbitrary cross-sensory mappings. Correspondences also share a further characteristic with amodal mappings in that they both support modality-independent representations (e.g., the correspondence between auditory pitch and each of visuospatial elevation and visual surface brightness support modality-independent notions of elevation and brightness, respectively). In light of these similarities, perhaps the cross-sensory correspondence between pitch and visuospatial elevation should be reclassified as an amodal association.

Arguing against such reclassification is the fact that acoustic frequency is unreliable and inaccurate as a cue to the spatial elevation of the source of a sound (e.g., Roffler & Butler, 1968). Indeed, as work on the ventriloquism effect reveals, when good visual information also is available it dominates in determining the perceived location of the source of the sound (see Alais & Burr, 2004). This is especially the case in relation to the perceived spatial *elevation* of simple sounds (see Butler & Belendiuk, 1977, concerning the situation with regard to broadband noise bursts). In short, therefore, auditory pitch is unable to function as a reliable, invariant, narrowly tuned cue to spatial elevation, and so is unsuitable for absolute mapping onto visuospatial elevation. On the contrary, and as noted already, cross-sensory correspondences in general, and the correspondence between auditory pitch and visuospatial elevation in particular, involve mappings that are primarily relative in nature (Chiou & Rich, 2012; Marks, 1987; Parise, 2015; L. Walker & P. Walker, 2015; P. Walker, 2016). For example, the same absolute pitch for a sound will map on to a higher or lower visuospatial location depending whether the alternative sound with which it appears is higher or lower in pitch than itself (Chiou & Rich, 2012; Marks, 1987). A similar situation occurs in relation to the correspondence between visual surface brightness and haptic size (L. Walker & P. Walker, 2015). In brief, the context-sensitive, relative mapping of cross-sensory features sets the pitch-elevation correspondence, and correspondences in general, apart from amodal cross-sensory mappings.

In reviewing the evidence for relative and absolute mapping in cross-sensory associations, L. Walker and P. Walker (2015) identify a precondition for the latter to be observed: Sensory information from the two domains being cross-referenced should refer to the same measurable feature of a stimulus, and should be spatio-temporally coincident (or close to such), consistent with the information originating from the same object. Though these authors did not have the notion of amodal cross-sensory associations in mind when reflecting on instances of absolute mapping, this precondition fits nicely with this notion. There is an important implication that follows from this. Specifically, L. Walker and P. Walker also discuss how the distinction between absolute and relative mapping relates to the nature of the features being mapped onto each other, and to the nature of the modality-independent feature emerging from their convergence. They note the generally held view that whereas absolute mapping involves sensory-perceptual levels of representation, relative mapping involves more abstracted, conceptual levels of representation. Certainly, coding something as being *heavier* or *brighter* than something else requires the complex and relatively abstract (conceptual) coding of stimuli. If true, then the conceptual nature of the feature representations involved in correspondences is a further characteristic distinguishing them from amodal mappings.

Though the idea that the newborns in our study had learned the pitch-elevation correspondence in the womb cannot be ruled out, it faces many challenges. Apart from anything else, fetuses will have little, if any, *visual* sense of the spatial location of a sound source in the external environment (e.g., only strong light from the red end of the spectrum penetrates maternal tissue). They are also unlikely to have any auditory sense of its spatial location in the external world. Either one of these limitations will preclude them from establishing any visuospatial-auditory mappings pertaining to relevant feature co-occurrences in the external environment. In brief, any co-occurrences in the external world that are supportive of the pitch-elevation correspondence are not available to the fetus. These obstacles apart, it is also the case that fetuses change the orientation of their bodies frequently, as do their mothers, making it especially difficult for them to encode the spatial elevation of anything in the external world. Indeed, in the third trimester of pregnancy, fetuses typically engage their heads down relative to their mother's body. If we assume their mothers maintain an upright position, and the fetus uses a head-centred spatial frame of reference (see Reid et al., 2017, for evidence that they do), then they risk learning the reverse of the pitch-elevation association they will encounter after birth (and that was tested in the present study). Some of these difficulties could be avoided if fetuses were able to adopt a spatial frame of reference grounded in the external world (e.g., gravity), because this would give them some independence from all the changes in their, and their mother's, bodily orientation. The benefit would be a seamless transfer to the same pitch-elevation association they will experience after birth. However, the authors know of no evidence that fetuses are able to adopt a spatial frame of reference anchored in the external world. With all these challenges to overcome, it will be fascinating to witness correspondences in general, and the pitch-elevation correspondence in particular, being learned in the womb.

It is also possible that our newborns learned the pitch-elevation correspondence in the first hours after birth (i.e., by eventually experiencing relevant feature co-occurrences in the external world). However, this possibility also faces several challenges. For example, relevant co-occurrences would need to be prominent in the newborns' environment, and the newborns would need to adapt to them quickly despite their probabilistic nature. But the typical environment to which newborns are exposed can be relatively limited in nature (in the UK and Italy at least), a problem that is compounded by the fact that newborns generally spend much of their time in a recumbent position and asleep. Were they to adopt a head-centred frame of reference to code the spatial elevation of a sound, being in a recumbent position would certainly confuse matters, and might even cause them to learn a pitch-elevation correspondence different from the one they will encounter subsequently (and different from the one tested in the present study). Again, newborns could avoid some of these problems by coding the spatial elevation of a sound using a frame of reference grounded in the environment (e.g., gravity). This would allow them to gain experience of the co-occurrence that is believed by some to underpin the correspondence between auditory pitch and visuospatial elevation (Parise, Knorre & Ernst, 2014).

Attempts to show that newborns must learn correspondences postnatally, on the basis of experiencing relevant natural feature co-occurrences, might do well to capitalize on the relatively brief period over which it would now appear newborns are capable of learning them. Newborns could be deliberately exposed to selected co-occurrences and later tested for their sensitivity to the correspondences these support. Alternatively, a complete inventory of the feature co-occurrences experienced by individual newborns could be compiled to see if this predicts their sensitivity to different correspondences. For the moment, however, it seems most likely that sensitivity to at least one correspondence is present at birth.

References

Alais, D. & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257-262.

Bahrick, L. E., Hernandez-Reif, M. & Flom, R. (2005). The development of infant learning about specific face-voice relations. *Developmental Psychology, 41,* 541-552.

Bahrick, L. E., Netto, D. & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development, 69,* 1263-1275.

Bidet-Ildei, C., Kitromilides, E., Orliaguet, J-P., Pavlova, M., & Gentaz, E. (2013). Preference for point-light human biological motion in newborns: Contribution of translational displacement. *Developmental Psychology,* doi: 10.1037/a0032956

Boltz, M. (2011). Illusory tempo changes due to musical characteristics. *Music Perception, 28,* 367-386.

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

Butler, R. A. & Belendiuk, K. (1977). Spectral cues utilized in the localization of sound in the median saggital plane. *Journal of the Acoustical Society of America, 61,* 1264-1269.

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception, 41,* 339-353. doi: 10.1068/p7161

Chiou, R., Stelter, M. & Rich, A. N. (2013). Beyond colour perception: Auditory-visual synaesthesia induces experiences of geometric objects in specific locations. *Cortex, 49,* 1750-1763.

Collier, W. G., & Hubbard, T. L. (2004). Musical scales and brightness evaluations: Effects of pitch, direction, and scale mode. *Musicae Scientiae, VIII,* 151-173.

Di Giorgio, E., Frasnelli, E., Salva, O. R., Luisa, S. M., Puopolo, M., Tosoni, D., NIDA-Network, Simion, F. & Vallortigara, G. (2016). Difference in visual social predispositions between newborns at low- and high-risk for autism. *Scientific Reports, 6, 26395;* doi: 10.1038/srep26395

Di Giorgio, E., Leo, I., Pascalis, O. & Simion, F. (2012). Is the face-perception system human-specific at birth? *Developmental Psychology, 48,* 1083-1090.

Di Giorgio, E., Lunghi, M., Simion, F. & Vallortigara, G. (2017). Visual cues of motion that trigger animacy perception at birth: The case of self-propulsion. *Developmental Science,* 20: n/a, e12394. doi:10.1111/desc.12394

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science, 24,* 613-621.

Eitan, Z. & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114,* 405-422.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10,* 1-12.

Fitch, W. T., & Hauser, M. D. (2002). Unpacking 'honesty': Vertebrate vocal production and the evolution of acoustic signals. In *Acoustic Communication,* edited by A.M. Simmons, R.R. Fay, and A.N. Popper. Springer: New York, *pp.* 65–137.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68,* 1191-1203. doi: 10.3758/BF03193720

Kussner, M. B. & Leech-Wilkinson, D. (2013). Investigating the influence of musical training on cross-modal correspondences and sensorimotor skills in a real-time drawing task. *Psychology of Music, 42, 448-469.*

Leo, I. & Simion, F. (2009). Newborns' Mooney-Face perception. *Infancy, 14,* 641-653.

Lewkowicz, D. J. & Minar, N. J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences: A comment on Walker et al. (2010). *Psychological Science, 25,* 832-834.

Ludwig, V. U., Adachi, I., & Matzuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *PNAS, 108,* 20661–20665.

Macchi Cassia, V., Valenza, E., Simion, F., & Leo, I. (2008). Congruency as a nonspecific perceptual property contributing to newborns' face preference. *Child Development, 79,* 807-820.

Marks, L. E. (1978). *The Unity of the Senses: Interrelations among the modalities.* New York: Academic Press.

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 384-394. doi: 10.1037/0096-1523.13.3.384

Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. *Developmental Science, 16,* 327-335.

Morton, E. S. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols, & J. J. Ohala (Eds), *Sound Symbolism,* pp. 348-365. New York: Cambridge University Press.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds), *Sound Symbolism,* pp. 325-347. New York: Cambridge University Press.

Parise, C. V. (2015). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research, 29,* 7-28

Parise, C. V., Knorre, K. & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *PNAS, 111,* 6104-6108.

Pena, M., Mehler, J., & Nespor, M. (2012). The role of audiovisual processing in early conceptual development. *Psychological Science, 22,* 1419-1421.

Perrott, D. R., Musicant, A., & Schwethelm, B. (1980). The expanding-image effect: The concept of tonal volume revisited. *Journal of Auditory Research, 20,* 43-55.

Peters, M. A. K., Balzer, J., & Shams, L. (2015). Smaller = denser, and the brain knows it: Natural statistics of object density shape weight expectations. *PLoS ONE, 10,* e0119794. doi: 10.1371/journal.pone.0119794

Pietraszewski, D., Wertz, A. E., Bryant, G. A., & Wynn, K. (2017). Three-month-old human infants use vocal cues of body size. *Proceedings of the Royal Society, Series B, 284:* 20170656

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donavon, T., & Reissland, N. (2017). The human fetus preferentially engages with face-like visual stimuli. *Current Biology,* http://dx.doi.org/10.1016/j.cub.2017.05.044

Roffler & Butler (1968) Localization of tonal stimuli in the vertical plane. *Journal of the Acoustical Society of America, 43,* 1260-1266.

Simion, F., Regolin, L. & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *PNAS, 105,* 809-813.

Simion, F., Valenza, E., Cassia, V. M., Turati, C. & Umilta, C. (2002). Newborns' preference for up-down asymmetrical configurations. *Developmental Science, 5,* 427-434.

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,* 333-338.

Spector, F. & Maurer, D. (2009). Synesthesia: A new approach to understanding the development of perception. *Developmental Psychology, 45,* 175-189.

Tarte, R. D. (1982). The relationship between monosyllables and pure tones: An investigation of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 21,* 352-360.

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

Walker, L. & Walker, P. (2015). Cross-sensory mapping of feature values in the size-brightness correspondence can be more relative than absolute. *Journal of Experimental Psychology: Human Perception and Performance, 42,* 138-150.

Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception, 41,* 1186-1192.

Walker, P. (2012b). Cross-sensory correspondences and naïve conceptions of natural phenomena. *Perception, 41,* 620-622.

Walker, P. (2016). Cross-sensory correspondences: A theoretical framework and their relevance to music. *Psychomusicology: Music, Mind, and Brain, 26,* 103-116.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010a). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*, 21-25.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2014). Preverbal infants are sensitive to cross-sensory correspondences: Much ado about the null results of Lewkowicz and Minar (2014). *Psychological Science, 25*, 835-836.

Walker, P., Francis, B. J., & Walker, L. (2010b). The brightness-weight illusion: Darker objects look heavier but feel lighter. *Experimental Psychology, 57,* 462-469.

Walker, P. & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception, 13,* 75-81.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex, 42,* 264-280.

Acknowledgements

The authors are deeply indebted to Professors Eugenio Baraldi, Head of Unità Operativa Complessa di Terapia Intensiva e Patologia Neonatale, and Giorgio Perilongo, Head of the Department of Health of Woman and Child of the University of Padova, for providing laboratory space and allowing us to test the newborns. We would also like to thank the nursing staff, the babies and their parents for their collaboration, and Valentina Angeli for assistance with newborn testing.

This study was carried out within the framework of the agreement between the University of Padova and the Azienda Ospedaliera of Padova (n° 91644645).

Author Contributions

G. B., S. D., F. S., & P. W. developed the study concept and designed the research. B. D. B. was instrumental in contacting the nurses and getting consent from the mothers. M. L. & F. S. conducted the research. P. W. analysed the data and drafted the manuscript. G. B., S. D., & F. S. provided critical revisions.