TARGET ARTICLE WITH COMMENTARIES AND RESPONSE

Staring us in the face? An embodied theory of innate face preference

Nick Wilkinson,[1] Ali Paikan,[1] Gustaf Gredebäck,[3] Francesco Rea[2] and Giorgio Metta[1,4]

1. iCub Facility, Istituto Italiano di Tecnologia, Genova, Italy
2. Robotics, Brain and Cognitive Sciences, Istituto Italiano di Tecnologia, Genova, Italy
3. Child and Baby Lab, Department of Psychology, University of Uppsala, Sweden
4. Centre for Robotics and Neural Systems, University of Plymouth, UK

For commentaries on this article see Smith and Jayaraman (2014) and Méary, Mermillod and Pascalis (2014).

Abstract

Human expertise in face perception grows over development, but even within minutes of birth, infants exhibit an extraordinary sensitivity to face-like stimuli. The dominant theory accounts for innate face detection by proposing that the neonate brain contains an innate face detection device, dubbed 'Conspec'. Newborn face preference has been promoted as some of the strongest evidence for innate knowledge, and forms a canonical stage for the modern form of the nature–nurture debate in psychology. Interpretation of newborn face preference results has concentrated on monocular stimulus properties, with little mention or focused investigation of potential binocular involvement. However, the question of whether and how newborns integrate the binocular visual streams bears directly on the generation of observable visual preferences. In this theoretical paper, we employ a synthetic approach utilizing robotic and computational models to draw together the threads of binocular integration and face preference in newborns, and demonstrate cases where the former may explain the latter. We suggest that a system-level view considering the binocular embodiment of newborn vision may offer a mutually satisfying resolution to some long-running arguments in the polarizing debate surrounding the existence and causal structure of newborns' 'innate knowledge' of faces.

Research highlights

- Provides a formally specified, mechanistic explanation for innate face preference.
- Extends the standard 'linear systems model' of newborn early vision to the binocular case.
- Demonstrates the potential importance of embodiment to cognition and perception.
- Lends support to the developmental systems perspectives on cognitive and behavioural inheritance.

Introduction

The empirical position is, to be sure, in agreement with the nativistic on a number of points–for example, that local signs of adjacent places on the retina are more similar than those farther apart and that the corresponding points on the two retina are more similar than those that do not correspond. (Helmholtz, 1878)

Face perception is one of the most intensively researched topics in psychology and neuroscience (Bruce & Young, 2012). One of the liveliest debates in the face perception literature concerns whether face perception is subserved by a domain specific module or domain general expertise (Kanwisher, McDermott & Chun, 1997; Gauthier, Tarr, Anderson, Skudlarski & Gore, 1999; Gauthier & Nelson, 2001; McKone, Kanwisher, & Duchaine, 2007) Even infants in their first minutes and hours of life appear to exhibit a special interest in face-like stimuli, lending support to the notion that faces enjoy 'special' status in visual perception (Goren, Sarty & Wu, 1975; Johnson, Dziurawiec, Ellis & Morton,

1991). These findings and their interpretation form the focus of the current theoretical contribution. In these studies, 'face-like' has been determined subjectively by experimenters, while 'special' is defined relative to the predictions of the *linear systems model* (LSM) (Banks & Salapatek, 1981; Banks & Ginsburg, 1985), a newborn version of the widely adopted *energy model* approach to front end visual processing (Morrone & Burr, 1988; Mante & Carandini, 2005).

The LSM mathematically describes the obligatory filter properties of the immature visual system, such as the resolution of rods in the retina, the quality of the ocular medium, and the neural point-spread function, to define the visibility of any given stimulus based on the average newborn *contrast sensitivity function* (Banks & Salapatek, 1976). Preferences are predicted to simply be proportional to visibility in this sense. On this view, the shape/configuration of the contrast in a stimulus should make no difference. The LSM represents the empiricist *sensory hypothesis*; that newborn preferences for particular stimuli are proportional to their match to general properties of an infant's sensory systems (Kleiner & Banks, 1987). However, newborn preferences for face-like stimuli appear to deviate from the predictions of the LSM (Goren *et al.*, 1975; Johnson *et al.*, 1991), because configuration is important even when visibility is controlled. This motivated Morton, Johnson and Maurer (1990), Morton and Johnson (1991) and Johnson (2005) to argue that the sensory hypothesis is insufficient, and propose the influential nativist *structural hypothesis* to explain this extraordinary 'newborn face preference':

> In general terms, the idea is that some of the visual preferences of the human infant are not merely determined by the extent to which the psychophysical properties of a stimulus match those of the infant's sensory channels. More specifically, we discuss evidence in support of the idea that infants possess some information about the structural characteristics of faces from birth...This contrasting view holds that the neonate brain contains innate information concerning the structure of faces. (Morton and Johnson, 1991, pp. 164–170)

Following up on their reproduction and extension of Goren *et al.* (1975), Johnson *et al.* (1991) conducted a third experiment examining face preferences over the first few months postnatal. They reported that the newborn face preference was no longer statistically significant after 4–6 weeks postnatal. Taken together, these results formed the evidential foundation of Conspec-Conlern, the influential *two process theory* of newborn face perception (Morton & Johnson, 1991). This theory proposes that an innate face detection device, termed *Conspec*, guides newborn orienting in the first weeks of life, ensuring a stimulus stream rich in facial patterns. This bootstraps the development of cortical specialization for face perception . These developing cortical mechanisms, collectively termed *Conlern*, later come to suppress Conspec and are responsible for ongoing face processing.

The current contribution revisits these seminal findings and their interpretation. Our proposal builds on the sensory hypothesis, but goes beyond the obligatory notion of bare visibility to the richer realm of selective attention; distinguishing between equally visible signals on the basis of some *relevance criteria*. In this sense, our approach aligns with the widely held view in developmental psychology that human newborns are innately attracted to faces, and indeed the 'like me' in general (Meltzoff & Brooks, 2001; Meltzoff, 2007; Johnson, 2006). The current contribution aims to extend the mechanistic conceptual framework and formal rigour of the LSM to accomodate the convincing evidence that newborns *selectively attend* to faces. Echoing critics of the nature–nurture dichotomy (Lehrman, 1953; Gottlieb & Kuo, 1965; Oyama, 2000; Lickliter & Berry, 1990; Spencer, Blumberg, McMurray, Robinson, Samuelson & Tomblin, 2009), we argue that the sensory and structural hypotheses are not as mutually exclusive as, Johnson and Morton (1991) imply in the above quote. An innate (present at birth, independent of stimulus specific experience) 'face preference' is both entirely plausible given current scientific knowledge of early visual physiology, and entirely compatible with reasonable empiricist approaches, notably Helmholtz (1878) and Mendelson, Haith and Gibson (1976). We suggest that a mutually satisfying resolution to 'one of the longest running debates in developmental psychology' (Johnson, 2005) may lie in closer consideration of newborn *embodiment* (Varela, Thompson & Rosch, 1991; Brooks, 1991).

We employ computational and robotic modelling to show how simple binocular integration can provide a crude but effective 'face preference'. Humanoid robotics enables the extension of computational neuroscience to ecologically embedded context, analogous to the *in-vitro* vs. *in-vivo* distinction in neuroscience. This facilitates proof-of-concept numerical simulations of neuro-ethological theories which depend fundamentally on embodiment and situation, as well as neural computations (Morse, Herrera, Clowes, Montebelli & Ziemke, 2011). In the following section, we first briefly review related empirical and theoretical work on NFP and newborn binocularity. In the subsequent section, we propose and define a potential alternative, the 'binocular correlation model'. The next section details the methods and results of the current attempt to qualitatively

reproduce the psychophysical results of Johnson *et al.* (1991). The final section discusses the relevance and scope of our theoretical findings.

Background and related work

Current ideas regarding the causes of newborn face preference

Three main ideas regarding the mechanisms underlying the extraordinary newborn response to faces have been suggested to date. The historically precedent and most influential theory assumes that a low resolution face template, termed Conspec, is provided by evolution (Morton & Johnson, 1991; Johnson, 2005). In the second experiment reported in Johnson *et al.* (1991), the preference for the schematic Face over the three blob Config stimulus was marginally insignificant ($p = .07$). (See Figure 4b). On this basis, Morton and Johnson (1991, p. 172) submitted that 'our experiments have not been able to distinguish between the face stimulus and the config stimulus, so the simplest hypothesis is that the specification of Conspec resembles [Config]'. [1] Conspec proposes that faces are a special class of stimuli for newborns.

The Conspec hypothesis claims explanatory power for a wide range of results (Morton & Johnson, 1991; Johnson, 2005; Rosa Salva, Farroni, Regolin, Vallortigara & Johnson, 2011), and has become part of any textbook discussion of innate predispositions and the 'biological' bases of cognitive development (Johnson & Morton, 1991; Spelke, 1998; Johnson, 2010). The validity of these claims is difficult to assess because the proposal remains fluid, and does not make strong predictions (Morton & Johnson, 1991; Johnson, 2005; Johnson, Grossmann & Farroni, 2008), limiting the scope of attempts to model and test the idea (Bednar & Miikkulainen, 2003; Balas, 2010).

A second perspective suggests that domain general spatial biases for stimulus properties such as top-heaviness (Simion, Macchi Cassia, Turati & Valenza, 2001; Simion, Valenza, Macchi Cassia, Turati & Umilta, 2002) and congruency (Macchi Cassia, Valenza, Simion & Leo, 2008) conspire to bias neonates' attention towards face-like stimuli. On this view, faces are not special to newborns (Simion *et al.*, 2001; Turati, 2004). The third

view suggests that prenatal learning and proprioceptive experience inform the development of internal representations of facial structure (Meltzoff & Moore, 1993; Slater, Quinn, Lewkowicz, Hayes & Brookes, 2003; Sugita, 2009; Pitti, Kuniyoshi, Quoy & Gaussier, 2013). We share many agreements with this latter perspective, but there is at least one important difference. In particular, we agree that facial embodiment is essential, and that face specialism emerges at the level of the behaving organism, rather than the brain alone. However, we argue that the morphology of the sensor array can be exploited directly to perceive configuration in the environment (Towal, Quist, Gopal, Solomon, & Hartmann, 2011); prior encoding of facial structure in internal representations is not a necessary prerequisite for the perceptual use of facial structure.

Motivations for a binocular correlation approach to NFP

Current models

The LSM forms the fulcrum of quantitative theoretical reference in the newborn visual preference literature (Gayl, Roberts & Werner, 1983). The current incarnation of the LSM (Banks & Salapatek, 1981) is insufficient to explain the results of Johnson *et al.* (1991), but this does not imply that a specialist innate face detection device exists inside the newborn brain. We believe that any theory of extra-ordinary NFP should build on the LSM by proposing formally defined, testable models which map to identifiable physiology.

The domain general biases alternative to the Conspec hypothesis (Turati, 2004) makes the point that 'face specific' (and indeed 'face-like') is an underdefined concept, but itself remains underspecified. As the data have required the proposal of ever more complex and designed (one might even say face-like) biases, the model has arguably become an alternative implementation of the Conspec idea (Johnson *et al.*, 2008). Nonetheless, the original critical point of the generic biases hypothesis is well motivated in our opinion.

However, we also agree with the Conspec camp that there is *something* tantalizingly facey about the newborn responses which have been observed. It may appear obvious that one is really forced to choose here. Fortunately, this intuition is mistaken. Bednar and Miikkulainen (2003) quite rightly pointed out that the innate structural information could reside originally in the input to the network rather than the network itself and the computations it performs. They proposed that the three blob pattern could be learned prenatally from genetically specified endogenous activity. However, they still directly build faceness into the system, by hand

---

[1] This analysis is not without problems. Firstly, it presupposes the existence of Conspec. Secondly, it reverses the null. Thirdly, there was no preference observed for the Config stimulus relative to controls. This *p* value of .07 indicates a 7% chance that the null assumption – Conspec exists and has a three blob form – is true given this comparison.

designing the training input as three blobs. There is no evidence to support their assumption that such patterns are dominant in endogenous activity in the prenatal visual system. Overall, the mechanisms underlying NFP remain largely unknown, motivating our new proposal.

Source of the configural transform

What is required is an input spatiality which is inherently facial, but due to naturalized properties of the infant, rather than hand designed by the creator. We suggest

Figure 1   *Note how the eyes of the models overlap in the stereo pairs, regardless of depth and translations. This occurs because the distance between the camera lenses is about the same as the distance between the models' eyes, and the cameras are aligned in parallel. Although the absolute size and location of the models' projections on the 2D visual array change with depth and translation, the relationship between the 3D volumes projecting to each sensor remains constant with parallel axes. With vergence, the relationship changes with depth according to Equation 3.*
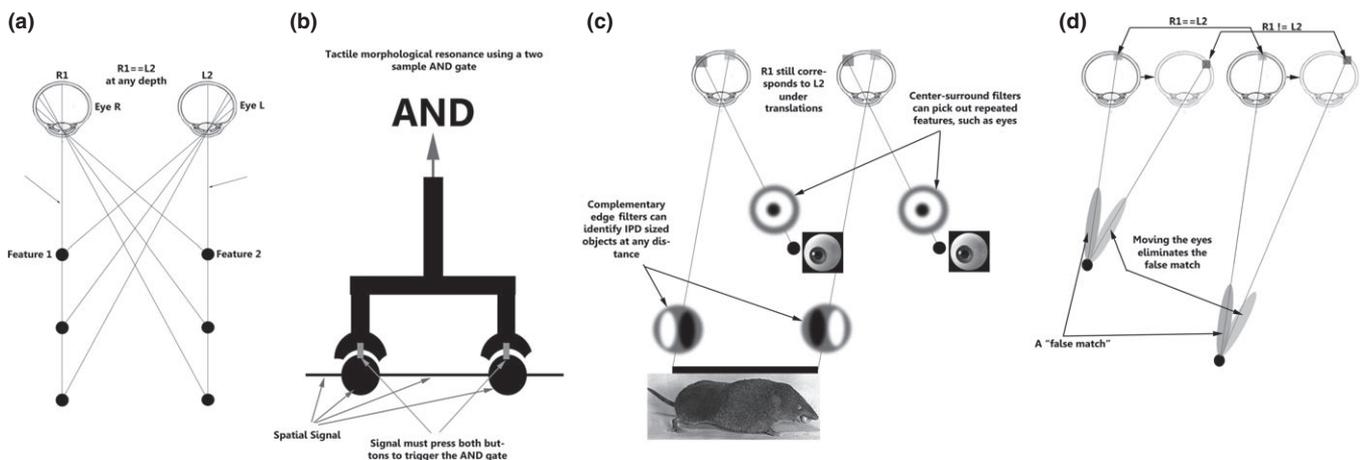
Figure 2   *(a) Conservation of distance over space. The relationship of Feature 1 to Eye R is the same as that of Feature 2 to Eye L, regardless of depth. (b) Schematic of morphological resonance as a physical AND gate. The spatial signal must press both input buttons to turn on the gate. (c) Correspondence is robust to translations. Different monocular filters can pick out different types of texture. (d) Texture occuring at different depths can give rise to ambiguities/false matches. These ambiguities are not robust to changes in perspective, so active visual movements can be used to eliminate false matches.*

that this input spatiality can be provided by the distribution in physical space of the sensors providing the input. Structural information about faces is directly present in the shift between the visual sensors, known as inter-pupillary distance (henceforth IPD). The required configural transform then consists in the large-scale neural architecture mapping these spatially offset inputs to the same origin in lateral geniculate nucleus (Swanson, Cowan & Jones, 1974; Kaas, Guillery & Allman, 1972), thereby *collapsing* IPD. Therefore the neural computation integrating the binocular samples can be uniform and local yet still be selective for (horizontal) configural relations, because it operates on inputs gathered from different spatial locations in the world. Figure 1 displays overlaid stereo images of three different faces at different depths and translations. This depicts the visual affordance (Gibson, 1961 1966, 1977) by which patterns of contrast typical of face-like stimuli induce correlated activity at corresponding retinal locations. Figure 2 depicts schematically how spatial resonance between sensor morphology and stimulus can give rise to these correlations. In essence, we propose that it is possible to use the body itself as a 'template' to match form in the environment, just as a neural connectivity pattern can act as a template for form on the retina or visual cortex.

A binocular correlation hypothesis for newborn visual attention

Correlated multimodal stimulation is widely thought to be an important cue to engage attention and trigger learning, even in newborns (Bahrick & Lickliter, 2000; Sai, 2005; Lewkowicz, Leo & Simion, 2010; Streri, Coulon & Guellaı̈, 2013). Less attention has been paid to correlated intramodal *bilateral* stimulation in the social context. Bilateral symmetry clearly plays an organizing role in vertebrate sensor distribution and integration. Tactile stimulation of one hand causes bilateral cortical activation at corresponding somatotopic points (Hansson & Brismar, 1999). Binaural correlation informs orienting and looking behaviour in neonates (Mendelson *et al.*, 1976; Jiang & Tierney, 1996; Furst, Bresloff, Levine, Merlob & Attias, 2004), and the foetus is capable of auditory orienting *in utero* (Voegtline, Costigan, Pater & DiPietro, 2013). From an aesthetic perspective, the 'sweet spot' region of binaural synchrony is manipulated by sound engineers to deliver the most enjoyable and engaging listening experience (Theile, 2000; Bauck, 2003), suggestive of a more general multimodal link between bilateral correlation, arousal and 'liking'.

We reasoned that binocular correlation might similarly engage newborn attention. The basic computational idea of the binocular correlation model (henceforth BCM) is that corresponding points on the binocular visual arrays which are salient *in both eyes* are amplified relative to those which are not. Given this, a certain level of 'face preference' would be a natural consequence (Figures 1 and 2). More specifically, our proposal extends the LSM model of newborn monocular vision (Banks & Salapatek, 1976 1981) to the binocular case, with the addition of a pointwise multiplication function which integrates the stereo pair.

Defining visual mechanisms of the BCM: monocular decorrelation, binocular correlation

The combined process can be summarized as local (monocular) decorrelation followed by global (binocular) correlation. Each monocular image is first decorrelated by convolution with a 'Difference of Gaussian' (DoG) kernel, to produce a 'contrast map'. This is a widely used computational technique for simulating antagonistic centre–surround connectivity in the nervous system (Birch, Mitra, Bangalore, Rehman, Young & Chatwin, 2010; Zuiderbaan, Harvey & Dumoulin, 2012) and can be particularly suitable for finding eyes in a visual scene (Yang, Atick, & Reid 1996). Measurements from cells in lateral geniculate nucleus (our proposed

Figure 3 *Images taken from the robot during operation. (a) The Face stimulus. (b) Superimposed stereo pair. Note how the eyes overlap. (c) Result of pointwise multiplication of the stereo pair. Note how overlapping regions of activity in the stereo pair are amplified relative to the rest of the signal. The red circle indicates the point of maximum salience in the map.*

neural substrate, see next section) show 'white' spectral characteristics consistent with decorrelation [143]. Kernel size was chosen to yield a peak sensitivity at about 0.1 cycles per visual degree, at the low end of the peak spatial frequency sensitivity observed in newborns (0.1–0.5 cpvd) (Slater & Sykes, 1977; Norcia, Tyler & Hamer, 1990; Atkinson, Braddick & French, 1979) to compensate for the necessary scaling up of the stimuli (see Experimental Set-Up below). Note that this is a minimal approximation to the LSM, not a precise instantiation.

Next, the contrast maps from the left and right eyes/cameras are multiplied pixelwise, outputting a single combined array we call the 'salience map'. See Figure 3 for a depiction of these stages for one of the stimuli employed. Formal definitions and further details are provided as Supporting Information to this article.

The BCM entails two basic neural requirements: (i) a large-scale neural architecture aligning the binocular signals retinotopically, and (ii) the multiplicative integration of the signals from binocularly corresponding points. In the following section, we assess the viability of these requirements given current neurophysiological evidence.

Neural substrate – gain control in the visual thalamus

Alignment of the binocular signal streams

Although pointwise multiplication of the stereo pair is computationally trivial, it is not physiologically trivial. Rather precise neuro-physiology is required to align the stereo streams in accurate correspondence. The layered structure of the thalamic *lateral geniculate nucleus* (henceforth LGN), first stage in the main visual pathway to cortex, provides just these conditions (Kaas *et al.*, 1972; Swanson *et al.*, 1974). LGN Comprises six layers. Connections from contralateral nasal retina project to layers 1, 4 and 6, while ipsilateral temporal retina projects to layers 2, 3 and 5. All of these layers are precisely in-reference with respect to retinal receptive fields. An active role for the thalamus in attention has long been theorized (Clark, 1932; Crick, 1984), and evidence supporting these suggestions is accumulating (Varela & Singer, 1987; Sillito, Jones, Gerstein & West, 1994; Sherman & Guillery, 2001; Sherman, 2007; Saalmann & Kastner, 2011).

Multiplicative stereo integration

Neural *gain control* can implement multiplicative interactions *in vivo* (Rothman, Cathala, Steuber & Silver, 2009). Gain control acts like an amplifier, or 'gate' for input signals. A gain of zero completely silences the signal. A gain of 1.0 reproduces the input signal as output, while a gain of 2.0 reproduces the signal and doubles its amplitude. Mutual gain control describes binaural interaction at the single cell level in mice (Xiong, Liang, Li, Mesik, Zhang, Polley, Tao, Xiao & Zhang, 2013), and is the basis of current models of binocular integration in adult humans (Ding & Sperling, 2006; Ding, Klein & Levi, 2013; Meese & Baker, 2011).

Gain control is widely thought to mediate attention (Aston-Jones & Cohen, 2005; Reynolds & Heeger, 2009; Feldman & Friston, 2010; Katzner, Busse & Carandini, 2011). and orientation (Sara & Bouret, 2012), and has been mechanistically linked to ascending projections from neuromodulatory hubs and the sympathetic nervous system (Aston-Jones & Cohen, 2005; Sara, 2009; Fuller, Sherman, Pedersen, Saper, & Lu, 2011). Mathematically, gain control has been formally equated with the modulation of Bayesian *precision* in probabilistic generative modelling (Feldman & Friston, 2010; Moran, Campo, Symmonds, Stephan, Dolan & Friston, 2013). The minimal model presented here is a much simplified mathematical approximation, not a detailed physiological model; our focus is on the role of organismic embodiment in structuring neural activity (Towal *et al.*, 2011) rather than the low-level details of the neural implementation. We have not included monocular (Carandini, Heeger & Movshon, 1997) or binocular (Moradi & Heeger, 2009) normalization.

Sensory gain control in LGN

LGN is strongly associated with visual gain control (Saalmann & Kastner, 2009 2011). LGN receives only about 10% of its input from retina, with the other 90% constituted in approximately equal proportions by cholinergic projections from the parabrachial nucleus of the brainstem, inhibitory control from the thalamic reticular nucleus, and feedback connectivity from layer 6 in striate cortex (Saalmann & Kastner, 2009). Recently, layer 6 in visual cortex has been shown to mediate gain control of superficial layers (Olsen, Bortone, Adesnik, & Scanziani, 2012; Vélez-Fort & Margrie, 2012). This extensive modulatory network effects gain modulation in LGN, thereby gating visual input to the cortex (Sherman, 2007; Saalmann & Kastner, 2011; Lien & Scanziani, 2013). In addition to this extensive feedback architecture, correlation of presynaptic neuronal activity facilitates postsynaptic firing (Huguenard & McCormick, 2007; Gotts, Chow & Martin, 2012; Friston, 2012), providing a feedforward means of gain control. Thus binocular correlation is very likely to enhance responses at thalamic efferents, at least in this feedforward manner. LGN also projects to the superior colliculus, providing a

fast subcortical pathway for closed loop visuomotor control (Saalmann & Kastner, 2009).

Stereo integration in newborns

Studies examining prestereoptic binocular vision in infants have produced conflicting results, and have not included neonatal subjects (Shimojo, O'Connell & Held, 1986; Brown & Miracle, 2003; Kavšek, 2013). The maturity of binocular gain control circuitry and normalization in human newborns is therefore unknown. In the rhesus macaque, considered a good model for the human visual system, binocular circuitry is quite mature at birth (Rakic, 1976; Horton & Hocking, 1996), and responses are limited by low monocular sensitivity rather than binocular immaturity (Chino, Smith, Hatta & Cheng, 1997). The BCM assumes that newborn binocular integration is qualitatively similar to that in adults, but does not prescribe the precise transform; many variations on mutual gain control are possible, and could generate robust face preferences. The most likely alternative to mutual gain control – simple weighted summation and thresholding – could generate a reduced level of face preference, according to our preliminary investigations, but this is quite dependent on the details of implementation (e.g. level of the threshold).

Overview

In summary, we propose that binocularly correlated stimulus energy will be selectively amplified at the thalamic level, providing the informational basis for selective attention to particular *configurations* of stimulus energy even though there is no 'configural processing' at the neural level. The physical enaction of selective attention is then implemented by the amplified downstream impact of these signals on motor efferents in superior colliculus (Hafed, Goffart, & Krauzlis, 2009) and predictive coding mechanisms in cortex (Rao & Ballard, 1999; Moran *et al.*, 2013). Not by chance, faces are a rich source of binocular correlation (see Figures 1 and 2), resulting in the emergence of an observable behavioural preference 'for' faces relative to many other stimuli. The putative involvement of amygdala in subcortical face processing (Johnson, 2005) and the well-established role of amygdala in affective regulation of plasticity and memory formation (McGaugh, 2004 2013) suggests a potential route from bilateral visual stimulation 'back down' to nervous arousal and thence to wider multimodal engagement, increased plasticity and ensuing social learning through interaction. Our ongoing work is exploring this line of enquiry.

Recreating newborn face preference studies with the iCub robot

At the psychophysical level, the BCM predicts that newborns 'prefer' binocularly correlated signals and that a certain level of inborn 'face preference' will drop out of this more general effect. Whether this contributes to any particular NFP finding is a case by case matter. Here we examine whether binocular correlation can explain the foundational evidence of inborn face preference in quasi face naïve newborns under an hour postnatal (Goren *et al.*, 1975; Johnson *et al.*, 1991). These results have been particularly important in establishing the innateness of NFP, as later studies have, almost without exception (Mondloch, Lewis, Budreau, Maurer, Dannemiller, Stephens & Kleiner-Gathercoal, 1999), used subjects with days of experience interacting with faces.

Methods

We implemented a model based on the BCM on the iCub, an open source humanoid robot designed to make available a standard platform for the implementation and testing of theory in perceptual and cognitive neuroscience (Metta, Natale, Nori, Sandini, Vernon, Fadiga, von Hofsten, Rosander, Lopes, Santos-Victor, Bernardino & Montesano, 2010). We then asked whether this model would reproduce the seminal NFP results of Johnson *et al.* (1991). As a control, we employed a monocular version of the model (i.e. just DoG filtering), approximating the LSM. The robot was isolated from the cluttered laboratory scene using a whitescreen, to model the plain rooms used for newborn studies. Stimuli were presented one at a time in the center of the robot's visual field. Illumination came from both internal lights and the windows, and was broadly similar across testing periods. We recorded the salience values for each stimulus averaged over three trials.

The format is a numerical demonstration of a theoretical model, rather than an empirical study. The model is basically deterministic (save for small environmental fluctuations and sensor noise). Very similar results are achieved on each trial. Thus the calculation of p values is not really appropriate in this context; the prior probability that the results are down to chance fluctuations is infinitesimal. Almost any systematic difference between stimuli could be rendered significant with sufficient trials. We suggest that a salience advantage of about 20% should be considered necessary to indicate a robust preference in practice. Two experiments from Johnson *et al.* (1991) were reproduced under the conditions just
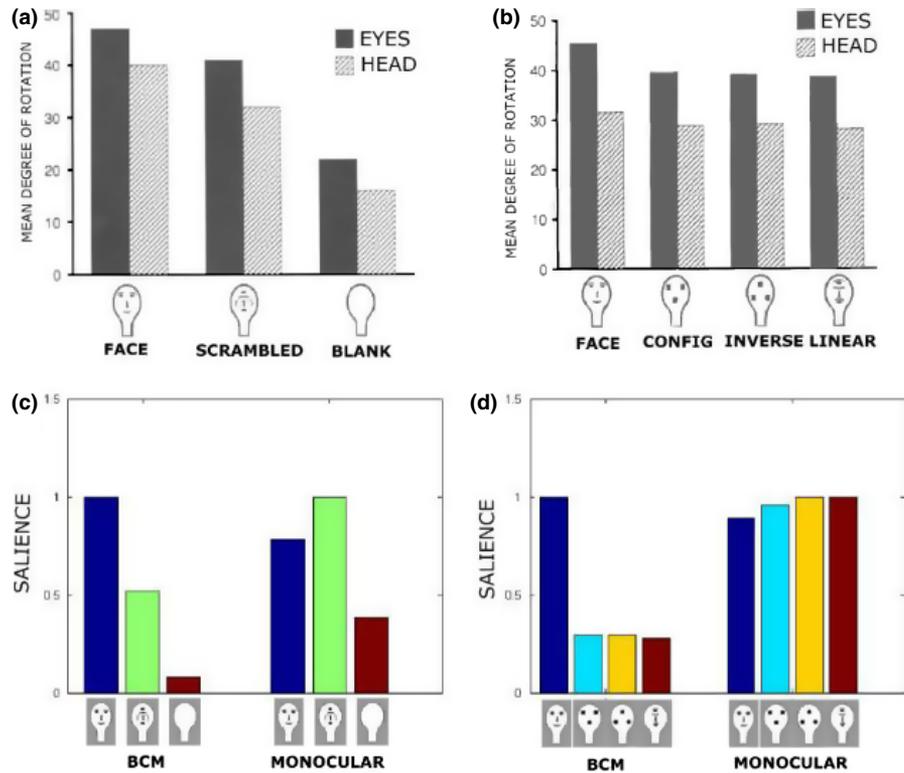
Figure 4   *Comparing the results from newborns with those generated by the BCM. (a) Results of first experiment on newborns. (b) Results of the second experiment on newborns. (c) Results generated by the BCM for the stimuli used in (a). (d) Results generated by the BCM for the stimuli used in (b). Graphs (a) and (b) reprinted from Cognition journal with permission from Elsevier, copyright 1991.*

described, the only difference between sessions being the stimuli employed (see Figure 4). In the third simulation, conditions were the same but we varied the vergence angle between the eyes of the robot, and examined the effects on the salience of the face-like stimulus.

Stimuli

We employed stimuli based on those used in Johnson *et al.* (1991). The stimuli were outline schematics on a white background, as depicted in Figure 4. The iCub's IPD is about 7 cm, whereas newborns have much smaller IPD around 4 cm (Pryor, 1969), so we scaled the stimuli by a factor of 7:4 from the reported dimensions (which were 'head-sized', interpreted as about 17–18 cm), to create roughly the same relative size to the robot's face as the original stimuli would have had to a newborn face. The size of the resulting stimuli was such that they just fitted on an A3 sheet of paper (297×420 mm). At the presentation distance of 30 cm, the stimuli subtended a visual angle of approximately 50 degrees horizontally and 70 degrees vertically.

Results

Experiment 1.1: A basic face preference

Johnson *et al.* (1991) reported a pattern of preferences as depicted in Figure 4a. Goren *et al.* (1975) reported the same pattern of preferences for these stimuli. The control monocular processing alone (just DoG filtering) ranks Scrambled over Face over Blank. It is clear that the BCM generates the same pattern of preferences observed in newborns. Thus while the monocular control (an approximation to the LSM) cannot explain this result, the correlations between those monocular signals can.

Experiment 1.2: Introducing the Config stimulus

In their second experiment, Johnson *et al.* (1991) introduced new stimuli to investigate the stimulus properties necessary to elicit the preference. In particular the Config stimulus, consisting of three high contrast blobs for eyes and mouth areas, was introduced. The pattern of their results is depicted in Figure 4b. The Face stimulus
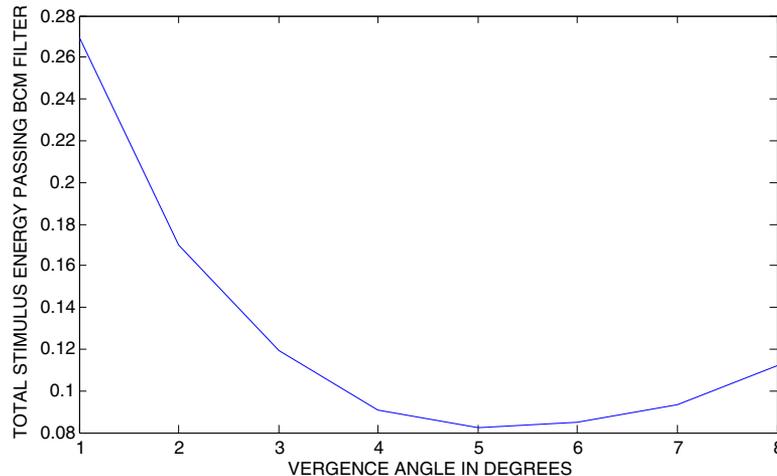
Figure 5   *As vergence angle increases from zero to around five degrees, the "salience" (total stimulus energy passing the BCM filter) rapidly decreases, due to the loss of binocular correlations caused by the "misalignment" of the visual sensor array.*

was preferred, whilst responses to Config, Inverted and Linear were almost identical. Monocular DoG filtering alone makes little distinction between the stimuli. The BCM clearly generates the same pattern of preferences observed in newborns. It should be noted that these particular Config stimuli had the eyes positioned too far apart to induce much binocular correlation. Other three blob stimuli with closer spaced eyes (such as have been used in later studies, e.g. Valenza, Simion, Cassia & Umilta, 1996) would generate stronger correlations and hence be more salient to the BCM.

Experiment 1.3: The decline of face preference

Experiment 3 in Johnson *et al.* (1991) examined the development of preference for schematic faces over the first 5 months of life. They found that face preference declines to trend level in the second month postnatal, and attributed this result to the suppression of the subcortical Conspec mechanism by developing, face expert cortical mechanisms termed Conlern (Morton & Johnson, 1991). In our final experiment, we investigated an alternative causal structure. The decline of NFP in month two coincides with the beginnings of vergence movements in infants (Thorn, Gwiazda, Cruz, Bauer & Held, 1994). Vergence changes ocular alignment, and hence the form of stimuli which will induce binocular correlations. Our third simulation varied the vergence angle of the robot to test whether this would result in a loss of salience for the Face stimulus. The results are displayed in Figure 5. As vergence increases from zero (parallel binocular alignment), the salience of the Face stimulus decreases rapidly, because the stimulus ceases to cause much binocular correlation.

To understand this effect in more detail, examine Figure 3. The binocular overlap between the left and right eyes of the stimulus is not perfect even here with no vergence, because the IPD of the Face stimulus is wider than that of the iCub robot, but there is still enough overlap to cause quite strong binocular correlations in the contrast maps. As vergence increases, this overlap further decreases until the stimulus' eyes no longer overlap at all, then increases again when the images align under proper binocular fixation. As the convergent component becomes more dominant in the sample population over time, averaged face preference could be expected to drop, at some point crossing whatever significance boundary is imposed by the chosen tests.

Discussion

Experiments 1.1 and 1.2 show that the BCM produces salience preferences which match Johnson *et al.*'s (1991) findings of 'face preference' in newborns. Experiment 1.3 illustrates a potential role for the development of vergence in the decline of NFP. However, the current version of the BCM described here does not perfectly reproduce the results of Experiment 3 in Johnson *et al.* (1991). In particular, the Scrambled stimulus causes widespread binocular correlations across many vergence angles, and is consistently rated as the most salient stimulus by the BCM. This suggests a need to revise the model.

A wide spread of activity in the salience map could dampen local responses via *divisive normalization*, increasingly thought to be a canonical neural computation (Carandini & Heeger, 2011) and known to operate monocularly in visual cortex (Carandini *et al.*,

1997; Busse, Wade & Carandini, 2009) and interocularly (Moradi & Heeger, 2009). Divisive normalization can be summarized as the division of local activity level by the sum of global activity level over some relatively large region. A wide spread of activity will increase the total energy reflected in the divisor, and therefore the absolute value of local activity will be reduced relative to the case where activity is highly concentrated in a small area. In the case of orienting, it may be that the absolute value of the strongest point of salience is an important determining factor, rather than just the total amount of salience over a wide region corresponding to the whole stimulus.

General discussion

We have shown how and why binocular integration could generate an 'innate face preference'. We make no claims as to causal exclusivity, indeed we are actively researching other potential contributors to visual-facial attention in developing newborns. See for example Wilkinson, Metta and Gredebäck (2011). Overall, we argue that there are five main benefits which yield the BCM preferable to the Conspec model. Firstly, the BCM is specified to a reasonable level of detail and can therefore generate specific predictions which could falsify it, and/ or lead to incremental improvements.

The BCM predicts that the configural preferences of *individual* newborn subjects will correlate *in real time* with the extent to which a stimulus causes synchronous activity at corresponding points on both retinae, subject to prefiltering of the monocular signals through the LSM. This may require some refinement. For example, normalization and thresholding effects (Carandini *et al.*, 1997; Carandini & Heeger, 2011) might imply that a few high amplitude spatial correlations will be more salient than a broad spread of spatial correlations. This question can be addressed experimentally. Also, the static LSM may be insufficient to deal with movement. We are extending the model to include normalization and temporal filtering in ongoing work.

This prediction could be tested by recruiting modern screen-based eye-tracking equipment to newborn experiments. Given knowledge of an individual subject's binocular visual perspectives, the salience predictions of the BCM can be compared to the subject's behaviour, potentially in real time. It is more tricky to make classical preference predictions that one stimulus will be preferred to another on average over numerous subjects, because a stimulus impacts differently upon the senses of each individual depending on her visual allometry and alignment, making eye tracking essential.

One potentially interesting stimulus comparison would be a Config style stimulus with the 'eye blobs' distributed around 4–6 cm apart (i.e. close to newborn IPD) versus the same stimulus enlarged 3 or 4 times, at 3 or 4 times the distance, such that monocular retinal image projections are preserved, but binocular relations change. Barring some observable compensatory ocular vergence gymnastics, the BCM predicts that the face-sized stimulus will be more salient, while all existing explanations would predict no distinction, as they are based on the monocular projection. A positive result would concurrently support a role for proto-binocular vision in the perception of 'size constancy' (Slater, Mattock & Brown, 1990). The practical and ethical difficulty of accessing face naïve subjects is a secondary problem. We anticipate that binocular correlation effects will still be strong enough to robustly identify in days-old babies, though one would expect to see further components arising increasingly with postnatal experience.

The definition of Conspec has suffered from semantic diffusion over time, explicitly becoming synomymous with the empirical trends which motivated it (Johnson *et al.*, 2008; Farroni, Johnson, Menon, Zulian, Faraguna & Csibra, 2005; Johnson, 2005). As Farroni *et al.* (2005) explain in their introduction, 'Conspec was initially defined as being a mechanism that "contains structural information concerning the visual characteristics possessed by conspecifics", but it has become more generally used to refer to an infant's disposition to direct their attention and sense of belonging to other human beings'. To the extent that Conspec has come to mean 'newborn face preference', the BCM may be viewed as a well-specified and plausible implementation of the Conspec idea. However, the BCM is quite specifically *not* based on an innate internal representation of facial structure; it is based on facial structure itself. It relies on generic binocular circuitry, not a specialist module. Neither is it necessary to invoke specific selective pressures 'for' NFP to explain its evolution; the facial similarity of conspecifics is largely guaranteed by the axiom of heredity, rather than optimized via selective processes. Thus in contrast to the adaptationist perspective dominating the literature (Slater & Kirby, 1998; Spelke, 1998; Johnson, 2005; Pascalis & Kelly, 2009), NFP might even emerge as an evolutionary 'exaptation' (S. Gould & Lewontin, 1979; S. J. Gould, 1997).

Secondly, the BCM generates the same qualitative pattern of newborn results reported in Johnson *et al.* (1991). The Conspec model can be assumed to predict results in conflict with the empirical data in one crucial case. The schematic Face stimulus was preferred to controls, but the subjects paid no special attention to the

three blob Config stimulus (see Figure 4). Neither were any inversion effects observed. These results suggest that newborns orient preferentially to face-like stimuli, but that this is *not* because they are using a three blob template. Misreporting of these results in the secondary literature may have contributed to the widespread impression that they provided evidence supporting the Conspec model. Johnson (2005) reported that Config was preferred to Inverse and Linear (see text box to Figure 2 of that review). Johnson (2005), Johnson *et al.* (2008) and Farroni *et al.* (2005, p. 768–769, 335 and 17245 rsp) all state that 'the early experiments with newborns indicated that a stimulus with three high-contrast blobs corresponding to the approximate location of the eyes and mouth might be sufficient to elicit the newborn preference'. However, Johnson *et al.* (1991, p. 10), concluded that 'our results suggest that a face like configuration of elements alone is insufficient'.

Thirdly, the BCM is biologically plausible and maps to known physiology, enabling the integration of NFP with fast growing wider knowledge of human perceptual and social development. Some form of binocular correlation underlies the normative function in all developmental models of stereopsis (see for example Franz & Triesch (2007); Solgi & Weng (2009); Wang & Shi (2010); Hunt, Dayan, & Goodhill (2013)). Mutual gain control is the current best candidate for binocular integration in adults (Ding & Sperling, 2006; Meese, Georgeson & Baker, 2006). The visual thalamus provides ideal conditions for the required gain control mechanisms (Saalmann & Kastner, 2009) and plausible links to the ongoing, intertwined development of perception, action and social skills (Saalmann & Kastner, 2011).

Fourthly, the assumptions of the BCM are more parsimonious and plausible than those of Conspec. The BCM requires the assumption that newborn human binocular integration is similar to that of human adults and of newborn primates. Conspec requires the assumption of a strong and continuous selective pressure 'for' NFP, universal fixation of a specific genetic adaptation to this pressure, and the ability of genes to code for detailed neural connectivity. Conspec is proposed as an evolutionarily ancient vertebrate adaptation (Johnson, 2005 2006; Rosa Salva *et al.*, 2011). This assumes the preservation of the *phenotypic outcome* of this particular set of alleles despite large changes in the genotype, up to and including speciation. Even low levels of genetic hysteresis could cause a dependence on a genetically encoded Conspec to become a problematic evolutionary constraint, as potentially advantageous mutations could be rendered disadvantageous by collaterally destabilizing the epigenetic pathway from Conspec genotype to Conspec phenotype. The apparent simplicity of Conspec

unravels into a deeply complicated story when these necessary assumptions are taken seriously.

Finally, the BCM is integrative with respect to existing hypotheses and conceptual dichotomies. In the terminology of the NFP debate, the BCM is both a 'sensory' and a 'structural' hypothesis, because it is based on sensory structure. It is both 'domain general' and 'face specific'. On one hand, functionality emerges from generic, local neural transforms which have no intrinsic relation to faces or any other subjective category of stimulus. On the other hand, the embodiment of the eyes on the face lends a certain face specificity to the visual system as a whole, which may be characterized to a first approximation by IPD. Indeed, the BCM is deeply face specific, as it depends fundamentally on the shape of the infant's face, whereas Conspec is ultimately template specific, as it depends upon the specification of a three blob representation. The BCM is both 'innate' and 'developing'. The basic ingredients are present at birth and require no visual experience of faces. However, the BCM is based on observable, ontogenic properties of the phenotype, which are subject to developmental emergence and change. Naturally the prenatal development of faces and binocular circuitry is canalized by genetic factors and has evolved over phylogenic timescales under any number of selective pressures, but reference to phylogeny, natural selection and DNA is unnecessary to the mechanistic explication of the BCM.

If it is accepted that having a face can qualify as having information about faces, then Morton and Johnson's (1991) originally controversial 'structural hypothesis' – that infants are born with some information about facial structure – is observably true. Further, it does not at all contradict the 'sensory hypothesis' that newborn preferences can be explained by the match between the stimulus and general properties of the infant's sensory channels.

Limitations and scope

We have presented a theoretical argument based on existing data, with a proof-of-concept computational model embodied on the iCub humanoid. Whether newborn humans do indeed manifest robust preferences for binocular correlations can only be addressed empirically. Although the computational mechanisms employed here remain conjectural and are much simplified approximations to the complexities of human vision, the facilitating affordance exists independently of the details of the implementation, and so our findings do relate directly to the newborn case via the modelling of the embodiment and situation of the newborn visual system. Our main point, which we hope has been made convincingly, is that structural information about faces is

directly available to newborns via binocular integration, and by implication, that 'knowledge' does not necessarily correspond to internal duplication. During the review and editorial process, concerns were expressed that the BCM has a limited scope due to our focus on the <1 hour postnatal population, and is incompatible with the extensive observations of inversion effects (Valenza *et al.*, 1996) and 'top-heaviness' effects (Macchi Cassia, Turati & Simion, 2004). Therefore some clarification on these issues is worthwhile.

The BCM is not at all intended as operating only in the first hours of life. Our proposal is that binocular correlation will be a continuous influence on infant visual behaviour, which develops into binocular cross-correlation (i.e. disparity processing) in its mature form (Filippini & Banks, 2009; Kara & Boyd, 2009). Whether the *functional relationship* between binocular correlation and face detection continues to exist in stereoptic vision is beyond the scope of this paper. This might be a neonatal ontogenic exaptation only. However, there are some reasons to suspect a longer term functional relationship, which we are exploring in ongoing work. Either way, the downstream effects of early attention to faces on the developmental trajectory of an infant and her social context will last a lifetime (Johnson & Morton, 1991; Johnson, 2003).

We do not dismiss inversion/top-heavy effects, but we do feel that they can be approached incrementally. The basic BCM is isotropic in the vertical dimension and so cannot in itself explain inversion and vertical asymmetry effects, but it is entirely compatible with extensions which could. For example, adding a simple bias favouring salience in the upper visual field might help to explain

some results. If there were findings in 1–3-day-old infants which were *inconsistent* with the BCM, this would be a problem for the model. However, inversion/top-heavy effects are not inconsistent with the BCM, and there is certainly no reason to assume that all face effects observed in neonates result from the same monolithic mechanism.

With respect to the relevance of the BCM to other neonate preference findings in days-old babies, a few points are worth mentioning. Preferences for horizontally oriented patterns over vertical ones (Slater & Sykes, 1977; Farroni, Valenza, Simion & Umilta 2000) may reflect the tendency of horizontal patterns to generate more binocular correlation than vertical ones (see Figure 6, stimuli reprinted from Farroni *et al.* (2000)). Perception of 'size constancy' in newborns (Slater *et al.*, 1990) might rely on binocular correlation information, as the physical size of an object will influence binocular correlation patterns across depth (see Figure 2).

The contrast polarity specifism of face effects which has been observed (Farroni *et al.*, 2005) can be accommodated simply by discarding negatively signed contrast at the monocular filtering stage (we implemented this in the model described here). At a mechanistic level, this could conceivably result from a predominance of binocular connectivity between ganglionic 'off-cell' pathways in comparison to 'on-cell' pathways (Bloomfield & Miller, 1986), though we are not aware of any direct evidence for the existence of such neural organization. Alternatively, this result might reflect the subjects' 3 days of experience with faces prior to testing. Finally, face specific effects which have been observed in eye patched infants in the temporal visual hemisphere (Simion, Valenza, Umilta & Barba, 1998) clearly cannot be explained by binocular
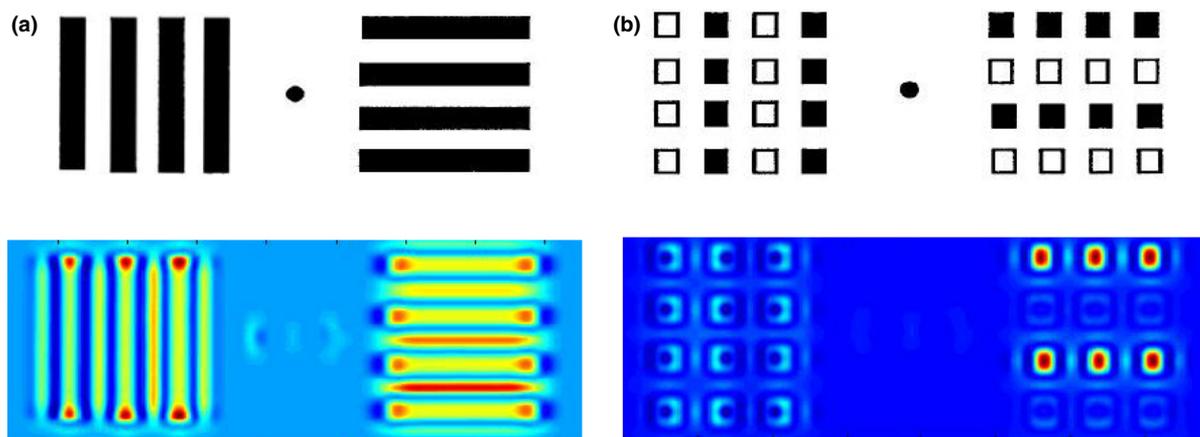
Figure 6   *Horizontal gratings (a) and patterns of contrast (b) will tend to generate more binocular correlation than vertical ones. Individual and inter-trial differences in IPD, visual alignment and contrast sensitivity will all affect individual levels of correlation in real time. For example, if the binocular shift equals the frequency of vertical gratings, then those gratings will cause a lot of binocular correlation. Note these images were generated on a computer, they are not from the robotic model. Stimuli images reprinted with permission from Pion Ltd, London www.pion.co.uk Copyright 2000.*

correlations, but they could be based on days of postnatal visual experience with faces. The absence of face specific effects in the nasal field of view could be down to the absence of binocular information.

Future development

Ongoing development is extending the model to broader 'like me' perception via spatiotemporal resonance and intersensory correlation in wider modalities, behaviour, simple learning and dynamical scenarios. We are extending the static LSM to include spatiotemporal filtering and predictive coding (Rao & Ballard, 1999; Bastos, Usrey, Adams, Mangun, Fries & Friston, 2012) using 'adaptive oscillators' (Buchli, Righetti & Ijspeert, 2008), and to local lateral connectivity (Stettler, Das, Bennett, & Gilbert, 2002). Weak local coupling in oscillator networks can give rise to complex dynamics which have been employed for visual processing (e.g. Chen & Wang 2002). We have previously found the stimulus-induced formation of spiral wave dynamics in oscillator networks to be useful for face detection (Wilkinson *et al.*, 2011). There is extensive evidence that LGN supports oscillatory activity (Huguenard & McCormick, 2007). Lateral connectivity might also help to explain newborn visual preferences for curves, shape and more complex contextual factors (Fantz & Miranda, 1975; Slater, Von der Schulenburg, Brown, Badenoch, Butterworth, Parsons & Samuels, 1998). Our long-term aim is biologically plausible modelling of early social and perceptuomotor development, and exploration of the potential downstream consequences of pathological perturbations to normal development at this crucial formative stage.

Conclusions

Overall, the major contribution of this work is to show that an inborn face sensitivity in human newborns can be accounted for by a simple extension of the standard linear systems model of newborn pattern vision to the binocular case. The newborn visual system is intrinsically well tuned to faces, because its front end sensors are mounted on the infant's face. A tuning bias towards conspecifics is by definition embedded in the sampling regime implied by sensory morphology. Our proposal here is that binocular correlations attract newborn visual attention, and that a certain level of 'face preference' will fall out of this more general effect, but we do not contend that the current version of the BCM can account for newborn face perception in its entirety. Other mechanisms may well be contributing to the multifaceted behavioural observations collectively termed 'newborn face preference'. The scientific method can only discover what these mechanisms are by proposing well-specified, falsifiable theories and then testing their predictions. If the BCM is wrong, it will not survive the testing of its predictions. If it does survive such tests, we will have a powerful new tool to understand newborn visual attention. Our hope is that publication of the current contribution will help to motivate collaborations with labs experienced in newborn research, to enable this testing process. Should such testing corroborate our embodied approach, we further hope that the result will be a wider consideration of the whole organism, not just the brain (or the genes), in the causal structure of 'instinctive' social behaviour.

References

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience*, **28**, 403–450.

Atkinson, J., Braddick, O., & French, J. (1979). Contrast sensitivity of the human neonate measured by the visual evoked potential. *Investigative Ophthalmology and Visual Science*, **18** (2), 210–213.

Bahrick, L.E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, **36** (2), 190–201.

Balas, B. (2010). Using innate visual biases to guide face learning in natural scenes: a computational investigation. *Developmental Science*, **13** (3), 469–478.

Banks, M., & Crowell, J. (1993). Front-end limitations to infant spatial vision: examination of two analyses. In K. Simons (Ed.), *Early visual development: Normal and abnormal*, (pp. 91–116) Oxford: Oxford University Press.

Banks, M., & Ginsburg, A. (1985). Infant visual preferences: a review and new theoretical treatment. In H.W. Reese (Ed.), *Advances in child development and behavior* (Vol. **19**, pp. 207–246). Greenwich, CT: JAI.

Banks, M., & Salapatek, P. (1976). Contrast sensitivity function of the infant visual system. *Vision Research*, **16** (8), 867–869.

Banks, M., & Salapatek, P. (1981). Infant pattern vision: a new approach based on the contrast sensitivity function. *Journal of Experimental Child Psychology*, **31** (1), 1–45.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K.J. (2012). Canonical microcircuits for predictive coding. *Neuron*, **76** (4), 695–711.

Bauck, J. L. (2003, 14 October). Loudspeaker array for enlarged sweet spot. Google Patents. (US Patent 6,633,648)

Bednar, J., & Miikkulainen, R. (2003). Learning innate face preferences. *Neural Computation*, **15**, 1525–1557.

Birch, P., Mitra, B., Bangalore, N. M., Rehman, S., Young, R., & Chatwin, C. (2010). Approximate bandpass and frequency response models of the difference of gaussian filter. *Optics Communications*, **283** (24), 4942–4948.

Bloomfield, S. A., & Miller, R. F. (1986). A functional organization of on and off pathways in the rabbit retina. *Journal of Neuroscience*, **6** (1), 1–13.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, **47** (1), 139–159.

Brown, A. M., & Miracle, J. A. (2003). Early binocular vision in human infants: limitations on the generality of the superposition hypothesis. *Vision Research*, **43** (14), 1563–1574.

Bruce, V., & Young, A. (2012). *Face perception*. London: Psychology Press.

Buchli, J., Righetti, L., & Ijspeert, A. (2008). Frequency analysis with coupled nonlinear oscillators. *Physica D: Nonlinear Phenomena*, **237** (13), 1705–1718.

Busse, L., Wade, A. R., & Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron*, **64** (6), 931–942.

Carandini, M., & Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, **13** (1), 51–62.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the primary visual cortex. *Journal of Neuroscience*, **17** (21), 8621–8644.

Chen, K., & Wang, D. (2002). A dynamically coupled neural oscillator network for image segmentation. *Neural Networks*, **15** (3), 423–439.

Chino, Y., Smith III, E., Hatta, S., & Cheng, H. (1997). Postnatal development of binocular disparity sensitivity in neurons of the primate visual cortex. *Journal of Neuroscience*, **17** (1), 296–307.

Clark, W. L. G. (1932). The structure and connections of the thalamus. *Brain*, **55** (3), 406–470.

Crick, F. (1984). Function of the thalamic reticular complex: the searchlight hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, **81** (14), 4586–4590.

Ding, J., Klein, S. A., & Levi, D. M. (2013). Binocular combination in abnormal binocular vision. *Journal of Vision*, **13** (2), 14.

Ding, J., & Sperling, G. (2006). A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America*, **103** (4), 1141–1146.

Fantz, R. L., & Miranda, S. B. (1975). Newborn infant attention to form of contour. *Child Development*, **46**, 224–228.

Farroni, T., Johnson, M., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). New borns' preference for face-relevant stimuli: effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (47), 17245–17250.

Farroni, T., Valenza, E., Simion, F., & Umilta, C. (2000). Configural processing at birth: evidence for perceptual organisation. *Perception*, **29** (3), 355–372.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, **4**.

Filippini, H. R., & Banks, M. (2009). Limits of stereopsis explained by local cross-correlation. *Journal of Vision*, **9** (1), 8.

Franz, A., & Triesch, J. (2007). Emergence of disparity tuning during the development of vergence eye movements. In *Development and Learning, 2007. ICDL 2007. IEEE 6th international conference on* (pp. 31–36).

Friston, K. (2012). Predictive coding, precision and synchrony. *Cognitive Neuroscience*, **3** (3–4), 238–239.

Fuller, P., Sherman, D., Pedersen, N. P., Saper, C. B., & Lu, J. (2011). Reassessment of the structural basis of the ascending arousal system. *Journal of Comparative Neurology*, **519** (5), 933–956.

Furst, M., Bresloff, I., Levine, R. A., Merlob, P. L., & Attias, J. J. (2004). Interaural time coincidence detectors are present at birth: evidence from binaural interaction. *Hearing Research*, **187** (1), 63–72.

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology*, **11** (2), 219–224.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, **2** (6), 568–573.

Gayl, I., Roberts, J., & Werner, J. (1983). Linear systems analysis of infant visual pattern preferences. *Journal of Experimental Child Psychology*, **35** (1), 30–45.

Gibson, J. (1961). Ecological optics. *Vision Research*, **1** (34), 253–262.

Gibson, J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

Gibson, J. (1977). The concept of affordances. In R. Shaw & J. Bransford(Eds.), *Perceiving, acting, and knowing* (pp. 67–82). New York: Wiley.

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

Gottlieb, G., & Kuo, Z. (1965). Development of behavior in the duck embryo. *Journal of Comparative and Physiological Psychology*, **59** (2), 183–188.

Gotts, S., Chow, C., & Martin, A. (2012). Repetition priming and repetition suppression: a case for enhanced efficiency through neural synchronization. *Cognitive Neuroscience*, **3** (3–4), 227–237.

Gould, S., & Lewontin, R. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, **205** (1161), 581–598.

Gould, S. J. (1997). The exaptive excellence of spandrels as a term and prototype. *Proceedings of the National Academy of*

Sciences of the United States of America, **94** (20), 10750–10755.

Hafed, Z. M., Goffart, L., & Krauzlis, R. J. (2009). A neural mechanism for microsaccade generation in the primate superior colliculus. *Science*, **323** (5916), 940–943.

Hainline, L., Turkel, J., Abramov, I., Lemerise, E., & Harris, C. (1984). Characteristics of saccades in human infants. *Vision Research*, **24** (12), 1771–1780.

Hansson, T., & Brismar, T. (1999). Tactile stimulation of the hand causes bilateral cortical activation: a functional magnetic resonance study in humans. *Neuroscience Letters*, **271** (1), 29–32.

Helmholtz, H. (1878). The facts of perception (originally published 1878). In R. M. Warren & R. P. Warren, *Helmhotz on perception: Its physiology and development*. (pp. 205–231). New York: Wiley (1968).

Horton, J. C., & Hocking, D. R. (1996). An adult-like pattern of ocular dominance columns in striate cortex of newborn monkeys prior to visual experience. *Journal of Neuroscience*, **16** (5), 1791–1807.

Horwood, A. (2003). Too much or too little: neonatal ocular misalignment frequency can predict later abnormality. *British Journal of Ophthalmology*, **87** (9), 1142–1145.

Huguenard, J. R., & McCormick, D. A. (2007). Thalamic synchrony and dynamic regulation of global forebrain oscillations. *Trends in Neurosciences*, **30** (7), 350–356.

Hunt, J. J., Dayan, P., & Goodhill, G. J. (2013). Sparse coding can predict primary visual cortex receptive field changes induced by abnormal visual input. *PLoS Computational Biology*, **9** (5), e1003005.

Jiang, Z. D., & Tierney, T. S. (1996). Binaural interaction in human neonatal auditory brainstem. *Pediatric Research*, **39** (4), 708–714.

Johnson, M. (2003). Functional brain development in infants: Elements of an interactive specialization framework. *Child Development*, **71** (1), 75–81.

Johnson, M. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, **6** (10), 766–774.

Johnson, M. (2006). Biological motion: a perceptual life detector? *Current Biology* **16** (10), R376–R377.

Johnson, M. (2010). *Developmental cognitive neuroscience*. Oxford: Wiley.

Johnson, M., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of faces and its subsequent decline. *Cognition*, **40** (1–2), 1–19.

Johnson, M., Grossmann, T., & Farroni, T. (2008). The social cognitive neuroscience of infancy: illuminating the early development of social brain functions. *Advances in Child Development and Behavior*, **36**, 331–372.

Johnson, M., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford: Blackwell.

Kaas, J., Guillery, R., & Allman, J. (1972). Some principles of organization in the dorsal lateral geniculate nucleus. *Brain Behavior and Evolution* **6**, 268–282.

Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, **17** (11), 4302–4311.

Kara, P., & Boyd, J. D. (2009). A micro-architecture for binocular disparity and ocular dominance in visual cortex. *Nature*, **458** (7238), 627–631.

Katzner, S., Busse, L., & Carandini, M. (2011). Gabaa inhibition controls response gain in visual cortex. *Journal of Neuroscience*, **31** (16), 5931–5941.

Kavšek, M. (2013). Infants' responsiveness to rivalrous gratings. *Vision Research*, **76**, 50–59.

Kleiner, K. A., & Banks, M. (1987). Stimulus energy does not account for 2-month-olds' face preferences. *Journal of Experimental Psychology: Human Perception and Performance*, **13** (4), 594–600.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparitive studies on the external morphology of the primate eye. *Journal of Human Evolution*, **40** (5), 419–435.

Lehrman, D. (1953). A critique of Konrad Lorenz's theory of instinctive behavior. *Quarterly Review of Biology*, **28** (4), 337–363.

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: new borns match nonhuman primate faces and voices. *Infancy*, **15** (1), 46–60.

Lickliter, R., & Berry, T. (1990). The phylogeny fallacy: developmental psychology's misapplication of evolutionary theory. *Developmental Review*, **10** (4), 348–364.

Lien, A. D., & Scanziani, M. (2013). Tuned thalamic excitation is amplified by visual cortical circuits. *Nature Neuroscience*, **16**, 1315–1323.

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

Macchi Cassia, V., Valenza, E., Simion, F., & Leo, I. (2008). Congruency as a nonspecific perceptual property contributing to newborns' face preference. *Child Development* **79** (4), 807–820.

Mante, V., & Carandini, M. (2005). Mapping of stimulus energy in primary visual cortex. *Journal of Neurophysiology*, **94** (1), 788–798.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience* **27**, 1–28.

McGaugh, J. L. (2013). Making lasting memories: remembering the significant. *Proceedings of the National Academy of Sciences of the United States of America* **110** (Supplement 2), 10402–10407.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences,* **11** (1), 8–15.

Meese, T. S., & Baker, D. H. (2011). Contrast summation across eyes and space is revealed along the entire dipper function by a 'Swiss cheese' stimulus. *Journal of Vision*, **11** (1), 23.

Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision*, **6** (11), 7.

Meltzoff, A.N. (2007). Like me: a foundation for social cognition. *Developmental Science*, **10** (1), 126–134.

Meltzoff, A. N., & Brooks, R. (2001). "Like" me as a building block for understanding other minds: bodily acts, attention, and intention. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 171–191). Cambridge, MA: MIT Press.

Meltzoff, A. N., & Moore, M. K. (1993). *Why faces are special to infants on connecting the attraction of faces and infants ability for imitation and cross-modal processing*. New York: Springer.

Mendelson, M., Haith, M., & Gibson, J. (1976). The relation between audition and vision in the human newborn. Monographs of the Society for Research in Child Development, **41** (4, Serial No. 167).

Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., von Hofsten, C.., Rosander, K., Lopes, M., Santos-Victor, J., Bernardino, A., & Montesano, L. (2010). The icubhumanoid robot: an open-systems platform for research in cognitive development. *Neural Networks*, **23**(89), 1125–1134.

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face perception during early infancy. *Psychological Science*, **10** (5), 419–422.

Moradi, F., & Heeger, D.J. (2009). Inter-ocular contrast normalization in human visual cortex. *Journal of Vision*, **9** (3), 13.

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., & Friston, K. J. (2013). Free energy, precision and learning: role of cholinergic neuromodulation. *Journal of Neuroscience*, **33** (19), 8227–8236.

Morrone, M. C., & Burr, D. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London Series B, Biological Sciences*, **235**, 221–245.

Morse, A.F., Herrera, C., Clowes, R., Montebelli, A., & Ziemke, T. (2011). The role of robotic modelling in cognitive science. *New Ideas in Psychology*, **29** (3), 312–324.

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

Morton, J., Johnson, M., & Maurer, D. (1990). On the reasons for newborns' responses to faces. *Infant Behavior and Development*, **13** (1), 99–103.

Norcia, A., Tyler, C., & Hamer, R. (1990). Development of contrast sensitivity in the human infant. *Vision Research*, **30** (10), 1475–1486.

Olsen, S. R., Bortone, D. S., Adesnik, H., & Scanziani, M. (2012). Gain control by layer six in cortical circuits of vision. *Nature*, **483** (7387), 47–52.

Oyama, S. (2000). *The ontogeny of information: Developmental systems and evolution*. Durham, NC: Duke University Press.

Pascalis, O., & Kelly, D. (2009). The origins of face processing in humans: phylogeny and ontogeny. *Perspectives on Psychological Science*, **4** (2), 200–209.

Pitti, A., Kuniyoshi, Y., Quoy, M., & Gaussier, P. (2013). Modeling the minimal newborn's intersubjective mind: the visuotopic-somatotopic alignment hypothesis in the superior colliculus. *PLoS ONE*, **8** (7), e69474.

Pryor, H. B. (1969). Objective measurement of interpupillary distance. *Pediatrics* **44** (6), 973–977.

Rakic, P. (1976). Prenatal genesis of connections subserving ocular dominance in the rhesus monkey. *Nature*, **261**, 467–471.

Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, **2**, 79–87.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, **61** (2), 168–185.

Rosa Salva, O., Farroni, T., Regolin, L., Vallortigara, G., & Johnson, M. (2011). The evolution of social orienting: evidence from chicks and human newborns. *PLoS ONE*, **6** (4), e18802.

Rothman, J., Cathala, L., Steuber, V., & Silver, R. (2009). Synaptic depression enables neuronal gain control. *Nature*, **457** (7232), 1015–1018.

Saalmann, Y. B., & Kastner, S. (2009). Gain control in the visual thalamus during perception and cognition. *Current Opinion in Neurobiology*, **19** (4), 408–414.

Saalmann, Y. B., & Kastner, S. (2011). Cognitive and perceptual functions of the visual thalamus. *Neuron*, **71** (2), 209–223.

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.

Sara, S.J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, **10** (3), 211–223.

Sara, S.J., & Bouret, S. (2012). Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron*, **76** (1), 130–141.

Sherman, S.M. (2007). The thalamus is more than just a relay. *Current Opinion in Neurobiology*, **17** (4), 417–422.

Sherman, S.M., & Guillery, R.W. (2001). Exploring the thalamus. Access Online via Elsevier.

Shimojo, S., Bauer, Jr, J.B., O'Connell, K.M., & Held, R. (1986). Pre-stereoptic binocular vision in infants. *Vision Research*, **26** (3), 501–510.

Sillito, A., Jones, H., Gerstein, G., & West, D. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, **369** (6480), 479–82.

Simion, F., Macchi Cassia, V., Turati, C., & Valenza, E. (2001). The origins of face perception: specific versus non-specific mechanisms. *Infant and Child Development*, **10** (1-2), 59–65.

Simion, F., Valenza, E., Macchi Cassia, V., Turati, C., & Umilta, C. (2002). Newborns' preference for up–down asymmetrical configurations. *Developmental Science*, **5** (4), 427–434.

Simion, F., Valenza, E., Umilta, C., & Barba, B. (1998). Preferential orienting to faces in newborns: a temporal.nasal asymmetry. *Journal of Experimental Psychology: Human Perception and Performance*, **24** (5), 1399–1405.

Slater, A., & Kirby, R. (1998). Innate and learned perceptual abilities in the newborn infant. *Experimental Brain Research*, **123** (1), 90–94.

Slater, A., Mattock, A., & Brown, E. (1990). Size constancy at birth: newborn infants' responses to retinal and real size. *Journal of Experimental Child Psychology*, **49** (2), 314–322.

Slater, A., Quinn, P. C., Lewkowicz, D. J., Hayes, R., & Brookes, H.(2003). Learning of arbitrary adult voice-face pairings at three months of age. In O. Pascalis, A. Slater, O. Pascalis, & A. Slater (Eds.), *The development of face processing in infancy and early childhood: Current perspectives* (pp. 69–78). Hauppauge, NY: Nova Science Publishers

Slater, A., & Sykes, M. (1977). Newborn infants' visual responses to square wave gratings. *Child Development*, **48** (2), 545-554.

Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development*, **21** (2), 345–354.

Solgi, M., & Weng, J. (2009). Developmental stereo: emergence of disparity preference in models of the visual cortex. *Autonomous Mental Development, IEEE Transactions, on* **1** (4), 238–252.

Spelke, E. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, **21** (2), 181–200.

Spencer, J., Blumberg, M., McMurray, B., Robinson, S., Samuelson, L., & Tomblin, J. (2009). Short arms and talking eggs: why we should no longer abide the nativist.empiricist debate. *Child Development Perspectives*, **3** (2), 79–87.

Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, **36** (4), 739–750.

Streri, A., Coulon, M., & Guellaï, B. (2013). The foundations of social cognition: studies on face/voice integration in newborn infants. *International Journal of Behavioral Development*, **37** (2), 79–83.

Sugita, Y. (2009). Innate face processing. *Current Opinion in Neurobiology*, **19** (1), 39–44.

Swanson, L. W., Cowan, W.M., & Jones, E. G. (1974). An autoradiographic study of the efferent connections of the ventral lateral geniculate nucleus in the albino rat and the cat. *Journal of Comparative Neurology*, **156** (2), 143–163.

Theile, G. (2000). Multichannel natural recording based on psychoacoustic principles. Paper presented at the Audio engineering society convention 108.

Thorn, F., Gwiazda, J., Cruz, A. A., Bauer, J. A., & Held, R. (1994). The deveopment of eye alignment, convergence, and sensory binocularity in young infants. *Investigative Ophthalmology and Visual Science*, **35** (2), 544–553.

Towal, R., Quist, B., Gopal, V., Solomon, J., & Hartmann, M. (2011). The morphology of the rat vibrissal array: a model for quantifying spatiotemporal patterns of whisker-object contact. *PLoS Computational Biology*, **7** (4), e1001120

Turati, C. (2004). Why faces are not special to newborns. *Current Directions in Psychological Science*, **13** (1), 5–8.

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

Varela, F. J., & Singer, W. (1987). Neuronal dynamics in the visual corticothalamic pathway revealed through binocular rivalry. *Experimental Brain Research*, **66**, 10–20.

Varela, F. J., Thompson, E. T., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. The MIT Press.

VéLlez-Fort, M., & Margrie, T. W. (2012). Cortical circuits: layer 6 is a gain changer. *Current Biology*, **22** (10), R411–R413.

Voegtline, K. M., Costigan, K. A., Pater, H. A., & DiPietro, J. A. (2013). Near-term fetal response to maternal spoken voice. *Infant Behavior and Development*, **36** (4), 526–533.

Wang, Y., & Shi, B. E. (2010). Autonomous development of vergence control driven by disparity energy neuron populations. *Neural Computation*, **22** (3), 730–751.

Wilkinson, N., Metta, G., & Gredebäck, G. (2011). Modelling the face-to-face effect: sensory population dynamics and active vision can contribute to perception of social context. *IEEE International Conference on Development and Learning (ICDL), Frankfurt am Main*, 1–6.

Xiong, X. R., Liang, F., Li, H., Mesik, L., Zhang, K. K., Polley, D. B., Tao, H.W., Xiao, Z., & Zhang, L. I. (2013). Interaural level difference-dependent gain control and synaptic scaling underlying binaural computation. *Neuron*, **79** (4), 738–753.

Yang, D., Atick, J., & Reid, R. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *Journal of Neuroscience*, **16** (10), 3351–3362.

Zuiderbaan, W., Harvey, B. M., & Dumoulin, S. O. (2012). Modeling center surround configurations in population receptive fields using MRI. *Journal of Vision*, **12** (3), 10.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

A. Formal description.
B. Eye movements.
C. Defining salience.