Word Counts:
Abstract: 59
Main Text: 6992
References: 1158

Entire Text: 8286

Title: Beyond Neonatal Imitation: Aerodigestive Stereotypies, Speech Development and Social Interaction in the Extended Perinatal Period

Authors: Nazim Keven and Kathleen A. Akins

Author Note
Nazim Keven, Department of Philosophy, Bilkent University.
Address: Department of Philosophy, Bilkent University, Cankaya/Ankara 06800 Turkey.
Email: nazimkeven@bilkent.edu.tr

Kathleen Akins, Department of Philosophy, Simon Fraser University.
Address: Simon Fraser University, 4604 Diamond Building, Burnaby, B.C. Canada V5A 1S6.
Email: kathleea@sfu.ca

Abstract: In our target article, we argued that the positive results of neonatal imitation are likely to be by-products of normal aerodigestive development. Our hypothesis elicited various responses on the role of social interaction in infancy, the methodological issues about imitation experiments and the relation between the aerodigestive theory and the development of speech. Here we respond to the commentaries.

R1.0 Introduction
We would first like to thank all of the commentators for their insightful replies and the time spent to formulate them. As we looked through the commentaries, most of the topics raised fell into three (often overlapping) categories: The role of social interaction in the development of imitation, both in human and Old World primates more generally; the correct methodological constraints on past imitation experiments and on our own aerodigestive theory and; the relation between the aerodigestive theory and the development of speech. In writing our response, we first focus on the origins of speech because it was central to many commentaries and while it was the least developed subject in our article, this topic, the evolutionary and developmental origins of speech, best highlights how a detailed description of early mechanisms of respiration and ingestion can fit with other aspects of development such as the role of social interaction. We will thus start with with the interaction between aerodigestive and speech development (Sections 2.0-2.4) and then use this first section on speech to illustrate and bolster our responses to the two other general subjects of criticisms: methodological issues (Sections 3.0-3.5) and social interactivity (Sections 4.0-4.2).

R2.0 The Origins of Speech
Many of the commentators (**Mayer et al., Choi et al., Buck, Meltzoff, Murray et al.**) asked about the origins of speech and its relation to aerodigestive function. The aerodigestive theory has clear implications for the development of speech. On our theory, the neonate does not arrive into the world with a set of innate, multimodal, cortical representations. Instead, neonatal behaviour begins with subcortical oscillators for repetitive behaviours, orofacial and otherwise, behaviours initiated and driven by arousal mechanisms. By birth, the stereotypies of aerodigestion have been woven together into the first sensorimotor sequences of the human body. These motor runs—e.g. breathing, swallowing, peristaltic motions of the tongue and esophagus—are themselves periodic events, controlled by networks of oscillators and tempered, even in utero, by multimodal sensory feedback. There are no cortical motor commands prior to birth, no feed forward predictive encodings that await confirmation or error signals. Nor are the patterns of somatosensory feedback that occur as a natural result of these oscillatory motor sequences represented *qua* the predictive 'results' of a given motor 'command'. Instead, as motor learning progresses, ever more complex, multimodal feedback is integrated into these oscillatory networks, a process that yields systems that are responsive in real time to the vicissitudes of a dual system for eating and breathing. Somehow, from this unlikely starting point, speech begins—and our commentators have rightly expressed curiosity about this bootstrapping process or outright doubt that this is possible.

R2.1 What are the evolutionary origins of human speech?

It seems clear that, in the target article, the authors have stumbled into a robust discourse about the evolutionary origins of speech. **Mayer and colleagues** provide comprehensive analyses with evidence that "there are at least some core speech movements which are direct ontogenetic adaptations of pre-existing digestive movements". A different, though structurally similar proposal for the ontogenetic adaptation of speech from pre-existing movements can be found in the work of Ghazanfar and colleagues, derived from MacNeilage's view (Ghazanfar et al. 2012; Shepherd et al. 2012, Borjon, Takahashi, Cervantes, & Ghazanfar, 2016; Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009), that speech has evolved from rhythmic facial expressions to which vocalizations have been added. **Murray et al.** emphasize that these mouth movements were long ago coopted into early infant-parent dynamic interaction.

While we are not experts on the evolution of speech, a few things seem clear. Our aerodigestive theory is not meant to be an evolutionary theory of *speech*, but rather is meant to explain why neonatal imitation is unlikely to occur given the facts of aerodigestive development. However, our theory is compatible with both sides of this debate, with the evolution of speech from either fetal aerodigestive behaviours or orofacial behaviours such as yawning, blinking, scowling, smiling, etc. (Dai & Hata, 2006; Kanenishi, Hanaoka, Noguchi, Marumo, & Hata, 2013; Kurjak, Azumendi, Andonotopo, & Salihagic-Kadic, 2007; Sato et al., 2014; Yigiter & Kavak, 2006). In our target article, we concentrated on aerodigestive behaviours on the belief that TP/R fits neatly into the class of aerodigestive behaviours. Apart from the neonatal imitation literature, and unlike lip-smacking among New World Primates, there is little to suggest that TP/R is a universal, affiliative, human behaviour (cf. **Murray et al**.) That said, facial expressions of the fetus and neonate have a developmental trajectory that is parallel to aerodigestive development, i.e. from individual stereotypies to sensorimotor sequences. In a set of experiments, Reissland and colleagues (Reissland, Francis, & Mason, 2012, 2013; Reissland, Francis, Mason, & Lincoln, 2011) showed that from 24 to 36 wGA these individual action units begin to coalesce into 'Gestalts' of emotional expressions, such as a 'happy face' or a 'cry face'. Prior to 24 weeks, single action units predominate facial expressions, whereas by 36 weeks, 85% of action units co-occur with 2 to 4 other units. Just as aerodigestive stereotypies coalesce into motor runs by 36

wGA, coherent facial expressions emerge over the same time period. Moreover, just as suckling and respiration are 'practiced' in utero without, e.g. air to breath, the 'pain face' of a fetus at 36 wGA occurs spontaneously and independently of any (visible) harmful event. So, aerodigestive behaviours and facial expressions share the same developmental trajectory. Thus our theory is consistent with *either or both* evolutionary theories.

.

R2.2 Is the aerodigestive development consistent with early, pre-linguistic behaviour of the infant?   We know that prior to speech acquisition, infants younger than 4 months of age begin to vocalize. Thus according to our view, infants must be able to learn such behaviours prior to gaining cortical control of articulatory structures. **Meltzoff** argues that infants younger than 4 months "produce diverse cooing sounds, which require tongue movements markedly different from suckling and tongue stereotypies." Thus, if "(i)nfants cannot control their tongues prior to 4 months of age" and "all tongue movements are purported to be the stereotypic thrust/retraction involved in suckling", then the aerodigestive theory must be false.

Our claim is that orofacial stereotypies develop via 'practice' prior to birth and are then incorporated into complex, sensorimotor sequences through networks of interacting oscillators. So both *learning* and tongue *control* occur before and after birth, the result of the developing oscillatory networks in the brainstem. Importantly, tongue thrust is a primitive *reflex* not a stereotypy, a difference that makes a difference. Unlike primitive reflexes, infant stereotypies are highly variable, e.g. small or large (just over the lip line or far beyond it) and in all directions (to the right or the left, or straight down the midline). On our view, stereotypies that continue to occur alone, independently of sensorimotor runs, organize cortical motor space through the somatosensory and proprioceptive feedback from the full range of possible movement. Thus TP/R explores the deformation space of tongue movement, across the full range of protrusive tongue movements. This variation provides a bridge between aerodigestive and articulatory behaviours. For example, **Mayer et al.** suggest that tongue-bracing contacts during swallowing are a subset of the tongue-bracing contacts in speech. Stated differently, the motor activation space of tongue-brace during swallow falls primarily within the activation space of tongue-brace for speech. If tongue-brace is a stereotypy—and it is—this transition is not mysterious. The stereotypy has already explored the sensorimotor space of tongue-brace prior to speech learning. Note that we are not claiming Mayer's view is necessarily true. Our claim is that sensorimotor learning occurs subcortically in early human development hence some articulatory movements need not require cortical input to be learned. Which tongue movements require cortical input depends upon the *type* of tongue movement. Ballistic movements such as catching a drip of ice cream as it escapes from the cone probably require cortical input to learn or initiate such directed tongue movements. This is our claim in the target article.

R2.3 How might human speech develop out of neonatal stereotypies?

Neither of the authors is a specialist in language development but the commentators' questions (cf. **Murray et al.** and **Choi et al.**) sparked our interest in the recent literature on speech development and oscillatory entrainment. If perinatal behaviour is a function of subcortical networks of CPGs, and if mature speech also involves repetitive, rhythmic movements, then perhaps interaction between parents and infants form coupled systems of oscillators, i.e. parental speech entrains the pre-linguistic behaviour of infants. Perhaps the best example of this research project is found in the work of Asif Ghazanfar and his colleagues who characterize human speech

as an inherently multimodal capacity that has evolved from facial expressions/stereotypies (i.e. the MacNeilage view) (Ghazanfar et al. 2012; Shepherd et al. 2012, Borjon, Takahashi, Cervantes, & Ghazanfar, 2016; Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). Their experiments suggest that at least three types of pre-linguistic learning is explained in terms of coupled oscillation modulated by speaker/listener arousal, namely cooperative vocal adjustment at a distance, the maturation of turn-taking, and the development of the 'phee' call in infant marmosets. Several of these experimental results are directly applicable to the human case.

First, human infants might learn conversational turn-taking in much the same way as infant marmosets. Ghazanfar and colleagues (2008) reported that infant marmosets, during the first post-natal days, respond to recordings of their own voices and parental cries with the same frequency of response and response time. This suggests the inability of the infant to distinguish between its own voice and the voices of others. By two months of age, infant marmosets attained the adult temporal pattern of vocal turn-taking, with ever decreasing response time to the parental voice. The authors explain this reduction in terms of the vocal entrainment of the infant's call by the parent.   In the human case, the dorsal cochlear nucleus receives somatosensory feedback from the face and parts of the vocal tract, thus a rudimentary mechanism to identify one's own vocalizations. Auditory stimulation with vocal tract/facial proprioceptive feedback is 'my voice; auditory stimulation without feedback equals 'someone else's voice'. Once a human infant differentiates her own voice from another speaker's, auditory entrainment can begin.

Second, self-entrainment might explain the importance of infant cooing. Cooing, the production of vowel-like sounds, involves large facial movements—the pursing of lips ('oo'), a wide-open jaw ('ah'), lip retraction ('ee') and so on. Such expressions are not unknown to the infant. Many of these expressions can be mapped directly onto the stereotypies that comprise the newborn's behavioural repertoire or fall within the variations seen in the first stages of stereotypy acquisition. Fagan (2014) reported that infants deaf from birth make less frequent vocalizations of all kinds. After early cochlear implantation, they resumed a normal pattern of pre-linguistic vocalizing. However, deaf newborns did not resume normal rates of crying and other vocal signals of distress. Fagan concludes that infant cooing/babbling 'are primarily motivated by auditory feedback'. To put this slightly differently, the infant's own vocalizations might create a self-sustaining training-cycle—of vocal production followed by both auditory and somatosensory encodings. As long as arousal by self-vocalization is speech sound specific, i.e. causes the vocalization of multiple 'vowels', learning paired associations between somatosensory (S1) and auditory (A1) encodings will occur. This suggestion may explain why 4 month-old infants cannot judge auditory-visual mismatches when prevented from making articulatory movements. As long as the categorization of *speech sounds* is primarily somatosensory—or if somatosensory categorization is required to entrain speech production—an infant will not be able to classify phonemes or differentiate between them without self-production. Nor would an infant be able to learn a new phoneme without first trying to reproduce it and thus see *what it feels like*.

Third, entrainment may explain the development of audio-visual matching between the voice and lip movement of a speaker with lip, the capacity that underlies the McGurk effect (Chandrasekeran et al. 2011). The 2 month-old infants look longer at matches between facial movements and speech sounds (Patterson & Werker, 2003); at 4 months, infants reverse this correlation, paying more attention to *mis*matches (Kuhl & Meltzoff, 1984). Intuitively, this looks like an initial period of multimodal learning followed by the ability to perceive exceptions based upon learned audiovisual associations. In speech, the motion of the lips and the resultant auditory signal are both rhythmic events with the same periods of oscillation, synchronized in space and time. Thus temporally synchronized auditory and visual signals, matched in periodicity, indicate a common source/speaker of these signals. But from where does this visual information about

dynamic faces come? We know that A1 and it adjacent areas contain a high percentage of multimodal (audiovisual) cells in adult primates. But at birth, the visual system has yet to begin processing dynamic patterns of visual stimuli, a necessary requirement for training up A1 cells that associate facial movements and their respective sounds. In the superior colliculus, the alignment of audiovisual maps is also experience dependent and will not occur until about 4 months of age.

One plausible route for this association is by way of the pulvinar, a division of the thalamus. In adult primates, human and otherwise, dynamic facial expressions are matched to speech through input from visual area MT in the posterior bank of the superior temporal sulcus (STS) (Ghazanfar et al. 2008). If cortical vision developed hierarchically, one would expect area MT to lag behind V1 in maturity. But directional sensitivity arises in V1 and MT *at the same time*, around post-natal 4 – 7 weeks. The key here may lie in the pulvinar (Kaas 2015). Newborn primates have a *transient* pathway from retina to pulvinar and then from pulvinar to visual associative area STS including area MT (Warner et al. 2012). So visual motion information via the pulvinar could be critical to the formation of multimodal, audiovisual cells in auditory cortex. The timing here is highly suggestive. If direction selectivity begins by 4-7 weeks, infants in the Patterson and Werker study (2003) would have had, on average, a few weeks of associative training and thus the basis for preferential looking.

R3.0 Methodological Issues and Assumptions

R3.1 The cross-target methodology: Should we trust it?

Several commentators argue that neonatal imitation can still be a viable theory when a wider array of data is taken into account (**Simpson et al. , Vincini et al., Meltzoff**). This evidence comes largely from experiments designed to test neonatal imitation using a methodology that has changed very little after the first wave improvements. When an experimental methodology continues to produce a tangle of positive and negative results over a 35-year period, this is deeply worrisome. It suggests that the methodology is somehow flawed and that its continued use will not produce definitive results in the future. Witness here the competing results of two recent, careful studies by Oostenbroek et al. (2016) and Simpson, Murray, Paukner, and Ferrari (2014).

In the target article, we attempted to set this issue aside in order to develop a positive theory. But ours is not a theory of how or why neonatal imitation occurs: It is meant to *explain away* the positive results of neonatal imitation experiments, not to vindicate them. As the commentaries demonstrate, we must address some issues of procedure and statistical analysis in the neonatal imitation work. Given the constraints of space we will mention one example that have relation to the aerodigestive mechanisms.

A standard part of the cross-target methodology is the 'burst-pause procedure' which in Meltzoff and Moore (1989) is justified as follows: "In previous work with newborns it was reported that attention and responsivity were maximized if adult gesturing was alternated with an interval in which the adult remained passive… (Meltzoff & Moore, 1983a)". If we look back at Meltzoff and Moore (1983), the authors reported that: "We found in preliminary work that a constant demonstration of the target gesture was not maximally effective in eliciting imitation. Therefore, in our design the experimenter alternated between the presentation of the gesture and a passive face. We are not certain why our burst-pause procedure is the more powerful, but we can suggest three possibilities". The three possibilities then given were that (a) it gives infants several periods of time over which to organize their motor response; (b) the burst-pause paradigm nicely

'bookended' the demonstrated gesture, thereby accentuating what was to be imitated, or; (c) the paradigm mimicked the give and take of conversational turn-taking, thus encouraging infant response. All of these options assume that imitation occurs and that the question about this methodology is *why it increases imitation.*

The correct question to ask is *why* the new methodology changes infant behaviour and whether those factors are related to imitation or to extraneous factors. For example, it could be that in addition to increasing attention, the methodology also increases arousal. Or perhaps the methodology increases negative affect and thus responsiveness. In the Still Face Effect, an adult who stares without facial expression or movement causes anxiety in infants as measured by fussing or crying, increased heart rate, skin conductance and vagal tone (Bertin & Striano, 2006; Ham & Tronick, 2006; Moore & Calkins, 2004; Striano, 2004). This effect begins in infants between 1 and 1.5 months of age (Bertin & Striano, 2006; Bigelow & Power, 2012), just around the age of testing (Meltzoff & Moore, 1992; Meltzoff & Moore, 1994). Yet imitation experiments use what amounts to a Still Face stimulus as a control condition. We now also know that neonates prefer their mother's face to that of a stranger (Bartrip, Morton, & Schonen, 2001) and to prefer a new face over a previously viewed face or a non-communicative face (Cecchini, Baroni, Di Vito, Piccolo, & Lai, 2011). So multiple aspects of the burst-pause paradigm are linked to negative affect. Finally, as we saw from the research on marmoset turn-taking, the third option may well have hit the nail on the head. Six-week old infants may already have learned turn-taking behaviour through interaction with their parents, what amounts to a 'call-and-response' sequence in which behaviour is inhibited during the 'call' and the inhibition is released when the static gesture ends, thus producing *more* gestures thereafter. But this need not involve imitation *per se*, merely a move towards conversational turn-taking or a reflexive 'wait and listen/see' disposition that reduces noise while in the listening phase. Because the authors assumed that the burst-pause paradigm increased *imitation*, they did not investigate *any* hypotheses about *why* "attention and responsivity were maximized" by the new paradigm. In our opinion, it is time to go back and look carefully at the assumptions incorporated into the cross-target paradigm.

R3.2. If your explanation is in terms of arousal, what explains the differential response?

As a number of our commentators have pointed out, the primary empirical question is not whether arousal increases the production of infant stereotypies, i.e. the base rate of production, but whether there is evidence for a differential response—more MO in response to modeled MO not TP, more TP in response to TP not MO (**Vincini et al., Simpson et al., Meltzoff**). As **Simpson et al.** assert, "differential imitation in neonates is incompatible with aerodigestive or arousal driven explanations". Moreover, the search for differential responses should not focus on TP alone, given that one needs to show a pattern of differential responses across neonatal behaviours.

Let's take the easiest question first, why we have focused primarily on TP. We *do* think the case of TP/R is special. It is the only gesture that has consistently garnered more positive than negative results in imitation experiments for which there is no ready, alternative explanation. In fact, TP/R is the most commonly modelled gesture across studies; approximately 85% of studies investigating imitation in neonates up to 6 weeks of age have included this gesture (Oostenbroek et al. 2013). Many reports of a differential response to other gestures have failed on replication, even in Meltzoff and Moore's own experiments (i.e. MO in the 1994 experiments). Moreover, TP/R is the standard comparison condition for the imitation of mouth opening, another gesture often cited as known to elicit imitation. So, if TP/R is not imitated, than a large and convincing body of evidence for the existence of neonatal imitation vanishes. All things being equal, such a result would probably spell the end of the NI research project. A vindication of the TP/R results, in contrast, would require the further investigation of differential responses to other gestures.

That said, our central worry about the criterion of differential imitation concerns the comparison class and the statistical analysis of the cross-target experiments. As **Kennedy-Costantini et al.** point out in their commentary, a recent longitudinal study of neonatal imitation at ages 1, 3, 6 and 9 weeks (Oostenbroek et al., 2016) reports that "….(infants) were just as likely to produce the gestures in response to control models as they were to matching models." By looking at a limited, cross-section of the data, the researchers were able to reproduce the same positive effects as reported in earlier studies. They conclude that failure to include adequate control conditions or test infants across multiple time points in previous studies has resulted in the false impression that infants selectively copy tongue protrusions, thereby perpetuating the idea that newborn imitation exists. For our part, given that we see the 'gestures' of infants as neonatal stereotypies, this makes a good deal of sense. Because we can actually categorize and then count up the kinds of neonatal stereotypies, just as Thelen did (Thelen, 1981), we have the correct comparison class in hand. So at least for neonates, the question cannot be "tongue protrusion or mouth-opening". This doesn't represent the statistical landscape. Of course, no one knows whether these results will themselves stand the test of time. But one possibility—the one on which we are betting—is that there will be no positive results at all. Hence no explanation will be needed.

If we are wrong, and the positive results prove robust, then we must give an alternative explanation. At present, we do not have a single reason, but we do have several we think are worth pursuing. For example, there is the intriguing result in Meltzoff and Moore (1992) that MO duration is longer than TP duration. When the authors tested MO using a dynamic stimulus, MO duration increased; it was twice as long on average than when the static stimulus was used. However, a dynamic presentation of TP produced only slightly more frequent TP's than in response to the static display (which were not timed only for duration). This difference in response suggests that the causal mechanisms of MO and TP differ. Therein could lie a tale of how arousal could differentially affect the rate and duration of TP and MO: Perhaps arousal/apprehension inhibits MO and increases TP. At this point in time, we have no clear explanation of *why* this should occur. But note that arousal plays a crucial role in Ghazanfar's examples of speech development as well, both in how infant marmosets learn to adjust the intensity of their voices as a function of listener distance (Choi et al. 2015) and in how parental calls, timed to infant vocalization, produce mature phee calls in the infant over time (Ghazanfar and Zhang 2016). These events occur beyond the neonatal period, of course, so they are only illustrative of how the context of arousal differentiates its causal effects.

We have also wondered whether Anisfeld (1996) and others might be correct in their explanation that the higher rate of MO in the mouth opening condition compared to its rate in the tongue protrusion condition, may be explained as a by product of infants' TP/R responses. TP/R and MO/C co-vary with each other: Increase in one results in a decrease in the other. Given that overall level of oral activity stays roughly constant in the two conditions, the rise in the rate of TP/R in the tongue protrusion condition seems to reduce the rate of MO/C in this condition. The higher rate of MO in the mouth opening condition compared to it's rate in the tongue protrusion condition may then be due to the lowering of the MO/C rate in the tongue protrusion condition, not to it's raising in the mouth opening condition. For instance, in Meltzoff and Moore (1983) the rates of MO/C and TP/R are about the same in the mouth opening condition, whereas the rate of TP/R is substantially higher than the rate of MO/C in the tongue protrusion condition. So what may rise and fall under different conditions is actually the rate of TP/R. More generally, we worry that even neonatal arousal is not a homogeneous state. Here the presupposition is that increasing arousal is like turning up the speed dial on an oscillating fan: It makes everything go faster. But there is no reason why this should be so, that the effects of arousal should be undifferentiated. Why, exactly, given the complexity of the neonatal brain, *must* this be so?

R3.3 Is our argument for the aerodigestive theory a form of Occam's Razor?

Several commentators argue that we should apply Occam's Razor to decide between the imitation and aerodigestive theories (**Beisert et al., Jones**). Originally, Occam's Razor was the dictum that we ought not to multiply explanatory entities beyond those needed.   As the old medical dictum goes: If you see hoof prints, look for horses not unicorns. Certainly part of our argument has followed Occam's Razor understood this way. We know that sub-cortical oscillators control the central aerodigestive behaviours of early infancy; neurodevelopmental research also suggests that the maps of primary visual cortex, S1 and M1 are still highly immature at birth. So taken together, if infants respond differentially to modelled tongue protrusion with more frequent tongue protrusions (than to other gestures), we ought not to look for the complex representational structures required by the hypothesis of opaque imitation, i.e. visual representations of facial gestures, somatosensory patterns of self-produced facial gestures, cortical motor commands that produce TP, plus the various mechanisms of mapping and association that yield genuine imitation. There is no independent evidence for the existence of such structures or abilities. Thus we should train our attention on the kinds of processes that we know to exist—or think are likely to exist—at birth in human infants. This would be an argument of the classic form.

Still, when dealing with the massive complexity of the human brain—and here we include the neonatal brain as well—we rarely have conclusive facts in hand of the sort that would support a simple version of Occam's Razor. Instead, what we suggest is more akin to inference to the best explanation than Occam's Razor. The question is this: What if we were to stand back from the specific and highly contested results of neonatal imitation experiments and look instead at the distance traveled over 35 years of neurodevelopmental research, in psychophysics, neurophysiology and neuroanatomy. What picture emerges? We know that neural mitosis and migration begins via chemical cues and/or via the use of nearby cells for pathfinding; we also know that prior to environmental input, Ca2+ activity influences every aspect of neural development from cell mitosis, migration, arborization, transmitter expression, axonal growth— and that once transducers are in place and functioning, development will continue based upon patterned environmental stimulation. We also know that this process of neural scaffolding can involve 'two steps forwards, one step back"— the expression and re-expression of transmitters, the growth of temporary structures such as the subcortical plate and of transient neural pathways, connections areas that will disappear when no longer needed. Although this developmental process proceeds simultaneously in multiple systems, by birth the human cortex is still immature. The alternating columns in V1 that segregated visual input by eye have yet to form, an essential organizational structure for stereoscopic vision. While the subcortical motor system is well developed at birth, the cerebrospinal system has yet to establish functional connections between M1 and the spinal gray matter, a process that will require protracted development after birth.

It is this cumulative picture that seems at odds with the existence of the infrastructure, at birth, required for cross-modal or amodal information transfer. Whether one supposes that such structures are innate or learned, the results of genetic transcription or neural activity, there is no reason to think that the neonatal brains 'comes with' such resources. It is also at odds with the kind of intentional explanations of neonatal behaviour often given by researchers. There is no *recognition* that your gesture is like mine, or *probing* a model to reveal model identity (Meltzoff & Moore 2002) or *understanding* that you are a thing 'like me' (Meltzoff, 2007). So either we 'deflate' what is normally meant by, e.g. a motor command (for there are no motor commands in the cerebrospinal system of the neonate) or we look towards the kinds of mechanisms that we know to exist. We then ask how the neonatal brain bridges the seemingly vast gap between non-intentional and intentional cognitive processes. **Zappettini** asks whether we, the authors, are fond of deflationary or 'killjoy' accounts, of accounts that take *prima facie* cognitive tasks and explain them in non-cognitive terms. We deny that we have a *predilection* for killjoy accounts. We are

not killjoys by nature. Rather, we agree with **Booth, Beisert et al.,** and **Provine** that we should avoid cognitively rich interpretations when there are more plausible cognitively lean interpretations available.

R3.4. Is Neonatal Imitation connected to later social skills?

Simpson et al. argue that neonatal imitation is connected to later social skills. If there is a connection between neonatal imitation and later social skills, first and foremost there should be a connection with the later imitation skills. However, many studies failed to find a connection between neonatal imitation and later imitation (Jacobson 1979; Abravanel & Sigafoos 1984; Fontaine 1984; Heimann, Nelson, & Schaller 1989; Kugiumutzakis 1999). On the contrary, neonatal imitation drops out after three months only to re-appear after 6 months. Among researchers who accept neonatal imitation as a fact, it is controversial whether this 'drop-out' is significant. Does the imitation of tongue-protrusion simply end abruptly at 2-3 months? Or does the infant merely move on to other forms of interaction with adults and resume a different repertoire of imitative behaviors a few months later? This is Meltzoff and Moore's explanation of the phenomenon of 'drop-out'.

As we have shown in the target article, that mammalian aerodigestion develops in two phases: (1) from the onset of isolated orofacial movements *in utero* to the post-natal mastery of suckling at 3 months after birth, and; (2) thereafter, from preparation to the mastery of mastication and deglutition of solid foods. This division in the maturation of the mammalian aerodigestive system has important consequences for the question of neonate imitation drop-out. Suppose, now, that tongue protrusion *qua* spontaneous neonatal behaviour itself ends between two and three months after birth. And suppose it does so *because* the developmental phase of which it is but one part comes to an end as a whole. This fact would suggest that the 'imitation' of tongue protrusion does not end because the infant looses interest in copying oro-facial gestures, but because spontaneous tongue protrusion itself declines as this first phase of aero-digestive maturation draws to a close. This fact—that an aero-digestive developmental stage, involving a period of spontaneous tongue protrusion, coincides with the period during which neonates 'imitate' tongue protrusion—is highly significant. This coincident phase makes it more plausible that the increase in neonate tongue protrusion in the experimental setting is the result of some extraneous cause—e.g. general arousal in the face of an interesting stimulus. Proof of an independent but coincident developmental phase, then, again raises the spectre that we have mistaken a spontaneous behaviour for an imitative one.

R3.5 Given that your theory posits a dynamical system of aerodigestive CPG's, why did you dismiss imitation via entrainment in the target article? What about mirror neurons?

One of the central features of the cross-target methodology is the alternation of each static display of gesture modeling with a period of static neutral face during which the response of the infant is recorded. During both periods, the model works to minimize any differences in presentation, paying particular attention to inadvertent social signals that might cue the production of stereotypies and/or create experimental artefacts. So the experiment seems almost perfectly designed to *preclude* the dynamic entrainment of the infant's behaviour by the model's behaviour. There is no entrainment without oscillatory activity and oscillatory activity is conspicuously absent from the standard cross-target methodology. This is why in the target article we dismissed entrainment as an explanation of the reported NI results.

In retrospect, a more nuanced answer is possible. There is one stage in the static cross-target paradigm that is dynamic, namely at the temporal boundary between the presentation of the static

gesture and the model's transition to the static face, e.g. when retraction of the tongue marks the completion of static TP modeling or after one full period of TP/R oscillation. Whether this has any effect on the neonate's behaviour is an open question. But entrainment theory could explain *why* the burst-pause methodology seems to promote imitation as well as why the imitative effects of gesture modelling are so weak (because there is only one very slow period of oscillation. That said, confirming a dynamical theory of behavioural entrainment would require a quite different experimental set-up, one that uses dynamic stimuli and records neonatal behaviour concomitantly.

As to mirror neurons, we agree entirely with **Fitch's** commentary. Were it not for the discovery of neonatal imitation in macaques, it would have been hard to explain why adult macaques, a species notoriously lacking in robust imitative behaviour, had mirror neurons at all. So neonatal imitation in macaques has served to bolster claims that mirror neurons underlie imitative behaviours in primates more generally. We do not wish to deny the existence of mirror neurons. But we are sceptical of any claim that mirror neurons explain the NI experimental results or that mirror neurons must be present in neonatal macaques/humans *because* of the NI experimental results. Independent evidence of either claim is needed before proceeding down that explanatory path (cf. Ferrari 2012). Still, contra **Leisman**, it is hard to imagine that mirror neurons exist in the human neonate. Mirror neurons require a functional, mirror neuron network. If AIM is unlikely to be true given the immaturity of the neonatal human cortex, then the same arguments apply to the cortical network/s required to drive mirror neurons—and thus to the existence of mirror neurons in the perinatal period.

R4.0 Social Engagement and Infant Automatons

R4.1 Social Engagement. A number of our commentators have stressed the importance of social interaction for infant development (**Aitken, Buck, Desseilles, Libertus et al., Murray et al., Simpson et al.**). We are entirely in agreement with this view and, more generally, with the interactive nature of human infants. At birth, infants come into the world entirely dependent upon the caretaking of adults. Infant survival requires the constant attention of their caregivers. This makes care as important to human infants as normal physiological development, e.g. of a functioning aerodigestive system at birth. But looking beyond bare survival, the extended period of post-natal motor development has the consequence that *what* is learned during this period rests heavily on parental interaction. Parents (or caregivers) facilitate the lion's share of the infant's interaction with the world prior to the attainment of goal-oriented action. For the neonate, being carried, coddled, cuddled, changed, fed, bathed, bounced and generally responded to 'in words or deed' are the rich events that foster immediate infant learning. Moreover, this learning involves multiple dimensions, the social and emotional no less than learning within the standard sensory and motor domains. Infants learn to be soothed by touch and voice, 'read' the prosody of human speech and the emotional 'temperature' of their social environments, to make eye contact and visually explore a human face, distinguish between the friend and 'foe' (the Still Face effect), and to relish in human interaction. Last but not least, they learn how to *engage* their caregivers—or, as my (Akins) mother-in-law used to say "how to run a household from the cradle." The inert, non-interactive infant is an infant at risk. But such infants will also learn *far less* about the world in general and about human social relations (cf. **Casartelli & Parma**).

Infant-maternal[1] bonding is thus essential for normal development. In the target article we mentioned the multitude of ways in which this occurs (See Section 7.3). These processes arise in the neonate alone (e.g. the olfactory recognition of the mother's colostrum at birth), the mother alone (e.g. oxytocin release, increased sensitivity to the infant's cry) and through interaction between the two (e.g. kangaroo care, the coordination of mother and infant heart beat). Social interaction begins immediately at birth—turning towards the mother's voice, visually exploring her face, making facial expressions such as smiles or grimaces—and gradually grows more sophisticated. We suggest that this interaction 'works'—creates and maintains a bond—at least in part because of the automatic human propensity for intentional interpretation. We see in that now-famous, first Pixar video, two lamps (one a large anglepoise lamp and the other a small gooseneck) interacting *as* a mother and child. They play, are watchful, talk, remonstrate and even sulk, cycling through the gamut of parent-child interactions. Indeed, we cannot help but see them as persons despite knowing that lamps cannot have intentional states (or have children for that matter). Our interactions with the newborn are no less intentionally infused; we see, and cannot help but see, the crying, grimacing and smiling of a human infant as *actions*. This is not an argument that we think infants are vegetative automatons, as **Aitken** would have us say. Rather, in practical terms, our own propensity to see intentional states is so deeply engrained that we see intentional behaviour even when our perceptions conflict with what we know. E.g. we see ur the neonate smiling even though we know that the transition to the social smile takes ~8 weeks after birth to develop. Add infant-adult turn-taking to the mix once the infant can distinguish self-from other- produced voices—i.e. remaining silent/inactive when the mother acts/speaks and then continuing activity when the mother stops—and any infant becomes entrancing to his or her caregivers (or at least to those disposed to be entranced at the outset). The net result of this and other such processes is maternal bonding, a state essential to the infant in all respects. In other words, the interactive baby has little need for a complex capacity for imitation. Nor is this possible, as **Campos and Neito** say, "their perceptive and attentional capacities (Volpe 2008), and face processing and intersensory processing abilities (Morton & Johnson 1991; Johnson et al 2015; Bahrick et al 2004; Lewkowicz 2014), among others, are too weak yet." The infant who reacts, in whatever way, to our attention reaps the benefits that advocates of neonatal imitation so often invoke.

R4.2 Infants as Automatons. A related misconception of our view is that we are throw-backs to the bad old days of psychological behaviorism. Or worse, we see infants as automatons that exhibit reflexive behaviours without variation or learning. To quote **Aitken** (who quotes Polani and Keith), we are tainted with the view that '(t)he newborn infant may be described as a tonic animal with oropharyngeal automatisms and neurovegetative mechanisms.' This is *not* our view although we would be hard pressed to say exactly what it is like to be a neonate. One misconception concerns our talk of stereotypies, a term we have adopted from the literature and which distinguishes between neonatal reflexes and stereotypies. Few sensorimotor neuroscientists talk in terms of reflexes anymore, at least not in the classical Sherrington sense, with the exception of the patellar reflex arc in children and adults. Indeed, neonatal 'reflexes' are notoriously difficult to evoke, requiring an experienced clinician. In any event, neonatal stereotypies are not invariant or 'released' by specific stimuli. On the contrary, our suggestion is that so-called neonatal stereotypes are useful precisely because they are *not* invariant. Precipitated by general arousal, their constant variation serves to explore the full-range of sensorimotor space and, as **O'Sullivan and Caldwell** point out, the stereotypic actions are likely to play a significant role in the development of associations between sensory and motor representations of the same

---

1   Bonding can occur with any caregiver, of course. Indeed, men can 'learn' to lactate. But are side-stepping the issue of what forms of bonding are possible, by concentrating on newborn-maternal bonding.

behaviour. Infant stereotypies occur precisely when the somatosensory and motor cortices are developing functional connections within and between the M1 and S1 regions, as well as developing the functionality of the corticospinal tract.   Of course, it is an essential part of our view that proprioceptive and motor learning begins in utero. But in just the way that 'suckling' in utero is a faint facsimile of actual suckling after birth, sensorimotor activity in utero fails to replicate the physics (and freedom) of terrestrial locomotion after birth. It's a whole new ball game, as they say, once on land. So for this reason, it is not a bad idea to 'wire' the sensorimotor system 'in place', at least for any species with a large and complex sensorimotor repertoire. At least, *a priori*, that would seem to be the case.

R5.0 Conclusion
The aerodigestive theory is built upon wide-ranging experimental results, from the neurophysiology of mammalian aerodigestive, sensory and motor systems, the practices of pediatric clinical neurology, the neurochemistry of activity-dependent neural processes, and developmental psychology more generally. It situates the gestures at issue within a known class of fetal/infant behaviours, rhythmic movements, but also within the known processes of early neural development. In contrast, after 40 years of investigation, the putative mechanisms of neonatal imitation remain oddly disconnected from the disciplines of human development. From where we stand, at least, the onus is now on the proponents to explain where this system of imitation resides and how it functions.

REFERENCES

Abravanel, E., & Sigafoos, A. D. (1984). Exploring the presence of imitation during early infancy. *Child development, 55*(2), 381-392.

Anisfeld, M. (1996). Only tongue protrusion modeling is matched by neonates. *Developmental Review, 16*(2), 149-161.

Bartrip, J., Morton, J., & Schonen, S. (2001). Responses to mother's face in 3-week to 5-month-old infants. *British Journal of Developmental Psychology*, *19*(2), 219-232.

Bertin, E., & Striano, T. (2006). The still-face response in newborn, 1.5-, and 3-month-old infants. *Infant Behavior and Development*, *29*(2), 294-297.

Bigelow, A. E., & Power, M. (2012). The effect of mother–infant skin-to-skin contact on infants' response to the Still Face Task from newborn to three months of age. *Infant Behavior and Development*, *35*(2), 240-251.

Borjon, J. I., Takahashi, D. Y., Cervantes, D. C., & Ghazanfar, A. A. (2016). Arousal dynamics drive vocal production in marmoset monkeys. *Journal of Neurophysiology, 116*(2), 753-764. doi: 10.1152/jn.00136.2016

Cecchini, M., Baroni, E., Di Vito, C., Piccolo, F., & Lai, C. (2011). Newborn preference for a new face vs. a previously seen communicative or motionless face. *Infant Behavior and Development*, *34*(3), 424-433.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS computational biology, 5*(7), e1000436. doi: 10.1371/journal.pcbi.1000436

Choi, J. Y., Takahashi, D. Y., & Ghazanfar, A. A. (2015). Cooperative vocal control in marmoset monkeys via vocal feedback. *Journal of Neurophysiology, 114*(1), 274-283. doi: 10.1152/jn.00228.2015

Dai, S.-Y., & Hata, T. (2006). Four-dimensional sonographic assessment of fetal facial expression early in the third trimester. *94*(2), 108-113. doi: 10.1016/j.ijgo.2006.05.004

Fagan, M. K. (2014). Frequency of vocalization before and after cochlear implantation: dynamic effect of auditory feedback on infant behavior. *Journal of experimental child psychology, 126*, 328-338. doi: 10.1016/j.jecp.2014.05.005

Fontaine, Roger. 1984. "Imitative Skills between Birth and Six Months." *Infant Behavior and Development* 7 (3): 323–33.

Ghazanfar, A. A., Chandrasekaran, C., & Logothetis, N. K. (2008). Interactions between the Superior Temporal Sulcus and Auditory Cortex Mediate Dynamic Face/Voice Integration in Rhesus Monkeys. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *28*(17), 4457–4469. http://doi.org/10.1523/JNEUROSCI.0541-08.2008

Ghazanfar, A. A., Takahashi, D. Y., Mathur, N., & Fitch, W. T. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology : CB*, *22*(13), 1176–1182. http://doi.org/10.1016/j.cub.2012.04.055

Ghazanfar, A. A., & Zhang, Y. S. (2016). The autonomic nervous system is the engine for vocal development through social feedback. *Current Opinion in Neurobiology, 40*, 155-160. doi: 10.1016/j.conb.2016.07.016

Ham, J., & Tronick, E. D. (2006). Infant Resilience to the Stress of the Still-Face. *Annals of the New York Academy of Sciences*, *1094*(1), 297-302.

Heimann, M., Nelson, K. E., & Schaller, J. (1989). Neonatal imitation of tongue protrusion and mouth opening: methodological aspects and evidence of early individual differences. *Scandinavian journal of psychology, 30*(2), 90-101.

Jacobson, S. W. (1979). Matching behavior in the young infant. *Child development, 50*(2), 425-430.

Kanenishi, K., Hanaoka, U., Noguchi, J., Marumo, G., & Hata, T. (2013). 4D ultrasound evaluation of fetal facial expressions during the latter stages of the second trimester. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 121*(3), 257-260. doi: 10.1016/j.ijgo.2013.01.018

Kaas, J. H. (2015). Blindsight: post-natal potential of a transient pulvinar pathway. *Current Biology : CB*, *25*(4), R155–7. http://doi.org/10.1016/j.cub.2014.12.053

Kugiumutzakis, G. (1999). Genesis and Development of early infant mimesis to Facial and vocal Models. In L. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 36-59). New York, N.Y.: Cambridge University Press.

Kuhl, P. K., & Meltzoff, A. N. (1984). The Intermodal Representation of Speech in Infants. *Infant Behavior and Development, 7*(3), 361-381. doi: 10.1016/S0163-6383(84)80050-8

Kurjak, A., Azumendi, G., Andonotopo, W., & Salihagic-Kadic, A. (2007). Three- and four-dimensional ultrasonography for the structural and functional evaluation of the fetal face. *American journal of obstetrics and gynecology, 196*(1), 16-28. doi: 10.1016/j.ajog.2006.06.090

Meltzoff, Andrew N., and M. Keith. 1989. "Imitation in Newborn Infants: Exploring the Range of Gestures Imitated and the Underlying Mechanisms." *Developmental Psychology* 25 (6): 954–62.

Meltzoff, Andrew N. 2007. "'Like Me': A Foundation for Social Cognition." *Developmental Science* 10 (1): 126–34. doi:10.1111/j.1467-7687.2007.00574.x.

Meltzoff, Andrew N, and M. Keith Moore. 1983. "Newborn Infants Imitate Adult Facial Gestures." *Child Development* 54 (3): 702–9.

Meltzoff, Andrew N., and M. Keith Moore. 1992. "Early Imitation within a Functional Framework: The Importance of Person Identity, Movement, and Development." *Infant Behavior and Development* 15 (4): 479–505.

Meltzoff, A., and M.K Moore. 1994. "Imitation, Memory, and the Representation of Persons." *Infant Behavior and Development* 17 (March): 83–99.

Meltzoff, A. N., & Moore, M. K. (2002). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, *25*(1), 39-61.

Moore, G. A., & Calkins, S. D. (2004). Infants' vagal regulation in the still-face paradigm is related to dyadic coordination of mother-infant interaction. *Developmental Psychology*, *40*(6), 1068.

Oostenbroek, Janine, Virginia Slaughter, Mark Nielsen, and Thomas Suddendorf. 2013. "Why the Confusion around Neonatal Imitation? A Review." *Journal of Reproductive and Infant Psychology* 31 (4): 328–41.

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Slaughter, V. (2016). Comprehensive Longitudinal Study Challenges the Existence of Neonatal Imitation in Humans. *Current biology : CB, 26*(10), 1334-1338. doi: 10.1016/j.cub.2016.03.047

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental science*.

Reissland, N., Francis, B., & Mason, J. (2012). Development of Fetal Yawn Compared with Non-Yawn Mouth Openings from 24–36 Weeks Gestation. *PLoS ONE, 7*(11), e50569. doi: 10.1371/journal.pone.0050569

Reissland, N., Francis, B., & Mason, J. (2013). Can healthy fetuses show facial expressions of "pain" or "distress"? *PLoS ONE, 8*(6), e65530. doi: 10.1371/journal.pone.0065530

Reissland, N., Francis, B., Mason, J., & Lincoln, K. (2011). Do Facial Expressions Develop before Birth? *PLoS ONE, 6*(8), e24081. doi: 10.1371/journal.pone.0024081

Sato, M., Kanenishi, K., Hanaoka, U., Noguchi, J., Marumo, G., & Hata, T. (2014). 4D ultrasound study of fetal facial expressions at 20-24 weeks of gestation. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 126*(3), 275-279. doi: 10.1016/j.ijgo.2014.03.036

Shepherd, S. V., Lanzilotto, M., & Ghazanfar, A. A. (2012). Facial muscle coordination in monkeys during rhythmic facial expressions and ingestive movements. *Journal of Neuroscience*, *32*(18), 6105–6116. http://doi.org/10.1523/JNEUROSCI.6136-11.2012

Simpson, Elizabeth A., Lynne Murray, Annika Paukner, and Pier F. Ferrari. 2014. "The Mirror Neuron System as Revealed through Neonatal Imitation: Presence from Birth, Predictive Power and Evidence of Plasticity." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 369 (1644): 20130289. doi:10.1098/rstb.2013.0289.

Striano, T. (2004). Direction of Regard and the Still-Face Effect in the First Year: Does Intention Matter?. *Child Development*, *75*(2), 468-479.

Thelen, E. (1981). Rhythmical behavior in infancy: An ethological perspective. *Developmental Psychology, 17*(3), 237-257. doi: 10.1037/0012-1649.17.3.237

Warner, C. E., Kwan, W. C., & Bourne, J. A. (2012). The early maturation of visual cortical area MT is dependent on input from the retinorecipient medial portion of the inferior pulvinar. *Journal of Neuroscience*, *32*(48), 17073-17085.

Yigiter, A. B., & Kavak, Z. N. (2006). Normal standards of fetal behavior assessed by four-dimensional sonography. *J Matern Fetal Neonatal Med, 19*(11), 707-721. doi: 10.1080/14767050600924129