From: N. Eilan, R. McCarthy, & B. Brewer (Eds), Spatial representation: Problems in philosophy and psychology, (pp. 219-235). Cambridge, MA: Blackwell, 1993

9
Molyneux's babies: Cross-modal perception, imitation and the mind of the preverbal infant

Andrew N. Meltzoff

Gareth Evans (1985) contemplated novel variants of the Molyneux problem to illuminate certain questions in epistemology. I and others have used variants of the Molyneux problem to address questions about the nature of the infant mind. When we study infants instead of blind adults, the questions cannot be verbal; we must infer infant perception and thought through their actions.

Imagine that the blind man is presented with a sphere to explore by touch. A sphere and cube are placed before him and sight is bestowed. In this modified example we do not ask the man to use a verbal label to designate which object he has just touched; we simply ask that he point towards or even to look longer at that object. The subject's response in this instance is nonverbal, but it raises most of the critical issues of the original case.

I will discuss cases in which we draw inferences about the classical epistemological problems of the Molyneux example from patterns of actions such as these. Our subjects were not sight-recovered patients, but young infants. There are parallels between the newly-sighted man and an infant. Like a blind man, a newborn infant has not visually inspected objects and has not had a chance to associate visual and tactual experiences of the same object. The original Molyneux problem stimulated many innovative studies of the blind. Had Molyneux posed his question using a newborn infant fresh from the womb, this might well have stimulated careful research with infants a century or two earlier than it became popular. Accelerated growth in the field of genetic psychology in turn may have aided philosophers. The information garnered from sight-recovered patients and cited in the philosophical literature is notoriously variable, as might be expected given this rare, and neurologically rather bizarre, population. An intact, normally-developing brain may be more useful as a touchstone for theorizing.

Focusing on newborn infants and not blind adults raises a whole family of new Molyneux-like questions. In fact, infants turn out to have far more sophisticated abilities to co-ordinate information from different modalities than we would ever have expected. While Molyneux wondered whether information about an object derived from touch and vision might be related prior to associative experience, it did not occur

to him to ask whether proprioceptive information and visual information concerning one's own body movements in space, or speech sounds and mouth movements, might be intrinsically related prior to such experience. Yet the behaviour of actual infants suggests that such cross-modal connections come into play quite early in development, and perhaps are available at birth. I will argue that the behaviour of young babies raises puzzles analogous to the ones raised by the Molyneux problem. One of the deepest and most intriguing of these is the problem of imitation.

1 The Imitation Problem

Imitation involves the form of an action in behavioural space, not the shape of an external object, a sphere or cube. Molyneux might have posed the following Imitation Problem:

> Suppose a blind man can perform simple body movements, such as mouth opening and closing; he can identify the movements when he produces them and can produce them on demand. Suppose then that an actor is placed before the blind man and the blind man is made to see. The actor silently opens his mouth. Can the newly-sighted man, without being allowed to touch the actor, imitate the actor's gesture by opening his own mouth?

Locke, Molyneux, Berkeley and others who answered the original query negatively would also answer the Imitation Problem negatively: there is no way for the subject to know that in order to produce a certain visual spectacle he must move his body in a particular way. There is no immediate equivalence between the mouth-opening-as-seen and the mouth-opening-as-done.

The Imitation Problem has not, to my knowledge, ever been tested with newly-sighted patients. However, we do know that one-year-olds can imitate facial gestures. Unfortunately, this does not provide us with the data relevant to the Imitation Problem. Although infants cannot now see their faces while they are imitating, they may have seen themselves in the past in mirrors. Mirrors are a tool for making the invisible face visible. Mirrors provide a tutorial in 'connecting' action-as-seen with actions-as-done. Mirror-experienced one-year-olds are not like newly-sighted men. Fortunately, developmental psychologists may easily locate mirror-naive infants. In fact, the relevant population is one that is naive to this whole set of potential mediators, not just mirrors. At about one year of age infants begin to reach out and touch their mother's mouth and then touch their own, thus providing tactual comparisons. But with appropriately young and inexperienced infants, we can test the Imitation Problem directly. Such infants can produce the relevant motor movements with their own faces, but have never seen them. They can see the relevant acts of others, but have never felt them. Will they be able to imitate? If so, what does this tell us about the organization of the infant mind?

2 Imitation in infancy: The phenomena and initial suggestions about supramodal perception

Meltzoff and Moore (1977) discovered that 12- to 21-day-old infants, who by all reports are mirror-naive, could successfully imitate a variety of facial acts (Figure 9.1).

Figure 9.1 Photographs of 12- to 21-day-old infants imitating facial gestures shown to them by an adult.

We showed imitation of four body actions: lip protrusion, mouth opening, tongue protrusion and sequential finger movement. These particular gestures helped evaluate the specificity of the imitative response. Infants responded differently to two movements of the same body part (mouth opening v. lip protrusion) and also to two body parts producing the same general movement (lip protrusion v. tongue protrusion). This suggested that infants were matching particular *acts*, not just activating a certain region of their body (lips) or producing very generally-defined movements in space (protrusions).

We also found that young infants could imitate from memory. A pacifier was put in infants' mouths as they watched the display so that they could only observe the adult demonstration but not duplicate the gestures. After the infant observed the display, the adult assumed a passive-face pose and only then removed the pacifier. Infants were then given a 2.5-minute period in which to respond, during which the adult maintained this passive face regardless of the infant's response.[1] Even with this pacifier technique, the infants imitated the two displays. Moreover, infants did not produce exact matches early in the response period. The first responses of the infants were often with the correct body part but an approximation of the adult's act. Infants would move their tongues but not produce full tongue protrusions. Infants appeared to home in on the detailed match, gradually correcting their responses over successive efforts to more exactly correspond to the details of the display. The adult was sitting with a passive face all this time, thus the infant was comparing his or her motor performance against some sort of internal model or representation of what had been seen.

The report of neonatal facial imitation surprised developmental psychologists because it did not fit with classical theories of infancy. However, the basic phenomenon of early imitation has now been demonstrated in more than a dozen studies, by independent investigators using different designs; apparently early imitation is a cross-cultural phenomenon: positive results have been reported in the USA, Britain, France, Switzerland, Sweden, Israel and rural Nepal (see Meltzoff, 1990a, for a review). The fact that neonates will duplicate certain basic acts performed by an adult is now well-established, though the explanation for this fact is still unclear.

There are several psychological mechanisms that might underlie this behaviour. The hypothesis I favour is that imitation is based on infants' capacity to register equivalences between the body transformations they see and the body transformations they only feel themselves make. On this account early imitation involves a kind of cross-modal matching. Infants can, at some primitive level, recognize an equivalence between what they see and what they do. We might imagine that there is something like an 'act space' or very primitive and foundational 'body scheme' that allows the infant to unify the perceptual and action systems into one framework. In this view, although the infant's own facial gestures are invisible to them, they are not unperceived, for even unseen body movements can be monitored by proprioception (O'Shaughnessy, 1980, and ch. 13 of this volume).

Learning theorists could argue that all this is unnecessary. The subjects were 12 to 21 days old. Perhaps they had been trained to imitate during the first weeks of life. Infants could be conditioned to poke out their tongues to a ringing sound, or to an adult tongue protrusion. Perhaps the conditioning of a few oral gestures is part of the natural interaction between mother and baby. To resolve the point, Meltzoff and Moore (1983) tested 40 newborns in a hospital setting. The average age of the sample was 32 hours old. The youngest infant was only 42 ***minutes*** old. The results showed that the newborns imitated both of the gestures shown to them, mouth opening and tongue protrusion. One further study showed that newborns also imitated a non-oral gesture, head movement (Meltzoff and Moore, 1989). We can infer that the capacity to imitate certain gestures is innate. Following some of the ideas advanced by Bower (1982), I would like to argue that imitation is just one manifestation of a larger capacity, a supramodal perceptual system, that can be tapped by other tasks. The next study pursued this point.

3 Asking infants Molyneux's question: Tactual–visual object perception

Bryant showed that six- to 12-month-old infants could recognize, by sight, an object that they had previously explored by touch alone (Bryant, Jones, Claxton and Perkins, 1972). This does not address the classic issue, however, because infants in the second half-year of life regularly reach out and tactually explore objects that they see; they also bring objects that are in their hands before their eyes for visual inspection. Through such simultaneous bi-modal exploration of objects, infants may have learned to associate particular tactual impressions with particular visual sensations.

Meltzoff and Borton (1979) designed a test to evaluate tactual–visual cross-modal perception in much younger infants, before such learning experiences were likely. The average age of the subjects was 29 days at the time of test.[2] How can we induce these

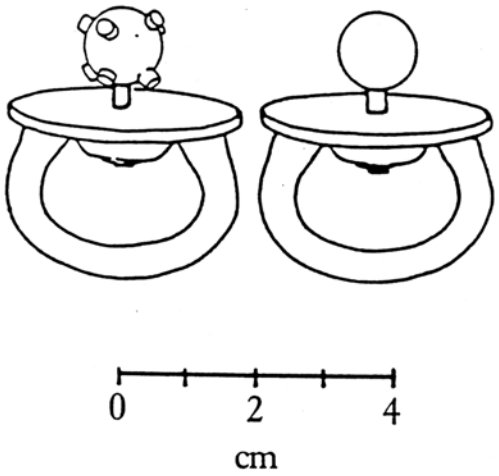

Figure 9.2 Shapes used to pose Molyneux's question to preverbal infants.

young babies to explore a shape by touch; and then how can they indicate to us which shape they felt?

Pacifiers were modified so that mouth-sized shapes could be mounted on them (Figure 9.2). The tactual shapes used in the test were a small sphere and a sphere with nubs. The pacifier was cupped in the experimenter's hand and slipped into the infants' mouths without them seeing it. Most infants were quite happy to suck on the object, rolling it around on the tongue and furrowing their brows, as if the tactual exploration of the novel object was of some interest. They were allowed to feel the object for 90 seconds; it was then slipped out of the mouth, unseen. Each infant was randomly assigned to an experimental condition, half the infants were given one shape to explore and half the other shape.

All infants were then shown two visual objects, side by side, that were the same shape as the two tactual objects. We measured infants looking to both objects during a fixed response period. If infants can relate shapes-as-felt to shapes-as-seen, then the object they looked at should vary as a function of tactual condition. The hypothesis was that infants would systematically look longer at the shape they had felt. The results showed just that: of the 32 infants tested, 24 fixated the shape matching the tactual object longer than the non-matching shape, which differs significantly from the 16 v. 16 split expected by chance alone. The mean percentage of total fixation time directed to the matching shape was 71.8 per cent, which was also significantly different from chance.

The infants in this study were not literal newborns, but one-month-olds. On purely logical grounds we cannot exclude the possibility that they may have learned (what for them are arbitrary) associations or connections between visual features and tactual features in these initial days of life, associations that then were used in the experiment. However, what we know about the actual behaviour of young infants argues against such a notion. Such associative connections are said to be formed when an object was simultaneously seen and touched and hence the two sense impressions are connected in

the observer's mind. It is well established that infants this young do not engage in simultaneous visual and tactual explorations of objects. At this early age, they do not yet use their fingers to explore objects while they visually inspect them. The empirical literature shows that this sort of co-ordinated bi-modal exploration only begins to occur at about three to six months of age. It is true that infants suck objects during the first month, but it is, after all, impossible for them to look at the objects ***while*** they suck them. Nor do infants of this age visually inspect objects before they are inserted into the mouth. During feeding, infants' commonest visual impressions are those of the mother's face (Spitz, 1965). Based on pure association, most infants should come to believe that seen faces feel like nipples! There must be prior constraints, psychological biases based on spatial properties, on the impressions that can and cannot be 'associated'.

I have suggested that newborns, without associative experience, register the same information about the shape of the object even if it is picked up through two different modalities, touch and vision. Having perceived the form through one perceptual mode, they are familiar with it when it is presented to them in the new mode. In the present experiment, the neonates may have been particularly interested in the visual instantiation of the form, because it provided modality-specific information, such as colour, that was not available through touch. Hence, they looked significantly longer at the matched than the mismatched shape. The neonates in this experiment seem to act in a way that is compatible with Evans's (1985) hypothetical philosopher, 'V'.

4 Speech by ear, eye and mouth

Are material objects unique in being registered through more than one modality? Speech is typically considered to be an auditory phenomenon. The sounds of speech are, of course, auditory. They are not seen. But articulatory acts, the causes of speech sounds, can be seen. In this sense, speech is not uniquely auditory, but a polymodal phenomenon. That speech can be seen, at least in adults, is demonstrated by the fact that we can 'read' a person's lips and grasp what was 'said', even when there is no sound. At what age and by what mechanism does the human perceiver apprehend the correspondence between auditory speech and visual speech, between audition and articulation?

5 Relations between seen and heard speech: A perception task

Kuhl and Meltzoff (1982; 1984) presented four-month-olds with an infant-tailored lip-reading problem. We tested whether infants recognized that an /a/ vowel sound (as in 'pop') corresponded to one articulatory gesture and that an /i/ sound (as in 'peep') corresponded to another articulatory gesture. The infants were shown a film of two faces articulating the vowels: one face was articulating the /a/ vowel and the other the /i/ vowel. The two faces were life-sized and in colour. The faces were filmed and edited so that they would articulate in perfect temporal synchrony with one another. The auditory vowel sounds, either the /a/ or the /i/, were presented from a loudspeaker placed midway between the two faces. Each infant heard only one of the sounds (played repeatedly for a two-minute test period), but had the visual choice of two faces.

This set-up allowed us to rule out two possible bases that infants might use to detect

an auditory–visual match: The central placement of the loudspeaker ruled out any spatial cues; the temporal synchrony between the two faces ruled out temporal cues. The only way infants could solve this problem was by recognizing that the auditory /a/ corresponded to the articulation involving a wide-open mouth and the auditory /i/ to the articulation involving narrowed lips with the corners pulled back.

We posed this problem to 18- to 20-week-old infants. We reasoned that if infants could detect the correspondence between auditory speech and visual speech, they would look longer at the face that produced movements appropriate to the sound they heard. The hypothesis was strongly supported: infants listened intently to the sound and looked back and forth between the two faces, settling on the particular face that matched the sound they heard. Subsequent studies in our laboratory using other vowel sounds, /i/ and /u/, extended these basic effects. The experiments suggest that by 18 weeks of age infants recognize that /a/ sounds go with mouths that are open wide, /i/ sounds with mouths that have retracted lips and /u/ sounds with mouths whose lips are protruded and pursed.[3]

6 Relations between audition and articulation: A production task

The foregoing studies probed the infant's knowledge of auditory–articulatory links in a speech perception task. A more important, but deeply related skill, is the link between audition and articulation in speech production.

Humans around the world do not sound the same; the sound pattern of English is vastly different from German or Japanese. Early auditory experience is critical to the development of a particular phonology: growing up in a particular language environment indelibly influences the 'accent' one uses, even in adults who may not have been exposed to the original language for decades. At what age does auditory input begin to influence the sounds people make? Do young infants mimic the speech sounds they hear, adjusting their unseen articulators to match the auditory model?

The infants in our speech study provided a convenient way of addressing this question. Recall that half the infants were randomly assigned to hear the /a/ vowel and half the /i/ vowel. The formant frequencies of /i/ are spread widely apart, while /a/'s formants are close together in frequency.[4] The formants of the infants' sounds were measured by computer and the relevant values of the infants' formants were calculated. The results supported the imitation hypothesis. Infants hearing /i/ produced sounds that were more /i/-like, while infants hearing /a/ produced sounds that were significantly more /a/-like (Kuhl and Meltzoff, 1988).

On the basis of this work, we have suggested that in 18-week-olds the representation of speech is not limited to its auditory properties. Rather, speech representations, like the body transformations in facial imitation, are probably organized in a way that is not exclusively auditory, motor or visual, but instead is supramodal. This internal representation is such that an auditory signal can influence behaviour in two other modes. The data show that an auditory signal influences where infants look, causing them to look at a silent moving mouth that is phonetically equivalent to the sound they hear. The auditory signal also influences what infants say, causing them to move their mouths in a way that will result in an event that is equivalent to the one they hear. It seems likely that both these phenomena, cross-modal perception and vocal imitation, are linked by some common representation of speech.

7 Imitation and the roots of the notion of self

The next line of studies concerns the development of the notion of self in the preverbal infant, a topic intrinsically tied to imitation, cross-modal functioning and the co-ordination of perception and action. It is difficult to design studies to address the preverbal notion of self. In fact there is only one established experimental paradigm for examining self in infants: the mirror self-recognition paradigm first used with infants by Amsterdam (1972) and adopted by many since then. The procedure is simple. It involves putting rouge on the infant's forehead without its knowledge (usually as part of wiping its face). Then the child is put in front of a mirror. Infants older than about 18 months look at themselves in the mirror and then reach up and rub the mark on their foreheads. Infants younger than about 18 months make no such attempts, although they will rub off marks that are put on visible parts of their bodies, such as hands or arms. The inference that has been drawn by some psychologists (e.g. Kagan, 1984) is that at about 18 months of age a sense of self suddenly emerges which is linked to the emergence of language. Compatible with this, it is said, are the related findings that children give no indication of recognizing photographs of themselves before about 18 months.

It is clear that the mirror or photographic self-recognition tests only assess a narrow aspect of self: the recognition of visual features. A prior, developmentally more fundamental aspect of self concerns one's own movements and body postures. You may need mirror experience to learn that your face does not normally have a red mark on it or that your eyes are green. However, if the arguments about facial imitation are sound, you don't need visual experience in order to know what your own unseen body movements would look like. Visual instantiations of your own body movements can be directly related to the movements that are felt. In short, self-recognition based on static featural information is quite different from self-recognition based on spatio–temporal movement patterns, and I believe the latter provides the ontogenetic foundation for the former.

How can we begin to investigate infants' ability to recognize that seen human movements are 'like me' (or, 'like the movements that are felt', or even 'like those that are intended' – distinctions to which I shall return)? Several approaches are possible; we chose one in which an adult experimenter acted as a kind of 'social mirror' to the infant, reflecting back everything the baby did. We wanted to know if infants could recognize this self–other correspondence despite the absence of featural identity. We tested infants at 14 months of age, an age at which infants fail the mirror self-recognition test.

Three converging experiments were conducted. The first investigated whether or not infants at this age showed any interest in their own behaviour being reflected back to them by another person. The infants sat at a table, across from two adults who sat side by side. All three participants were provided with replicas of the same toys. Everything the infant (X) did with his toy was directly mimicked by one of the adults (X′). If X banged the toy three times on the table, X′ banged his toy three times on the table. It was as if X′ was tethered to the infant, a puppet that was under X's control. The second adult (Y) was not so tethered. This adult sat passively, holding the toy loosely in her hands on the table top.

We thought that if infants could detect that their actions were being matched, they

would prefer to look at X′ and also smile at him more. We also thought that infants would tend to test the relationship between the self and the imitating other by experimenting with it. For example, infants might modulate their acts by performing sudden and unexpected movements to check if the X′ was still shadowing them. Adults do this when they unexpectedly catch sight of themselves in a store video camera; they wave their arms or make a sudden movement to check whether the image on the screen follows suit.

Scorers watched a videotaped record of the experiment and noted which side the infant looked at and all instances of smiling and testing behaviour. The results showed that infants had a clear-cut preference for X′ over Y. Infants looked significantly longer at X′, there were more smiles directed toward X′ and infants directed more test behaviour at X′.

There are several alternative interpretations of these findings. One is that infants can recognize the self–other equivalence that is involved when an adult imitates them. Alternatively, infants may simply be attracted to any adult who actively manipulates a toy, without invoking any detection of action equivalence. Such a simple interpretation does not explain why infants would direct more 'testing' behaviour towards the imitating adult, but perhaps such behaviour is displayed to any active adult, whether or not the adult is mimicking the baby.

In a follow-up study, the general procedure was similar to the first study, but the control experimenter did not remain passive. Instead, this adult actively manipulated the toys. Furthermore, we wanted the adult not only to be active, but to do 'baby-like' things with the toys so that no preference for the imitating experimenter could be based solely on a differentiation of adult versus infantile actions. This was achieved by using a yoked control procedure. There were two TV monitors situated behind the infants and in view of the adults. One monitor displayed the actions of the current infant, live. The other monitor displayed the video record of the immediately preceding infant.

The job of each adult was to mimic one of the infants on TV. Both adults performed in perfectly infantile ways, but only one matched the perceiving infant. Could the infants recognize which adult was acting like they were and which was acting like another baby? The results again showed that infants looked longer at X′, the person who imitated *them*, smiled more often at him and, most importantly, directed more testing behaviour toward him.

These findings constrain the possible interpretations. The demonstrated effects cannot be explained as simple reactions to activity, for both experimenters were active. Nor can they be explained by saying that the infants recognized a generic class of baby-like actions, for both experimenters were copying the acts of babies. It would seem that the subjects are recognizing the relationship between the actions of the self and the actions of the imitating other.

How did the babies detect this relationship? Very broadly speaking, two kinds of information are available. The first is purely temporal contingency information. According to this alternative the infant need only detect that whenever he does *a* the adult does *b*. The infant need not detect that *a* and *b* are in fact structurally equivalent, only that they are temporally linked. A second alternative is that the infant can do more than recognize the temporal contingency. In particular, the infant may be able to recognize that the actions of the self and other have the same form: they are structurally equivalent.

To differentiate these alternatives, we used a design similar to the previous two.

However, in this study the purely temporal aspects of the contingency were controlled by having both experimenters act at the same time. This was achieved by having three predetermined pairs of target actions. Both experimenters sat passively until the infant performed one of the target actions on this list. If and only if the infant exhibited one of these target actions, both experimenters began to act. The imitating adult performed the infant's act, and the control adult performed the other behaviour that was paired with it from the predetermined target list.[5]

For example, whenever an infant shook a toy, the imitating adult also shook his toy, carefully shadowing the infant. The behaviour of the other adult was also under complete temporal control of the infant, but this adult performed a different type of action. Whenever the infant shook his toy, the control adult would slide his matched toy, also carefully shadowing the speed and duration of the infant's act. If the infant began waving his toy, both adults stopped acting in unison, because waving was not one of the 'target acts' to which they were programmed to respond. This design achieves the goal of having both the adults' actions contingent on the infant's. What differentiates the two experimenters is not the purely temporal relations with the acting subject, but the structure of their actions *vis-à-vis* the subject.

The results showed that the infants looked, smiled, and most importantly, directed more testing behaviour at the matching actor. Thus, even with temporal contingency information controlled, infants can recognize the structural equivalence between the acts they see others perform and the acts they do themselves. In that sense they have already begun to elaborate a notion of self. This sense of self consists of a kind of extended 'body scheme'; a system of body movements, postures and acts.

8 Implications for genetic psychology

These findings and other recent work with young infants (Bower, 1982; Butterworth, 1981; ch. 5 of this volume) alter the classical story of early psychogenesis. In the classical view, infants from birth to at least several months old have separate 'heterogeneous spaces' (Piaget and Inhelder, 1969, p. 15), a tactile space and a visual space and an auditory space that are then gradually co-ordinated as 'the child begins to grasp what he sees, to bring before his eyes the objects he touches, in short to co-ordinate his visual universe with the tactile universe' (Piaget, 1954, p. 13). The project for genetic psychologists was to trace how an infant starting from such a deficient initial state developed into the mature adult. However, such development may never need to occur, because the initial state is not as limited as we supposed.

It seems likely that the young infant is not limited to registering isolated bits of sense data, such as tactual impressions, retinal images and acoustic frequencies. There is probably no time in development in which infants are restricted to modality-specific fragments, sense scraps that are connected through empirical correlations. Instead, infants may represent the world more abstractly, in terms of objects and events that transcend a single sensory modality. These 'distal projections' are not the product of a long period of experiencing sense-data → sense-data correlations. More likely, the psychological world of the human newborn is populated by objects and events that can be accessed by more than one modality. When a young baby brings a round rattle before his eyes, he is probably not engaged in discovering what visual sensation is

associated with this particular tactual impression; he already knows that. Instead, he is fascinated by the additional modality-specific features (the rich colours, visual sheen and shadows that could not have been known by touch alone) of the abstract form that he already apprehended through touch.

This picture of the infant's world follows from the results of the the tactual–visual experiment in which we put an unseen shape in the infants' mouths. These infants, 29 days old at the time of the test, were too young to have had many experiences associating shapes-as-felt with shapes-as-seen. Nevertheless, the results showed that infants who were given a shape to feel would systematically seek out the matching shape by eye. The mouth cannot see, the eye cannot touch, yet information picked up by one modality directs the other. It is also important that the test was designed to tap memory and representation. The tactual object was not felt at the time that the visual objects were seen: rather, the shape was removed from the infant's mouth, and only at that point was the visual choice presented. The results demonstrate that young infants can relate a visual perception to the memory of the information that was picked up through touch.

The studies of neonatal imitation push the story a bit further. In this case the infants tested were truly newly sighted. The youngest infant in the study was just 42 minutes old. We can say with assurance that the capacity to imitate certain facial acts is truly an innate aspect of the human mind. When the newly sighted infant sees certain human gestures he or she can immediately mimic these acts. Such facial imitation entails cross-modal functioning: the infant can see the adult's actions, but he cannot see his own face; indeed has never seen his own face in his entire life. There is some primordial connection between our own acts and the acts we see others perform.

Two details about the findings of innate imitation are particularly noteworthy. First, as in the cross-modal case, memory and representation are involved. It is conceivable that infants might have been restricted to imitating only if the to-be-matched target was in sight at the time the infant action was performed. Imitation might be the result of a kind of 'perceptual-motor resonance'. But recall that we inserted a pacifier in the infant's mouth during the adult demonstration, and it was only after the adult had stopped gesturing and assumed a passive-face pose that the pacifier was removed. In this situation, the infant needed to bring his own *unseen* body movements into accord with a currently *unseen* target act; not a task for an organism confined to a here-and-now world of raw sense impressions.

Second, the imitative response does not appear to be a mindless reflexive reaction. Simple reflexive acts don't bridge temporal gaps. Consider an analogy: suppose an infant's eyelids were propped open and a loud sound was suddenly presented, one that normally would lead to a startled eye blink. A few seconds after the sound ceased, the eye lids are freed. Would the infant show a delayed blink reflex once freed to do so? Certainly not. Reflexes do not work like that: they are triggered by the presence of the stimulus, not by a memory of the stimulus. In contrast, infants can imitate after a delay and may be particularly motivated to respond once the adult has stopped gesturing, a point I will return to below.[6] Moreover, instead of imitation bursting forth in a stereotypic, fully-formed manner, the infants correct the response over successive attempts so that they more and more closely approximate the adult target. Simple, mindless reflexes do not involve correction or a homing in on a target.

In my view, early imitation is an intentional act, in the minimal sense that it involves

goal-directed matching of the target. The infant is trying to correct his or her motor performance, which may not be accurate for all sorts of reasons, so that it more accurately matches what he intends. Thus conceived, the newborn encodes the adult's act in neither exclusively visual nor exclusively motor terms. Instead, the newborn's representation is a modality-free description of the body transformation. This internal representation is the 'model' that directs the infant's actions and against which he can match his motor performance. Thus, infants compare the proprioceptive information from their own unseen body movements to their representation of the visually perceived model and sharpen their match over successive efforts. Similarly, imitative responses would not need to be 'tripped', 'released' or 'fired-off' in the presence of the model, but might be initiated from the infant's memory of what the adult had done.

This returns us to the observation that infants seem prompted to imitate once the adult has stopped. When the gesturing stops the infants are confronted with a mismatch between their current perception of the adult and their stored representation. The infant may generate a matching response in order to reinstate the absent event, to make it perceptually present again. Thus, the disappearance of the adult gesture gives the infant a cognitive problem to work through, the conflict between the world-as-represented and the here-and-now world present to the visual system. This mismatch or disequilibrium between perception and representation motivates the infant to act, and so to imitate.

Some of these notions also can be applied to the cross-modal speech effects. In classical theory, infant speech is considered an acoustic event in the province of the 'sense of hearing', processed along the eighth nerve. The new research indicates that as early as 18 weeks of age, speech is not purely a matter of hearing. Infants probably access the same underlying phonetic representation whether they see, hear or motorically produce speech. The distal entity of interest is not sense-specific, but rather a supramodal phonetic unit.

An important caveat is that the infants in the speech studies, unlike those in the imitation studies or the cross-modal object studies, were 18- to 20-weeks old. Moreover, unlike the infants in the other studies, these infants do have some experience with matches between auditory and articulatory events. They watch adults talk, and they babble and hear the results of their babbling. Accordingly, three ontogenetic accounts may be offered, with the third being of special interest.

First, the infants may simply have learned which articulatory gestures go with which sounds by simultaneously watching and listening to adults. This might reduce to associative learning. Second, the hypothesized supramodal phonetic units might be innately specified, inasmuch as all the world's languages draw from a pool of only about 100 phonemes. If so, infants should succeed on a cross-modal test using foreign-language phonetic contrasts that were never heard or seen in the infant's particular culture. Similarly, newborns might also demonstrate cross-modal auditory–visual perception. However, there is also an intriguing third alternative, namely, that the infants' self-produced babbling experience may play an important role. This interpretation is interesting because it ties together several of the phenomena discussed in this chapter.

In the babbling account, infants are conceived of as carefully monitoring their own vocal play during cooing and babbling. They 'feel' their articulatory movements through proprioception and can perceive the consequences of these articulatory efforts through

audition. Thought of in this manner, the seemingly aimless vocal play of young babies is actually a way of practising the basic act of speaking, practising the production of phonetic units at will.

How could this babbling experience help infants in the cross-modal situation? It could help only if infants can relate the speech acts they see the adult perform in the experiment to the auditory–articulatory events they produced themselves during babbling. This is a cross-modal generalization. The research indicates that infants may well be able to do this. Infants' ability to imitate visual gestures demonstrates that they can relate mouth movements they see to their own mouth movements. The mouth-opening movement in Meltzoff and Moore's imitation experiments is similar to the mouth opening used to produce /a/ in Kuhl and Meltzoff's cross-modal speech case. So there is a foothold on the articulatory side – infants may relate their own unseen speech mouth movements to those they see the adult perform. There is also a similar foothold on the auditory side. Kuhl's (1979; 1983) speech categorization work indicates that infants can recognize the equivalence between the vowels uttered across talkers, including those produced by children and adults, despite the differences in the actual frequencies of the sounds that are caused by the differences in the size of the vocal tract. It therefore is reasonable to suppose that the infants in our speech study can recognize equivalences between the vowels they hear in the experiment and their own previous vocalizations. To summarize: the knowledge gained during their own babbling may contribute to infants' ability to recognize the auditory–visual correspondences for speech when seen and heard on another's body.

The infants' use of cross-modal capacities as leverage in grasping self–other relations seems an avenue worth pursuing, for there may be both constitutive and ontogenetic connections (Meltzoff, 1990b). This point came more clearly into focus in the tests involving the adult imitation of the infant. We arranged a situation in which infants were presented with two adults, one who was pre-programmed to imitate the child and the other who systematically mismatched the infant's behaviour. The results showed that the infants acted in very special ways toward the particular other who matched the self. The infants devoted most of their visual attention to the imitating adult, smiled at him more, and also directed more of what we called 'testing' behaviour towards him.

I suggest that the infant recognizes the adult as acting 'like me', that the self–other correspondence is evident to the child. A closely related notion is that the infants perceived the causal relationship with the imitating other. The imitating adult is seen by the infant as being 'more under my control'. This sense of causality might follow from the fact that the imitating adult acts in a way that is not only temporally contingent on the infant (both adults do this equally), but also that the patterns of behaviour are spatially matched.[7] The point is, however, that infants' intense interest in the imitating other ultimately derives from the infants' perception of the self. It is the cross-modal spatio–temporal correspondence between the pattern 'out there' and the pattern of self action 'here' that gives the imitating adult special psychological salience.

Indeed, it is probably the infants' own exquisitely detailed perceptual mechanisms that allow them to differentiate the imitating adult as clearly 'non-self'. There are at least two pieces of information that can be used: the adult is seen to be located in a different spatial location than the one in which the self's actions are perceived to be (both visually and proprioceptively); and moreover, the imitating adult, no matter how

practiced a mime, does not provide a perfect match.[8] The results of the study show that infants at this age spend virtually all their time watching the movements of the non-selves rather than watching the movements of the self; for example, they are not very interested in the movements of their own arms as they shake the toy (again indicating there is not total confusion, non-differentiation). Yet between the two non-selves, infants prefer the one that is more like the self. This attention to 'non-self that is none the less like self' may not be an altogether bad recipe for psychological development.[9]

The imitation game provides a kind of tutorial in the world of the child. It seems possible that children this age differentiate physical and psychological causality. Physical causality in the ordinary world of middle-sized objects has both spatial and temporal characteristics, there is 'physical contact' between the cause and effect. In the imitation game the infant 'causes' the adult to move in a particular way, but there is no physical contact between baby and adult. The child may interpret the perception of cause and effect without physical contact as psychological control or even communication. Such an ascription might be natural for the child when the agent is self and the recipient is another like-me human. If so, then the imitation game provides a situation in which to explore the parameters of psychological contact and communication. Just as hitting objects and watching them bump provides opportunities for exercising and enriching the child's 'naive physics', the imitation game provides opportunities for the exercise and development of the child's 'naive psychology'.

I want to return to the earliest phase of infancy. Young infants are known to be fascinated with other human beings; no toy can compete with people during the early phases of life; people are the infant's favourite playthings. We may inquire why even the youngest of infants are so fascinated by human beings. A simple learning view might be that there is nothing intrinsically special about other people to young babies; the human figure begins to command attention as it becomes associated with primary pleasures like food, warmth and comfort. A traditional nativist view might hold that there are built-in preferences for certain visual properties of human beings; perhaps the face gestalt is innately attractive.

In contrast, I would like to suggest that infants find human beings interesting because they have a primitive ability to recognize that the distinctively human movements they see are like the movements that they feel themselves make. It is not only, perhaps not even primarily, the features of the human form – these lips, these eyes – that attract attention and give humans special meaning. Rather, it is the fact that the spatio-temporal patterns of human body movements are, in a sense, 'familiar', in some very primitive way reminding babies of themselves. Of all the things in the newborn's visual field, it is only other human beings that will have this fascinating trait; neither inanimate objects, nor even other animals, will match in quite the same way. It is plausible that the infants' own perception of the self, coupled with a capacity for recognizing cross-modal similarities, may lead them to feel a primordial kinship with their fellow human beings.

ACKNOWLEDGEMENTS

For the title and philosophical guidance I am especially grateful to Alison Gopnik; for help in putting questions to babies and interpreting their reactions I owe an enormous debt to Patricia Kuhl and M. Keith Moore. Naomi Eilan and Bill Brewer helped me to sharpen my thoughts on many of the problems discussed here. This work was supported by a grant from NIH (HD22514).

NOTES

1 Infants' sucking reflex took precedence over any tendency to imitate. They did not tend to open their mouths and let the pacifier drop out during the mouth display; nor did they tend to push the pacifier away with their tongues during the tongue display. It was only after the pacifier was removed that the response was inaugurated, sometimes after a long period of motor inactivity coupled with careful inspection of the experimenter. Thus, the technique was effective in disrupting imitation when the target was perceptually present.

2 Infants spend most of their first month sleeping; they are in an awake and alert state less than five hours a day, with their hands swaddled much of the time.

3 Precisely what aspect of the auditory signal is needed? In further studies Kuhl, Williams and Meltzoff (1991) systematically dissected the speech signal into elementary parts. The principal findings showed that the cross-modal performance was not supported if parts of the acoustic signal that comprise speech ('distinctive features'; Jakobson, Fant and Halle, 1969) were provided as the auditory stimulus instead of the whole phonetic unit. The broader theoretical inference is that infants' cross-modal perception of speech does not originate through a process that progresses from 'simple parts' to 'wholes', in which infants initially relate faces and voices on the basis of a simple acoustic feature, and then gradually build up a connection between the two that involves, on the auditory side, an identifiably whole speech stimulus. These findings thus provide an instance in which young infants are responsive to the wholes, the phonetic unit *per se*, rather than to isolated components (the distinctive features of speech).

4 Formant frequencies are an acoustic property of speech pertaining to the frequencies where energy is concentrated.

5 The three pairs of actions were: (a) shake = slide, (b) pound = poke and (c) touch mouth with toy = touch non-oral region on the head, neck or shoulders. These pairs were chosen from an extensive video review which showed that these were six common 'action schemes' of infants this age, and that the acts within each pair were similar.

6 We have performed studies in which the adult demonstrated the gesture and then assumed a passive-face pose without using a pacifier. The results show that many infants will watch the display with fascination and only begin to inaugurate the matching response after the adult has stopped.

7 Two other alternatives also bear mention, and I thank Naomi Eilan, Bill Brewer, James Russell and others of the Cambridge Spatial Representation Workshop for highlighting them for me. (1) The imitating adult might also be of special interest because he is fulfilling the infant's action intentions; whenever the infant wills there to be a toy-in-the-mouth, this occurs (actually two such events occur, one for the infant and one for the imitating adult). (2) Infants might prefer looking at behavioural synchrony, two people doing the same thing, without regard to one of the actors being the self. However, the second alternative has difficulty explaining the infants' testing behaviour, in which infants suddenly deviate from the behavioural synchrony.

8 Mirror reflections could well be puzzling to young infants because this duality is not so easily resolved. In terms of spatial location the image is clearly non-self; yet it moves *completely* under the self's control, a characteristic of self. The resolution is to infer a reflecting surface and virtual image of the self.

9 The imitation game is a form of preverbal communication between adult and infant. It plays an important role in early enculturation, because a social mirror (unlike a physical mirror) is both selective and interpretive in its reflections. Parents, as social mirrors, provide 'creative reflections' to their infants – reflections that capture aspects of the infant's activity, but then go on beyond it to read in intentions and goals to that behaviour. The infant waves an object, but the parent interprets this as waving in order to shake, and therefore waves intensely enough to shake the toy and produce a sound, which in turn leads the infant beyond his or her initial starting point. Similarly, actions that are potentially meaningful in the culture, will

be reflected back more often than others (Bruner, 1975; 1983). Social communication via the imitation game begins a long time before verbal communication.

REFERENCES

Amsterdam, B. 1972: Mirror self-image reactions before age two. *Developmental Psychobiology*, 5, 297–305.

Bower, T. G. R. 1982: *Development in infancy*, (2nd ed.). San Francisco: W. H. Freeman.

Bruner, J. S. 1975: From communication to language – a psychological perspective. *Cognition*, 3, 255–87.

Bruner, J. S. 1983: *Child's Talk: Learning to use language*. New York: Norton.

Bryant, P. E., Jones, P., Claxton, V. and Perkins, G. M. 1972: Recognition of shapes across modalities by infants. *Nature*, 240, 303–4.

Butterworth, G. 1981: *Infancy and Epistemology: An evaluation of Piaget's theory*. Brighton: Harvester Press.

Evans, G. 1985: Molyneux's question. In his *Collected Papers*. Oxford: Clarendon Press, 364–99.

Jakobson, R., Fant, C. G. M. and Halle, M. 1969: *Preliminaries to Speech Analysis: The distinctive features and their correlates*. Cambridge, Mass: MIT.

Kagan, J. 1984: *The Nature of the Child*. New York: Basic Books.

Kuhl, P. K. 1979: Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 66, 1668–79.

Kuhl, P. K. 1983: Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior & Development*, 6, 263–85.

Kuhl, P. K. and Meltzoff, A. N. 1982: The bimodal perception of speech in infancy. *Science*, 218, 1138–41.

Kuhl, P. K. and Meltzoff, A. N. 1984: The intermodal representation of speech in infants. *Infant Behavior & Development*, 7, 361–81.

Kuhl, P. K. and Meltzoff, A. N. 1988: Speech as an intermodal object of perception. In A. Yonas (ed.), *Perceptual Development in Infancy: The Minnesota Symposia on Child Psychology*. Hillsdale, NJ: Erlbaum, vol. 20, 235–66.

Kuhl, P. K., Williams, K. M. and Meltzoff, A. N. 1991: Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 829–40.

Meltzoff, A. N. 1990a: Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory in infancy. In A. Diamond (ed.), *The Development and Neural Bases of Higher Cognitive Functions. Annals of the New York Academy of Sciences*, 608, 1–31.

Meltzoff, A. N. 1990b: Foundations for developing a concept of self: The role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In D. Cicchetti and M. Beehgly (eds), *The Self in Transition: Infancy to childhood*. Chicago: University of Chicago Press, 139–64.

Meltzoff, A. N. and Borton, R. W. 1979: Intermodal matching by human neonates. *Nature*, 282, 403–4.

Meltzoff, A. N., Kuhl, P. K. and Moore, M. K. 1991: Perception, representation, and the control of action in newborns and young infants: Toward a new synthesis. In M. J. Weiss and P. R. Zelazo (eds), *Newborn Attention: Biological constraints and the influence of experience*. Norwood, NJ: Ablex, 377–441.

Meltzoff, A. N. and Moore, M. K. 1977: Imitation of facial and manual gestures by human neonates. *Science*, 198, 75–8.

Meltzoff, A. N. and Moore, M. K. 1983: Newborn infants imitate adult facial gestures. *Child Development*, 54, 702–9.

Meltzoff, A. N. and Moore, M. K. 1989: Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25, 954–62.
O'Shaughnessy, B. 1980: *The Will: A dual aspect theory*. Cambridge: Cambridge University Press.
Piaget, J. 1954: *The Construction of Reality in the Child*. New York: Basic Books.
Piaget, J. and Inhelder, B. 1969: *The Psychology of the Child*. New York: Basic Books.
Spitz, R. A. 1965: *The First Year of Life*. New York: International Universities Press.