Developmental Science 2:3 (1999), pp 306–317

REPORT

A computational and neuropsychological account of object-oriented behaviours in infancy

Denis Mareschal,[1,2] Kim Plunkett[1] and Paul Harris[1]

1. *Oxford University, UK*
2. *Washington Singer Laboratories, Exeter University, UK*

Abstract

Infants under 7 months of age fail to reach behind an occluding screen to retrieve a desired toy even though they possess sufficient motor skills to do so. However, even by 3.5 months of age they show surprise if the solidity of the hidden toy is violated, suggesting that they know that the hidden toy still exists. We describe a connectionist model that learns to predict the position of objects and to initiate a response towards these objects. The model embodies the dual-route principle of object information processing characteristic of the cortex. One route develops a spatially invariant surface feature representation of the object whereas the other route develops a feature blind spatial–temporal representation of the object. The model provides an account of the developmental lag between infants' knowledge of hidden objects and their ability to demonstrate that knowledge in an active retrieval task, in terms of the need to integrate information across multiple object representations using (associative) connectionist learning algorithms. Finally, the model predicts the presence of an early dissociation between infants' ability to use surface features (e.g. colour) and spatial–temporal features (e.g. position) when reasoning about hidden objects. Evidence supporting this prediction has now been reported.

Newborns possess sophisticated object-oriented perceptual skills (Slater, 1995) but the age at which infants are able to reason about *hidden* objects remains unclear. Using *manual search* to test infants' understanding of hidden objects, Piaget concluded that it is not until 7.5–9 months that infants understand that hidden objects continue to exist because younger infants do not successfully reach for an object hidden behind an occluding screen (Piaget, 1952, 1954). More recent studies using a violation of expectancy paradigm have suggested that infants as young as 3.5 months do have some understanding of hidden objects. These studies rely on non-search indices such as surprise instead of manual retrieval to assess infant knowledge (e.g. Baillargeon, Spelke & Wasserman, 1985; Baillargeon, 1993). Infants *watch* an event in which some physical property of a hidden object is violated (e.g. solidity). Surprise at this violation (as measured by increased visual inspection of the event) is interpreted as showing that the infants know (a) that the hidden object still exists, and (b) that the hidden object maintains the physical property that was violated (Baillargeon, 1993). The nature and origins of this developmental lag between understanding the continued existence of a hidden object and searching for it remains a central question of infant cognitive development.[1]

The lag cannot be attributed to a delay in manual control because infants as young as 4.5 months reach for a moving visible object, and by 6 months can reach around or remove an occluding obstacle (Von Hofsten, 1980, 1989). Nor can it be attributed to immature

[1] Some recent studies have called into question the Baillargeon, Spelke and Wasserman (1986) findings on methodological grounds (e.g. Bogartz, Shinskey & Speaker, 1997). Although these studies pose a serious challenge to the interpretation of the original Baillargeon *et al.* work, numerous other studies (e.g. Baillargeon, 1993) continue to argue for precocious object knowledge when infants are tested using a surprise-based measure.

Address for correspondence: Denis Mareschal, Centre for Brain and Cognitive Development, Department of Psychology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK; e-mail: d.mareschal@bbk.ac.uk

planning or problem-solving abilities because infants have been shown to solve problems involving identical or more complex planning procedures (Baillargeon, 1993; Munakata, McClelland, Johnson & Siegler, 1997).

Clues may be found in recent work on cortical representation of visual object information. Anatomical, neurophysiological and psychophysical evidence points to the existence of two processing routes for visual object information in the cortex (Ungerleider & Mishkin, 1982; Van Essen, Anderson & Felleman, 1992; Goodale, 1993; Milner & Goodale, 1995). Although the exact functionality of the two routes remains a hotly debated question, it is generally accepted that they contain radically different kinds of representations. The dorsal (or parietal) route processes spatial–temporal object information, whereas the ventral (or temporal) route processes object feature information.

Cells in the dorsal stream encode information consistent with the presence of multiple spatial representations such as location in a continuously updated, body-centred frame or in a gaze dependent frame (Hietmen & Perrett, 1993; Rizzolati, Riggio & Sheliglia, 1994). Because objects move and disappear it is necessary to anticipate their hidden position to act effectively on them (Hietmen & Perrett, 1993). Although cells in the parietal cortex appear to be involved in tracking moving, visible objects, many also continue to respond during a brief occlusion (Newsome, Wurtz & Komatsu, 1988). Cells in other parts of the dorsal stream code the relative motion and size changes that accompany looming (Rizzolati *et al.*, 1994) as well as the object size, shape and orientation information needed for accurate reaching and grasping (Jeannerod, 1988). Some cells associated with visual fixation and reaching fire independently of whether the target object is desired or not (Rolls *et al.*, 1979). Thus, some spatial–temporal components are coded even if an object is undesired and reaching does not occur.

Cells in the ventral route have complementary properties to those in the dorsal route. They are sensitive to the figural and surface properties of objects as well as finer grained external form and shape information used to identify objects (Takane, 1992; Milner & Goodale, 1993). Many have large retinal receptive fields. Although they can process detailed feature information they lose much of the spatial resolution on the retina, effectively developing spatially invariant feature representations of objects. As information progresses downstream away from the retina, more cells respond to complex feature clusters. These characteristics optimize the recognition of objects, scenes and individuals with enduring features rather than transient changes in the visual array. Such transformation-invariant representations could support recognition memory (Milner & Goodale, 1995). The responsiveness of cells in the ventral stream can be modulated by the prior occurrence of a stimulus (Goodale, 1993) suggesting that some kind of feature memory trace remains.

The dorsal and ventral routes both project into the frontal lobes (Goodale, 1993). As a whole, the frontal lobes play a crucial role in learning what responses are appropriate given an environmental context (Passingham, 1993). They have been closely tied to the development of planning and underlie the execution of voluntary actions, particularly in the context of manual search by human infants (Diamond, 1991).

Voluntary retrieval such as manual search for an occluded object must involve the integration of spatial–temporal information concerning the location of the occluded object with surface feature information concerning its identity. The surface feature information is required to decide whether an object is desired or not, and spatial–temporal information is required to direct the response. Furthermore, the cortical representation of these two types of information must be sufficiently well developed for accurate integration to occur. We suggest that early in development only visible objects offer the degree of representational precision needed to support an accurate integrated response because cell activations diminish when a target is no longer visible.

This suggests an explanation for the developmental lag in manual retrieval. The lag occurs whenever it is necessary to integrate two potentially imprecise sources of information: (i) spatial–temporal information about the location of the occluded object and (ii) featural information about the identity of the occluded object. This explanation predicts that tasks requiring access to only one imprecise source of information or tasks that are performed with a visible object will not result in a developmental lag. In contrast, any task that calls for the integration of cortically separable representations will fail unless performed with a visible object or with precise cortical representations. This account does not attribute the lag to any difficulties the infant might encounter in attempting to remove or circumvent the occluder in manual retrieval tasks. In addition, the lag does not depend on the response modality. Instead, it arises from information processing considerations associated with voluntary, object-directed behaviours. Surprise reflex responses, which may subsequently be manifested by an increased inspection time or spontaneous visual search behaviours, can be elicited by access to only one of the object representations.

We constructed a connectionist computational model to explore this architecture and to investigate its developmental implications.

The model

Figure 1 shows the model in schematic outline. It consists of a modular architecture. Each functional module is enclosed by a dotted line. Some units are shared by two modules (e.g. the 75 hidden units are shared by the response integration and trajectory prediction networks) and serve as a gateway for information between the modules. In accordance with the neurological evidence reviewed above, spatial–temporal information about objects in the world is processed independently of feature information. Information enters the network through a two-dimensional retina homogeneously covered by feature detectors. It is then concurrently funnelled into one pathway that processes the spatial–temporal history of the object and another pathway that develops a spatially invariant feature representation of the object.

The retina consists of a 4 × 25 cell grid (Figure 2). Each cell contains four feature detectors responding to different properties (e.g. light/dark, high/low contrast, hot/cold, soft/hard). If a projected object image overlaps with a grid cell, the cell's feature detectors take on the value +1.0 if the feature is present and −1.0 if the feature is absent (cf. Treisman & Sato, 1990). Cells on which the object image is not projected are quiescent and take on the value 0.0. An occluding screen is also

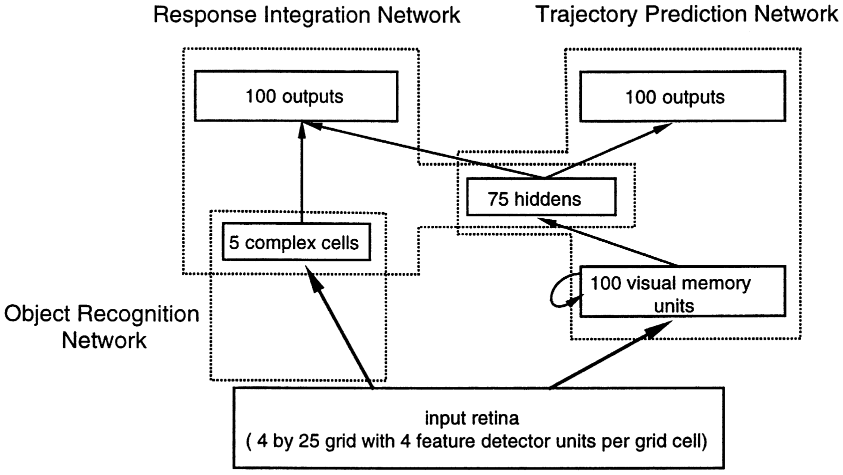
Response Integration Network
Trajectory Prediction Network
100 outputs
100 outputs
75 hiddens
5 complex cells
100 visual memory units
Object Recognition Network
input retina
(4 by 25 grid with 4 feature detector units per grid cell)

Figure 1 *Schema of the modular network architecture.*

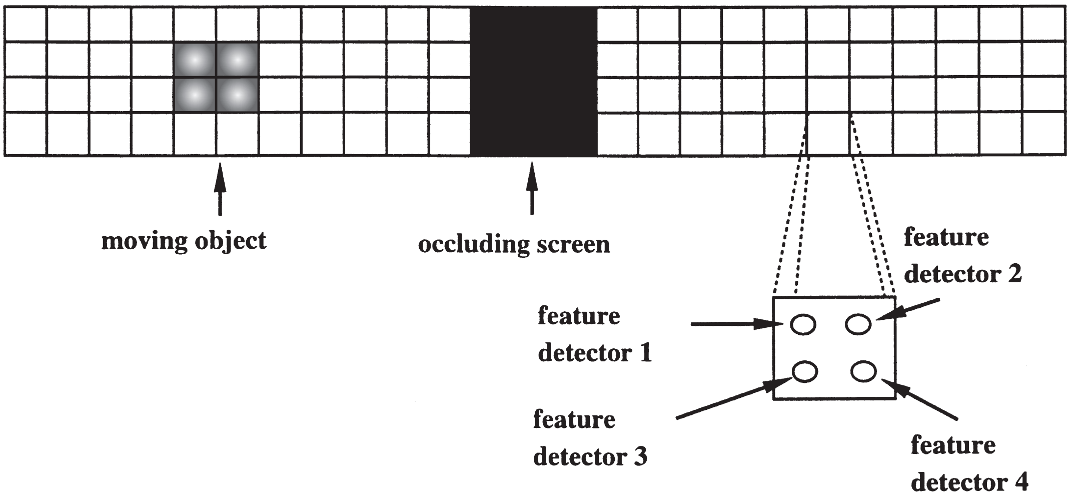

Figure 2 *Schema of the retinal structure.*

projected onto the retina. The retinal cells corresponding to the screen's image have a constant activation of 1.0.

The network experiences four different objects with correlated features (i.e. $\{-1\ 1\ -1\ 1\}$, $\{-1\ 1\ 1\ -1\}$, $\{1\ -1\ 1\ -1\}$, $\{1\ -1\ -1\ 1\}$). All object images are 2×2 grid cells large. For each object presentation, an object moves once back and forth across the retina, either horizontally or vertically. All horizontal movements across the retina involve an interim occluding event whereas vertical movements across the retina can result in either non-occluding or partially occluding events. Completely occluded vertical movements are never observed because the occluder height is identical to the height of the retina. At any specific time step there are four possible next positions for the object: up, down, left, or right. Predicting the next object position can only be resolved by learning to attend to the trajectory of the object.

The object recognition module generates a spatially invariant representation of the object by using a modified version of the unsupervised learning algorithm developed by Foldiak (Foldiak, 1991, 1996). This algorithm belongs to the family of competitive learning algorithms. Initially, a bank of five complex cells is fully and randomly connected to all feature detectors. The algorithm exploits the fact that an object tends to be contiguous with itself at successive temporal intervals. Thus, two successive images will probably be derived from the same object. At the end of learning each complex cell becomes associated with a particular feature combination wherever it appears on the retina.

Learning is constrained by three parameters: the learning rate ε determining the scale of any weight changes, the range around 0.0 of initial random weight values, and the memory δ determining the proportion of new activation used to update a complex cell's activation. Setting the activations of the losing units in the competitive phase to a small negative value $(-\beta)$ instead of 0.0 greatly increased the stability of the representations under continued training. The following parameter values were used: $\delta = 0.1$, $\beta = 0.02$, learning rate $\varepsilon = 0.001$, and weight range $= 0.05$. The impact on learning of varying these parameters is discussed in detail in Mareschal (1997).

The trajectory prediction module uses a partially recurrent, feedforward network trained with the back-propagation learning algorithm. All back-propagation networks in the model used the following parameter values: learning rate $\varepsilon = 0.1$ and momentum $\alpha = 0.3$. At each time step, information about the position of the object on the retina is extracted from the 100 retinal grid cells and mapped onto the visual memory layer. The retinal grid cells with which the object image overlaps become active (+1.0) whereas the other cells remain inactive (0.0). The network is trained to predict the next instantaneous, retinal position of the object. The prediction is output onto a bank of 100 units coding position in the same way as the inputs into the module. The network has a target of +1.0 for those units corresponding to the next object position and 0.0 for all other units.

All units in the visual memory layer have a self-recurrent connection (fixed at $\mu = 0.3$). The resulting spatial distribution of activation across the visual memory layer takes the form of a comet with a tail that tapers off in the direction from which the object has come. The length and distinctiveness of this tail depend on the velocity of the object. The information in this layer is then forced through a bottleneck of 75 hidden units to generate a more compact, internal re-representation of the object's spatial–temporal history. As there are no direct connections from the input to the output, the network's ability to predict the next position is a direct measure of the reliability of its internal object representation. We interpret the response of the trajectory prediction network as a measure of its sensitivity to spatial–temporal information about the object.

The output of the response integration network corresponds to the infant's ability to coordinate and use the information it has about object position and object identity. This network integrates the internal representations generated by other modules (i.e. the feature representation at the complex cell level and spatial–temporal representation in the hidden unit layer) as required by a retrieval response task. It consists of a single layered perceptron whose task is to output the same next position as the prediction network for two of the objects, and to inhibit any response (all units set to 0.0) for the other two objects. This reflects the fact that infants do not retrieve (e.g. reach for) all objects. In general, infants are not asked or rewarded for search. The experimental set-up relies on *spontaneous* search by the infant. Some objects are desired (e.g. sweet) whereas others are not desired (e.g. sour). Heightening the desirability of an object (e.g. by providing the infant with a prior opportunity to play with the object) has been shown to elicit more search in manual retrieval tasks (Harris, 1971). Any voluntary retrieval response will necessarily require the processing of feature information (to identify the object as a desired one) as well as trajectory information (to localize the object). Related arguments about the need to coordinate 'what' and 'where' information in object directed tasks have been presented elsewhere (e.g. Prazdny, 1980; Leslie, Xu, Tremoulet & Scholl, 1998).

The model embodies the basic architectural constraints on visual cortical pathways revealed by contemporary neuroscience: an object-recognition network that develops spatially invariant feature representations of objects, a trajectory-prediction network that is blind to surface features and computes appropriate spatial–temporal properties even if no actions are undertaken towards the object, and a response module that integrates information from the first two networks for use in voluntary actions. We suggest that surprise can be modelled by a mismatch between the information stored in an internal representation and the new information arriving from the external world. More specifically, in the trajectory prediction module, surprise occurs when there is a discrepancy between the predicted reappearance of an object from behind an occluder and its actual reappearance on the retina. In the object recognition module, surprise occurs when there is a discrepancy between the feature representation stored across the complex units and the new representation produced by the new image.

Model performance

Object localization

The trajectory prediction network learns very quickly to predict an object's next position when it is visible. However, the hidden unit representations that are developed persist for some time after the object has disappeared and allow the network to keep track of the object even when it is no longer directly perceptible. Figure 3 shows a graphic representation of the network's ability to predict the next position of an occluded object. The left-hand column shows what is projected onto the retina once feature information has been removed. The right-hand column shows the corresponding object position predicted by the trained trajectory network. The rows (from top to bottom) correspond to successive time steps. This network has seen 30 000 presentations of randomly selected objects moving back and forth in random directions at a fixed speed.

Both the screen and the object are projected onto the retina. The network correctly predicts the next position of the object even when the object is occluded by the screen and not directly perceptible. At $t = 0$, the object is about to disappear behind the occluding screen. At all subsequent time steps, the network correctly predicts that the object will have moved over one position. Note especially step $t = 3$ for which the direct perceptual information available to the network is *exactly the same* as at $t = 2$, in that only the occluding screen is visible. The network is able to predict the subsequent reappearance of the object, taking account of how long it has been behind

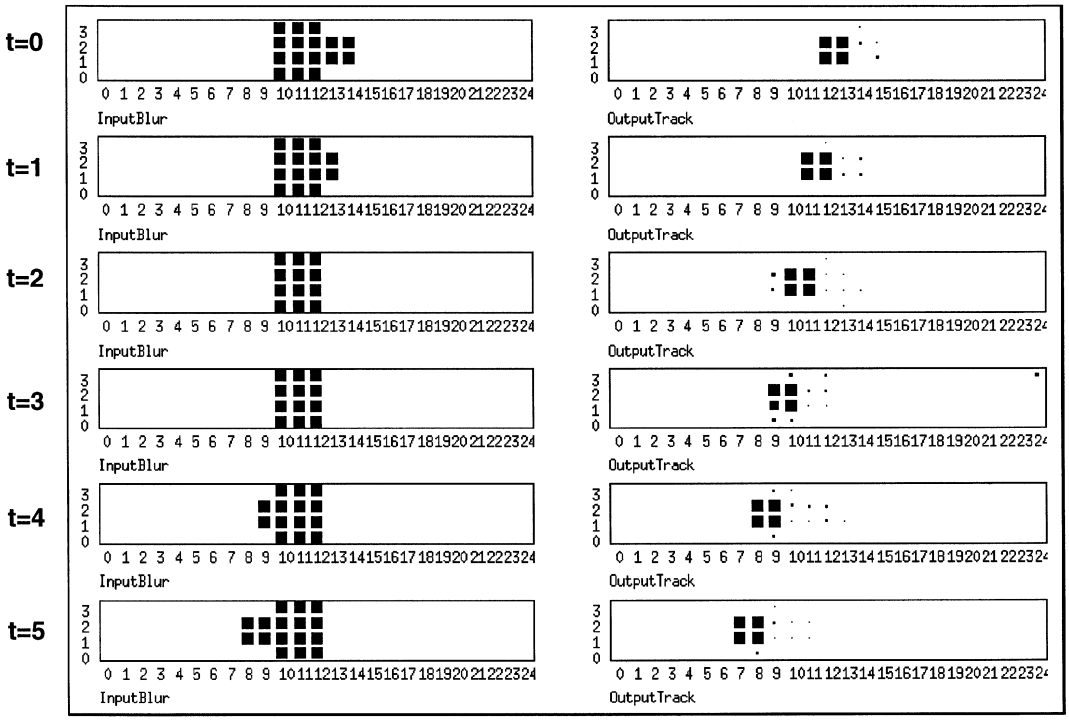

Figure 3 *Network tracking of an occluded object. The indices 0 through 5 down the left of the figure index successive time steps.*

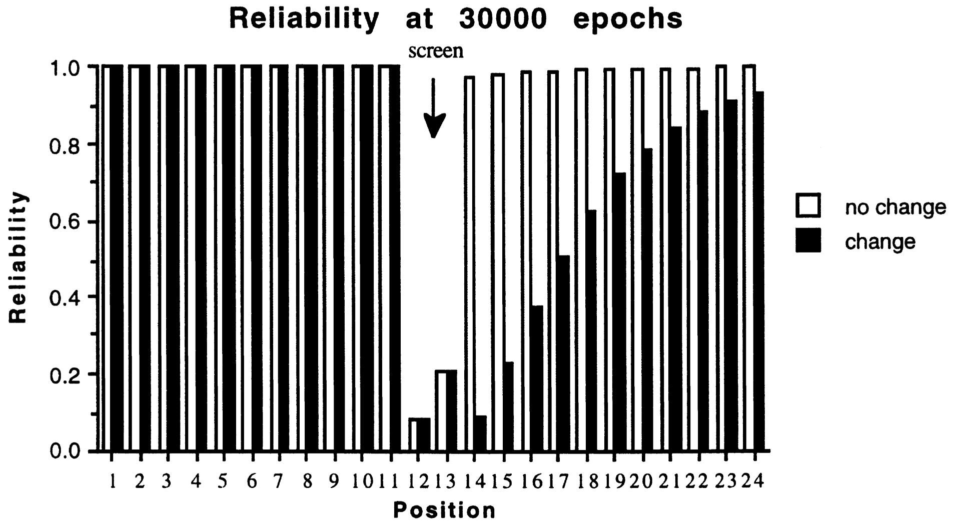

Figure 4 *Reliability of feature representation across complex cells.*

the screen[2]. Moreover, as found with infants (Muller & Aslin, 1978), the network's ability to track an occluded object depends on the length of the occluding screen: the longer the screen, the worse the performance.

Feature monitoring

The object recognition network also maintains a representation of the features of the object that persist beyond direct perception. Figure 4 shows the reliability of the internal feature representation developed across the complex cells. The reliability is computed as the dot product between the existing activation across the complex cells and the new activation pattern produced by the incoming feature input. It represents how similar the stored feature representation is to the new feature representation. Each of the columns represents the feature reliability as the object moves horizontally through the 24 positions for which the object image falls entirely on the retina. Positions 12 and 13 correspond to the object being fully hidden behind the screen. The white columns show performance when there are no changes in features whereas the black columns show the reliability when the object is surreptitiously changed behind the screen. When there are no changes, the reliability drops while the object is behind the screen (since there is no perceptual evidence with which to assess the internal representation) but recovers immediately when the object reappears. However, when the object is surreptitiously changed, there is a delayed recovery in reliability. This reflects the fact that the new object features are different from those that are stored in the recognition module's internal representations. The rate of recovery is directly related to the similarity between the new object features and the original object features. Effectively, delayed recovery corresponds to a surprise reaction.

Developmental lag in retrieval responses

The model was designed to test the hypothesis that the developmental lag between voluntary retrieval and surprise-based indices arises from the difference in the integration demands of the two tasks. Network responses when presented with an unoccluded desired object, an occluded desired object, and an occluded undesired object are depicted in Figure 5. The reliability of a module is computed as 1 − (sum-of-squares error of outputs) averaged over the output units and patterns involved in the event. Because the networks begin with random weights, the initial (untrained) output activations are also random. The initial network response is to turn off almost all output units. This results in an immediate increase in reliability (decrease in error) but it only reflects a blanket inhibition of output activity (including some cells which should be active). Hence, this stage of learning does not reflect the acquisition of position-specific knowledge. To normalize for this, the plotted reliabilities are linearly scaled to range between 0.0 and 1.0 with the origin of the scale (the baseline) corresponding to the reliability value obtained when all output units are turned off. Any increase in reliability above this origin corresponds to an increase in the

[2] Requiring the networks to predict the exact position of the hidden object makes explicit the richness of the spatial–temporal information encoded by the hidden units. This family of networks performs equally well on a task requiring them simply to predict the reappearance of the object from behind the screen.

ability to predict the object's next position. The baseline reliability value was 0.863 since on average about 86% of the units will be silent in producing an accurate response.

Figure 5 (a) shows the average network performance ($n = 10$) on both the position prediction and retrieval tasks when presented with an unoccluded, desired object. We interpret network behaviour by assuming that a threshold of reliability over and above the previously mentioned baseline level is required to control an accurate prediction/response. Consider the case where this threshold is set to 0.8. At this level, it can be seen from Figure 5 (a) that the network learns very quickly (within 1000 epochs) not just to predict the position of the desired object but also to produce an appropriate retrieval response.

When the object is occluded the network's behaviour is very different (Figure 5 (b)). Predictive localization and retrieval responses are initially equally poor. The internal representations are not adequately mature to support *any* reliable response. However, the reliability of tracking develops faster than that of retrieval. By 10 000 epochs the prediction response has achieved the requisite level of reliability whereas the retrieval response does not achieve this level until approximately 20 000 epochs. In other words, the network replicates the well-established finding that infants exhibit a developmental lag between successful predictive tracking of an occluded object and successful retrieval of an occluded object.

Of course, the threshold level we have used to interpret Figures 5 (a) and 5 (b) is only one among a range of possible values. Nevertheless, Table 1 shows that the success/failure of prediction and response at 15 000 epochs remains stable across a range of threshold levels. In particular, this pattern of results remains stable within the band 0.725–0.850. Below this range, virtually no developmental lag is observed between prediction and response for occluded objects. Above this level, the lag between prediction and response is not abolished even after 30 000 epochs.

The output required for retrieval of a desired, occluded object is identical to that required for predictive localization. Moreover, both sets of output units receive exactly the same information from the hidden units about the spatial–temporal history of the object. The two modules differ only in that the retrieval response module must *also* integrate information coming from the object recognition module. Thus, the developmental lag in the network arises from the added task demands of integrating information concerning the location and identity of an *occluded* object.

An advantage of modelling is that we can test this hypothesis directly using a manipulation which would

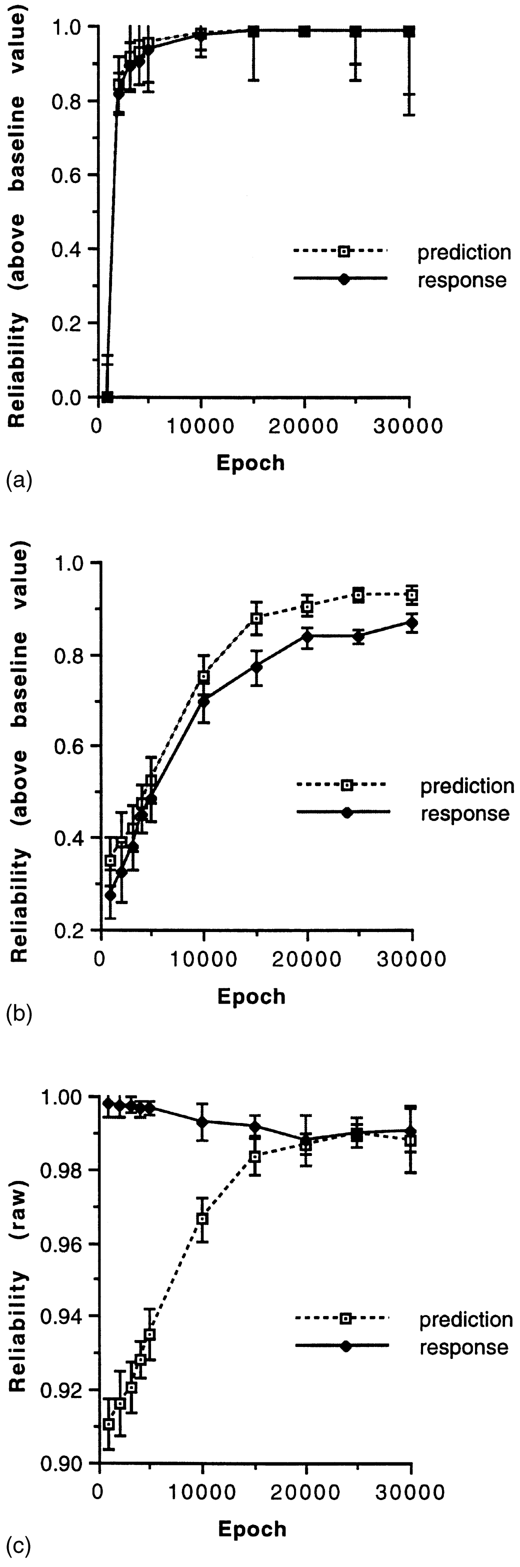

Figure 5 *Network performance on tracking and responding to (a) a desired unoccluded object, (b) a desired occluded object, and (c) an undesired occluded object. Standard deviations are also plotted.*

Table 1 *Reliability levels relative to a threshold criterion*

	Unoccluded desired object		Occluded desired object	
Threshold value	Prediction	Response	Prediction	Response
	5000 epochs			
0.975	N	N	N	N
0.950	Y	N	N	N
0.925	Y	Y	N	N
0.900	Y	Y	N	N
0.875	Y	Y	N	N
0.850	Y	Y	N	N
0.825	Y	Y	N	N
0.800	Y	Y	N	N
0.775	Y	Y	N	N
0.750	Y	Y	N	N
0.725	Y	Y	N	N
0.700	Y	Y	N	N
0.675	Y	Y	N	N
0.650	Y	Y	N	N
0.625	Y	Y	N	N
0.600	Y	Y	N	N
0.575	Y	Y	N	N
0.550	Y	Y	N	N
0.525	Y	Y	N	N
0.500	Y	Y	Y	N
	15 000 epochs			
0.975	Y	Y	N	N
0.950	Y	Y	N	N
0.925	Y	Y	N	N
0.900	Y	Y	N	N
0.875	**Y**	**Y**	**Y**	**N**
0.850	**Y**	**Y**	**Y**	*N*
0.825	**Y**	**Y**	**Y**	**N**
0.800	**Y**	**Y**	**Y**	**N**
0.775	**Y**	**Y**	**Y**	**N**
0.750	**Y**	**Y**	**Y**	**N**
0.725	**Y**	**Y**	**Y**	**N**
0.700	Y	Y	Y	Y
0.675	Y	Y	Y	Y
0.650	Y	Y	Y	Y
0.625	Y	Y	Y	Y
0.600	Y	Y	Y	Y
0.575	Y	Y	Y	Y
0.550	Y	Y	Y	Y
0.525	Y	Y	Y	Y
0.500	Y	Y	Y	Y
	30 000 epochs			
0.975	Y	Y	N	N
0.950	Y	Y	N	N
0.925	**Y**	**Y**	**Y**	**N**
0.900	**Y**	**Y**	**Y**	**N**
0.875	**Y**	**Y**	**Y**	**N**
0.850	Y	Y	Y	Y
0.825	Y	Y	Y	Y
0.800	Y	Y	Y	Y
0.775	Y	Y	Y	Y
0.750	Y	Y	Y	Y
0.725	Y	Y	Y	Y
0.700	Y	Y	Y	Y
0.675	Y	Y	Y	Y
0.650	Y	Y	Y	Y
0.625	Y	Y	Y	Y
0.600	Y	Y	Y	Y
0.575	Y	Y	Y	Y
0.550	Y	Y	Y	Y
0.525	Y	Y	Y	Y
0.500	Y	Y	Y	Y

Notes: Y and N code reliabilities that have or have not exceeded the threshold respectively. Bold marks threshold values for which a developmental lag arises.

not be possible with infants. If the lag is due to the need for information integration concerning the location and identity of an occluded object, then it should disappear on a task that does not require such integration. Undesired objects do not require information integration because it suffices to attend only to the identity representation in order to elicit an appropriate response. An inhibitory output can then be emitted which does not require any spatial–temporal information. Figure 5 (c) shows the network's performance when it is presented with an undesired object. Here, raw reliabilities are plotted because the correct response is to turn all output units off. The network learns to inhibit any attempt at retrieval because it can ignore information from the spatial–temporal channel even though it is still learning to predict the object's position. In summary, inspection of Figure 5(c) shows as predicted that the developmental lag disappears on tasks not requiring integration of information across modules.

Implications of model performance

The model is successful in demonstrating how the requirement to integrate information across two object representations in a voluntary retrieval task can lead to a developmental lag relative to performance on surprise tasks that only require access to either spatial–temporal information concerning an occluded object or surface feature information accessed separately. Early mastery of surprise tasks that claim to show the coordination of position and feature information (e.g. Baillargeon, 1993) have – on close scrutiny – provided evidence only for the use of positional information in conjunction with size or volume information. Both size and volume are spatial dimensions that are encoded by the dorsal route requiring access to only a single cortical route. Note that early surprise responses can arise from feature violations, from spatial–temporal violations and even from both types of violation arising concurrently and independently, but not from a violation involving the *integration* of feature and spatial–temporal information concerning an occluded object. The model predicts that infants will show a developmental lag not just on manual search tasks but also on surprise tasks that involve such integration.

Data supporting this prediction have recently become available. Infants fail to use surface feature information to individuate and enumerate objects that move behind and out from a screen (Simon, Hespos & Rochat, 1995; Xu & Carey, 1996). In these studies, infants watched two different objects move in and out (one at a time) from behind an occluder. The screen was subsequently removed to reveal either one or two objects. Young infants consistently ignored surface feature information and relied on spatial–temporal cues when assessing the number of objects behind the occluder as indexed by fixation time. Using a similar paradigm to Xu and Carey (1996), Wilcox (1997, submitted) systematically varied (one at a time) the features by which pairs of objects differed when they appeared from behind the occluding screen. She found that at 4.5 months infants will use shape and size information to individuate objects, but only at 7.5 months will they use surface texture information, and not until 11.5 months do they use colour to individuate objects. Note that shape is not a cortically separable object feature as it is processed in both the dorsal and ventral routes. Thus, an infant relying on the dorsal representation only can still access both shape and position information simultaneously. The age at which surface information (e.g. texture and colour) is used *in conjunction with* spatial–temporal information to monitor the number of hidden objects behind an occluder corresponds to the age at which infants begin to succeed at manual retrieval tasks (i.e. 7.5–9.5 months). This confirms the model prediction that infants should also show a developmental lag on surprise tasks that involve integration across cortically separable representations.

As noted earlier, the developmental lag in the model is not caused simply by the need to integrate spatial–temporal and featural information. The same integration demands are present when the network is required to respond to a desired, visible object. However, no lag is observed in this condition. Consistent with the model, infants reach accurately for moving visible objects as young as 4.5 months of age. In such cases, information is directly available in the perceptual array.

The developmental lag for occluded objects arises as a natural consequence of the associative learning process. Internal object representations developed over the complex cells and the hidden units persist when the object passes behind the screen, but decay with time. Hence, activation levels drop when the object is occluded. The learning algorithm updates network weights in proportion to the *sending unit's activation* level. For an identical error signal, the weight updates are smaller when the object is hidden given the lower activation of the sending units. Consequently, it will take longer to arrive at an equivalent level of learning for hidden compared with visible objects. This outcome is not unique to the learning algorithm used in the current model; it will arise in any learning mechanism that updates weights in proportion to the sending unit activation, providing a clear example of how developmental behaviours are constrained by micro-level mechanisms.

The model also predicts that infants will show an ability to respond to a conjunction of spatial–temporal and surface feature information when faced with unoccluded objects prior to their ability to use spatial–temporal information only when faced with an occluded object. Consider again Figures 5 (a) and 5 (b). Even though prediction reliability for occluded objects develops faster than retrieval reliability for occluded objects (Figure 5 (b)), prediction reliability for occluded objects develops slower than retrieval reliability for unoccluded objects (compare Figures 5 (a) and 5 (b)). We know of no empirical data that currently bear on this prediction and therefore offer it as a way of falsifying the model. One way to test this prediction would be to use a procedure initially developed to test infant memory for visual compounds (Cohen, 1973; Burnham, Vignes & Ihsen, 1988). Infants are habituated to two objects (e.g. a blue square at location A and a red square at location B). Infants are then tested with a stimulus in which the components are identical to those in the familiarization phase, but the compounds have been changed (e.g. a red square at location A and a blue square at location B). The model predicts that infants would dishabituate to a change in surface–feature/spatial–temporal compounds when presented with objects that are never occluded prior to their ability to use spatial–temporal information alone when faced with an occluded object.

Munakata *et al.* (1997) describe a connectionist model of infant object permanence behaviours. They argue that object representations develop gradually through interactions with an environment. Their model consists of a single-route network with a layer of hidden units, trained using back-propagation to predict the reappearance of an occluded object when a screen moves away. Although our model is also a connectionist model (and therefore also argues for graded object representations that develop through interactions with the environment) it differs significantly from the Munakata *et al.* model in both structure and performance.

The most significant difference is that the model described in this paper posits a representational dissociation between surface feature information and spatial–temporal information. This assumption is based on experimental findings suggesting that a similar functional dissociation exists in the cortex. The representational dissociation is at the heart of our account of the developmental lag. We suggest that it is the need to integrate information across the separate representations (coupled with an associative learning mechanism in which the connection weights are adjusted in proportion to the sending unit activation) that produces a developmental lag between infants' surprise and retrieval responses when faced with occluded objects. Munakata *et al.* account for the developmental lag by lowering the learning rate in the reaching portion of their network and delaying the training of the reaching module. This leads to the incorrect prediction of a developmental lag in the presence of visible as well as invisible objects.

The developmental origins of dual-route processing in the cortex remain a hotly debated issue in the neurosciences (e.g. Johnson, 1996). The separate pathways could constitute hardwired innate constraints on learning or they could emerge through competition for learning resources (e.g. Jacobs, Jordan & Barto, 1991). The current model is uncommitted as to whether the initial representational dissociation is present at birth or develops through learning. The critical assumption is that a cortical dissociation is present prior to the age at which infants begin to demonstrate knowledge of hidden objects.

In the current model, the trajectory prediction network is the module that determines the fastest possible rate of development of the response module. This is because it is slower to develop than the feature recognition module. Rueckl, Cave and Kosslyn (1989) have described a computational model of 'what' versus 'where' visual processing in which they argue that the 'what' module (analogous to our feature recognition module) is the slowest to develop because object identity is 'more difficult' to compute than object location. However, studies using visual event related potentials (ERPs) from infants engaged in object-related tasks suggest that the ventral pathway matures before the dorsal pathway, implying that the processing of surface feature information matures before the processing of spatial–temporal information (Johnson, 1998). Note that whatever module develops last will not change the main conclusions of the model concerning the origins of the developmental lag.

As in any modelling endeavour, a number of simplifications have been built into the model. The most important of these relate to the learning environment of the model. The current model only ever experiences the occluding screen in the same location. Testing the model with the screen in a different location would lead to a breakdown in performance. In contrast to this, humans can generalize their knowledge of occlusion in one location to that in another location. However, humans live in a much richer environment than the current model. Similarly, an enrichment of the network's learning environment (i.e. by giving it examples of occlusion in different locations) would also result in appropriate generalization. A second simplification is that the network only ever experiences one object at a time. Simultaneously processing information about multiple objects remains a central issue of current neural network research (e.g. Mozer, 1991; Shastri & Ajjanagadde, 1993). This is one direction in which future modelling should be directed.

It is important to note that this model does not capture the richness of *all* the behaviours that fall under the broad label of 'object permanence'. Indeed, we have intentionally avoided using the term object permanence to avoid confusion with a more mature concept of object-hood (Harris, 1983) and simple object-directed behaviours in infancy. This model deals exclusively with trajectory prediction, feature monitoring and retrieval responses. Nevertheless, these are constituents of an object concept on which the infant must build in order to develop a more mature level of competence.

Finally, it is also worth noting the close match between the performance of this model and adult neuropsychological data. There are documented cases of patients with ventral stream damage but an intact dorsal stream who suffer from a kind of visual form agnosia (they are unable to recognize objects based on shape information alone) and yet are able to reach accurately and even catch objects (Goodale, Milner, Jakobson & Carey, 1991; Milner & Goodale, 1995). After training, the feature recognition module of the model could be damaged in a way that does not affect its ability to respond with a targeted reach (Mareschal, 1997). As discussed above, shape (or form) can be encoded down both pathways so damage of the ventral stream shape processing does not interfere with shape processing in the dorsal stream.

In summary, we propose that changes in infant object-directed behaviours are determined by the developing reliability of multiple object representations encoding distinct aspects of visual object information. Connectionist models provide the tools to explore how neuropsychological organization constrains cognitive development. Models make explicit predictions that allow the testing of such neuropsychological accounts of behaviour. In particular, this model has suggested that performance on tasks that require the integration of cortically separable representations in the presence of occluded objects will be delayed compared with tasks that do not require such integration. This prediction has been independently confirmed for colour and texture information (Wilcox, 1997, submitted). We would predict that other surface features (such as faces) that are coded only in the ventral stream but not the dorsal stream should show similar dissociations. Conversely, a retrieval response (such as a reach) that was not modulated by a feature-based decision should occur at an earlier age than one that was modulated by a decision. In this case, spatial–temporal information would be sufficient to guide the response and no integration would be needed.

Acknowledgements

This research was funded in part by the Medical Research Council (UK) and the FCAR (Quebec).

References

Baillargeon, R. (1993). The object concept revisited: new directions in the investigation of infants' physical knowledge. In C.E. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 265–315). London: Lawrence Erlbaum.

Baillargeon, R., Spelke, E.S., & Wasserman, S. (1986). Object permanence in 5-month-old infants. *Cognition*, **20**, 191–208.

Bogartz, R.S., Shinskey, J.L, & Speaker, C.J. (1997). Interpreting infant looking: the event set × event set design. *Developmental Psychology*, **33**, 408–422.

Burnham, D.K., Vignes, G., & Ihsen, E. (1988). The effect of movement on infants' memory for visual compounds. *British Journal of Developmental Psychology*, **6**, 351–360.

Cohen, L.B. (1973). A two-process model of infant visual attention. *Merrill-Palmer Quarterly*, **19**, 157–180.

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey & G. Gelman (Eds), *The epigenesis of mind: Essays on biology and cognition* (pp. 67–110). Hillsdale, NJ: Lawrence Erlbaum.

Foldiak, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, **3**, 194–200.

Foldiak, P. (1996). Learning constancies for object perception. In V. Walsh & J. Kulikovski (Eds), *Visual constancies: Why things look as they do*. Cambridge: Cambridge University Press.

Goodale, M.A. (1993). Visual pathways supporting perception and action in the primate cerebral cortex. *Current Opinion in Neurobiology*, **3**, 578–585.

Goodale, M.A., Milner, A.D., Jakobson, L.S., & Carey, D.P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, **349**, 154–156.

Harris, P.L. (1971). Examination and search in infants. *British Journal of Developmental Psychology*, **52**, 469–473.

Harris, P.L. (1983). Infant cognition. In P.H. Mussen (Ed.), *Handbook of child psychology*, Vol. 2 (4th edn, pp. 689–781). New York: Wiley.

Hietmen, J.J., & Perrett, D.I. (1993). Motion sensitive cells in the macque superior temporal polysensory area 1: Lack of response to the sight of the monkey's own limb movement. *Experimental Brain Research*, **93**, 117–128.

Jacobs, R.A., Jordan, M.I., & Barto, A.G. (1991). Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. *Cognitive Science*, **15**, 219–250.

Jeannerod, M. (1988). *The neural and behavioural organization of goal-directed movements*. Oxford: Oxford University Press.

Johnson, M.H. (1996). *Developmental cognitive neuroscience*. Oxford: Blackwell.

Johnson, M.H. (1998). ERP studies of functional brain development in infants. *Proceedings of the 26th British Psychophysiology Society*.

Leslie, A.M., Xu, F., Tremoulet, P.D., & Scholl, B.J. (1998). Indexing and the object concept: developing 'what' and 'where' systems. *Trends in Cognitive Sciences*, **2**, 10–18.

Mareschal, D. (1997). Visual tracking and the development of object permanence: a connectionist enquiry. Unpublished Doctoral Dissertation, Oxford University.

Milner, A.D., & Goodale, M.A. (1993). Visual pathways to perception and action. *Progress in Brain Research*, **95**, 317–337.

Milner, A.D., & Goodale, M.A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Mozer, M.C. (1991). *The perception of multiple objects: A connectonist approach*. Cambridge, MA: MIT Press.

Muller, A.A., & Aslin, R.N. (1978). Visual tracking as an index of the object concept. *Infant Behavior and Development*, **1**, 309–319.

Munakata, Y., McClelland, J.L., Johnson, M.N., & Siegler, R.S. (1997). Rethinking infant knowledge: towards an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, **104**, 686–713.

Newsome, W.T., Wurtz, R.H., & Komatsu, H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. II: Differentiation of retinal from extraretinal inputs. *Journal of Neurophysiology*, **60**, 604–620.

Passingham, R.E. (1993). *The frontal lobes and voluntary action*. Oxford: Oxford University Press.

Piaget, J. (1952). *The origins of intelligence in the child*. New York: International Universities Press.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Prazdny, S. (1980). A computational study of a period of infant object-concept development. *Perception*, **9**, 125–150.

Rizzolati, G., Riggio, L., & Sheliglia, B.M. (1994). Space and selective attention. In C. Umilta & M. Moscovitch (Eds), *Attention and performance,* Vol. XV: *Conscious and non-conscious information processing* (pp. 231–265). Cambridge, MA: MIT Press.

Rolls, E.T., Perrett, D., Thorpe, S.J., Puerto, A., Roper-Hall, A., & Maddison, S. (1979). Response of neurons in area 7 of the parietal cortex to objects of different significance. *Brain Research*, **169**, 194–198.

Rueckl, J.G., Cave, K.R., & Kosslyn, S.M. (1989). Why are 'what' and 'where' processed by separate cortical systems? A computational investigation. *Journal of Cognitive Neuroscience*, **1**, 171–186.

Shastri, L., & Ajjanagadde, V. (1993). From simple association to systematic reasoning: a connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioural and Brain Sciences*, **16**, 417–494.

Simon, T.J., Hespos, S.J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn. *Cognitive Development*, **10**, 253–269.

Slater, A. (1995). Visual perception and memory at birth. *Advances in Infancy Research*, **9**, 107–162.

Takane, K. (1992). Inferotemporal cortex and higher visual functions. *Current Opinion in Neurobiology*, **2**, 502–505.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Psychological Reveiw*, **16**, 459–478.

Ungerleider, L.G., & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, & R.J.W. Mansfield (Eds), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Essen, D.C., Anderson, C.H., & Felleman, D.J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science*, **255**, 419–423.

Von Hofsten, C. (1980). Predictive reaching for moving objects by human infants. *Journal of Experimental Child Psychology*, **30**, 369–382.

Von Hofsten, C. (1989). Transition mechanisms in sensorimotor development. In A. de Ribaupierre (Ed.), *Transition mechanisms in child development: The longitudinal perspective* (pp. 223–259). Cambridge: Cambridge Univeristy Press.

Wilcox, T. (1997, April). 4.5 and 7.5-month-old infants' use of shape, color, and size when reasoning about object identity. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Washington, DC.

Wilcox, T. (submitted). Object individuation: Infants' use of shape, size, pattern, and color.

Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, **30**, 111–153.

Received: 7 October 1998
Accepted: 18 February 1999