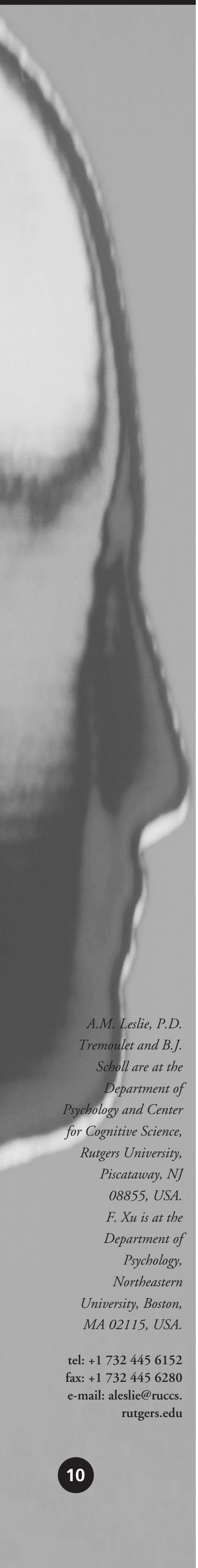
feature maps *Biol. Cybern.* 43, 59–69
40 Comon, P. (1994) Independent component analysis, a new concept? *Signal Process.* 36, 287–314
41 Bell, A.J. and Sejnowski, T.J. (1997) in *Advances in Neural Information Processing Systems* (Vol. 9) (Mozer, M.C., Jordan, M.I. and Petsche, T., eds), pp. 831–837, MIT Press
42 Pham, D.T. and Garat, P. (1997) Blind separation of a mixture of independent sources through a quasi-maximum likelihood approach *IEEE Trans. Signal Process.* 47, 1712–1725
43 Amari, S., Cichocki, A. and Yang, H.H. (1996) in *Advances in Neural Information Processing Systems* (Vol. 8) (Touretzky, D.S., Mozer, M.C. and Hasselmo, M.E., eds), pp. 757–763, MIT Press
44 Makeig, S. *et al.* (1997) Blind separation of auditory event-related brain responses into independent components *Proc. Natl. Acad. Sci. U. S. A.* 94, 10979–10984
45 Breiman, L. (1996) Bagging predictors *Machine Learning* 24, 123–140
46 Freund, Y. and Schapire, R.E. (1997) A decision-theoretic generalization of on-line learning and an application to boosting *J. Comput. Syst. Sci.* 55, 119–139
47 Leisch, F. and Hornik, K. (1997) in *Advances in Neural Information Processing Systems* (Vol. 9) (Mozer, M.C., Jordan, M.I. and Petsche, T., eds), pp. 522–528, MIT Press



Indexing and the object concept: developing 'what' and 'where' systems

Alan M. Leslie, Fei Xu, Patrice D. Tremoulet and Brian J. Scholl

The study of object cognition over the past 25 years has proceeded in two largely non-interacting camps. One camp has studied object-based visual attention in adults, while the other has studied the object concept in infants. We briefly review both sets of literature and distill from the adult research a theoretical model that we apply to findings from the infant studies. The key notion in our model of object representation is the 'sticky' index, a mechanism of selective attention that points at a physical object in a location. An object index does not represent any of the properties of the entity at which it points. However, once an index is pointing to an object, the properties of that object can be examined and featural information can be associated with, or 'bound' to, its index. The distinction between indexing and feature binding underwrites the distinction between object individuation and object identification, a distinction that turns out to be crucial in both the adult attention and the infant object-concept literature. By developing the indexing model, we draw together two disparate sets of literature and suggest new ways to study object-based attention in infancy.

A.M. Leslie, P.D. Tremoulet and B.J. Scholl are at the Department of Psychology and Center for Cognitive Science, Rutgers University, Piscataway, NJ 08855, USA.
F. Xu is at the Department of Psychology, Northeastern University, Boston, MA 02115, USA.

tel: +1 732 445 6152
fax: +1 732 445 6280
e-mail: aleslie@ruccs.rutgers.edu

> 'We readily suppose an object may continue individually the same, though several times absent from and present to the senses; and ascribe to it an identity, notwithstanding the interruption of the perception, whenever we conclude, that if we had kept our eye or hand constantly upon it, it would have conveyed an invariable and uninterrupted perception.'
>
> David Hume, *A Treatise of Human Nature*, 1740.

There is a long-standing view that the notion of objecthood is one of the fundamental structures of human thought[1–6]. Physical objects are a major focus of human attention in the first year of life, and structure visual attention in adults[7,8]. We present a new theory of the 'object-concept' in infancy, drawing inspiration from ideas developed in the study of adult visual attention. According to our framework, a key component of object cognition is an internal representation which functions as an 'index' to a physical object in the world. Just as a finger that points at something conveys no information about the nature of what it points at, so too an 'object index', in our account, is an entirely abstract representation that conveys no information about the properties of the object involved.

Objects and indexes

Although making best use of our limited processing resources demands selective attention[9,10], it is likely that attention can span more than a single object at a time.

 1364-6613/98/$19.00 PII: S1364-6613(97)01113-3

Without this, it would be hard to compare, or grasp a relation between, two or more objects. Our cognitive design may allow a compromise in which a small set of objects can be simultaneously attended to. In fact, we will suggest that our object indexing mechanism is limited to four indexes.

Indexing forms the basis for the infant's object concept through its role in the individuation, identification and enumeration of physical objects. Individuation establishes the notions 'single object' and 'more than a single object', while identification establishes the notion 'self-same one'. Existing evidence (reviewed below) suggests that infants have a limited ability to enumerate sets of 'simultaneously individuated' objects. We hypothesize that object indexing is the source of the limit on infant enumeration.

According to our model, the infant brain can simultaneously track a small set of physical objects, assigning to each object a 'pointer' from small set of mental indexes. Because an index does not provide information about the properties of the object at which it points, property information has to be associated with or 'bound' to the index.

The 'object concept' as a mechanism of attention

According to Piaget[6], infants cannot represent objects unless they are currently perceived; consequently, occlusion annihilates objects for infants. Bower[11,12] argued instead that infants' responses to appearing and disappearing objects reflect immature attempts to assign, or to re-assign, numerical identities to the objects. Studies of visual attention in infancy have supported Bower's general view. Even young infants continue to attend to occluded objects[13], to attribute properties to hidden objects[14], to apprehend physical objects as cohesive bounded three-dimensional volumes that trace continuous paths in space and time[4,5], to attribute causal roles to objects in dynamic interactions[15–17], and to represent the numerosity of small sets of objects[18].

We analyze the development of the 'object concept' as the development of mechanisms of object-based attention. We first introduce our notion of object indexing, and review key ideas from the adult visual attention literature that inform our model; we then apply the model to recent infancy findings.

Object indexing

The central idea in our framework is the notion of an 'object index'. We hypothesize a number of properties.

First, an object index is a mental token that functions as a pointer to an object.

Second, an index does not inherently represent any of the properties or features of the object pointed at. A finger can point at an object but the finger itself does not say whether the object is red or green, what shape it has, or what kind of object it is. Likewise, an object index is an entirely abstract representation of what it points at. If featural or property information is to accompany the object representation, such information must be specifically associated with or 'bound' to the index.

Third, object indexing is a mechanism of selective attention and is therefore resource-limited: available indexes form a small fixed set. Although we are unsure how many indexes infants have available, we believe that it is not more than four. We hypothesize that object indexes play an important role in infant sensitivity to the specific number of objects in small sets.

Fourth, indexes are assigned to objects primarily by location. Although indexes are assigned by location, they are not assigned to the locations themselves but to the objects in the locations.

There are four basic principles of index assignment: (1) a distinct object can attract only a single index; (2) once assigned, an index sticks to the object even as the object moves through different locations in space; (3) distinct indexes may be assigned to objects if and only if the objects occupy distinct locations in space; (4) because there is a small number of indexes, they need to be reused. An already assigned index must be de-assigned before it can be used with a new object.

If locational information is unavailable or ambiguous, then index assignment can be determined by property information. This might not be possible early in development.

The idea of an object-indexing process that picks out an individual object by location, but that disregards featural properties of the object, recalls two sets of ideas from the visual attention literature. First are recent theories of object-based visual attention mechanisms, most notably, Pylyshyn's 'FINST' theory (see Box 1) and Kahneman and Treisman's 'object file' theory (see Box 2). Second is the distinction between the 'what' and the 'where' systems of visual processing[19–23]. There is a body of evidence suggesting that featural information and locational information are processed, to a large extent independently, by anatomically distinct circuits in the brain. Featural information is processed mainly in circuits linking primary visual cortex, through extrastriate cortex, to the inferior temporal cortex, while information about the location of visual objects is processed mainly in a stream running from striate to parietal cortex.

We assume that the object indexing system allows a small number of objects to be indexed simultaneously. The indexes provide immediate access to the objects' locations, even though the objects may be in motion. In the case of occluded objects, the indexes allow access to the approximate location of the objects (e.g. they point to 'somewhere behind the occluder'). We also assume that the initial assignment of indexes to objects, that is, object individuation, is determined on the basis of the locations objects occupy. Finally, the indexes by themselves do not encode information about the nature of the objects they index (e.g. they do not provide featural information). For all these reasons, we assume that the 'where' system plays a fundamental role in object indexing (Fig. 1A).

Studies of adult visual attention have also uncovered relevant properties of the 'what' system. For example, a visual object that differs by a single feature or 'texton' from its background (e.g. a green circle among red circles) can be detected rapidly, pre-attentively, without effort, and independently of the number of distractors by a parallel search process. This is called visual 'pop-out'[24–30]. By contrast, the detection of a target that is defined by a conjunction of features (e.g. a green circle among green crosses and red circles) or by the absence of a feature (e.g. an O among Q's[31])

Box 1. The 'FINST' visual indexing theory

Visual attention imposes a limitation on our capacity to process visual information, but it is not clear at the outset what the correct units are for characterizing this limitation. It was traditionally argued, or assumed, that attention simply restricts various types of visual processing to certain areas of the visual field – in the manner of a spotlight[a,b] or a zoom-lens[c]. It has recently been demonstrated, however, that there must also be an object-based component to visual attention, in which attentional limitations are characterized in terms of the number of pre-attentively defined discrete objects that can be simultaneously processed[d,e].

Several recent theories have attempted to explain how objects are individuated and continuously represented by processes of visual attention (e.g. Treisman's object file theory; Box 2). Pylyshyn's 'FINST' theory of visual indexing[f,g,h] complements these other theories by attempting to describe the mechanism by which the earliest object-based representations are realized.

In order to detect even simple geometrical properties among the elements of a visual scene (e.g. being inside, or being collinear), Pylyshyn[f,g] argues that the visual system must be able to simultaneously reference, or 'index', multiple objects. Similarly, although focal attention can be scanned in a prescribed direction until it finds objects, it cannot focus on a particular object unless the location of that object has already been indexed. Pylyshyn's model is based on visual indexes that can be pre-attentively assigned to various items in the visual field, and serve as a means of access to those items for higher-level processes that allocate focal attention. In this regard, they function like pointers in a computer data structure; they reference certain items in the visual field (identifying them as distinct objects), without themselves accessing the properties of those objects.

The visual indexes in Pylyshyn's theory were historically called FINSTs ('Fingers of INSTantiation') because physical fingers work in an analogous way: 'Even if you do not know anything at all about what is located at the places that your fingers are touching, you are still in a position to determine such things as whether the object that finger number 1 is touching is to the left of or above the object that finger number 2 is touching… [T]he access that the finger contact gives makes it inherently possible to track a particular token, that is, to keep referring to what is, in virtue of its historical trace, the same object'[f].

FINSTs can be assigned to objects in the visual field regardless of their spatial contiguity (in contrast to a spotlight model), with the following restriction: the architecture of the visual system provides only about four FINSTs. Furthermore, FINSTs are sticky; if an indexed item in the visual field moves, the FINST moves with it. FINSTs bestow a processing priority to the indexed items, and processes of focal attention can thus be applied to the (possibly moving) items without first engaging in a serial search for them (Ref. i; J.M. Intriligator, PhD thesis, Harvard University, 1997).

Several different experimental tasks have been used to adduce support for the FINST indexing framework, including evidence from multiple-object tracking, subitizing, visual search, and the line-motion illusion[g,h]. Of particular relevance here is the recent demonstration that the FINST indexing system recognizes and makes allowances for occlusion, suggesting that the indexes operate on perceptual representations rather than retinally-based representations[j].

References

a Eriksen, C.W. and Hoffman, J.E. (1972) Temporal and spatial characteristics of selective encoding from visual displays *Percept. Psychophys.* 12, 201–204

b Posner, M., Snyder, C. and Davidson, B. (1980) Attention and the detection of signals *J. Exp. Psychol.* 109, 160–174

c Eriksen, C.W. and St James, J.D. (1986) Visual attention within and around the field of focal attention: a zoom lens model *Percep. Psychophys.* 40, 225–240

d Kanwisher, N.G. and Driver, J. (1992) Objects, attributes, and visual attention: which, what, and where *Curr. Dir. Psychol. Sci.* 1, 26–31

e Egeth, H. and Yantis, S. (1997) Visual attention: control, representation, and time course *Ann. Rev. Psychol.* 48, 269-297

f Pylyshyn, Z.W. (1989) The role of location indexes in spatial perception: a sketch of the FINST spatial index model *Cognition* 32, 65–97

g Pylyshyn, Z.W. (1994) Some primitive mechanisms of spatial attention *Cognition* 50, 363–384

h Pylyshyn, Z.W. *et al.* (1994) Multiple parallel access in visual attention *Can. J. Exp. Psychol.* 48, 260–283

i Pylyshyn, Z.W. and Storm, R. (1988) Tracking multiple independent targets: evidence for a parallel tracking mechanism *Spatial Vis.* 3, 179–197

j Scholl, B.J. and Pylyshyn, Z.W. (1997) Tracking multiple items through occlusion: clues to perceptual objecthood *Technical Report No. 42*, Rutgers University Center for Cognitive Science

requires the serial scrutiny of each potential target, is relatively effortful, and requires a search time that increases linearly with the number of distractors.

The integration of featural information with object location information appears to take a number of steps and is not achieved in a unitary process. Some of these processes are parallel and pre-attentive, while others operate serially and require attention or 'scrutiny'[24,32]. In object file theory, featural information is added at a later stage to an already active object representation[33].

An important issue is the relation between 'perceptual objecthood' (the focus of current theories of visual attention) and 'conceptual objecthood' (the traditional focus of infant studies) (see Box 3). In the concept of the physical object, judgments of object identity are based on both spatiotemporal and featural grounds (e.g. the notion of an object 'sortal'). If a ball and a cup take turns appearing from an occluder and disappearing behind it again, we judge that there are two distinct objects rather than one (whose features change). This judgment is based on featural differences alone because the two objects are never seen simultaneously to occupy different locations. Object indexing must therefore allow purely featural information to drive index assignment. Nevertheless, spatiotemporal information is primary for object indexing; only when such information is ambiguous (e.g. because of occlusion) does featural information determine individuation. We need to introduce a contrast between individuating-by-location and individuating-by-feature. Infant responses to spatiotemporally ambiguous situations, which we discuss below, suggest that individuation-by-feature might develop somewhat later in infancy. In order to capture the later

Box 2. The 'object file' theory of visual indexing

Like Pylyshyn's 'FINST' visual indexing theory (Box 1), Kahneman and Treisman's object-file theory[a–c] attempts to describe the nature of object-based representations of visual attention. (Note that we assume, as do Kahneman and Treisman, that visual indexes and object files are both parts of a single indexing system.)

One traditional idea in psychology is that visual stimuli are identified as objects when their visual projections activate semantic representations in long-term memory (LTM). Visual experience, on this view, consists simply in various shifting patterns of this type of LTM activation. Treisman and her colleagues call this the 'display-board model of the mind', and note that it has a number of distressing shortcomings[c]. It appears to be the case, for instance, that objects can be perceived and tracked through space even when they remain unidentified. Furthermore, when objects are initially mis-identified, and later correctly recognized, there is still never any doubt that the object involved was the same object: 'Two identical red squares in successive fields may be perceived as distinct objects if the spatial/temporal gap between them cannot be bridged, but the transformation of frog into prince is seen as a change in a single visual object'[c]. Identification *of* a particular object, in other words, is distinct from identification *as* an object in the first place.

Kahneman and Treisman have argued that an intermediate level of representation is needed to mediate this latter task. In their theoretical framework, this role is played by object files. According to their theory, attending to an object in the visual field causes a temporary representation called an object file to be created. This object file stores information about the object's properties, and this information is continually updated when the world changes. Each time an object's properties change, the new state of the object is compared with the previous state of the object file. If these two states are similar enough, then the object is seen as continuous, and the object file is updated appropriately. If the two states are dissimilar enough, however, then the previous object file decays, and a new object file is opened to represent the 'new' object.

Initially, when an object file is first 'opened', it contains only spatiotemporal information about the object's location. Later in on-line processing, however, further featural properties (e.g. the object's color or shape) may be added to the file. In this way, object files are quite distinct from LTM representations. They store information about objects' properties, but the files as wholes are addressed only by location, and will quickly decay once the corresponding objects leave the visual field. Object files do, however, survive motion and apparent motion[c], while the question of occluded motion has not been studied.

References

a Kahneman, D. and Treisman, A. (1984) in *Varieties of attention* (Parasuraman, R. and Davies, D.R., eds), pp. 29–61, Academic Press

b Treisman, A. (1988) Features and objects: the 14th Bartlett memorial lectures *Q. J. Exp. Psychol.* 40, 201–237

c Kahneman, D., Treisman, A. and Gibbs, B.J. (1992) The reviewing of object files: object-specific integration of information *Cognit. Psychol.* 24, 174–219

system, we need to modify the model in Fig. 1A (see Fig. 1B).

We listed above a number of reasons why location information plays a fundamental role in our indexing model of object individuation. The decisive reason for this design property may be learnability[34]. Some kinds of objects can undergo radical featural changes over time while retaining their identity (e.g. seedlings to trees, artefacts under repair, rotating objects). Tracking identity over featural change probably allows children to learn facts about particular object kinds.

The distinction between 'what' and 'where' neural systems may underpin the functional distinction between, respectively, the identification and individuation of perceptual objects[20]. The exact relationship between neural systems and functional models of attention, however, remains unclear. Nevertheless, the developing integration of these neural systems[35] should be a prime suspect in accounting for changes in infant object cognition. An important step in investigating this relationship is to construct a functional model of developing object attention in infants.

A model of infant object indexing

We align hitherto separate theoretical frameworks by distilling key ideas from adult object-based attention into a model of the infant object concept. Our model's basic claim (Fig. 2) is that the individuation of a physical object entails the assignment of a mental index that points at the object. Assignment is primarily determined on spatiotemporal grounds ('individuation by location'). Featural and, more generally, property information influences the indexing process in two distinct ways: *individuation*-by-feature and *identification*-by-feature.

In our model, visual features, such as color and shape, are registered in a 'feature map'[25,26]. The feature map provides indexing with two types of output. The first type of output simply tells indexing whether or not salient new features have appeared. The 'old/new' message from the feature map signals detection of a feature difference without

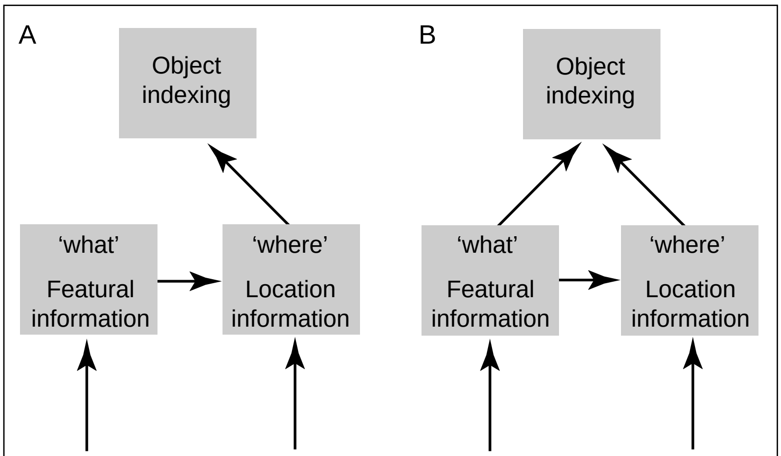


Fig. 1 Models of object indexing. (A) Objects are indexed by location only; there is no way to individuate objects by feature alone. **(B)** Object indexing by feature as well as by location. Dual-route indexing might develop later in infancy than indexing by location alone.

Box 3. The relationship between perceptual and conceptual objecthood

One general issue raised by our approach concerns the relationship between perceptual objecthood and the concept of 'physical object'. In the infancy literature 'physical objects' are countable, persistent three-dimensional volumes with mechanical properties[a,b]. In the adult attention literature, the 'objects' studied are almost always transitory, two-dimensional texture elements produced on a cathode ray tube. What relationship should one expect between these two notions of objecthood? One possibility is that perceptual objecthood develops 'out of' conceptual objecthood. Another is that the two develop independently. A third possibility is that both literatures study the same mechanisms that, though adapted through evolution for cognizing 'real' physical objects, nevertheless operate seamlessly on the transitory entities that flit across a phosphor screen. Object indexing provides a framework for addressing this issue.

Can we simply identify object indexes with either FINSTs or object files? There are reasons to think not, so far as these theoretical entities are currently understood. Firstly, both FINSTs and object files serve to individuate perceptual episodes, whereas object indexes individuate objects as physical objects. The reappearance of a physical object from behind a screen might form a new perceptual episode – but it is not a new physical object. Secondly, in object indexing, a distinct object attracts a single index. But FINSTs can be assigned such that a given object attracts one FINST while a part of that object attracts a second FINST (when binding the variables of the predicate 'is-part-of')[c]. Thirdly, FINSTs operate after parallel pre-attentive visual processes but before serial focal attention to provide 'the attentional focus information about where the item is in the retinal image' (Ref. d, p. 72, but see Ref. j, Box 1). Object indexes in our framework continue to track objects that are completely occluded. In such cases, the item does not appear on the retina. In occlusion, the locational access that the index provides might be only an approximate region. Perhaps more important than occlusion for distinguishing FINSTs and object indexes will be the time scale of indexing through occlusion. If FINSTs decay rapidly following occlusion, object indexes will have to endure for relatively long periods of time.

A reappearing, previously indexed object should pull its index with it. The 'stickiness' of the index is needed to track the numerical identity of the object. However, stickiness is not a complete solution. Consider the situation where one of two previously indexed objects behind a screen reappears. Which one? Without consulting featural information, it is not possible to tell. Object file theory has never specified that featural information alone can cause the opening of an object file. In our model, featural information must be able to influence index assignment. However, this might be a property of a relatively mature processing system and thus occur later, in a developmental sense.

References

a Leslie, A.M. (1994) in *Mapping the mind: domain specificity in cognition and culture* (Hirschfeld, L. and Gelman, S., eds), pp. 119–148, Cambridge University Press

b Spelke, E.S., Phillips, A. and Woodward, A.L. (1995) in *Causal cognition: a multidisciplinary debate* (Sperber, D., Premack, D. and Premack, A.J., eds), pp. 44–78, Clarendon Press

c Pylyshyn, Z.W. (1989) The role of location indexes in spatial perception: a sketch of the FINST spatial index model *Cognition* 32, 65–97

d Trick, L.M. and Pylyshyn, Z. (1994) Cueing and counting: Does the position of the attentional focus affect enumeration? *Visual Cognit.* 1, 67–100

indicating what that difference is. This limited information is sufficient to support individuation decisions. What it will not do is support identification judgments because these depend upon specifying the features. The second type of output permits identification-by-feature: specific featural information is obtained from the feature map and bound to a particular index. Feature binding occurs after index assignment (individuation), if it occurs at all.

Consider the distinction between individuation and identification: an observer watches as different objects are, one by one, removed from a bag, briefly shown, then replaced in the bag. The observer is then asked two questions: an individuation question, 'How many objects were shown?', and an identification question, 'What objects were shown?'. The first question requires that the observer count the objects. Because they are shown one at a time, the observer must increment the count only if a new feature is perceived (so the same object is not counted twice). The observer can then discard featural information and still know the number of objects shown. Remembering that there were six objects is easier than remembering what each object was (a ball, a pencil, …) or what color and shape each had. The latter information, however, must be remembered in order to identify the objects and answer the second question. To answer the first question, featural information is used only to individuate the objects and is then discarded; to answer the second question – to identify the objects – featural information must be retained.

We capture the distinction between object individuation and identification in terms of feature binding. The assignment of 'raw' indexes alone will suffice to individuate objects. But if these objects are later to be identified, then appropriate featural information must be associated with each index.

The distinction between individuation and identification echoes the contrast drawn by Sagi and Julesz (Ref. 36, p. 620) between feature detection and feature identification. Sagi and Julesz showed that adults could rapidly count (subitize) targets that differed from a large array of distractors by a single feature (in this case, orientation). Subjects were considerably slower to discriminate whether or not the targets differed from each other. Furthermore, whereas the amount of time subjects took to count targets was not dependent on their number, the time required to discriminate among them was. Sagi and Julesz concluded that target individuation and target identification require different mechanisms. Although processing times in Sagi and Julesz's study and in the studies of infants we examine below differ by two orders of magnitude, a similar distinction between individuation and identification seems relevant to both.

While drawing upon theories of adult visual attention, our framework is distinct: (a) in a number of respects related to the requirements of indexing physical objects as opposed to perceptual objects (Box 3); and (b) in its focus on the development of these mechanisms.

Key findings on infants' object cognition

We now discuss some recent studies of infants from the point of view of indexing theory. All the studies used measures of infant looking-times to expected and unexpected outcomes.

A seminal study by Spelke[4,37] found that young infants' individuation judgments were influenced by spatiotemporal continuity (Fig. 3). We can interpret these findings as indexing by location. Take Spelke's first condition: as the first object appears, it is assigned an index. The index sticks to the object as it moves along, disappearing and reappearing from behind each of the screens in turn (Fig. 3A, i). A single object attracts a single index. The test phase, in which the single object is seen again, concurs with indexed expectations and has little novelty (Fig. 3B, i). The two-object test (Fig. 3B, ii) requires the infant to assign a new index (by location) to the second object and consequently attracts additional attention[30]. In the discontinuous condition (Fig. 3A, ii), the first appearance of the object attracts an index that, again, sticks as the object disappears behind the screen. But in this condition the object does not reappear. Instead, another object appears from behind the second screen. Because the first index still points behind the first screen and has not traversed the gap, a new index must be assigned to the second object. Now the infant has two indexes active, which translates into an expectation for two objects. When a single object is shown in the test, the infant has two indexes active with only one of them pointing at something. The infant looks for the 'vanished' object.

Recently, Xu and Carey[34] showed infants pairs of objects by removing and replacing them from behind a screen. The objects are placed in the infants' view either two at a time (spatial condition) or alternating, one at a time (temporal condition). The objects always differ by kind (e.g. a shoe and a cup). Following familiarization, the screen is removed revealing either only one of the objects previously shown, or both objects.

In the spatial condition, 10-month-old infants looked longer when the screen revealed a single object. However, in the temporal condition, the infants looked equally at the revelation of one and two objects. They appeared unable to infer that a shoe and a cup must be distinct objects, unless they saw both objects together at the same time. When shown the cup and shoe at different times, they did not infer the presence of two distinct objects. Slightly older infants, at 12 months, successfully inferred two objects under both conditions.

Apparently, Xu and Carey's younger infants individuated only by location. Indexing theory, drawing on independently motivated notions, provides a ready explanation. Because objects are indexed by location, seeing two objects in different locations at the same time forces the assignment of two indexes: therefore two individual objects are inferred. However, the index does not automatically carry featural information. When only one object at a time is in view, only one index is assigned. The featural differences across successive appearances might be registered and remembered in the

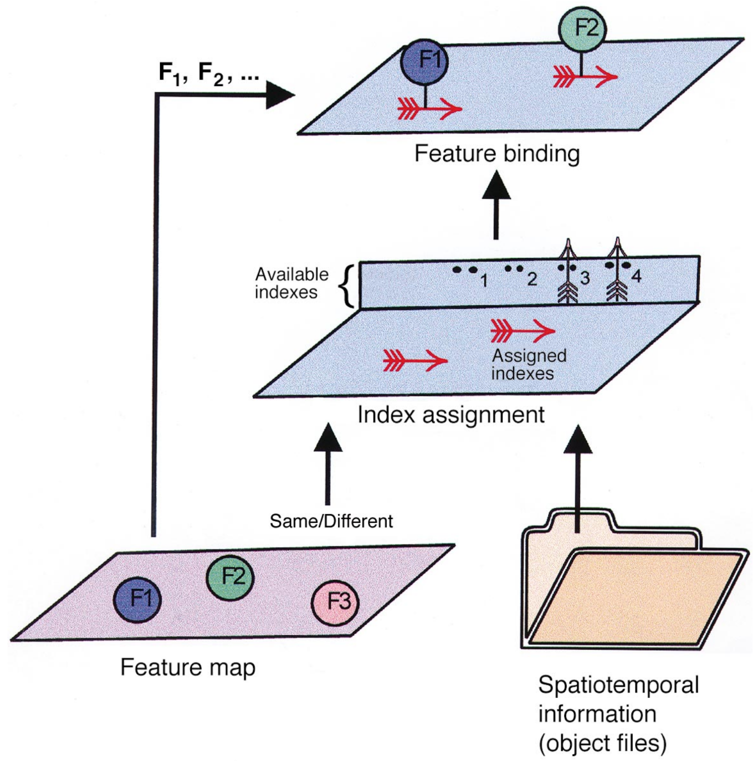


Fig. 2 The mature object indexing system can individuate (assign indexes) either by location or by feature. The binding of featural information (F_1, F_2, etc.) to an index (denoted by red arrows) occurs after the index has been assigned, if it occurs at all. Although feature binding is not required for individuation (even for individuation by feature), it is required for identification of objects (by feature). We assume that there is only a limited number of indexes available (up to the subitizing limit).

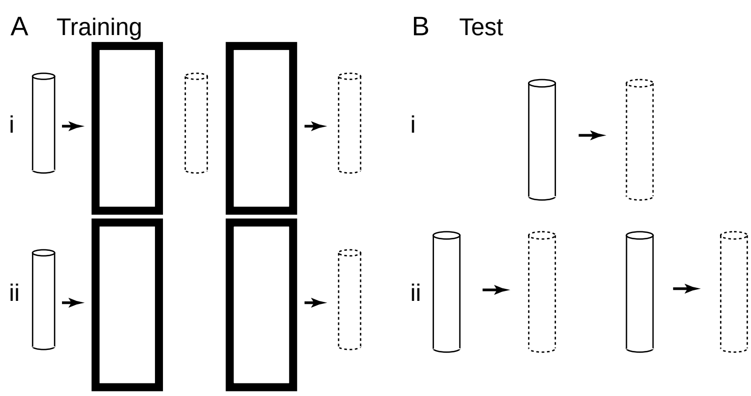


Fig. 3 The effect of spatiotemporal continuity on infant object individuation. (A) Training Infants were familiarized either with (i) Continuity; a display in which an object moved on a continuous trajectory behind two occluders, or (ii) Discontinuity; discontinuous trajectory in which it failed to appear in the gap. Following familiarization, during which looking times declined, infants were shown one of two test displays. **(B) Test** The first test display (i) showed the same object moving to and fro without occluders. The second display (ii) showed two objects moving to and fro without occluders. The infants trained on continuity look longer at the display showing two objects, suggesting that they perceived the original event as involving a single object. The infants trained on discontinuity looked longer at the one object display, reversing the preference found in the first condition. The failure of the object to appear in the gap apparently led these infants to assume that two distinct objects were involved (Modified from Ref. 37).

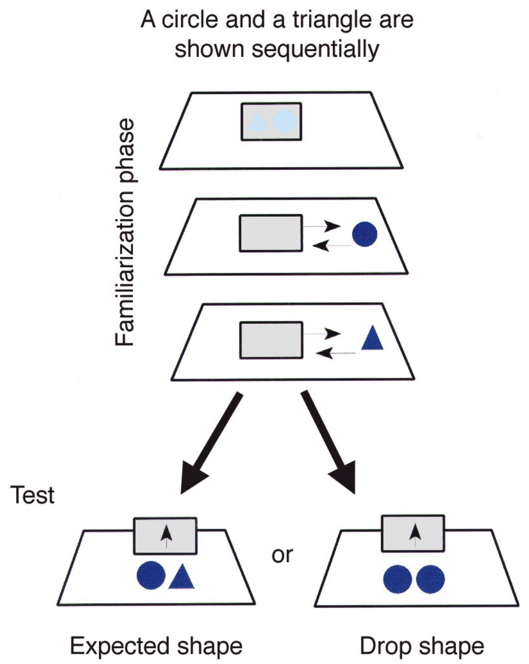


Fig. 4 Infant object indexing by shape. A group of 12-month-old infants was familiarized with an event in which a circle was removed from and replaced behind a screen, followed by a triangle removed from and replaced behind the screen. The two objects were thus shown in the same location but across different times. When the screen was removed following familiarization, the infant saw either the triangle and circle (expected) or two triangles (unexpected) or two circles (unexpected). Infants in the unexpected conditions ('drop shape') looked longer, indicating that they indexed two distinct objects during the familiarization period. By 12 months, infants will assign distinct indexes based on shape differences alone (Modified from Ref. 38).

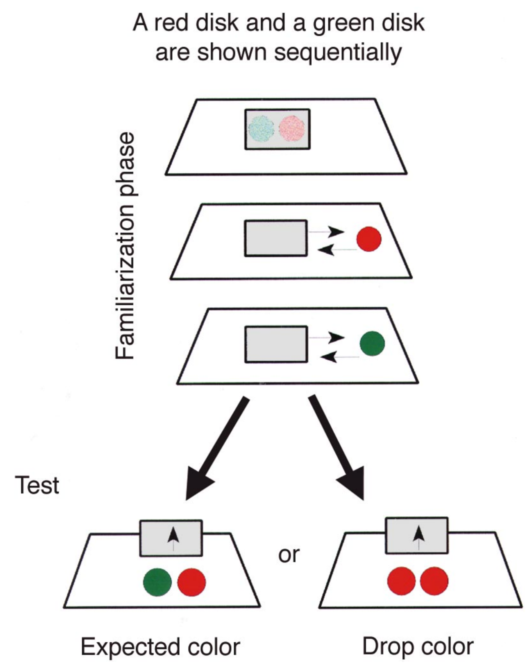


Fig. 5 Infant object indexing by color. Twelve-month-olds do not look longer at the unexpected 'drop color' event than at the 'expected color' event following familiarization to differently colored disks shown sequentially. Comparison with Fig. 4 shows that shape and color features are handled differently in the individuation process at this age (Modified from Ref. 38).

infant's feature map, but they do not force the assignment of distinct indexes. Under these conditions at 10 months, indexing appears to be driven by the 'where' and not by the 'what' system. By 12 months, however, the featural differences across time apparently do force assignment of a second index. One intriguing hypothesis is that the change between 10 and 12 months in these tasks reflects increased integration of ventral ('what') and dorsal ('where') neural systems.

Recent findings suggest that object individuation processes are still immature at 12 months[38]. Following sequential presentation of two differently shaped objects, infants looked longer when the screen revealed two objects of the same shape (Fig. 4). By 12 months, infants will assign distinct indexes based on shape differences alone. A second experiment, in which color rather than shape was manipulated, produced different results (Fig. 5). Following familiarization, infants did not look longer when the screen revealed two red disks (unexpected) compared with a red and a green disk (expected).

Although color is sometimes more salient than shape for infants[39], a further experiment examined the possibility that infants fail to attend to the color of physical objects (Fig. 6). In fact, the 12-month-olds did attend to color and used the color difference to individuate the objects: if repeatedly shown same-colored objects, infants expected one object; if shown differently colored objects, they expected two. However, in neither case (Figs 5 and 6) did the infants care whether the revealed objects were of the same or different colors.

Twelve-month-olds use color and shape features in subtly different ways. With sequential presentation, they establish a new object representation when their feature map registers a new shape and then bind that shape feature to the object index. Twelve-month-olds also assign a new object index when their feature map registers a new color property but the color feature is not so readily bound to the index. In the case of shape, 12-month-olds individuate and identify on the basis of shape, but apparently they only individuate, and do not identify, on the basis of color. This might be a useful strategy for learning about object kinds.

Studies of infant 'counting'

Infants are sensitive to the numerosity of perceptual items in two-dimensional patterns, at least so long as the number of items is small. By controlling for properties that are correlated with numerosity, such as the density, area and length of arrays, looking-time studies show that infants can reliably detect the difference between sets of two and three items[40–44], and can sometimes detect the difference between three and four[41,45].

Wynn[18] argues that five-month-old infants can count physical objects. Infants were familiarized with a single object (a doll) placed upon a small stage, and then occluded. A second doll was introduced and then added to the one behind the screen. The screen was removed to reveal either one (unexpected) or two (expected) objects. Infants looked longer at the single object than at the pair of objects. Further conditions familiarized infants to two objects on

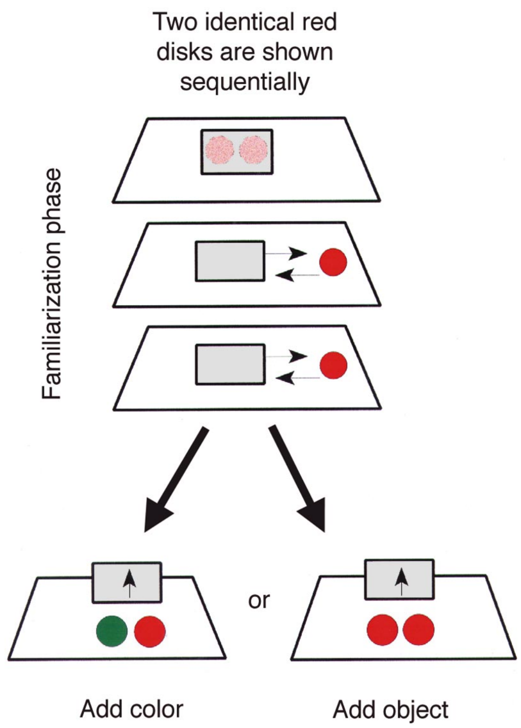


Fig. 6 Object indivduation, but not identification, by color. Infants were familiarized with two identical objects (e.g. red disks) shown individually and sequentially. Infants were then tested with one of two events when the screen was removed. In one condition ('add color'), the screen revealed two differently colored objects. If infants had not registered color, then the new color should not be noticed. In fact, infants did look longer at this event. In a second condition ('add object'), the screen revealed the two identical disks with which the infants had been familiarized. This event too attracted longer looking – about the same level as the 'add color' stimulus and significantly longer than in the expected color condition of the previous experiment (Fig. 5). Why did the two familiarized objects produce 'surprise' in the infants? Because the two familiarization objects did not differ featurally, there was no way the infants could have known they were being shown two objects alternately rather than one object repeatedly. The infants assumed conservatively that they were seeing only a single object and did not expect to see two (Modified from Ref. 38).

stage. After screening, one of the objects was removed from behind the screen and carried off stage. When the screen was lowered, infants saw either one object (expected) or two (unexpected). Again the infants looked longer at the unexpected outcome. These findings have been replicated and extended in other studies[46,47]. Wynn argues that results such as these show that infants possess numerical representations.

In replicating Wynn's results, Simon and colleagues[47] found that five-month-old infants looked longer at unexpected numerosity outcomes, but did not look longer when the identities of the objects were changed unexpectedly. Translating Wynn's and Simon *et al.*'s results into indexing theory, numerosity is tracked by the assignment of indexes to objects (specifically, indexing-by-location); true numerical representation may not be necessary[48]. At five months, infants will index-by-location without binding features; however, this prevents them from later identifying the objects by feature. Our model thus provides an explanation for Simon *et al.*'s finding that, to young infants numerosity violation is more 'surprising' than featural change. Data reviewed earlier suggested that infants at 10 months have limited abilities to individuate-by-feature[34], and that at 12 months, infants show limited ability to identify-by feature, even following individuation-by-feature[38].

Outstanding questions

- What is the relationship between perceptual and conceptual objecthood? Are the points of contact between the literature on adult object-based attention and that on the infant object concept simply analogies, or are the same mechanisms at work in both cases? If the latter, what precise role do these mechanisms play in conceptual development; for example, in the development of sortal concepts?
- Are the theoretical entities – object indexes, FINSTs, and object files – the same or different? One obstacle to identifying them is that these representations have been studied at very different time scales: seconds or tens of seconds versus tens or hundreds of milliseconds. Do these different time scales merely reflect different methodologies, or do they also have theoretical significance?
- We have hypothesized that the ability to individuate physical objects by location comes on-line earlier in development than the ability to individuate by feature, which in turn comes on-line earlier than the ability to identify by feature. Does this sequence map onto the neural development of 'what' and 'where' visual pathways, and if so, how?
- In order to track the numerosity of sets of objects, do infants simply employ object indexes and not 'true' numerical representations, or does object indexing merely provide a mechanism whereby numerical representations can be deployed?
- Our model assumes, for convenience, the notion of 'feature'. Under what circumstances do infants bind features to object representations? How does binding change with development? Are there different developmental patterns for different kinds of features (e.g. shape, size, color, other textures)? Can other kinds of properties (e.g. mechanical properties, such as rigidity, solidity, cohesiveness, perceived agency, or the property of belonging to a certain object kind), also be brought within our framework, or does the binding of these properties belong to a distinct level of cognitive architecture?

Conclusions

We have constructed a set of theoretical notions, each of which is independently grounded in adult visual attention, and have fashioned a model of the infant 'object concept'. Although we make specific proposals, our intention is not to take up a cast-iron stance on details. It is much too early for that. Instead, we have outlined a theoretical framework that can cast existing data in a new light and suggest new approaches to old questions. The classical idea of object representations as bundles of sensations, perceptual features, or properties of any kind, might be fundamentally mistaken. Instead, the heart of any object representation might be inherently abstract, a kind of mental pointing at a 'this' or at a 'that'.

Acknowledgements

We are particularly grateful to Susan Carey for stimulating our interest in these questions over the years, to Zenon Pylyshyn for helpful comments on an earlier draft, and to these and other members of the Princeton–Rutgers–NYU Object Group; especially Bela Julesz, Anne Treisman, Eileen Kowler, Thomas Papathomas, and Jacob Feldman for helpful discussions. This work was supported by NSF project grant SBR-9616342 awarded to A.M.L.

References

1 Chomsky, N.A. (1975) *Reflections on Language*, Pantheon
2 Hume, D. (1740) reprinted in *David Hume on Human Nature and*

Understanding (Flew, A., ed.) (1962), pp. 167–285, Collier Books
3 Quine, W.V. (1960) *Word and Object*, MIT Press
4 Spelke, E.S. (1988) in *Thought Without Language* (Weiskrantz, L., ed.), pp. 168–184, Clarendon Press/Oxford University Press
5 Spelke, E.S. (1994) Initial knowledge: six suggestions *Cognition* 50, 431–445
6 Piaget, J. (1955) *The Child's Construction of Reality*, Routledge & Kegan Paul
7 Duncan, J. (1984) Selective attention and the organization of visual information *J. Exp. Psychol.* 113, 501–517
8 Baylis, G. and Driver, J. (1993) Visual attention and objects: evidence for hierarchical coding of location *J. Exp. Psychol. Hum. Percept. Perform.* 19, 451–470
9 Broadbent, D.E. (1958) *Perception and Communication*, Pergamon Press
10 James, W. (1890) *Principles of Psychology*, Henry Holt
11 Bower, T.G.R. (1967) The development of object permanence: some studies of existence constancy *Percept. Psychophys.* 2, 74–76
12 Bower, T.G.R. (1982) *Development in Infancy* (2nd edn), W.H. Freeman & Co.
13 Baillargeon, R., Spelke, E.S. and Wasserman, S. (1985) Object permanence in five month old infants *Cognition* 20, 191–208
14 Baillargeon, R. (1986) Representing the existence and the location of hidden objects: object permanence in 6- and 8-month old infants *Cognition* 23, 21–41
15 Leslie, A.M. (1984) Infant perception of a manual pick-up event *Br. J. Dev. Psychol.* 2, 19–32
16 Leslie, A.M. (1995) in *Causal Cognition: a Multidisciplinary Debate* (Sperber, D., Premack, D. and Premack, A.J., eds), pp. 121–149, Oxford University Press
17 Leslie, A.M. and Keeble, S. (1987) Do six-month-old infants perceive causality? *Cognition* 25, 265–288
18 Wynn, K. (1992) Addition and subtraction by human infants *Nature* 358, 749–750
19 Maunsell, J.H.R. (1995) The brain's visual world: representation of visual targets in cerebral cortex *Science* 270, 764–769
20 Posner, M.I. and Presti, D.E. (1987) Selective attention and cognitive control *Trends Neurosci.* 10, 13–17
21 Sagi, D. and Julesz, B. (1985) 'Where' and 'what' in vision *Science* 228, 1217–1219
22 Ungerleider, L. and Mishkin, M. (1984) in *Analysis of Visual Behavior* (Ingle, D.J., Goodale, M.A. and Mansfield, R., eds), pp. 549–586, MIT Press
23 Van Essen, D.C. and Maunsell, J.H.R. (1983) Hierarchical organization and functional streams in the visual cortex *Trends Neurosci*. 6, 370–375
24 Julesz, B. (1991) Early vision and focal attention *Rev. Modern Phys.* 63, 735–772
25 Treisman, A. (1988) Features and objects: the 14th Bartlett Memorial Lecture *Q. J. Exp. Psychol.* 40, 201–237
26 Treisman, A. and Gelade, G. (1980) A feature integration theory of attention *Cognit. Psychol.* 12, 97–136
27 Palmer, J. (1995) Attention in visual search: distinguishing four causes of a set size effect *Curr. Dir. Psychol. Sci.* 4, 118–123
28 Wolfe, J.M. (1994) Guided search 2.0: a revised model of visual search *Psychonomic Bull. Rev.* 1, 202–238
29 Wolfe, J.M. (1996) in *Attention* (Pashler, H., ed.), University College London Press
30 Yantis, S. (1993) Stimulus-driven attentional capture *Curr. Dir. Psychol. Sci.* 2, 156–161
31 Treisman, A. and Souther, J. (1985) Search asymmetry: a diagnostic for preattentive processing of separable features *J. Exp. Psychol.* 114, 285–310
32 Julesz, B. (1994) *Dialogues on Perception*, MIT Press
33 Kahneman, D. and Treisman, A. (1984) in *Varieties of Attention* (Parasuraman, R. and Davies, D.R., eds), pp. 29–61, Academic Press
34 Xu, F. and Carey, S. (1996) Infants' metaphysics: the case of numerical identity *Cognit. Psychol.* 30, 111–153
35 Atkinson, J. (1993) in *Spatial representation: Problems in philosophy and psychology* (Eilan, N., McCarthy, R.A. and Brewer, B., eds), pp. 325–339, Blackwell
36 Sagi, D. and Julesz, B. (1984) Detection versus discrimination of visual orientation *Perception* 14, 619–628
37 Spelke, E.S. *et al.* (1995) Spatiotemporal continuity, smoothness of motion and object identity in infancy *Br. J. Dev. Psychol.* 13, 113–142
38 Leslie, A.M., Hall, D.G. and Tremoulet, P. (1997) Object-based attention in infancy *Technical Report No. 41*, Rutgers University Center for Cognitive Science
39 Coldren, J.T. and Colombo, J. (1994) The nature and processes of preverbal learning: implications from nine-month-old infants' discrimination problem solving *Monogr. Soc. Res. Child Dev.* 59, 4
40 Antell, S.R. and Keating, D. (1983) Perception of numerical invariance by neonates *Child Dev.* 54, 695–701
41 Starkey, P. and Cooper, R.G. Jr (1980) Perception of numbers by human infants *Science* 210, 1033–1035
42 Starkey, P.D., Spelke, E.S. and Gelman, R. (1983) Detection of numerical correspondences by human infants *Science* 222, 179–181
43 Starkey, P.D., Spelke, E.S. and Gelman, R. (1990) Numerical abstraction by human infants *Cognition* 36, 97–128
44 Strauss, M.S. and Curtis, L.E. (1981) Infant perception of numerosity *Child Dev.* 52, 1146–1152
45 Von Loesbrook, E. and Smitsman, A.W. (1990) Visual perception of numerosity in infancy *Dev. Psychol.* 26, 916–922
46 Baillargeon, R. (1994) How do infants learn about the physical world? *Curr. Dir. Psychol. Sci.* 3, 133–140
47 Simon, T.J., Hespos, S.J. and Rochat, P. (1995) Do infants understand simple arithmetic?: a replication of Wynn (1992) *Cognit. Dev.* 10, 253–269
48 Simon, T.J. (1997) Reconceptualizing the origins of number knowledge: a non-numerical account *Cognit. Dev.* 12, 349–372

Trends in Cognitive Sciences Questionnaire

I am delighted to reveal that the response to our recent questionnaire far exceeded our expectations. The results are now being processed and will help us to improve the content and style of the journal. Many thanks to all those who participated.

The five lucky winners, who will each receive a free one-year personal subscription to *Trends in Cognitive Sciences* are as follows:

F. Tong, Cambridge, MA, USA
R. Ferstl, Kiel, Germany
C. Senior, London, U.K
R-M. Liao, Taipei, Taiwan
A. Treisman, Princeton, USA

Congratulations to the lucky winners!