Preprint
August 30, 2010

Trajectory recognition as the basis for object individuation: A functional model of object file instantiation and object token encoding

Chris Fields

Apdo 363-4013, Atenas 20501, Costa Rica

fieldsres@gmail.com

Keywords: Object file; Object persistence; Visuomotor representation; Dorsal stream; Visual short-term memory; Binding; Categorization; Autism Spectrum Disorders

Abstract

The perception of persisting visual objects is mediated by transient intermediate representations, object files, that are instantiated in response to some, but not all, visual trajectories. The standard object file concept does not, however, provide a mechanism sufficient to account for all experimental data on visual object persistence, object tracking, and the ability to perceive spatially-disconnected stimuli as coherent objects. Based on relevant anatomical, functional, and developmental data, a functional model is developed that bases object individuation on the specific recognition of visual trajectories. This model is shown to account for a wide range of data, and to generate a variety of testable predictions. Individual variations of the model parameters are expected to generate distinct trajectory and object recognition abilities. Over-encoding of trajectory information in stored object tokens in early infancy, in particular, is expected to disrupt the ability to re-identify individuals across perceptual episodes, and lead to developmental outcomes with characteristics of autism spectrum disorders.

Introduction

It is now well accepted that the human ability to perceive the visual world as composed of discrete, persisting entities rests on the construction of intermediate visual representations, termed “object files” by Kahneman and Treisman (1984), that bind spatial and featural information to form “objects” that can be tracked as their apparent locations, sizes and surface features change through time (reviewed by Treisman, 2006; Scholl, 2007; Flombaum *et al.*, 2008). Instantiation of an object file is, therefore, what mechanistically distinguishes perceiving a persisting object at some location from perceiving a cluster of features of the local background at that location. Object files are standardly conceived of as containers, analogous to file folders, labeled by the current location of and containing the currently-bound features of a perceived object (Flombaum *et al.*, 2008). Treisman (2006) emphasizes that while object files as standardly conceived mediate the comparison of current to immediately-previous location and feature information, they do not maintain even brief histories of objects; current information over-writes and hence erases previous location and feature information when an object file is updated. As noted by Leslie *et al*. (1998), object files labeled by current location serve as "sticky indices" that point to but do not necessarily describe objects, thus capturing the function of "direct reference" defined by Pylyshyn (1989; 2009). Given sufficient attention, the contents of an object file can be written to long-term memory (LTM) as an "object token" representing a specific perceived instance of an object, with a pointer to the episodic memory representing the context in which the object appeared (reviewed by Treisman, 2006; Zimmer & Ecker, 2010). Such LTM-resident object tokens are taken to mediate the re-identification across perceptual episodes of familiar individual objects, as distinct from novel members of familiar categories of objects; they thus differ structurally and functionally from LTM-resident, feature-based category representations (Zimmer & Ecker, 2010).

While the functions subserved by the object file clearly require access to location information encoded by the dorsal visual stream and optionally carry feature information encoded by the ventral stream, a neurofunctional implementation of object files has yet to be proposed. Pylyshyn (2009) characterizes the individuation of objects as "primitive and nonconceptual" (p. 13) and as something that the early visual system is "'wired' to do" (p. 32), but offers no mechanism for how this task is accomplished. Flombaum *et al.* (2008) describe several "principles" that the object-individuation process appears to follow, including assurance of object cohesion and minimization of distance traveled and relative motion (p. 143), but offer no account of where or how these principles – actually heuristics – are implemented. On the purely functional level, neither the criteria that determine whether a spatiotemporal path is sufficiently continuous to indicate objecthood (Flombaum & Scholl, 2006; Flombaum *et al.*, 2008), nor the characteristic development of these criteria during infancy (reviewed by Gerhardstein *et al.*, 2009) are fully understood. The relationship between criteria for objecthood based on spatiotemporal continuity and the criterion of object cohesion, which is often taken to be equally primitive (Scholl, 2007; Baillargeon, 2008; Flombaum *et al.*, 2008), is also poorly understood, especially in cases in which coherently-moving but unbounded collections of objects, such as point-light walkers or schools of fish, are perceived as single objects. Finally, it is unclear how a representation that does not encode the history of an object can enforce observed criteria of spatiotemporal continuity or cohesion: it is not clear how a "compatible match" (Treisman, 2006) between current and immediately-previous states of a candidate object could be computed on a step-by-step basis without a stored representation of either average velocity or at least two previous locations. Hence while it has considerable heuristic value, it is difficult to regard the standard object file concept as providing an adequate functional model of visual object individuation.

Motivated by the early development of motion perception in infancy (Gerhardstein *et al*., 2010) and by recent work showing both that continuous motion can confer objecthood even in the absence of static distinguishability from background (Gao & Scholl, 2010) and that dorsal-stream visuo-motor networks downstream of the medial-temporal (MT) motion-detection area are involved ubiquitously in the interpretation of observed motions (reviewed by Gazzola & Keysers, 2009; Nassi & Callaway, 2009), the present paper proposes that the specific recognition of trajectories underlies and drives visual object individuation. On this proposal, the object file consists of location and feature information bound to one of a finite number of recognizable trajectories, and that all and only location-feature clusters so bound are individuated as objects. This proposal thus directly challenges the view that object files do not encode history. A functional model based on this proposal is developed, and shown to be not only consistent with available anatomical, cognitive, and developmental data, but also capable of organizing and explaining data that are not easily accommodated within the standard object file concept.

By incorporating trajectory information into the object file, the model proposed here raises two issues not explicitly dealt with by the standard object file concept. First, human beings can, at least after the first few months of infancy, individuate stationary objects based on segmentation and featural criteria alone. Second, LTM-resident object tokens represent objects largely independently of their trajectories in any particular observational episode; if they did not, they could not subserve their function of re-identifying objects across contexts. As will be discussed in detail below, the individuation of stationary objects requires a mechanism by which feature-driven categorization instantiates object files in a top-down fashion, while the encoding of sufficiently general object tokens requires a mechanism for the suppression of trajectory information prior to encoding. These mechanisms share a functional requirement that feature information dominates trajectory information during the process of binding a current object file to an LTM-resident object token or category. The existence of mechanisms to assure the dominance of feature information in LTM-resident representations raises the possibility of alternative developmental pathways in which the typical functioning of these mechanisms is disrupted, i.e. in which trajectory information dominates feature information in recorded object tokens and in categories generalized from them. It is shown that such an alternate developmental pathway would be expected to produce cognitive outcomes with many of the characteristics of autism spectrum disorders (ASD), including biological motion and other "whole object" perception deficits (reviewed by Simmons *et al.*, 2009), low central coherence (Happé & Frith, 2006), language-learning difficulties (reviewed by Tager-Flusberg *et al*., 2009), and obsession with repetitive motions (Baron-Cohen & Wheelwright, 1999). It is suggested that the "systemizing" cognitive style defined by Baron-Cohen (2002; 2008), which emphasizes attention to abstract forms and causal structure and is highly prevalent among scientists, mathematicians and engineers, may be a developmental outcome associated with relatively weak feature dominance in object token encoding.

Background: trajectories and object persistence in the context of the standard object file concept

Beginning with the classic studies of Burke (1952) that defined the tunnel effect, extensive experimental work has demonstrated that some, but not all, trajectories of simple geometric shapes indicate objecthood to cognitively-typical adult observers. A simple shape moving continuously across an uncluttered scene is individuated as an object, even if its surface features change, provided that its apparent size does not decrease to zero; this dominance of trajectory continuity over featural change is preserved even if the trajectory is totally occluded, provided the occlusion is brief (Scholl, 2007; Flombaum *et al.*, 2008). Trajectory continuity indicates objecthood even when the "object" cannot be distinguished from the background in a static scene (Gao & Scholl, 2010), indicating that trajectory continuity does not require the detection of static boundaries.   However, occlusion accompanied by kinematically-significant delays or trajectory changes disrupts perception of a persisting object. Flombaum and Scholl (2006), for example, show that the perception of object persistence is disrupted by 1) occlusion for one second in trajectories for which occlusion for 180 to 320 ms would be expected from the perceived motion; (2) an occluded horizontal shift in an observed horizontal trajectory; and (3) "implosion" prior to occlusion, even if the implosion does not shrink the apparent size of the imploding object to zero. Object persistence is similarly disrupted if an observed moving object appears to split into two identically-featured copies (Scholl, 2007). The principles of object cohesion and of minimizing distance traveled and relative motion (Flombaum *et al*., 2008) capture some general features of these observations, but not their kinematic details. The systematic dependence of the perception of objecthood on kinematic details of the observed trajectory in part motivates the widespread assumption that at least some components of "folk physics" are innately specified (e.g. Pinker, 1997; Baillargeon, 2008).

Multiple-object tracking (MOT) studies using simple geometric shapes with identical surface features indicate that trajectory continuity is sufficient to distinguish objects picked out as “targets” from distractors, even in the presence of occluders, provided that no more than four objects must be so identified (reviewed by Scholl, 2009). However, trajectory continuity in the MOT context is not sufficient to individuate the target objects from each other, indicating that while multiple identically-featured objects can be tracked, the final locations of individual objects cannot be reliably associated with their trajectories (Pylyshyn, 2004; Scholl, 2009). As in the case of single object tracking, implosion disrupts object persistence in MOT (Scholl, 2009); trajectories that cross or closely approach also disrupt tracking (Shim *et al*., 2008). If target objects have distinguishing features that allow individuation, MOT performance improves, but this improvement vanishes if targets and distractors share features (Makovski and Jaing, 2009a; 2009b). If multiple objects must be tracked along trajectories that terminate behind an occluder, feature information dominates trajectory information in object individuation (Hollingworth & Franconeri, 2009). As Hollingworth and Franconeri (2009) point out, in real-world MOT tasks such as freeway driving, featural information is critical in determining which currently-perceived object is the continuation of a some particular previously-observed object across few-second gaps in observation. Scholl (2009) interprets the failure of object individuation in MOT with identically-featured objects as indicating that MOT is carried out “in the present,” with only the current and immediately-previous locations of an object available for heuristic determinations of sameness or difference at each timestep, and with previous-location information “flushed” after a determination each made. Experiments with differently-featured objects, in which featural input contributes to individuation, have not assessed the ability of subjects to recall the starting points of or trajectories followed by each target object.

Adult-like abilities to individuate objects based on their trajectories develop progressively during infancy (Gredebäck & von Hofsten, 2007; Gerhardstein *et al*., 2010); however, the heuristics that appear to govern infants' perceptions of object persistence differ in some cases from those of adults. As do adults, 4 month old infants interpret occluded trajectories as continuous if the periods of occlusion are short; unlike adults, infants appear unable to perceive object persistence across large occluders even if the occlusion time is consistent with the object's observed velocity (Bremner *et al*., 2005). Again as do adults, 4 month old infants interpret an occluded horizontal shift in a horizontal trajectory as indicating a novel object (Bremner *et al*. (2007). However, Bremner *et al*. (2007) also show that both an occluded 90° “bounce” and an 18° diagonal trajectory briefly occluded by a vertical occluder are interpreted as discontinuous, but that a diagonal trajectory occluded by an occluder placed at 90° to the trajectory is interpreted as continuous. Predictive tracking of occluded horizontal trajectories based on observed velocity improves during the first year, but again like adults, infants interpret trajectories in which objects implode or disappear prior to brief occlusions as discontinuous, even if the object re-appears at a time and place consistent with its observed velocity (Bertenthal *et al.*, 2007). Skill in velocity-based predictive tracking is facilitated by repeated experience with specific trajectories at 4 months (von Hofsten *et al.*, 2007; Johnson & Shuwairi, 2009), and such experience-based learning is robust by 6 months (reviewed by Rakison, 2007). These observations argue against the notion that a single set of innately-specified trajectory-based heuristics is expressed in both infants and adults. As noted by Bremner *et al*. (2007), they also argue against either the innate specification of physically plausible trajectories, or the learning of physically plausible trajectories on the basis of unstructured observational experience, as explanations of the observed trajectory-based criteria for object persistence.

While the occlusion and MOT studies briefly reviewed above raise the questions of what trajectory-based criteria drive the perception of object persistence, and of how such criteria could be implemented, studies employing point-light displays, and in particular point-light walkers (Johansson, 1973; reviewed by Puce & Perrett, 2003; Blake & Shiffrar, 2007), raise the question of how boundaryless and hence non-cohesive objects are individuated and tracked through time. A point-light walker is effectively an identical-feature MOT display in which the perceptual task is to individuate a composite object by categorizing it on the basis of collective motion criteria alone. Even newborns display a preference for an upright point-light walker over an inverted one (Simion *et al.*, 2008). By 6 months, infants are able to extract overall trajectory information from such displays, indicating that they have successfully identified the point-light walker as a coherently moving object (Kuhlmeier *et al*., 2010). Adult recognition of point-light walkers as coherently moving objects requires less than 100 ms (Pavlova *et al*., 2006), comparable to the visual short-term memory (VSTM) consolidation times of adults (50 ms; Vogel *et al*., 2006) and infants older than 7.5 months (no more than 300 ms; Oakes *et al.*, 2006), and to adult unimodal binding times for components of an episodic event file (240 – 280 ms; Zmigrod & Hommel, 2010). Accounting for the human ability to perceive point-light walkers as persisting objects in the context of the standard object-file concept would require postulating either that object files corresponding to the individual point lights are organized into a coordinated multi-object representation in substantially less than half the time normally required for unimodal feature binding, or that a single object file labeled by a rapidly-computed overall object location somehow tracks all of the individual point lights simultaneously, in either case without the maintenance of history information. Neither of these options is consistent, at least *prima facie*, with the mirror-neuron based, global motion detection mechanisms typically invoked to explain such data (reviewed by Rizzolatti & Craighero, 2004; Dinstein *et al.,* 2007; Cattaneo & Rizzolatti, 2009).

In summary, the available data challenge the standard object file concept on several fronts. First, it is not clear how the complex requirements on trajectories that indicate the continuous motion of a persistent object could be computed from simple principles such as distance or motion minimization applied to current and immediately-previous locations alone. Second, it is unclear, without a specified implementation of the requirements on trajectories, how the developmental timecourse or the characteristic infant specializations of these requirements are to be explained. Third, it is fundamentally unclear, especially in the MOT context, what links the object file to the object: how a current location "labels" a nascent or newly-updated object file has never been explained. Scholl (2009) criticizes the notion that object files serve as object-specific indices in identical-feature MOT trials by pointing out that no data-driven mechanism to link the index to the object has been proposed; however, as Pylyshyn (2009) points out, the alternative notion that target objects are successfully tracked because they serve as attentional foci has no explanatory grip without an independent criterion of persistent objecthood. Finally, it is unclear how object files can represent complex, unbounded displays such as point-light walkers as coherently moving objects within the very short timeframes observed. The functional model outlined below addresses these issues by proposing that the object file is not an initially-empty container or a nondescriptive index to which current location and surface-feature information are bound, but rather is a specific trajectory, implemented by excitation of a specific, post-MT visuomotor network, to which current location and surface-feature information are bound. As will be shown, this model generates a wide variety of testable predictions, some of which are consistent with available data, while others of which remain to be tested.

Model: Specific trajectory recognition drives object individuation

Following the proposal of Rizzolatti and Matelli (2003) that the dorsal visual steam be conceptualized as comprising distinct dorso-dorsal action-guidance and a ventro-dorsal action-interpretation streams, it has become increasingly clear that motion information is processed in specific ways and for specific uses by a variety of cross-modulating but anatomically and functionally distinguishable areas of the superior temporal and posterior parietal cortices, with components of the superior parietal lobule (SPL) being particularly involved in visual tracking and components of the inferior parietal lobule (IPL) being particularly involved in visual target selection, object manipulation, and visuospatial attention (Nassi & Callaway, 2009). These post-MT motion analysis areas and the pre-motor areas with which they are reciprocally connected are consistently shown to be active in both observing and executing actions (Rizzolatti & Craighero, 2004; Dinstein et al., 2007; Cattaneo & Rizzolatti, 2009; Gazzola & Keysers, 2009). While the vast majority of studies have focused on the perception of actions and of biological motion, non-biological motions have been shown to activate "mirror" areas typically involved in biological motion perception in adults (Schubotz & van Cramon, 2004; Engel et al., 2007), and mirror system specificities have been shown to be reconfigurable by experience (Catmur et al., 2007; 2008). The ability of human beings to interpret simple linear motions of simple geometric shapes as intentional and hence biological (reviewed by Scholl & Tremoulet, 2000) suggests that, in the right contexts, nearly any motion can activate the human mirror system.

Motivated by the considerations outlined above, the present paper proposes three core hypotheses: 1) that *all* perceived motion activates and is interpreted by post-MT visuo-motor areas; 2) that one function implemented by these areas is the recognition of specific trajectories; and 3) that an activated trajectory representation is the "index" to which current location and surface feature data are bound to form an object file. These hypotheses, as elaborated below, define what will be referred to as the "specific trajectory recognition" (STR) model of object individuation. The STR model characterizes an object file as a transient, VSTM-resident co-activation and hence temporal binding of dorsal-stream trajectory and current location information with ventral-stream current shape and surface feature information. This model extends and modifies the standard object-file concept in three fundamental ways. First, it proposes a specific implementation of the indexing function, the "file folder" of the object file. Second, it adds history to the object file in the form of a trajectory to which the current location is attached. Third, it proposes that "legal" trajectories for persisting objects are not computed on a step-by-step basis from current and immediately-previous location information, but are rather specifically recognized as global features of an unfolding event. The STR model has six basic implications, as discussed in the six subsections that follow.

Trajectory recognition is limited, specific and hierarchical.

All trajectories begin as vectors in the topographic space defined by visual area V1. These vectors are correlated, correlated groups are bounded, and 2d speed is given depth to yield a 3d, segmented velocity map as an output from MT (reviewed by Born & Bradley, 2005; Kourtzi, 2008). A trajectory is effectively the history of a set of correlated segments of this map over some finite time. The rapid perception of point-light displays as coherently-moving objects indicates that these sets do not have to be compact and the segments contained within a set do not have to share a single 3d velocity. Multiple disjoint velocity segments moving at different speeds in different directions, such as the multiple lights of a point light walker, can be recognized as a "legal" trajectory.

The STR model requires that some, but not all, trajectories be specifically recognized by type in a position-invariant manner. Hence it requires the existence of a finite number of distinct local or distributed networks, each of which receives input from MT, that recognize paths of compact correlated velocity segments with specific curvilinear forms in the 3d space defined by MT. These "simple" trajectory recognition networks (TRNs) effectively recognize "legal" trajectories of compact, bounded objects characterized by a single velocity, such as a rolling ball or a colored disk in a MOT display. While the complete set of legal trajectories is not known, some trajectories are known not to be legal; for example, trajectories that recede to visual infinity and then re-appear (Flombaum et al., 2008).

Unless the spatiotemporal path of a correlated velocity segment in MT excites one or more simple TRNs, it will not be recognized as a legal trajectory, the corresponding localized cluster of features will not be bound to an object file, and perception of a persisting visual object will not be experienced. Trajectory-based object individuation as proposed by the STR model is, therefore, fundamentally distinct from the featural, segmentation, background-subtraction, or categorization-based object-individuation methods that are commonly employed in object-tracking software systems (reviewed by Yilmaz *et al.,* 2006).

The STR model further requires a finite hierarchy of distinct local or distributed networks that recognize specific correlations between the activities of simple TRNs. These "complex" TRNs effectively recognize "legal" trajectories of complex (i.e. articulated, fluid, or comprising multiple independently-moving parts) bounded objects as well as unbounded objects such as point-light walkers or the two components of a temporarily-occluded moving bar. It is assumed that simple TRNs encode trajectories as sequences of time points, with a time resolution on the order of the minimal VSTM consolidation time, i.e. 50 ms in adults (Vogel *et al*., 2006) and less than 300 ms in 7.5 month infants (Oakes et al., 2006). Detection of correlated activity among simple TRNs would, therefore, require greater than this time. As a matter of parsimony, it is assumed that TRNs incorporate effectively continuous local velocity labels as well as spatial labels at each time point, rendering the recognition of trajectories velocity-invariant (within the dynamic range of the label) as well as position-invariant.

***Fig. 1 about here.

Mirror neuron networks selective for specific manipulative actions such as grasping or swinging a hammer (Rizzolatti & Craighero, 2004; Culham & Valyear, 2006; Lewis, 2006) clearly satisfy the functional requirements of TRNs. At least some cells with mirror-like response to specific manipulative actions have response times to visual stimuli in the 50 – 100 ms range (Tkach *et al.*, 2007) or less than 200 ms (Mukamel *et al.*, 2010). Such mirror cells may be components of simple or complex TRNs, or of pre-motor systems downstream of TRNs. The specific association of SPL with visual tracking and grasping – a salient source of object movement, especially in infancy – suggests that simple TRNs may be components of or at least originate in this post-MT visuomotor area. The well-established role of areas of superior temporal sulcus (STS) in biological motion perception (Puce & Perrett, 2003; Rizzolatti & Craighero, 2004) suggests complex TRNs extend at least into this area; their extension into downstream pre-motor areas implicated in mirror function cannot be ruled out. Recent lesion and imaging studies suggest that complex motion detectors within STS may be specific to human-like biological motions (Pyles *et al*., 2007; Saygin, 2007), while more ventral areas, specifically inferior occipital sulcus (IOS), may be specific to complex but not human-like motions (Pyles *et al*., 2007). Networks within or functionally associated with STS have been shown to encode abstracted, viewpoint- and scale-independent representations of human-like trajectories (Jellema & Perrett, 2006; Grossman *et al*., 2010), as required of complex TRNs by the STR model.

By requiring that trajectory recognition be limited, specific and hierarchical, the STR model provides a framework for understanding how only some spatiotemporally continuous trajectories support the perception of object persistence, while also explaining how complex unbounded displays such as point light walkers can be perceived as single persisting objects. It resolves the question of how a local computation based on only current and immediately-previous locations could determine whether a trajectory is "legal" by not requiring such local computations. It explicitly predicts that fast-responding, specific TRNs exist downstream of MT for *every* form of trajectory that is quickly recognizable, without conscious cognition or the use of external tools such as drawings or calculations, by the primate brain. As will be considered in more depth below, it also provides a mechanism by which trajectories can be remembered, and objects categorized or re-identified as individuals based on their observed trajectories.

Motion perception and hence object individuation develop with visual and manipulative experience.

The STR model bases object individuation on trajectory recognition. It therefore implies that object individuation ability develops as trajectory recognition abilities develop. Infants display increasing sensitivity to and ability to discriminate between distinct motions during the first 6 months (Gerhardstein *et al*., 2009), a period during which grasping, manipulation and locomotion capabilities are rapidly developing (reviewed by Piek, 2006). While early studies have attributed motion detection abilities prior to 6 months of age primarily to the maturation of MT, infant abilities to recognize point light walkers as objects (Kuhlmeier *et al*., 2010), recognize actions such as grasping as intentional (Wellman *et al.*, 2008), and recognize repeated trajectories sufficiently accurately to perform predictive tracking of occluded objects (von Hofsten *et al.*, 2007; Johnson & Shuwairi, 2009) all indicate the involvement of post-MT networks by the middle of the first year. Mirror activity has been directly measured by electrophysiology at 8 months (Nyström *et al*., 2009) and 14 months (van Elk *et al*., 2008). Right posterior temporal activity associated with brief object occlusion but not with apparent disintegration has been measured by electrophysiology at 6 months (Kaufman *et al*., 2005). Given this developmental profile for trajectory recognition abilities, the STR model would predict that object individuation ability would begin to develop by 6 months, would significantly improve by 12 months, and would approach maturity during the second year. The close coupling of visuomotor with pre-motor networks would also predict that first-year infant experience with the manipulation of objects would facilitate the development of visual object individuation capabilities.

Four-month old infants in fact display non-adult biases in the perception of object persistence, including an apparent indifference to large changes in velocity (Bremner *et al.*, 2005) and a bias against diagonal and "bouncing" trajectories (Bremner *et al.*, 2007). By 12 months, performance on simple occlusion tasks approaches adult levels (Gredebäck & von Hofsten, 2007; Flombaum *et al.*, 2008; Gerhardstein *et al*., 2010), although the full range of adult-level occlusion heuristics (e.g. tall objects cannot be fully hidden by short occluders) develop more slowly (Baillargeon *et al*., 2006; Gredebäck & von Hofsten, 2007; Baillargeon, 2008). While the effects of experience with object manipulation on the perception of object persistence have not been tested directly, it has been demonstrated that manipulative experience facilitates the interpretation of trajectories as actions by 10 months (Sommerville & Woodward, 2005), that movement and visual attention are coupled at 3 months (Robertson & Johnson, 2009), and that object-manipulation experience facilitates visually-guided behaviors at 3 to 4 months (Lobo & Galloway, 2008). Direct tests of whether observed or performed manipulations affect performance of object persistence tasks, for example, of whether prior observations of a ball being bounced by 4 month olds would facilitate the perception of the "bouncing" trajectories employed by Bremner *et al*. (2007) as indicating persistence, would be interesting in this context.

By tying object individuation to trajectory recognition, the STR model predicts that, in the absence of background or categorical knowledge as discussed below, objects moving on unfamiliar or ambiguous trajectories would be perceived as non-persistent. The trajectories shown in Fig. 2, for example, would be expected to be perceived as non-persistent by 1 year olds, and with high velocities in the diagonal or occluded segments, even by adults. It also predicts that individual differences in the development of complex TRNs would generate detectable individual differences in the perception of object persistence among children or adults. Potentially-significant individual differences have typically been averaged over in existing studies.

***Fig. 2 about here.

The decay times of trajectory representations determine the maximum occlusion times for persisting objects.

The STR model predicts that the decay times of TRNs, not the decay times of motion segments in MT, determine the maximum occlusion time over which persistence is detected. The STR model predicts, therefore, that up to some saturation time, trajectories with longer observed durations prior to occlusion would survive occlusion longer than trajectories with shorter observed durations prior to occlusion. While the occlusion time is commonly treated as a variable to be manipulated in occlusion tasks, the observation duration prior to occlusion is not. If such a prior-observation duration effect is observed, measuring the duration at which it saturates would provide an indirect measure of the dynamic range of TRN activation.

Object individuation by segmentation and featural criteria is functionally and developmentally derivative.

Children and adults readily individuate stationary objects using segmentation and featural criteria, while young infants tend to rely on motion criteria alone for object individuation (Flombaum *et al.*, 2008). It is often assumed that the existence of objects as free-standing entities separate from the "background" of the world is an innately-specified foundational category (Treisman, 2006; Scholl, 2007; Baillargeon, 2008). While the STR model is not inconsistent *per se* with the innate encoding of a foundational category "free-standing object", by providing a mechanism for object individuation in the absence of an innate object category it suggests that no such category need exist. Moreover, the STR model implies that stationary objects of a given type can be individuated on the basis of segmentation and featural criteria alone only after experience with moving objects of that type: it implies that whether a particular stationary cluster of features should be individuated as an object, as distinct from a cluster of features of the local background, is something that must be learned. For example, the STR model implies that although infants appear to be innately capable of recognizing human faces, they are not capable of individuating an *object* – for example their mother – that has a face in the absence of its own motion or the motion of other objects sufficiently similar to it. Here "motion" is meant to include the discontinuous motions of popping suddenly into or out of view, as objects often do in experiments conducted using video displays. The STR model thus implies not only that a foundational category "free-standing object" is unnecessary, but that no such foundational category can specify what is to count as an "object" as distinct from a localized cluster of features of the background.

While the categorization abilities of infants from 3 months onwards and children following the onset of language have been extensively studied, it is not clear when the overarching category "object" becomes effectively deployable, and hence it is not clear at what age medium-sized segmented components of a scene become individuated as objects by default. Intermediate-level categories for objects common to the infant environment, as well as the salient higher-level categories "human", "animal" and "inanimate object" (including living things such as plants) are deployed early in infancy (reviewed by Rakison & Yermolayeva, 2010), but members of these categories often move or are moved in ordinary settings, and specific experiments to determine whether such categories could be formed in the absence of motion information have not been and perhaps could not be performed. Historical evidence, however, suggests that segmentation and distinctive features are insufficient for object individuation even in adults. In medieval European cosmology, for example, the stars were regarded as holes in a hard sphere separating Earth from luminous Heaven – that is, as features, not as individual, movable objects (reviewed by Abrams & Primack, 2001). Only in 20th century cosmology were the "fixed stars" recognized as moving objects. It is not clear that Earth's continents were regarded as objects as opposed to features, despite their obvious boundaries, until the 20th century development of plate tectonics. Historical investigation of the response of various cultures to "first contact" with entirely unfamiliar categories of objects may reveal similar evidence that what are now seen as objects have in the past or in different cultures been seen as features.

If segmentation and distinctive features are insufficient for object individuation as predicted by the STR model, a mechanism is required to instantiate an object file for a static feature cluster once feature clusters of that kind have been associated with "legal" trajectories and hence determined to indicate objects. Binding of a localized feature cluster to a category provides a straightforward mechanism to do this. Hence the STR model predicts that object files are constructed by two distinct processes and in two distinct temporal sequences. In particular, the STR model predicts that in the case of moving objects with recognized trajectories, object files are constructed prior to categorization by a bottom-up, data-driven process. In the case of static objects, the STR model predicts that object files are constructed as a consequence of categorization by a top-down, memory-driven process. Bottom-up object files are anchored by trajectory representations activated by the dorsal stream. Top-down object files are anchored by category representations activated by the ventral stream. The "null trajectory" that consists of staying in one place relative to the local background is, on this model, implemented by binding a categorized cluster of features continuously to a single location: stationary uncategorized feature clusters are expected, on the STR model, to be perceived as features of their local background, not as objects. The general category "object" emerges, on this model, as a relatively mature fall-back option for individuating a familiar, stationary cluster of features as an object, and hence imputing to it the possibility of motion.

Categories and object tokens encode abstracted trajectory information

If object individuation by segmentation and featural criteria is derivative from object individuation on the basis of recognized trajectories, object categories cannot be generalized on the basis of segmentation and featural criteria alone. Hence the STR model predicts that object categories are generalized from object tokens that contain trajectory information. Because in general the trajectories recorded for a given cluster of features will not be identical, category learning will in general involve abstraction of trajectory information.

Extensive data indicate that early-developing categories, including in particular the categories "human", "animal" and "inanimate object" and as a subset of the latter "self-propelled object", contain information specifying typical motions (Baillargeon, 2008; Rakison & Lupyan, 2008; Luo *et al.*, 2009). While such categories are often taken to be innately-specified (Karmaloff-Smith, 1995; Baillargeon, 2008), Rakison and Lupyan (2008) show that categories for animate and inanimate objects can be learned from examples that include feature and motion information, provided that object individuation is assumed. Expectations about typical trajectories – for example, an expectation of linear motion for inanimate objects and non-linear motion for animals and humans – are important components of these learned categories. Constraints on possible motions are, in general, important components of sortal categories in adults (reviewed by Xu, 2007).

Unlike categories, object tokens represent individual objects generalized across the episodes in which they occur (Zimmer & Ecker, 2010). An object may execute different trajectories in different episodes; hence object tokens must also abstract trajectory information. Re-identification of an object as the same individual as encountered in a previous perceptual episode must, therefore, involve generalization and hence loss of information about its specific trajectory. It is reasonable to suppose that such generalization occurs as a component of the binding process that links a current object file to an LTM-resident object token. Hence the STR model predicts that object-token binding, like categorization, involves loss of trajectory information.

Categorization and object-token binding suppress trajectory information in working memory.

The most straightforward mechanism for suppressing detailed trajectory information during the categorization or object-token binding processes is downward inhibition within the hierarchy of TRNs. If excitations of more general TRNs are assumed to inhibit the less-general TRNs immediately below them in the hierarchy, binding an object file to a category or object token that is a good feature match and hence has high amplitude would be expected to ripple inhibition downward through the TRN network, suppressing details of the trajectory anchoring the object file in favor of the abstracted trajectory information contained in the category or object token. This hypothetical mechanism is illustrated in Fig. 3. The STR model requires that this mechanism, or one with similar effects and timecourse, is implemented during the categorization and object-token binding processes.

***Fig. 3 about here.

Suppression of trajectory details by category binding provides an explanation for the inability of subjects in MOT trials to recall target trajectories that is anticipated by neither Pylyshyn (2004; 2009) nor Scholl (2009). The initial labeling of some of the disks in the MOT display as “targets” categorizes them with a familiar category – everyone knows what a “target” is – that the experimental instructions associate with transient blinks or a transiently-visible “T” feature. While the disks are in motion, activation of TRNs representing their trajectories is driven from the bottom up, and supports object tracking. Once the motion stops, such detailed trajectory information is suppressed by the continued, task-driven binding of the “target” category, with its general representation that inanimate “targets” tend to move along smooth curvilinear trajectories. Hence subjects would be expected to recall that the target disks moved along smooth trajectories, but not what those trajectories were. A similar failure of trajectory recall would be expected in a MOT experiment in which the disks moved along jerky, non-smooth trajectories; in this case, subjects would be expected to report that the disks appeared to be animate, not inanimate objects.

The encoding of uncategorized object tokens that then serve as “proto-categories” by supporting the re-identification of the tokened object as an individual present a special case of trajectory information suppression. Encoding of fully uncategorized object tokens would be expected only in early infancy, prior to the robust deployment of the general categories “animate” and “inanimate”, or in short-duration or high-noise perceptual situations in which neither of these general categories out-competes the other. In such situations, the STR model requires that featural information dominates trajectory information in the object token, i.e. that the object-token encoding process itself suppresses trajectory information relative to feature information. Object tokens in which trajectory information dominates feature information would be expected, as discussed in more detail below, to support re-identification of the encoded individual only if the current and previously-observed trajectories activated the same TRNs.

Summary of STR model and predictions

To summarize, the STR model modifies and extends the standard object file concept by proposing that object files are anchored by specifically recognized trajectories. The number of specific trajectories that human beings can recognize is limited; a coherently-moving cluster of features is recognized as a persisting object only if it follows a recognized trajectory. Because trajectory recognition is hierarchical, trajectories can be abstracted; trajectory abstraction allows the encoding of object tokens that represent individuals across episodes, and of categories that represent objects by types. Binding of an object file to either an object token or a category effectively replaces the currently-observed trajectory with the abstracted trajectory encoded by the object token or category, resulting in a loss of detailed trajectory information. Hence adults fail to recall trajectory details in MOT trials, and are, in general, notoriously poor at remembering the precise trajectories of objects in common experience. Trajectories that are highly salient or very familiar can, however, drive categorization or the re-identification of individuals in the absence of specific feature input: even infants can identify other humans by the limb trajectories of activities such as walking, and most people can identify familiar individuals by their gaits or characteristic gestures.

The STR model posits the existence of particular structures within the human, and by extension primate, post-MT visuo-motor systems: TRNs specific to position- and scale-invariant trajectories. It predicts that one or more TRNs are activated by any perceived motion. It predicts that TRNs encode motions as sequences of locations of coherent motion segments defined by MT, with a time resolution on the order of 50 ms in adults. It predicts that TRNs are arranged hierarchically; the lowest-level basic TRNs are expected to originate in SPL, while higher-level complex TRNs may be distributed across the parietal-temporal-frontal mirror network. It predicts that TRNs develop with perceptual and manipulative experience. Finally, the model predicts that downward inhibition in the TRN hierarchy is responsible for the loss of trajectory information on category or object token binding.

The STR model also makes a number of functional predictions. It predicts that human beings should find it difficult if not impossible to see “features” as moving, even if they are explicitly told to expect the features of an object to move; it predicts, in other words, that human observers will instantiate object files, and hence reify moving cluster of features as “objects” by default. It predicts that trajectory consistency across the initial few episodes of observation of a novel individual or category will facilitate object-token encoding or category formation. It predicts that, in a MOT context, the trajectories of objects identified as members of specific categories (e.g. ducks) will be recalled with greater detail than trajectories of objects identified as members of general categories (“targets”). It predicts that point-light walker recognition or MOT performance should be disrupted by instructions to attend to the trajectory of a particular point light or disk. Finally, it predicts that not just a few, but in fact the majority of possible spatiotemporally-continuous object trajectories should disrupt the perception of object persistence, especially in infancy and early childhood. It predicts, in other words, that the fact that objects of common experience follow relatively simple trajectories is neither an accident nor a consequence of fundamental physics, but rather reflects the existence of a relatively limited set of TRNs in the human visuomotor system.

By positing a mechanism based on specific recognition, the STR model raises the possibility of significant normal-range individual differences in trajectory recognition ability, and hence in object individuation. Experiments using point light displays of non-human motions, such as those of Engel *et al*. (2007) or Pyles *et al*. (2007), would be expected to yield a coherent range of abilities in the recognition and classification of trajectories within the cognitively typical, neurotypical population. The model also predicts that variations in the strength of downward inhibition in the TRN network, or in the balance between ventral-stream feature and dorsal-stream trajectory activation in event perception, will result in significant differences in the level of specificity with which trajectory information is encoded in object tokens and categories. Individuals with relatively high dorsal-stream activation levels would be expected, given suitably rich developmental experiences, to form higher-specificity TRNs, and to encode object tokens and categories with higher-specificity trajectory information. Such individuals would be expected to display higher than average interest in events involving similar trajectories, and higher than average tendencies to classify events by similarities among trajectories. A focus on kinematic and dynamic similarities over featural similarities between events is typical of physical scientists, and of "systemizers" (Baron-Cohen, 2002; 2008) in general. Assessments of non-biological motion perception and point light display object individuation in subjects classified by Systemizing Quotient (SQ) scores (Baron-Cohen et al., 2003) would be interesting in this regard.

Prediction: Over-encoding of trajectory information in object tokens produces an ASD-like developmental profile

A complex of differences from typical visual perceptual performance, including enhancements in the perception of detail and deficits in the perception of overall gestalt, are well-documented in ASD (reviewed by Behrmann *et al.*, 2006; Golarai *et al.*, 2006; Mottron *et al*., 2006; Simmons *et al*., 2009). In particular, both deficits and accurate but delayed functioning in the perception of biological motion, as executed by point light walkers, have been reported in ASD (Simmons et al., 2009). Recent studies employing point light arrays have demonstrated enhanced attention to apparently-causal but biologically-irrelevant correlations in 2 year olds with early diagnoses of ASD (Klin *et al.*, 2009), accurate but delayed biological motion detection with concomitant activation differences across a broad range of visuomotor areas in ASD adolescents and young adults (Freitag *et al.*, 2008), and accurate but delayed abilities in biological motion detection in ASD adults even in the presence of significant noise (Murphy *et al.*, 2009). As the STR model predicts significant individual variation in the processing of trajectory information, it is of interest to ask whether variations in the mechanisms proposed by the model, if taken to extremes, would produce outcomes typical of ASD.

As discussed above, the STR model predicts that trajectory information is encoded in object tokens, and that object-token re-instantiation involves top-down TRN activation. In neurotypical development, the trajectory information encoded by object tokens representing familiar individuals is abstracted during the process of category or previous object-token binding. Similarly, in neurotypical development, the trajectory information encoded by object tokens representing novel, uncategorized individuals is suppressed relative to featural information. Suppression of trajectory information in object tokens is critical to the use of object tokens for re-identification of individuals, and hence to the ability of featurally-similar object tokens to support category learning by inductive generalization as demonstrated (Rakison & Lupyan, 2008; see also Gopnik & Tenenbaum, 2007). Hence disruption of trajectory information suppression in object tokens representing uncategorized individuals would be expected to disrupt both individual re-identification and category formation.

The human beings most exposed to uncategorized individuals, and hence most vulnerable to a failure of trajectory information suppression during object token encoding, are infants who have yet to develop robust general categories such as "animate" and "inanimate". If an object token encoded by an infant during a particular perceptual episode *E* included overly-specific trajectory information, bound for example to a correctly-identified but novel face and facial expression, one would expect the infant to correctly re-identify the person observed as "the same individual" across episodes only if he or she exhibited the same motions that he or she exhibited in *E*; i.e. the infant's capability to re-identify the person based on featural similarity and abstracted trajectory information would be compromised. All individual objects are novel to infants on their first presentation, so an infant who typically over-encoded trajectory information in uncategorized object tokens would tend to encode multiple, overly trajectory-specific object tokens for individuals, and aberrant, overly trajectory-specific categories for types. Such overly trajectory-specific object tokens and categories would, in turn, not support the development of abstracted, viewpoint-invariant and individual-nonspecific TRNs. Hence an infant who typically over-encoded trajectory information in uncategorized object tokens would be expected to develop a complex of perceptual phenotypes including over-attention to trajectory details, difficulties with the re-identification of individuals across perceptual episodes, aberrant, trajectory-specific categories that cut across normal feature-based categories, and insensitivity to the general features of what would normally be regarded as classes of trajectories.

Individuals with ASD are in fact overly attentive to simple, repetitive, and specifically non-biological motions (Baron-Cohen & Wheelwright, 1999). Infants and children with ASD have difficulty perceiving point light walkers as objects and in particular as humans (Simmons *et al*., 2009; Klin *et al*., 2009); adolescents and adults with ASD exhibit delays in point light walker recognition that extend to the recognition of complex motions in general, and these differences correlate with differences in activation patterns across the visuomotor and mirror networks (Freitag *et al.*, 2008; Simmons *et al*., 2009). Children and adults with ASD have well documented difficulties with face perception that correlate with activation differences in the fusiform face area (FFA; Behrmann *et al.*, 2006; Golarai *et al.*, 2006); however, it is unclear whether these difficulties result from deficits in the recognition of faces *per se* as opposed to deficits in the identification of representations (e.g. photographs) of unfamiliar faces, or deficits in the ability to consistently recognize an individual person by their face. The STR model would predict that the latter deficit would be a contributing factor in face-recognition difficulties in ASD, and a significant underlying cause of the typical "mind-blindness" and associated social phenotypes of ASD (Baron-Cohen, 2002; Baron-Cohen *et al.*, 2003). Children and adults with ASD often exhibit extreme attention to details, overly-narrow categorization and a pervasive failure to grasp gestalt; this complex of phenotypes has been termed "weak central coherence" (Happé & Frith, 2006). The aberrant, trajectory-focused categories predicted by the STR model would be expected to cause over-attention to motion at the expense of features, and a pervasive inability to integrate or generalize coherently along featural dimensions, consistent with weak central coherence. Such an inability would, in turn, be expected to cause delayed and disorganized language learning, as is often observed in ASD (Tager-Flusberg *et al*., 2009). While the symptomatology of ASD is extraordinarily complex and single-mechanism accounts of its etiology have been unconvincing (Happé *et al*., 2006; Rajendran & Mitchell, 2007), these brief considerations do suggest that over-encoding of trajectory information in object tokens may contribute to the developmental outcomes characteristic of ASD.

Conclusions

The object file concept developed over the last three decades (Treisman, 2006; Scholl, 2007; Flombaum *et al*., 2008) suffers a number of difficulties: it is not clear how local computations with access only to the current and previous locations of an object could determine whether its trajectory is "legal" for indicating persistence; it is not clear how object files can be instantiated for disconnected sets of objects such as point light walkers; and it is not clear what happens to the precise trajectory information that enabled the perception of a persistent object when a permanent object token is encoded. By proposing that objects are only perceived as persistent if their trajectories are specifically recognized as legal by a hierarchical trajectory recognition network, the STR model resolves these difficulties, and provides a framework for interpreting both developmental and adult data on object persistence, MOT capabilities, and complex motion recognition. The STR model makes a variety of anatomical and functional predictions accessible to direct experimental tests.

As the mechanisms proposed by the STR model would be expected to vary in their relative specificities and efficiencies among individuals, the model predicts significant individual differences in the perception of both trajectories and object persistence. Systemizing as a cognitive style (Baron-Cohen, 2002; 2008) may result from a particular configuration of variable parameters of the STR model. Extreme variants in the relative strength of trajectory information encoding in object tokens may lead to pathology; in particular, over-encoding of trajectory information during infancy predicts, within the STR model, a complex of developmental outcomes strikingly consistent with those observed in ASD. If the STR model is confirmed, difficulties in the re-identification of individuals across episodes in which their perceived motions significantly vary would be expected to have value as an early indicator of ASD risk.

Statement regarding conflict of interest

The author states that he has no conflicts of interest relevant to the reported research.

References

Abrams, N. E. & Primack, J. R. (2001). Cosmology and 21st century culture. *Science 293*, 1769-1770.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science 3*(1), 2-13.

Baillargeon, R., Li, J., Luo, Y. & Wang, S. (2006). Under what conditions do infants detect continuity violations? In Y. Munkata and M. H. Johnson (Eds) *Processes of Change in Brain and Cognitive Development* (*Attention and Performance XXI*, pp. 163-188). New York: Oxford University Press.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences 2*(2), 248-254.

Baron-Cohen, S. (2008). Autism, hypersystemizing, and truth. *The Quarterly Journal of Experimental Psychology 61*(1), 64-75.

Baron-Cohen, S. & Wheelwright, S. (1999). 'Obsessions' in children with autism or Asperger syndrome: Content analysis in terms of core domains of cognition. *British Journal of Psychiatry 175,* 484-490.

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N. & Wheelwright, S. (2003). The systemizing quotient: An investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London B 358*, 361-374.

Behrmann, M., Thomas, C. & Humphreys, K. (2006). Seeing it differently: Visual processing in autism. *Trends in Cognitive Sciences 10*(6), 258-264.

Bertenthal, B. I., Longo, M. R. & Kenny, S. (2007). Phenomenal permanence and the development of predictive tracking in infancy. *Child Development 78*(1), 350-363.

Blake, R. & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology 58*, 47-73.

Born, R. T. & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience 28*, 157-189.

Bremner, J. G., Johnson, S. P., Slater, A., Mason, U., Foster, K., Cheshire, A. & Spring, J. (2005). Conditions for young infants' perception of object trajectories. *Child Development 76*(5), 1029-1043.

Bremner, J. G., Johnson, S. P., Slater, A., Mason, U., Cheshire, A. & Spring, J. (2007). Conditions for young infants' failure to perceive trajectory continuity. *Developmental Science 10*(5), 613-624.

Burke, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology 4*(3), 121-138.

Catmur, C., Walsh, V. & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current Biology 17,* 1527-1531.

Catmur, C., Gillmeister, H., Bird, G., Liepelt, R., Brass, M. & Heyes, C. (2008). Through the looking lass: Counter-mirror activation following incompatible sensorimotor learning. *European Journal of Neuroscience 28*, 1208-1215.

Cattaneo, L. & Rizzolatti, G. (2009). The mirror neuron system. *Archives of Neurology 66*(5), 557-560.

Culham, J. & Valyear, K. (2006) Human parietal cortex in action. *Current Opinion in Neurobiology 16*, 205-212.

Dinstein, I., Hasson, U., Rubin, N. & Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *Journal of Neurophysiology 98*, 1415-1427.

Engel, A., Burke, M., Fiehler, K., Bien, S. & Rosler, F. (2007). How moving objects become animated: The human mirror system assimilates non-biological movement patterns. *Social Neuroscience 3,* 368-387.

Flombaum, J. I. & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance 32*(4), 840-853.

Flombaum, J. I., Scholl, B. J. & Santos, L. R. (2008). Spatiotemporal priority as a fundamental principle of object persistence. In: B. Hood & L. Santos (Eds) *The origins of object knowledge* (pp. 135-164). Oxford University Press.

Freitag, C. M., Konrad, C., Häberlen, M., Kleser, C., von Gontard, A., Reith, W., Troje, N. F. & Krick, C. (2008). Perception of biological motion in autism spectrum disorders. *Neuropsychologia 46*, 1480-

1494.

Gao, T. & Scholl, B. J. (2010) Are objects required for object files? Roles of segmentation and spatiotemporal continuity in computing object persistence. *Visual Cognition 18*(1), 82-109.

Gazzola, V. & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analysis of unsmoothed fMRI data. *Cerebral Cortex 19,* 1239-1255.

Gerhardstein, P., Schroff, G., Dickerson, K. & Adler, S. A. (2009). The development of object recognition through infancy. In: B. C. Glenyn & R. P. Zini (Eds) *New Directions in Developmental Psychobiology* (pp. 79-115). Nova Science Publishers.

Golarai, G., Grill-Spector, K. & Reiss, A. L. (2006). Autism and the development of face processing. *Clinical Neuroscience Research 6,* 145-160.

Gopnik, A. & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science 10*(3), 281-287.

Gredebäck, G. & von Hofsten, C. (2007). Taking an action perspective on infants' object representations. *Progress in Brain Research 164*, 265-282.

Grossman, E. D., Jardine, N. L. & Pyles, J. A. (2010). fMR adaptation reveals invariant coding of biological motion on the human STS. *Frontiers in Human Neuroscience 4*, Article 15, 1-18 (DOI: 10.3389/neuro.09.015.2010).

Happé, F. & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders 36*(1), 5-25.

Happé, F., Ronald, A. & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience 9*(10), 1218-1220.

Hollingworth, A. & Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. *Cognition 113*(2), 150-166.

Jellema, T. & Perrett, D. I. (2006). Neural representations of perceived bodily actions using a categorical frame of reference. *Neuropsychologia 44,* 1535-1546.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics 14*, 201-211.

Johnson, S. P. & Shuwairi, S. M. (2009). Learning and memory facilitate predictive tracking in 4 month olds. *Journal of Experimental Child Psychology 102*(1), 122-130.

Kahneman, D. & Treisman, A. (1984). Changing views of attention and automaticity. In: R. Parasuraman & R. Davies (Eds) *Varieties of Attention* (pp. 29-61). New York: Academic Press.

Karmaloff-Smith, A. (1995). *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, MA: MIT Press.

Kaufman, J., Csibra, G. & Johnson, M. H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *Proceedings of the National Academy of Sciences USA 102*(42), 15271-15274.

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G. & Jones, W. (2009). Two-year-olds with autism orient to nonsocial contingencies rather than biological motion. *Nature 459*, 257-261.

Kourtzi, Z., Krekelberg, B. & van Wezel, R. J. A. (2008). Linking form and motion in the primate brain. *Trends in Cognitive Sciences 12*(6), 230-236.

Kuhlmeier, V. A., Troje, N. F. & Lee, V. (2010). Young infants detect the direction of biological motion in point-light displays. *Infancy 15*(1), 83-93.

Johnson, S. P. & Shuwairi, S. M. (2009). Learning and memory facilitate predictive tracking in 4-month olds. *Journal of Experimental Child Psychology 102*(1), 122-130.

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*(1), 10-18.

Lewis, J. (2006) Cortical networks related to human use of tools. *The Neuroscientist 12 (3)*, 211-231.

Lobo, M. A. & Galloway, J. C. (2008). Postural and object-oriented experiences advance early reaching, object exploration, and means-end behavior. *Child Development 79*(6), 1869-1890.

Luo, Y., Kaufman, L. & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology 58*(4), 441-486.

Makovski, T. & Jaing, Y. (2009a). Feature binding in attentive tracking of distinct objects. *Visual Cognition 17*(1-2), 180-194.

Makovski, T. & Jaing, Y. (2009b). The role of visual working memory in attentive tracking of unique objects. *Journal of Experimental Psychology: Human Perception and Performance 35*(6), 1687-1697.

Mottron, L., Dawson, M., Soulières, I., Hubert, B. & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders 36*(1), 27-43.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M. & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology 20*, 750-756.

Murphy, P., Brady, N., Fitzgerald, M. & Troje, N. F. (2009). No evidence for impaired perception of biological motion in adults with autistic spectrum disorders. *Neuropsychologia 47*, 3225-3235.

Nassi, J. J. & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nature Reviews Neuroscience 10*(5), 360-372.

Nyström, P., Ljunghammar, T., Rosander, K. & von Hofsten, C. (2009). Using mu rhythm perturbations to measure mirror neuron activity in infants. Preprint, http://www.robotcub.org/misc/review5/papers/09_Nystrom_etal.pdf, retrieved 19 August 2010.

Oakes, L. M., Ross-Sheehy, R. & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science 17*(9), 781-787.

Pavlova, M., Birbaumer, N. & Sokolov, A. (2006). Attentional modulation of cortical neuromagnetic gamma response to biological movement. *Cerebral Cortex 16*, 321-327.

Piek, J. P. (2006). *Infant Motor Development.* Champaign, IL: Human Kinetics.

Pinker, S. (1997). *How the Mind Works*. New York: Norton.

Puce, A. & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Proceedings of the Royal Society of London B 358*, 435-445.

Pyles, J. A., Garcia, J. O., Hoffman, D. D. & Grossman, D. D. (2007). Visual perception and neural correlates of novel 'biological motion'. *Vision Research 47*, 2786-2797.

Pylyshyn, Z. (1989). The role of location indices in spatial perception: A sketch of the FINST spatial-index model. *Cognition 32*(1), 65-97.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking (MOT): I. Tracking without keeping track of object identities. *Visual Cognition 11*, 801-822.

Pylyshyn, Z. (2009). Perception, representation, and the world: The FINST that binds. In: D. Dedrick & L. Trick (Eds) *Computation, Cognition, and Pylyshyn* (pp. 3-48). Cambridge, MA: MIT Press.

Rajendran, G. & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review 27*, 224-260.

Rakison, D. H. (2007). Fast tracking: Infants learn rapidly about object trajectories. *Trends in Cognitive Sciences 11*(4), 140-142.

Rakison, D. H. & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development 73*(1), 1-130.

Rakison, D. H. & Yermolayeva, Y. (2010). Infant categorization. *Wiley Interdisciplinary Reviews: Cognitive Science 1* (in press).

Rizzolatti, G. & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience 27*, 169-192.

Rizzolatti, G. & Matelli, M. (2003). Two different streams form the dorsal visual system: Anatomy and functions. *Experimental Brain Research 153*, 146-157.

Robertson, S. S. & Johnson, S. L. (2009). Embodied infant attention. *Developmental Science 12*(2), 297-304.

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain 130*, 2452-2461.

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind & Language 22*(5), 563-591.

Scholl, B. J. (2009). What have we learned about attention from multiple-object tracking (and vice versa)? In: D. Dedrick & L. Trick (Eds) *Computation, Cognition, and Pylyshyn* (pp. 49-77). Cambridge, MA: MIT Press.

Scholl, B. & Tremoulet, P. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences 4* (8) 299-309.

Schubotz, R. & van Cramon, D. Y. (2004). Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observations and imagery. *The Journal of Neuroscience 24 (24)*, 5467-5474.

Shim, W. M., Alvarez, G. A. & Jaing, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review 15*(2) 390-397.

Simion, F., Regolin, L. & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences USA 105*(2), 809-813.

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P. & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research 49*, 2705-2739.

Sommerville, J. A. & Woodward, A. L. (2005). Pulling out the intentional structure of action: The relation between action processing and action production in infancy. *Cognition 95*, 1-30.

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., Rice, M., Stoel-Gammon, C., Wetherby, A. & Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language and Hearing Research* 52, 643-652.

Tkach, D., Reimer, J. & Hatsopoulos, N. G. (2007). Congruent activity during action and action observation in motor cortex. *Journal of Neuroscience 27*(48), 13241-13250.

Treisman, A. (2006). Object tokens, binding and visual memory. In H. D. Zimmer, A. Mecklinger & U. Lindenberger (Eds.) *Handbook of binding and memory: perspectives from cognitive neuroscience* (pp. 315-338). Oxford: Oxford University Press.

van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C. & Bekkering, H. (2008). You'll never crawl alone: Neurophysiological evidence for experience-dependent motor resonance in infancy. NeuroImage 43(4), 808-814.

Vogel, E. K., Woodman, G. F. & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance 32*(6), 1436-1451.

von Hofsten, C., Kochukhova, O. & Rosander, K. (2007). Predictive tracking over occlusions by 4 month old infants. *Developmental Science 10*(5), 625-640.

Wellman, H. M., Lopez-Duran, S., LaBounty, J. & Hmilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology 44*(2), 618-623.

Xu, F. (2007). Sortal concepts, object individuation, and language. *Trends in Cognitive Sciences 11*(9), 400-406.

Yilmaz, A., Javed, O. & Shah, M. (2006). Object tracking: A Survey. *ACM Computing Surveys 38*(4), Article 13, 1-45 (DOI: 10.1145/1177352.1177355).

Zimmer, H. D. & Ecker, U. K. H. (2010). Remembering perceptual tokens unequally bound in object and episodic tokens: Neural mechanisms and their electrophysiological correlates. *Neuroscience & Biobehavioral Reviews 34*(7), 1066-1079.

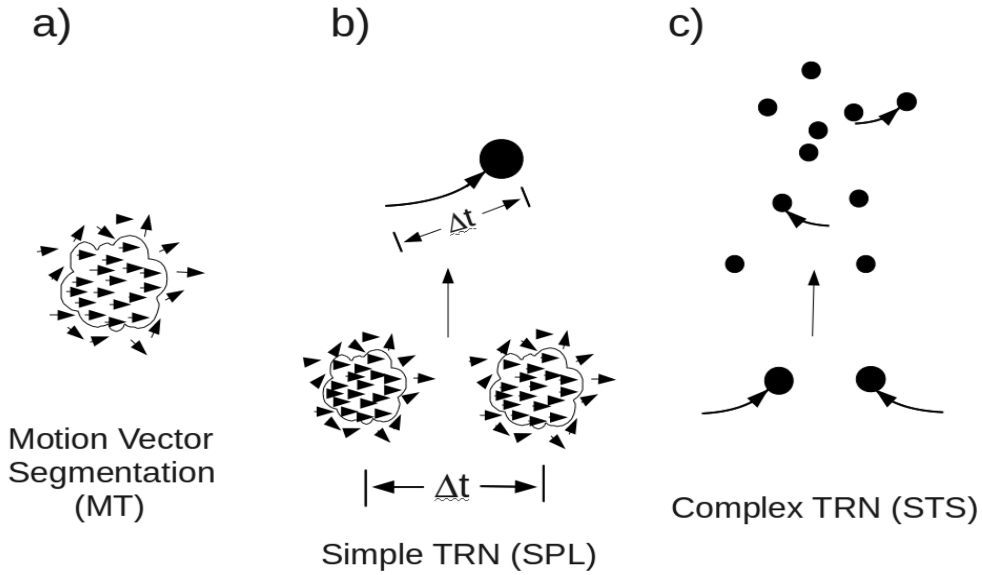
Zmigrod, S. & Hommel, B. (2010). Temporal dynamics of unimodal and multimodal feature binding. *Attention, Perception and Psychophysics 72*(1), 142-152.

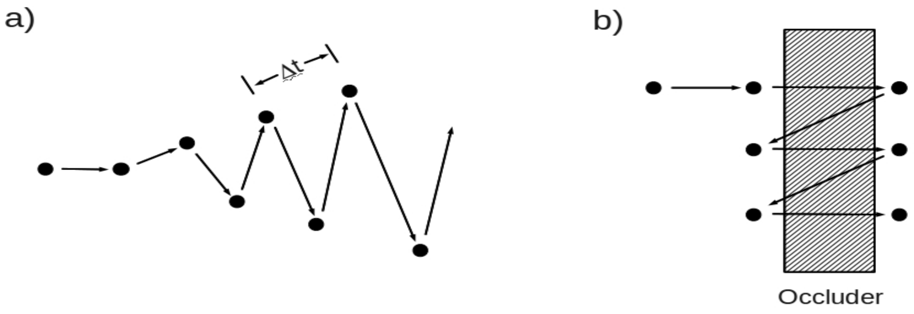
Figure Captions

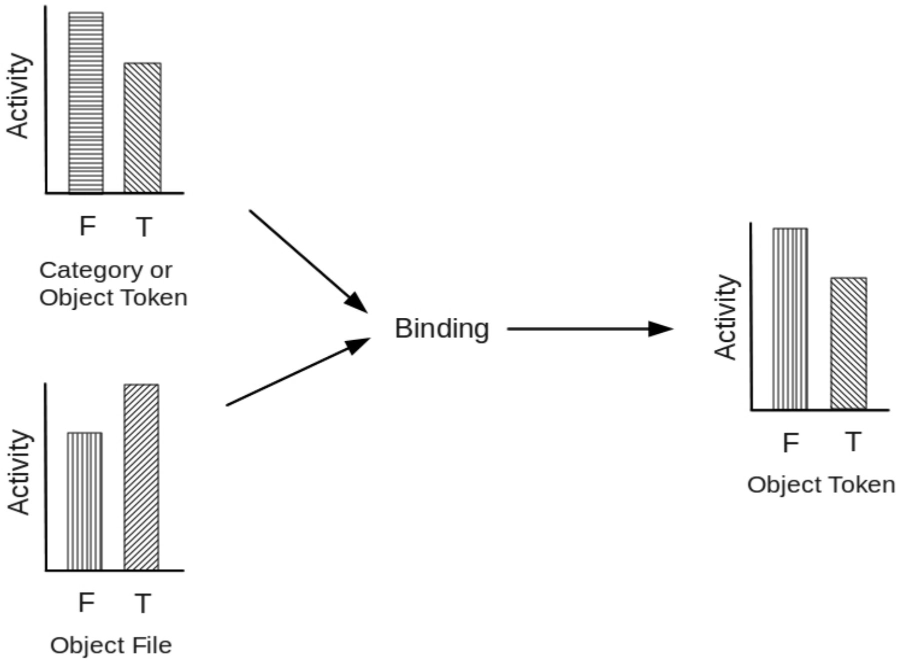
Fig. 1: Schematic representation of the specific trajectory recognition (STR) model of object individuation. a) Compact areas of coherent velocity are segmented in MT. b) “Simple” trajectory recognition networks (TRNs) respond to specific paths of coherent velocity segments in the MT-defined space, with a time resolution of Δt on the order of the VSTM consolidation time. c) “Complex” TRNs respond to specific correlations between simple TRN activity.

Fig. 2: Trajectories predicted to be perceived as violating object persistence by infants and, at high velocity, by adults. a) a zig-zag trajectory with velocity increasing to maintain a constant value of Δt between extreme points, which is predicted to appear as a single object splitting into two objects that follow divergent paths. b) an occluded zig-zag, which is predicted to appear as two objects falling down opposite sides of the occluder.

Fig. 3: Schematic representation of trajectory information suppression by binding to an LTM-resident category or object token. The observed trajectory is replaced in the newly-bound object token by the abstracted trajectory encoded by the LTM-resident category or object token, while the observed features in the newly-bound object token are amplified relative to the trajectory. F = Feature information; T = Trajectory information.


a)
b)
c)
Δt
Δt
Motion Vector
Segmentation
(MT)
Simple TRN (SPL)
Complex TRN (STS)


a)
Δt
b)
Occluder


Activity
F
T
Category or
Object Token
Activity
F
T
Object File
Binding
Activity
F
T
Object Token