



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, Costa Rica

Edited by:

Laurence T. Maloney, Stanford University, USA

Reviewed by:

Laurence T. Maloney, Stanford University, USA
Jacqueline M. Fulvio, University of Minnesota, USA

***Correspondence:**

Chris Fields, Apdo 363-4013, Atenas 20501, Costa Rica.
e-mail: fieldsres@gmail.com

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 continuously existing objects. Based on relevant anatomical, functional, and developmental data, a functional model is constructed that bases visual object individuation on the recognition of temporal sequences of apparent center-of-mass positions that are specifically identified as trajectories by dedicated “trajectory recognition networks” downstream of the medial–temporal motion-detection area. This model is shown to account for a wide range of data, and to generate a variety of testable predictions. Individual differences in the recognition, abstraction, and encoding of trajectory information are expected to generate distinct object persistence judgments and object recognition abilities. Dominance of trajectory information over feature information in stored object tokens during early infancy, in particular, is expected to disrupt the ability to re-identify human and other individuals across perceptual episodes, and lead to developmental outcomes with characteristics of autism spectrum disorders.

Keywords: object file, object persistence, visuomotor representation, dorsal stream, visual short-term memory, binding, categorization, 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 what mechanistically distinguishes perceiving a persisting object at some location from perceiving a cluster of features of the local background at that location; in decision-theoretic language, an object file implements a prior probability of unity that the referenced object is persistent through time, and hence capable in principle of motion. 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 (unless they contain featural information) 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 and 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 and 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. Baillargeon (2008) similarly characterizes the assumption that perceived objects persist through time as an innate default assumption of human beings. Flombaum et al. (2008) note that judgments of object persistence depend critically on the perception of spatiotemporal continuity and describe several additional “principles” that object persistence judgments appear 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 an apparent spatiotemporal path

is sufficiently continuous to indicate objecthood (Flombaum and 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 recent history of an object can enforce inferred 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 can 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 persistence judgments or object individuation.

Motivated by the early development of motion perception in infancy (Gerhardstein et al., 2009) and by recent work showing both that continuous motion can confer objecthood even in the absence of static distinguishability from background (Gao and Scholl, 2010) and that dorsal-stream visuomotor networks downstream of the medial-temporal (MT) motion-detection area are involved ubiquitously in the interpretation of observed motions (reviewed by Gazzola and Keysers, 2009; Nassi and Callaway, 2009), the present paper proposes that the specific recognition of trajectories underlies and drives visual object persistence and hence object individuation. On this proposal, an 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 persistent 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 clusters of features as persisting 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 subservise 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 and Frith, 2006), language-learning difficulties (reviewed by Tager-Flusberg et al., 2009), and obsession with repetitive motions (Baron-Cohen and 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 persistent objecthood to cognitively typical adult observers. A simple shape moving continuously across an uncluttered scene is individuated as a persisting object, even if its surface features change, provided that its apparent size does not decrease toward 0; 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 persisting objecthood even when the “object” cannot be distinguished from the background in a static scene (Gao and 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, even if perceived features (other than location) do not change; hence human beings do not always follow “Leibniz’s Law” of the identity of indiscernibles, even in situations where only one object of a given kind is present. Flombaum and Scholl (2006), for example, show that the perception of object persistence is disrupted by (1) occlusion for 1 s in trajectories for which occlusion for 180–320 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 0. 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), even in paradigms in which most of the objects move coherently in straight lines (Ma and Huang, 2009). 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,b). If multiple objects must be tracked along trajectories that terminate behind an occluder, feature information dominates trajectory information in object individuation (Hollingworth and Franconeri, 2009). As Hollingworth and Franconeri (2009) point out, in real-world MOT tasks such as free-way 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 (i.e., persistence) or difference (i.e., replacement by a distinct object) at each timestep, and with previous-location information “flushed” after each same/different determination is 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 and von Hofsten, 2007; Gerhardstein et al., 2009); 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 and Shuwairi, 2009), and such experience-based trajectory 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 and Perrett, 2003; Blake and Shiffrar, 2007), raise the question of how boundary-less 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 and 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 and Craighero, 2004; Dinstein et al., 2007; Cattaneo and 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 time-frames observed. The functional model outlined below addresses these issues by proposing that the object file is neither an initially empty container nor a non-descriptive 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 stream be conceptualized as comprising distinct dorso-dorsal action-guidance and 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 and 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 and Craighero, 2004; Dinstein et al., 2007; Cattaneo and Rizzolatti, 2009; Gazzola and 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 and 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 and Tremoulet, 2000) suggests that, in the right contexts, nearly any motion can activate the human mirror system.

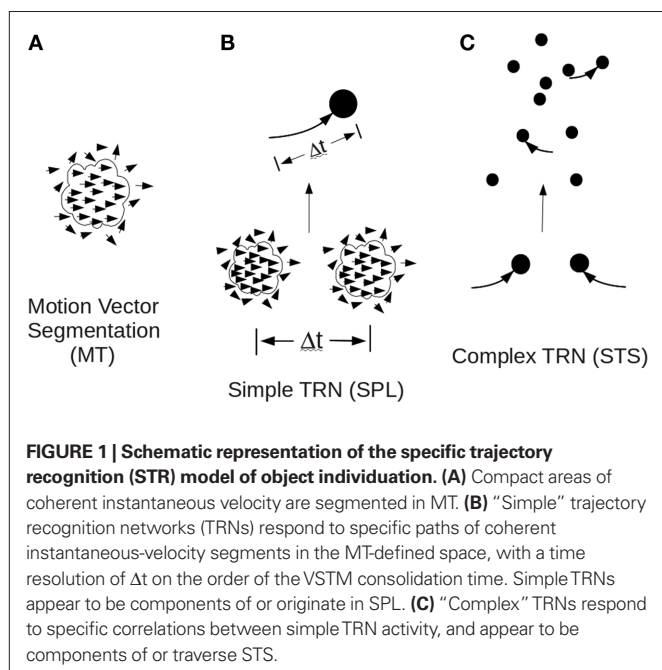
Motivated by the considerations outlined above, the present paper proposes three core hypotheses: (1) that *all* perceived motion activates and is interpreted by post-MT visuomotor areas; (2) that one function implemented by these areas is the recognition of specific trajectories; and (3) that an activated 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 persistence judgments and 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 recognized 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. Compact sets of correlated vectors are bounded, and 2d velocity is given depth to yield a 3d, segmented, instantaneous velocity map as an output from MT (reviewed by Born and Bradley, 2005; Kourtzi et al., 2008). A “trajectory” is a time-sequence of apparent 3d positions of the center (in case of a rotating object, the center-of-mass) of a set of correlated segments of this MT-encoded instantaneous-velocity segment map over some finite time. The rapid perception of point-light displays as coherently moving objects indicates that, at least in the case of visual perception by human beings, 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 instantaneous-velocity segments moving at different speeds in different directions, such as the multiple lights of a point-light walker, can be recognized by human beings as a set of correlated segments; hence such a correlated set can be considered to have a trajectory. A consequence of this definition is that species or even individuals within a species that implement different criteria for the recognition of correlations between instantaneous-velocity segments will recognize distinct patterns of motion as “trajectories.”

The STR model requires that some, but not necessarily all, trajectories be specifically recognized by their curvilinear forms in a position- and scale (i.e., subtended-angle) invariant manner. This requirements has two parts, as illustrated in **Figure 1**. First, the STR model 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 instantaneous-velocity segments with specific curvilinear forms in the 3d space defined by MT (**Figure 1B**). These “simple” trajectory recognition networks (TRNs) effectively recognize trajectories of compact, bounded objects characterized by a single overall instantaneous velocity at each time point, such as a rolling ball or a colored disk in a MOT display. While the complete set of simple recognizable trajectories is not known, some trajectories are known not to be recognizable by cognitively typical humans; 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



simple trajectory, the corresponding localized cluster of features will not be bound to create an object file, and perception of a persisting visual object will not be experienced.

The second requirement of the STR model is that a finite hierarchy of distinct local or distributed networks recognize specific correlations between the activities of simple TRNs. These “complex” TRNs effectively recognize trajectories of complex (i.e., articulated, fluid, rotating, vibrating, 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.

Mirror neuron networks selective for specific manipulative actions such as grasping or swinging a hammer (Rizzolatti and Craighero, 2004; Culham and 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 and Perrett, 2003;

Rizzolatti and 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, particularly 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 and Perrett, 2006; Grossman et al., 2010), as required of complex TRNs by the STR model.

On the STR model, a visually perceived “object” is a coherent set of instantaneous-velocity segments, the recent trajectory of which is recognizable by either a simple or a complex TRN, and an “object file” is a transient binding of the current location and any active features of a coherent set of instantaneous-velocity segments to the recognized trajectory. A segmented set of correlated instantaneous-velocity vectors that moves along a trajectory that is not recognized by a simple or complex TRN – for example, a segmented set of vectors that appears to disappear and then re-appear – is not perceived as an object and no object file is formed. Trajectory-based object individuation as proposed by the STR model is fundamentally distinct from and prior to the featural, segmentation, background-subtraction, dynamical-systems estimation (e.g., Kalman filter), or categorization-based object-individuation methods that are commonly employed in object-tracking software systems (reviewed by Yilmaz et al., 2006). Such methods all assume that objects are persistent by definition, and that the task of individuation is to distinguish persistent objects from each other. The TRN model provides an implementation for this fundamentally qualitative prior assumption that what is perceived continues to exist as “the same thing” over the duration of the observation. In the language of statistical models of perception and action (reviewed by Maloney and Zhang, 2010), it sets the prior probability of objecthood at or near unity if a trajectory is recognized, and at or near zero if one is not. Decisions about how to act with respect to what is perceived, including how to categorize it, are made on the basis of this trajectory-recognition based assumption that what is perceived either is or is not a persisting object.

The STR model does not require that object persistence judgments or object individuation be computed purely bottom-up; indeed as discussed in Section “Object Individuation by Segmentation and Featural Criteria is Functionally and Developmentally Derivative,” it predicts that the individuation of static objects based on segmentation and feature criteria is computed top-down. What is required is that an object file be created before it can be categorized by binding to a stored category representation or object token. The STR model also does not require that only a single TRN be activated in any particular case. Trajectories can be ambiguous, and top-down modulation may lead to spurious TRN activation. In such a case multiple TRNs may be activated with similar levels of activation simultaneously, and an ambiguous object file binding features to multiple active TRNs and hence multiple perceived motions may be created. What the model does require is that any object have at least one recognized trajectory. Perceivers may, on the STR model, be highly uncertain about how an object has moved

in the recent past or how it will move in the immediate future, but they are predicted to be rarely if ever uncertain about whether what is perceived is a persisting object.

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 indicative of objecthood 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 human 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; hence it 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 and 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 predicts that object individuation abilities begin to develop by 6 months, significantly improve by 12 months, and approach maturity during the second year. Given the close coupling of visuomotor with pre-motor networks, it also predicts that infant experience with the manipulation of objects facilitates the development of visual object individuation capabilities.

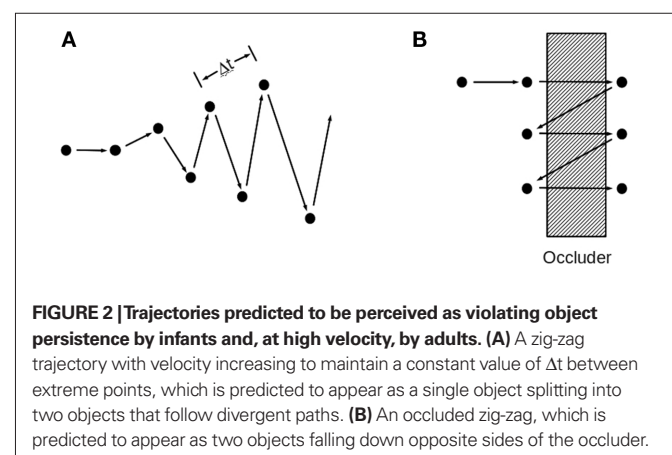
Four-month-old infants in fact display non-adult biases in the perception of object persistence, including an apparent indifference to large changes in velocity (Bremner et al., 2005) and a bias against diagonal and “bouncing” trajectories (Bremner et al., 2007). By 12 months, performance on simple occlusion tasks approaches adult levels (Gredebäck and von Hofsten, 2007; Flombaum et al., 2008; Gerhardstein et al., 2009), 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 and 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 and Woodward, 2005), that movement and visual attention are coupled at 3 months (Robertson and Johnson, 2009), and that object-manipulation experience facilitates visually guided behaviors at 3–4 months (Lobo and Galloway, 2008). Direct tests of whether observed or performed manipulations affect performance of object persistence tasks could indicate both the extent to which object classification modulates the continuing perception of object permanence, and the extent to which trajectories can be associated with multiple objects across object categories. For example, prior observations of a ball – something already recognized as persistent – being bounced by 4-month-olds may facilitate the perception of the “bouncing” trajectories employed by Bremner et al. (2007), in which the “bounce” occurs behind an occluder, as indicating persistence. Such facilitation would indicate a transfer of a recognized trajectory previously associated with a ball to a different kind of object, a moving dot on a screen: a kind of transfer across categories that occurs routinely in adults.

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 **Figure 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, but can potentially be assayed quantitatively by trajectory-prediction experiments (e.g., Maloney and Zhang, 2010).

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 duration of observation 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 onward 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 and 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 for ethical reasons 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 and Primack, 2001). Only in twentieth 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 twentieth 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 recognized trajectories and hence determined to indicate persistent 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. Transcranial magnetic stimulation or high-resolution evoked potential measurements may be able to distinguish the top-down versus bottom-up activation time courses for the attribution of objecthood predicted by the STR model in the contrasting cases of static observation versus observation while in 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 both abstraction of trajectory information and the association of multiple abstracted trajectories with a given category. The extent of abstraction of trajectories encoded by object categories is expected to be at least that implemented

by TRNs, i.e., the position, orientation, and scale of trajectories will always be abstracted, while the duration of a trajectory can be expected to be abstracted within category-specific limits.

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 and 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 persistence and individuation are assumed. Expectations about typical trajectories – for example, an expectation of linear motion for inanimate objects and erratic or 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 and Ecker, 2010). An object may execute different trajectories in different episodes; hence object tokens must also encode abstracted specifications of multiple observed or inferred possible trajectories. 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 previously observed 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 **Figure 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.

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

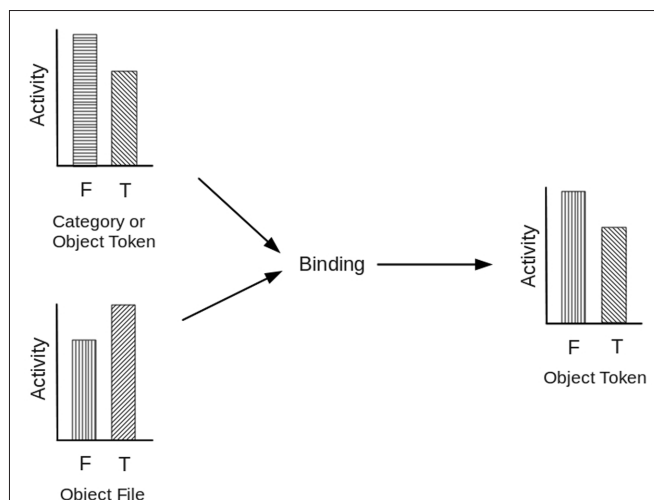


FIGURE 3 | Schematic representation of trajectory information suppression by binding to an LTM-resident category or object token. The observed trajectory is replaced in the newly bound object token by the abstracted trajectory encoded by the LTM-resident category or object token, while the observed features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

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. It posits the existence of particular structures within the human, and by extension primate, post-MT visuomotor 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 encoded by MT, with a time resolution on the order of 50 ms in adults. It predicts that TRNs are arranged hierarchically; the lowest-level simple 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 clusters 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 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: OVERLY SPECIFIC 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 and Lupyan, 2008; see also Gopnik and 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 categorized 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 encoded overly specific 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 complex TRNs. Hence an infant who typically encoded overly specific 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 be regarded by neurotypicals as classes of similar trajectories.

Individuals with ASD are in fact overly attentive to simple, repetitive, and specifically non-biological motions (Baron-Cohen and Wheelwright, 1999). Infants and children with ASD have difficulty perceiving point-light walkers as objects and in particular as humans (Klin et al., 2009; Simmons 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 (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 and 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 and Mitchell, 2007), these brief considerations do suggest that overly specific encoding of trajectory information in object tokens may contribute to the developmental outcomes characteristic of ASD.

CONCLUSION

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 indicates object 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 incorporating abstracted trajectory information is encoded. By proposing that objects are only perceived as persistent if their trajectories are specifically recognized by a hierarchical TRN, 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 biases in the recognition, abstraction and encoding of trajectory information. Extreme variants in the relative strength of trajectory information encoding in object tokens may lead to pathology; in particular, overly specific 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, infant 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.

ACKNOWLEDGMENTS

The comments of two anonymous referees were of value in clarifying the presentation.

REFERENCES

- Abrams, N. E., and 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. *Perspect. Psychol. Sci.* 3, 2–13.
- Baillargeon, R., Li, J., Luo, Y., and Wang, S. (2006). “Under what conditions do infants detect continuity violations?” in *Processes of Change in Brain and Cognitive Development: Attention and Performance XXI*, eds Y. Munkata and M. H. Johnson (New York: Oxford University Press), 163–188.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends Cogn. Sci.* 2, 248–254.
- Baron-Cohen, S. (2008). Autism, hyper-systemizing, and truth. *Q. J. Exp. Psychol.* 61, 64–75.
- Baron-Cohen, S., Richler, J., Bisarya, D., Guranathan, N., and Wheelwright, S. (2003). The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 361–374.
- Baron-Cohen, S., and Wheelwright, S. (1999). ‘Obsessions’ in children with autism or Asperger syndrome: content analysis in terms of core domains of cognition. *Br. J. Psychiatry* 175, 484–490.
- Behrmann, M., Thomas, C., and Humphreys, K. (2006). Seeing it differently: visual processing in autism. *Trends Cogn. Sci.* 10, 258–264.
- Bertenthal, B. I., Longo, M. R., and Kenny, S. (2007). Phenomenal permanence and the development of predictive tracking in infancy. *Child Dev.* 78, 350–363.

- Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73.
- Born, R. T., and Bradley, D. C. (2005). Structure and function of visual area MT. *Annu. Rev. Neurosci.* 28, 157–189.
- Bremner, J. G., Johnson, S. P., Slater, A., Mason, U., Cheshire, A., and Spring, J. (2007). Conditions for young infants' failure to perceive trajectory continuity. *Dev. Sci.* 10, 613–624.
- Bremner, J. G., Johnson, S. P., Slater, A., Mason, U., Foster, K., Cheshire, A., and Spring, J. (2005). Conditions for young infants' perception of object trajectories. *Child Dev.* 76, 1029–1043.
- Burke, L. (1952). On the tunnel effect. *Q. J. Exp. Psychol.* 4, 121–138.
- Catmur, C., Gillmeister, H., Bird, G., Liepelt, R., Brass, M., and Heyes, C. (2008). Through the looking lass: counter-mirror activation following incompatible sensorimotor learning. *Eur. J. Neurosci.* 28, 1208–1215.
- Catmur, C., Walsh, V., and Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Curr. Biol.* 17, 1527–1531.
- Cattaneo, L., and Rizzolatti, G. (2009). The mirror neuron system. *Arch. Neurol.* 66, 557–560.
- Culham, J., and Valyear, K. (2006). Human parietal cortex in action. *Curr. Opin. Neurobiol.* 16, 205–212.
- Dinstein, I., Hasson, U., Rubin, N., and Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *J. Neurophysiol.* 98, 1415–1427.
- Engel, A., Burke, M., Fiehler, K., Bien, S., and Rosler, F. (2007). How moving objects become animated: the human mirror system assimilates non-biological movement patterns. *Soc. Neurosci.* 3, 368–387.
- Flombaum, J. I., and Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: facilitated change detection for persisting objects. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 840–853.
- Flombaum, J. I., Scholl, B. J., and Santos, L. R. (2008). "Spatiotemporal priority as a fundamental principle of object persistence," in *The Origins of Object Knowledge*, eds B. Hood and L. Santos (Oxford: Oxford University Press), 135–164.
- Freitag, C. M., Konrad, C., Häberlen, M., Kleser, C., von Gontard, A., Reith, W., Troje, N. F., and Krick, C. (2008). Perception of biological motion in autism spectrum disorders. *Neuropsychologia* 46, 1480–1494.
- Gao, T., and Scholl, B. J. (2010). Are objects required for object files? Roles of segmentation and spatiotemporal continuity in computing object persistence. *Vis. Cogn.* 18, 82–109.
- Gazzola, V., and 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. *Cereb. Cortex* 19, 1239–1255.
- Gerhardstein, P., Schroff, G., Dickerson, K., and Adler, S. A. (2009). "The development of object recognition through infancy," in *New Directions in Developmental Psychobiology*, eds B. C. Glenyn and R. P. Zini (Hauppauge: Nova Science Publishers), 79–115.
- Golarai, G., Grill-Spector, K., and Reiss, A. L. (2006). Autism and the development of face processing. *Clin. Neurosci. Res.* 6, 145–160.
- Gopnik, A., and Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Dev. Sci.* 10, 281–287.
- Gredebäck, G., and von Hofsten, C. (2007). Taking an action perspective on infants' object representations. *Prog. Brain Res.* 164, 265–282.
- Grossman, E. D., Jardine, N. L., and Pyles, J. A. (2010). fMR adaptation reveals invariant coding of biological motion on the human STS. *Front. Hum. Neurosci.* 4:15. doi: 10.3389/fneuro.09.015.2010.
- Happé, F., and Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord.* 36, 5–25.
- Happé, F., Ronald, A., and Plomin, R. (2006). Time to give up on a single explanation for autism. *Nat. Neurosci.* 9, 1218–1220.
- Hollingsworth, A., and Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. *Cognition* 113, 150–166.
- Jellema, T., and 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. *Percept. Psychophys.* 14, 201–211.
- Johnson, S. P., and Shuwairi, S. M. (2009). Learning and memory facilitate predictive tracking in 4-month-olds. *J. Expt. Child Psychol.* 102, 122–130.
- Kahneman, D., and Treisman, A. (1984). "Changing views of attention and automaticity," in *Varieties of Attention*, eds R. Parasuraman and R. Davies (New York: Academic Press), 29–61.
- Karmaloff-Smith, A. (1995). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.
- Kaufman, J., Csibra, G., and Johnson, M. H. (2005). Oscillatory activity in the infant brain reflects object maintenance. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15271–15274.
- Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., and Jones, W. (2009). Two-year-olds with autism orient to nonsocial contingencies rather than biological motion. *Nature* 459, 257–261.
- Kourtzi, Z., Krekelberg, B., and van Wezel, R. J. A. (2008). Linking form and motion in the primate brain. *Trends Cogn. Sci.* 12, 230–236.
- Kuhlmeier, V. A., Troje, N. F., and Lee, V. (2010). Young infants detect the direction of biological motion in point-light displays. *Infancy* 15, 83–93.
- Leslie, A. M., Xu, F., Tremoulet, P. D., and Scholl, B. J. (1998). Indexing and the object concept: developing 'what' and 'where' systems. *Trends Cogn. Sci.* 2, 10–18.
- Lewis, J. (2006). Cortical networks related to human use of tools. *Neuroscientist* 12, 211–231.
- Lobo, M. A., and Galloway, J. C. (2008). Postural and object-oriented experiences advance early reaching, object exploration, and means-end behavior. *Child Dev.* 79, 1869–1890.
- Luo, Y., Kaufman, L., and Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cogn. Psychol.* 58, 441–486.
- Ma, J. M., and Huang, W. (2009). No capacity limit in attentional tracking: evidence for probabilistic inference under a resource constraint. *J. Vision* 9, 1–30.
- Makovski, T., and Jaing, Y. (2009a). Feature binding in attentive tracking of distinct objects. *Vis. Cogn.* 17, 180–194.
- Makovski, T., and Jaing, Y. (2009b). The role of visual working memory in attentive tracking of unique objects. *J. Exp. Psychol. Human Percept. Perform.* 35, 1687–1697.
- Maloney, L. T., and Zhang, H. (2010). Decision-theoretic models of visual perception. *Vision Res.* 50, 2362–2374.
- Mottron, L., Dawson, M., Soulières, I., Hubert, B., and Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *J. Autism Dev. Disord.* 36, 27–43.
- Mukamel, R., Ekstrom, A. D., Kaplan, J., Jacoboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Curr. Biol.* 20, 750–756.
- Murphy, P., Brady, N., Fitzgerald, M., and 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., and Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* 10, 360–372.
- Nyström, P., Ljunghammar, T., Rosander, K., and von Hofsten, C. (2009). Using mu rhythm perturbations to measure mirror neuron activity in infants. Preprint, retrieved on August 19, 2010 from http://www.robotcub.org/misc/review5/papers/09_Nystrom_etal.pdf
- Oakes, L. M., Ross-Sheehy, R., and Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychol. Sci.* 17, 781–787.
- Pavlova, M., Birbaumer, N., and Sokolov, A. (2006). Attentional modulation of cortical neuromagnetic gamma response to biological movement. *Cereb. 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., and Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 435–445.
- Pyles, J. A., Garcia, J. O., Hoffman, D. D., and Grossman, D. D. (2007). Visual perception and neural correlates of novel 'biological motion'. *Vision Res.* 47, 2786–2797.
- Pylyshyn, Z. (1989). The role of location indices in spatial perception: a sketch of the FINST spatial-index model. *Cognition* 32, 65–97.
- Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking (MOT): I. Tracking without keeping track of object identities. *Vis. Cogn.* 11, 801–822.
- Pylyshyn, Z. (2009). "Perception, representation, and the world: the FINST that binds," in *Computation, Cognition, and Pylyshyn*, eds D. Dedrick and L. Trick (Cambridge, MA: MIT Press), 3–48.
- Rajendran, G., and Mitchell, P. (2007). Cognitive theories of autism. *Dev. Rev.* 27, 224–260.
- Rakison, D. H. (2007). Fast tracking: infants learn rapidly about object trajectories. *Trends Cogn. Sci.* 11, 140–142.
- Rakison, D. H., and Lupyan, G. (2008). Developing object concepts in infancy: an associative learning perspective. *Monogr. Soc. Res. Child Dev.* 73, 1–130.
- Rakison, D. H., and Yermolayeva, Y. (2010). Infant categorization. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 894–905.
- Rizzolatti, G., and Craighero, L. (2004). The mirror neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Rizzolatti, G., and Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Exp. Brain Res.* 153, 146–157.

- Robertson, S. S., and Johnson, S. L. (2009). Embodied infant attention. *Dev. Sci.* 12, 297–304.
- Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain* 130, 2452–2461.
- Scholl, B., and Tremoulet, P. (2000). Perceptual causality and animacy. *Trends Cogn. Sci.* 4, 299–309.
- Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind Lang.* 22, 563–591.
- Scholl, B. J. (2009). “What have we learned about attention from multiple-object tracking (and vice versa)?” in *Computation, Cognition, and Pylyshyn*, eds D. Dedrick and L. Trick (Cambridge, MA: MIT Press), 49–77.
- Schubotz, R., and van Cramon, D. Y. (2004). Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observations and imagery. *J. Neurosci.* 24, 5467–5474.
- Shim, W. M., Alvarez, G. A., and Jaing, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychon. Bull. Rev.* 15, 390–397.
- Simion, F., Regolin, L., and Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proc. Natl. Acad. Sci. U.S.A.* 105, 809–813.
- Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., and Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Res.* 49, 2705–2739.
- Sommerville, J. A., and 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., and Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *J. Speech Lang. Hear. Res.* 52, 643–652.
- Tkach, D., Reimer, J., and Hatsopoulos, N. G. (2007). Congruent activity during action and action observation in motor cortex. *J. Neurosci.* 27, 13241–13250.
- Treisman, A. (2006). “Object tokens, binding and visual memory,” in *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience*, eds H. D. Zimmer, A. Mecklinger, and U. Lindenberger (Oxford: Oxford University Press), 315–338.
- van Elk, M., van Schie, H. T., Hunnius, S., Vesper, C., and Bekkering, H. (2008). You’ll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. *Neuroimage* 43, 808–814.
- Vogel, E. K., Woodman, G. F., and Luck, S. J. (2006). The time course of consolidation in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1436–1451.
- von Hofsten, C., Kochukhova, O., and Rosander, K. (2007). Predictive tracking over occlusions by 4 month old infants. *Dev. Sci.* 10, 625–640.
- Wellman, H. M., Lopez-Duran, S., LaBounty, J., and Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Dev. Psychol.* 44, 618–623.
- Xu, F. (2007). Sortal concepts, object individuation, and language. *Trends Cogn. Sci.* 11, 400–406.
- Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: a Survey. *ACM Comput. Surv.* 38, 1–45. doi: 10.1145/1177352.1177355.
- Zimmer, H. D., and Ecker, U. K. H. (2010). Remembering perceptual tokens unequally bound in object and episodic tokens: neural mechanisms and their electrophysiological correlates. *Neurosci. Biobehav. Rev.* 34, 1066–1079.
- Zmigrod, S., and Hommel, B. (2010). Temporal dynamics of unimodal and multimodal feature binding. *Atten. Percept. Psychophys.* 72, 142–152.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 December 2010; accepted: 17 March 2011; published online: 25 March 2011.

Citation: Fields C (2011) Trajectory recognition as the basis for object individuation: a functional model of object file instantiation and object-token encoding. *Front. Psychology* 2:49. doi: 10.3389/fpsyg.2011.00049

This article was submitted to *Frontiers in Perception Science*, a specialty of *Frontiers in Psychology*.

Copyright © 2011 Fields. This is an open-access article subject to an exclusive license agreement between the authors and *Frontiers Media SA*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.