Keeping track of objects while exploring a spatial layout with partial cues: Location-based and deictic direction-based strategies

Nicolas J. Bullot, Jacques Droulez

Keeping track of objects while exploring a spatial layout with partial cues: Location-based and deictic direction-based strategies

Nicolas J. Bullot [a, b, *] and Jacques Droulez [a]

[a] Laboratoire de Physiologie de la Perception et de l'Action, Collège de France, 11 place Marcelin Berthelot, 75005 Paris, France.

[b] Institut Jean Nicod, CNRS-EHESS-ENS, 1 bis avenue Lowendal, 75007 Paris, France.

Abstract

A growing interdisciplinary literature has been concerned with object cognition and its relation to deictic/indexical reference mechanisms. This literature postulates that explanation of many phenomena in perception requires appeal to interactions between agent and environment, and to the ways that sensory-motor information connects to cognition through deictic (or demonstrative) reference. In this paper we report a study of the ability to keep track of distal targets in informationally impoverished situations – in particular when only the direction of these objects is known and not their locations. This study introduces a new experimental paradigm called the Modified Traveling Salesman Problem. This task requires subjects to visit once and only once *n* invisible targets in a 2D display, using a virtual vehicle controlled by the subject. Subjects can only see the directions of the targets from the current location of the vehicle, displayed by a set of directional segments that can be viewed inside a circular window surrounding the vehicle. Two conditions were compared. In the "*allocentric*" condition, subjects see the vehicle move across the screen and change orientation under their command. The "*egocentric*" condition is similar except for how the information is provided: the position and orientation of the vehicle icon remains fixed at the center of the screen and only target directions, as indicated by the directional segments, change as the subject "moves" the vehicle and changes its orientation relative to the objects (but not relative to the screen). The unexpected finding is that this task can be performed, in either condition, for up to 10 targets. We consider two types of strategies that might be used, "*location-based*" strategies and "*deictic direction-based*" strategies. Location-based strategies rely on spatial memory and attempt to infer the locations of all the targets. Direction-based strategies rely on a deictic frame of reference and focus on the directional segments themselves, keeping track of the ones that represent already-visited or to-be-visited targets. A number of observations suggest that the direction-based strategy was used, at least for larger numbers of targets, which is consistent with the deictic approach. According to our hypothesis, keeping track of the directional segments requires the use of deictic strategies for tracking segments and associating them with their status in the task – given by current status predicates *Visited*(x) or *Not-visited*(x) – perhaps using visual indexes (Pylyshyn, 2001), deictic pointers (Ballard et al., 1997), or object files (Kahneman et al, 1992).

Keywords: Deictic Reference, Location, Direction, Identification, Predicate Ascription, Visual Inferences, Object Perception, Spatial Memory, Indexicality.

* Corresponding author. Laboratoire de Physiologie de la Perception et de l'Action, Collège de France, 11 place Marcelin Berthelot, 75005 Paris, France. Tel. +33-1-4427-1624 ; fax. +33-1-4427-1382.

E-mail address: nicolas.bullot@college-de-france.fr (N. J. Bullot)

Internet website: http://www.objectcognition.net/NJB/

0 Roadmap of the article

In *section 1* of this article, we will explain the general interest of studying deictic (or demonstrative) reference in cognitive science. We will contrast two approaches to what perception is. The first views perception as a process of "reconstruction" in which an internal model of the perceived world is constructed. The second grounds perception in situated cognition, deictic reference and indexical representation and leaves much of the world as a source of perceptual information that can be consulted as needed. In *section 2*, we will describe a new experimental paradigm called the Modified Traveling Salesman Problem (MTSP). The aim of this paradigm is to distinguish these two distinct views of how vision and the other senses manage to maintain a connection with the perceived world. MTSP may also be viewed in the context of research on memory for object locations (Burgess, Jeffery, & O'Keefe, 1999; Morris, Nunn, Abrahams, Feigenbaum, & Recce, 1999; Posma & De Haan, 1996; Shelton & McNamara, 2001a) and visual inferences (Allwein & Barwise, 1996; Campbell, 2002; Evans, 1982; Gattis & Dupeyrat, 2000; Stenning, 2002; Ullman, 1984). We have applied the paradigm to study the use of mental indexicals linking memory representations with distal objects and to analyze subjects' ability to keep track of distal targets while moving among them with limited information concerning their locations (i.e., with only the relative direction of the targets). Finally, in *section 3*, we present the findings and propose an analysis based on the distinction between two types of strategies, location-based strategies and direction-based deictic strategies.

1 Perception grounded in deictic reference and indexical representation

1.1.1 The constructivist approach of perception

Classical experiments in psychophysics investigate the abilities of perceptual systems to recover information from sensory signals according to a *constructivist* view. This approach views the perception of a scene as the reconstruction of an internal representation or model of properties of the distal objects. According to this constructivist approach, the human capacity to perceive objects is defined entirely in terms of an internal model/representation of the physical (i.e. intrinsic, spatial) properties of the distal object – for instance, its 3D details and parts. The constructivist approach comes in a variety of subtypes, which differ in the properties or relations thought of as being 'modeled' in perceptual representations. Classical examples have been analyzed in research on vision (Ballard, Hayhoe, Pook, & Rao, 1997; Churchland, Ramachandran, & Sejnowski, 1994; Pylyshyn, 2000).

A basic constructivist problem in vision science has been the so-called "inverse problem" – or the problem of reconstructing a 3D layout from the 2D retinal stimulus; see for example Palmer (1999: 23-24). Marr's proposal (Marr, 1982) belongs to this approach. Related research in visual recognition of objects is also committed to constructivist models. As a consequence of this implicit hypothesis, many psychophysical paradigms focus on the relation between stimulus – or object-intrinsic – characteristics, and their internal representation. More generally, such approaches have not been concerned with the relation between the perceiver's strategies and inferences (e.g. goals, sub-goals, plans, visual inferences) and perceptual processing. In a related example, the pictorialist view of mental imagery is a constructivist approach, in the sense that it postulates internal mental displays that preserve fundamental spatial or topological relationships between represented elements (Kosslyn, 1994; Pylyshyn, 2002).

The constructivist approach in vision has been the target of several objections. We do not intend to present these objections in detail. Our goal is simply to refer to problems that are solved in an alternative approach grounded in the theory of deictic reference. On the falsification standpoint, empirical evidence against the view that the visual system constructs a complete (three-dimensional) model of the scene has been outlined by several authors (Hayhoe, Bensinger, & Ballard, 1998; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Horowitz & Wolfe, 1998; Karn & Hayhoe, 2000; O'Regan & Noë, 2001; Pylyshyn, 2001, 2003; Rensink, O'Regan, & Clark, 1997; Simons, 2000). Moreover, it is worth remarking that these objections directed against the analyses of 3D internal reconstructions could be adapted to other versions of the constructivist approach that are based on 2D internal reconstructions (e.g., Clark, 2000: 100; Humphreys, 1999: 175-176; O'Regan, 1992: 464-471; Pylyshyn, 2002). On the theoretical standpoint, the idea of an elaborate scene model, has been challenged by the emergence of alternative approaches, called active (Findlay & Gilchrist, 2001, 2003), animate (Ballard, 1991), deictic (Agre, 1997; Ballard et al., 1997), embodied (Clark, 1999), interactive (Churchland et al., 1994) or situated (Brooks, 1999; Pylyshyn, 2000) vision and cognition, that take advantage of the observer's actions in order to minimize the number of object-intrinsic properties that must be encoded in the (visual) representation. In these works, there has been a lasting discussion of various shortcomings of constructivist views, such as the need for a homunculus for examining the internal model – or seeing the 'internal screen' (O'Regan, 1992), 'Cartesian Theater' (Dennett, 1991, 2001), 'internal image' (Pylyshyn, 2003), or 'internal 3D model' (Ballard, 1996; Churchland et al., 1994).

1.1.2 Perception grounded in deictic reference

The theory of deictic reference[1] can be thought of as an alternative approach to that championed by the constructivists. This approach relies on the appeal to perceptual demonstratives (or mental indexicals) and deictic routines that allow the visual system to *keep track of*, and *reason about* a *limited* number of distal objects without re-constructing a detailed model of these distal objects in the mind or brain. This parsimony explains the growing interest in cognitive science for new models based on deictic reference.

Zenon Pylyshyn (2001; 2003) has tried to give a unified view of this new trend. He has also used an experimental methodology for studying these deictic capabilities, which had already been studied in the philosophy of mind and language. Though deictic abilities have been studied previously in philosophy of mind and language, Pylyshyn introduced the use of experimental methodology to their investigation. Pylyshyn (2001; 2003) argued that classical theories of object perception, or of representation in general, cannot give satisfactory explanations of the way representations connect with the real world, because these theories do not clearly analyze the interactions between the active perceiver and his environment. According to Pylyshyn, we need to explain how the subject can connect sensory-motor information with conceptual knowledge (descriptive or propositional representations) about the environment. To do that we must study the use of mental indexicals, or deictic representations, which serve to "directly" connect token objects with mental representations *or* with certain actions that may be performed on them.[2] Traditionally in philosophy and semantics, a representation – such as a thought, a belief, or an utterance – is said to be indexical [3] if its referent depends on the particular context in which it occurs. Pylyshyn proposes to deploy this

[1] For purposes of our present discussion, we regard the phrases "deictic reference" and "demonstrative reference" as synonyms.

[2] Pylyshyn (1989; 2001) has suggested that deictic reference in the visual system is closely related to a mechanism, called a "visual index" or "FINST" (from FINgers of INSTantiation), which links individual object tokens with information upon which beliefs and actions can be based. A visual index serves like a "pointer" from a (conceptual) representation of an object to an actual object in the scene.

[3] The concept of 'indexical' has been traditionally used in the philosophy of language and semantics in order to refer to terms like pronouns 'I', 'my', 'you', demonstrative pronouns like 'that', 'this', or adverbs like 'here', 'now', 'tomorrow'. The classical observation relative to the use of these terms is that : (i) the referent depends on the context of use; (ii) their meaning provides a rule which determines the referent in terms of certain aspects of the context (Kaplan, 1989, 490); (iii) 'true' demonstratives – as David Kaplan called them in his seminal article *Demonstratives* (Kaplan, 1989, 490) – need an associated demonstration (typically: a pointing) in order to determine their referent (this means that the linguistic rules which govern the use of the Kaplanian true demonstratives are not sufficient to determine their referent in all contexts). Since then, indexical representations have been studied in many domains, from semantics and philosophy of language (e.g., Dokic, 1996; Lewis, 1979; Lyons, 1977; Perry, 2000, 2001; Recanati, 1993) to artificial intelligence models of action and planning (e.g., Agre, 1997: 230-34; Lespérance & Levesque, 1995).

notion of indexicality in developing a theory of vision and visual attention. What he calls "indexical reference," or "demonstrative reference," therefore applies not only to linguistic reference to the object, but also to the visuo-motor capacity of perceivers. It allows them to individuate and act upon objects by keeping track of them in the course of active perception. This extension is in the spirit of several recent works concerned with perceptual reference (Ballard et al., 1997; Ballard, Hayhoe, Salgian, & Shinoda, 2000; Campbell, 2002; Clark, 2000; Hayhoe et al., 2003; Land, Mennie, & Rusted, 1999; Scholl, 2001).

The notion of a deictic *perceptual* reference is a particular direct mode of presentation of (or reference to) an object *x*. It occurs when one is currently perceiving the object *x* (e.g., a cat, a cup, a person) and needs to refer to it in perception, thought or communication based on that perception. A deictic reference to an object therefore requires a particular *episodic* encounter with that object. Hence, since early analyses (e.g., Kaplan, 1989; Russell, 1911, 1956), deictic (or *de re*) modes of presentation are usually contrasted with descriptive ones (Bach, 1987; Clark, 2000: 131-136; Evans, 1982: 143-45; Pylyshyn, 2000: 199; 2001: 129-130; Recanati, 1993: 97-118). For instance, consider these two sentences: (*i*) "This [*pointing in the direction of x*] is *P*."; and (*ii*) "The *D* is *P*.", wherein *D* is a description satisfied by *x*. Only (*i*) is traditionally conceived of as a deictic reference to object *x*, because it is assumed to refer to a particular token of an object and to ascribe the predicate *P* to it without the mediation of an explicit conceptual description.

Deictic reference is frequently studied in terms of the following distinct aspects: (1) the "informational link" with a token individual, and (2) the binding of a predicate to a token individual which allows one to draw perceptual inferences. The first aspect concerns the fact that deictic reference is the capacity to maintain cognitive *access* to an individual object by the means of the body's sensors and effectors, and by selection in focal attention. Thus, it relies essentially on the binding of sensory-motor processes to distal objects. This general characteristic is implicitly endorsed by most of the analyses in philosophy.[4] Likewise, recent psychological theories have studied the *link* between cognitive states (e.g., content of working memory, intention detection, and communication) and various sensory-motor primitives whereby deictic reference can be achieved. These sensory-motor primitives include eye fixations[5], visual and attentional tracking[6], pointing gestures[7], hand grasping[8] or gestures that help the anchoring or learning of language[9] and reasoning[10]. In particular, according to several authors (Ballard et al., 1992; Ballard et al., 1997: 725; Findlay & Gilchrist, 2003: 145-47), eye fixations are involved in *deictic strategies* that use an external frame of reference centered at the fixation point as a pointer to information. For example, an object is usually grasped by first looking at it and then directing the hand to the center of the fixation coordinate frame (Jeannerod, 1988; Milner & Goodale, 1995).

---

[4] Traditional philosophical theories of demonstrative identification (Evans, 1982; Kaplan, 1989; Strawson, 1959) and of *de re* thoughts (Bach, 1987; Burge, 1977; McDowell, 1984; Segal, 1989) emphasize the epistemic importance of linking sensory discrimination with conceptual representation. In such a framework, it is generally assumed that situated reference is realized by the capacity to make sensory-motor discriminations – see Clark (2000, 130-163) for a philosophical overview. Following Strawson (1959: 18-20) and Evans (1982: 121, 146), Campbell (2002) makes the further claim that selective attention might be a crucial mechanism for accomplishing deictic reference.

[5] Many recent experimental works have studied eye fixations in the framework of deictic reference theory (see for instance Ballard, Hayhoe, Li, & Whitehead, 1992; Ballard et al., 1997; Findlay & Gilchrist, 2003; Hayhoe, 2000; Hayhoe et al., 2003; Henderson, 2003; Land & Furneaux, 1997; Land & Hayhoe, 2001; Land et al., 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1996; Triesch, Ballard, Hayhoe, & Sullivan, 2003).

[6] See for example the works of Pylyshyn et al. (e.g., Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999), Logan (1995), Yantis (1992).

[7] Pointing gestures have been studied in the literature on joint attention, language development and communication (Baldwin, 1993; Baron-Cohen, 1997; Bruner, 1983; Butterworth & Grover, 1990; Desrochers, Morissette, & Ricard, 1995).

[8] See for example the works of Jeannerod (1988) and Goodale and Milner (Milner, 1999; Milner & Goodale, 1995).

[9] See for example the review of Baldwin (1993).

[10] See for example the works of Goldin-Meadow et al. (Garber & Goldin-Meadow, 2002; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001) and Kirsh (1995a; 1995b).

As for the second aspect, deictic reference relies on the binding of external targets to (cognitive) predicates, which can then take part in *inferential* processes. This condition is accepted in various analyses. For instance, the study of deictic identification is closely related to the study of predicative structures, particularly in the work of Strawson (1959), Evans (1982), Campbell (2002) – and recently Hurford (in press). Thus, deictic identification can be conceived of as relying on mechanisms that make it possible to represent facts about particular individual things in using property or predicate ascription. We could represent a cognitive predicate by the expression "$F(x_1, \ldots x_n)$". In this notation, "$(x_1, \ldots x_n)$" is a set of n deictic variables or indexes linked to individual objects and F is a conceptual or non-conceptual predicate linked with each of these objects (Ballard et al., 1997; Evans, 1982; Hurford, in press; Miller & Johnson-Laird, 1976; Pylyshyn, 2001).

1.1.3 From perceptual tracking to keeping track cognitively: visual tracking and visual inferences

From these characteristics, one can conclude that a system performing deictic reference has to solve computational problems related to the cognitive *tracking of token individuals*. Typically, the deictic system has to keep the (bi-directional) informational link with an individual target while incrementally constructing an episodic representation *of* the target (Dretske, 1981; Evans, 1982; Pylyshyn, 2001: 129-131). "To keep the informational link" refers here to the ability to pick out a token individual x (that is, to select by focal attention) and keep a cognitive contact with it through space-time, in spite of x's property changes. The link is necessary to preserve the ability to distinguish x from any other individual and to preserve the connection between x and a particular predicate or "tag". From a computational standpoint, keeping the informational link requires solving "binding problems" such as the problem of "the incremental encoding of information on each token object" (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 2000; Strawson, 1959: 34-6; Treisman, 1996; Ullman, 1984), and managing spatio-temporal coordination so as to maintain the object-predicate connection and avoid erroneous predicate ascription. With regard to the indexical representation, this tracking ability has an explanatory role since the correctness value (or semantic value) of the related episodic representation depends on tracking success or failure.

With respect to behavioral research in vision, several studies have been published on the ability to track visual objects (e.g., Blaser, Pylyshyn, & Holcombe, 2000; Pylyshyn & Storm, 1988; Scholl & Pylyshyn, 1999). However, understanding deictic reference requires more than the study of perceptual tracking in isolation. That is, it also requires an understanding of how perceptual tracking of a token object is carried out *and* how this relates to *cognitively keeping track of* the object when it leaves the sensory field (e.g., Ballard et al., 1997; Graziano, Hu, & Gross, 1997; Strawson, 1959: 18-23, 32-6). We frequently have to *keep cognitive contact* with several distal objects around us even though we may not be able to see or perceive them – for instance when we have to remember recently seen or used objects (McNamara, 2003; Posma & De Haan, 1996). Hence, when perceptual information about distal objects becomes impoverished or completely absent, the capacity to keep cognitive contact requires additional capacities, such as memory, mental imagery, communication or reasoning.

As a result, we should distinguish (1) the *perceptual* tracking of a set of n targets (where targets remains in the receptive fields of sensory systems) and (2) *cognitively* keeping track of these targets (e.g., by memory, communication and reasoning). The latter may depend on spatial memory of distal target location and identity – the capacity of automatic spatial updating of locations and directions (Farrell & Thomson, 1998, 1999; Rieser, 1989) and keeping track of distal targets by recall of their partially-specified locations and identity (Attneave & Farrar, 1977; Attneave & Pierce, 1978; Milner & Goodale, 1995: 88-91; Posma & De Haan, 1996; Shelton & McNamara, 2001a). It may also depend on the capacity to draw perceptual inferences about *the* distal target *objects*, on the basis of perceptual indexicals and episodic object-based representations.[11] The use of visual indexicals ("pointers" or "indexes") serves to link thoughts and memories with distal token individuals (Ballard et al., 1997; Pylyshyn, 2001).

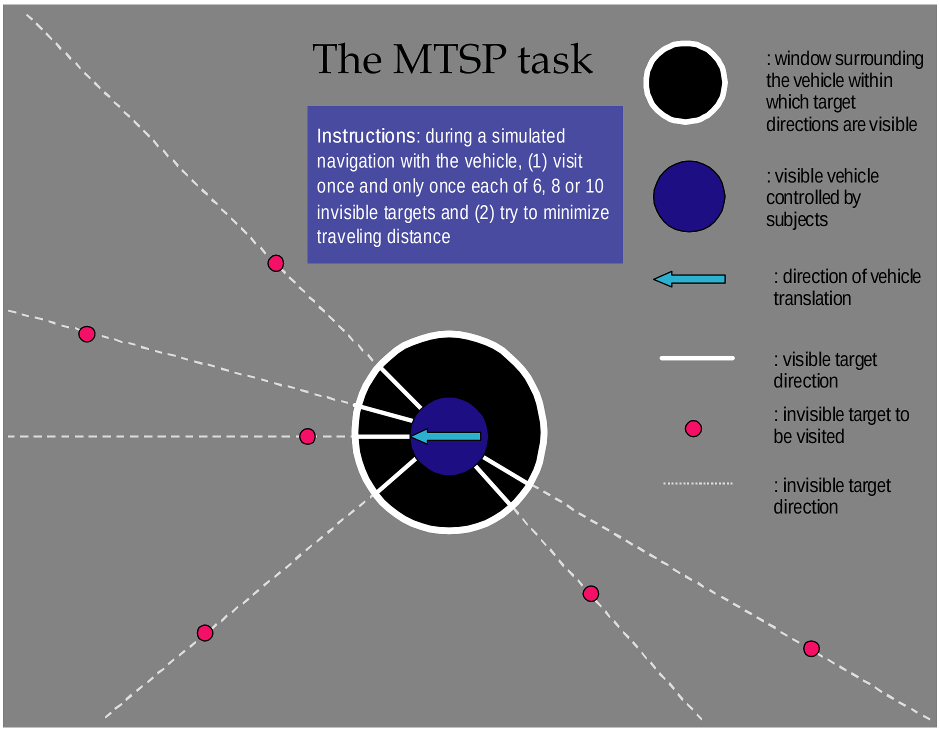
In the present work, we study how the mind and brain can bind these different capacities and address the following two questions. First, how do we keep track of target objects when information about the objects' location is partial or indeterminate? Second, how many stationary targets can one keep track of while moving among them, when only directions of the targets (and an indication of when the targets are actually encountered) are known? We studied these two questions using a new paradigm, the Modified Traveling Salesman Problem (MTSP).

2 Modified Traveling Salesman Problem (MTSP)

2.1 Principles and structure of the MTSP paradigm

2.1.1 Principles, goals, instructions

The experimental paradigm, which we call the Modified Traveling Salesman Problem (hereafter MTSP), serves to study the capacity to keep track of distal targets during a simulated navigation. More precisely, it designed to study how subjects can *keep cognitive contact* [12] with a set of *n* invisible targets knowing only their direction with respect to the vehicle and whether they had been contacted in the course of the vehicle's travels (Figure 1). This task was inspired by the Traveling Salesman Problem (hereafter TSP) – i.e. find the shortest path for joining a set of *n* points (Michie, Fleming, & Oldfield, 1968; Verblunsky, 1951) – which has been studied in various domains including psychophysics (MacGregor, Ormerod, & Chronicle, 1996; MacGregor, Ormerod, & Chronicle, 1999; MacGregor, Ormerod, & Chronicle, 2000; Ormerod & Chronicle, 1999).



[11] About object-based perceptual inferences, see for instance Ullman (1984), Ballard et al. (1997), Campbell (2002: 84-113), Hodgson et al (2000), Pylyshyn (2003), Stenning (2002: 54-92).
[12] In the MTSP task, "keep cognitive contact" means to keep a cognitive record of history of contacts with targets and/or with their location.

Figure 1 The standard MTSP (in "allocentric" condition according to the taxonomy explained below): Schematic Displays shown to subjects during the MTSP experiment (background is presented here in grey instead of black in the original display). The blue arrow indicates the vehicle frontal direction, which is its direction of translation. White lines point towards the invisible targets (6 in this case, visible here in red). (Demonstrations – small movies – of the experiment are available at on the Internet Site of the first author.)

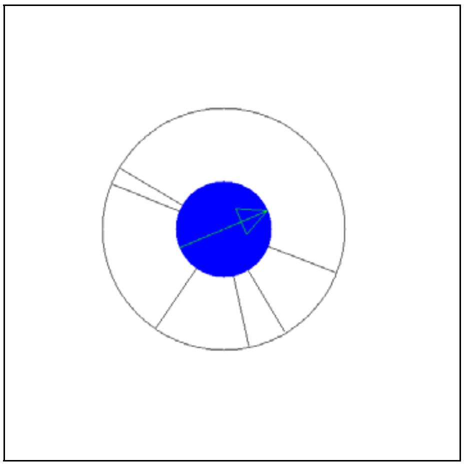


Figure 2 Screenshot of the MTSP display (not to scale, segment and background colors have been inverted). The image comes from a trial with six targets, in the allocentric condition (the arrow can move inside the circular blue disk). One can see the six white directional segments which point toward each target location.

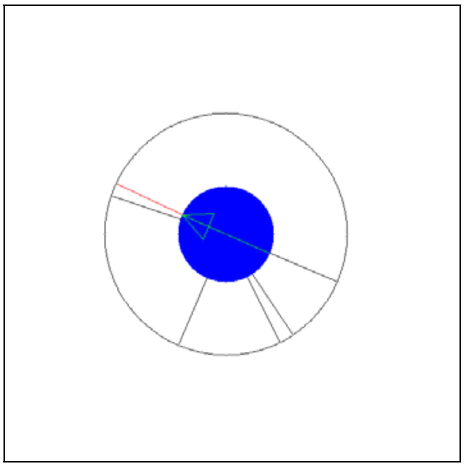


Figure 3 Detail of a screenshot of the MSTP display (not to scale, segment and background colors have been inverted): when a target is hit, the segment that was followed changes its color from white to red. This image comes from a trial with 6 targets, in the allocentric condition.

The goal of the MTSP task is to reach each target by controlling the displacements of a small circular vehicle displayed on a computer screen. A set of *n* stationary targets (x_1, x_2, … x_n) have to be visited in sequence. Subjects are instructed to (1) move the vehicle so as to visit each invisible target *once and only once* (primary goal), and (2) try to minimize the distance traveled (secondary goal). The specific constraint in the MTSP task is that *only* the *directions* of targets from the current vehicle location are shown – indicated by directional segments displayed inside a circular window surrounding the vehicle (see Figure 1). These directional segments are referred to in the article by the symbol 's_n' and will be referred to by particular names such as

s_1, s_2, … s_n.[13] The directional segments are all identical in appearance. Each can be distinguished, or picked out, only on the basis of its orientation and motion, i.e., on the basis of its spatio-temporal trajectory. In order to reach a token target, the subject typically follows a particular directional segment. When the target is hit, the segment being followed changes its color (see Figure 3). Only this color change provides the crucial information that one has *reached* a target. In order to complete the task, subjects may draw different *visual inferences* by scrutinizing the motion of the directional segments, which co-vary with their motor commands. Our goal was to study these visual inferences.

2.1.2 Deictic strategies in the MTSP task and current-status predicates

The challenge of the MTSP task is to keep track of the *status* of each target (whether or not it has been visited) in a context where targets are invisible and directional segments are identical in appearance. Hence, subjects cannot directly pick out targets either by feature or by location. They must nonetheless *distinguish* each target from the other targets in order to avoid missed or multiple visits. Two questions are raised by this task. First, given the restriction that prevents visual tracking, and given subjects' memory limitations, how can subjects individuate and *keep track* of each target? Second, what strategies are available for solving the MTSP task? Our approach of these questions is related to the theory of deictic reference (Ballard et al., 1997; Campbell, 2002; Pylyshyn, 2001, 2003). Below we suggest a way of experimentally studying possible strategies by examining performance in two different conditions of the MTSP task, which we refer to as the "egocentric" and the "allocentric" condition. Our hypothesis is that the MTSP task can be performed with deictic strategies involving visual inferences about invisible targets. Deictic strategies – in the sense of Ballard et al. (1997: 725), also called informally "do-it-where-I'm-looking" strategies – use an external frame of reference centered at the fixation point as a "pointer" to information, that can rapidly moved to different locations and be used for the execution of motor or cognitive programs. This kind of strategy has been documented as having many advantages. It allows computational and memory economy (either in cognitive programs or in sensory-motor routines), and simplification of cognitive programs (Ballard et al., 1997, 739) by the means of "minimal memory strategies" (Ballard et al., 1997: 731-34), or a reduction of computational complexity (Agre, 1997; Agre & Chapman, 1987; Hendriks-Jansen, 1996). For instance, visual fixations have been described (Ballard et al., 1997; Hodgson et al., 2000) as a deictic mechanism which allows information to be recovered in due time without requiring the memorization or representation of 3D geometrical structures of objects.

Because targets cannot be visually selected (since they remain invisible), directional segments themselves may be the target of focal attention. For instance, a subject could use deictic references such as suggested by the following sentence, using the demonstrative 'this':

> This segment [*eye fixation and focal attention directed at a token directional segment*] points toward the second target to be visited.

Using such indexical representations requires the capacity to link the tracked segments with a "tag" or a current status predicate. This point connects with one of the interests of deictic references, namely that they can be used to bind the referent to conceptual predicates (Garnham, 1989; Hurford, in press; Logan & Sadler, 1996: 494; Miller & Johnson-Laird, 1976; Pylyshyn, 2001: 174). Predicate-argument link is a binding capacity that may require deictic variables. Presumably, once a system has a set of indexed deictic variables, it can predicate or "tag" attributes to these variables.[14] In the MTSP task, counting visits and avoiding revisits may

[13] In this article, according to the terminology of semantics, a symbol such as 's_1' is a proper name for a particular directional segment. The name 's_1' refers to one and only one token individual (in this case, the directional segment s_1 on the display). This notation is used only for convenience. Of course, subjects in MTSP do not possess any proper name for identifying particular directional segments (see below). According to our hypothesis (see below), instead of proper names, they use indexical representations for identifying directional segments and targets.

[14] Ascribed predicates or "tags" are not necessarily properties of the reference of the deictic variable: it can be attributes related to the relation between the deictic system and its target-reference, like the current-status predicate described in this section.

involve the ascription two types of current-status predicates. The first type involves predicates of ordering and counting, such as:

First(s), *Second*(s), *Third*(s), *N*(s).

The second type involves status predicates with respect to the navigation goal, such as:

Visited(s), or *Not-visited*(s).

Both types of predicates convey information about context-dependant states of affair that cannot be *directly* visible in the display. When an embodied subject ascribes a current status predicate to a particular directional segment, she focuses on the relation between the selected segment and the current step of her action. Such an ascription is therefore context-dependant at several levels. It depends on the selected token individual, for which it is correct or incorrect (with respect to the subject's ongoing action). Moreover, its value is related to a specific time of performance: the ascription of the current-status predicate can be correct at a time t_1 but incorrect at a time t_2.

2.1.3 Visual inferences based on "location-based" and deictic "direction-based" strategies

The MTSP task allows investigation of the strategies subjects use in order to keep track of targets in a spatial layout with partial cues. Strategies are linked with specific types of visual information and visual inference based on deictic references. Generally speaking, in the absence of featural information, two types of strategies might be used for solving the MTSP task: "location-based" strategies and "direction-based" strategies. Each type of strategy might work in a global or grouping variety, meaning that there are four logically possible strategies. Each of these four logically possible strategies is described below, in *Table 1*, as a function of their respective frame of reference, relevant cues and memory requirements.

A central characteristic for this taxonomy is that location-based and direction-based strategies use distinct *frames of reference* and *distinct cues*. Location-based strategies work to infer the allocentric locations of the set of targets (which is the complete set in the global variety) and to compute a mental map of the n target locations. Location-based strategies rely thus on an allocentric frame of reference, with respect to which target locations can be inferred and memorized. For instance, by controlling the movement of the vehicle, subjects can gather information and draw inferences about spatial properties related to directional segments and targets. These inferences primarily include the segments' orientation and angular velocity as a function of the vehicle position and movement. Subject can also estimate the relative distance, and perhaps even the time-to-contact of a particular target x, on the basis of the angular velocity of the directional segment pointing toward x. The inference rule might be something like:

> If this segment [*selection of* s_1] which points towards the first target [*i.e., target* x_1] moves faster than the others [*i.e.,* s_2, s_3, … s_6], then the vehicle is closer to the first target than to the others [*i.e.,* x_2, x_3, … x_6].[15]

In addition, target locations are known precisely when they are encountered, since they are located where the vehicle is standing at each time the color of a directional segment changes. As a consequence, in location-based strategies, the subject has to memorize the vehicle allocentric location when this change occurs (relevant cue (i) for the location-based strategies). Since they rely on the recall of n target locations, location-based strategies puts emphasis on visuo-spatial memory capacities, and do not emphasize the vehicle as a visual target-object for securing a frame of reference.

In contrast, the principle of direction-based strategies consists in the visual tracking of n directional segments, of the ones that represent to-be-visited or already-visited targets, so as to keep the dynamic link between each directional segment and its correct current-status predicate.

[15] In MTSP, if Θ is the angle between straight-ahead and the direction of the segment pointing to some target, and if the vehicle is moving at a constant velocity v, then the closer the target object is the faster will the angle be changing. Although this does not tell the observer how far the vehicle is from the target, it does provide some guidance about the relative distance to various targets (for small non-zero Θ).

In contrast with location-based strategies, direction-based strategies are “deictic” in the sense that they use deictic frames of reference, centered on the gaze fixation point or relative to the current focus of attention (Ballard et al., 1997) – fixations and attention being used as a “pointer” to relevant information. In the MTSP task, direction-based strategies are deictic strategies since the focus of attention has to split exclusively over the *directional segments* themselves, without wondering about the target locations. In contrast to the global location-based strategy, one can carry out this strategy without keeping target locations in memory because one only needs to bind each selected segment with its current-status predicate. Typically, direction-based strategies aim first at a color change of a token directional segment (target contact), update the current-status predicate (from *Not-Visited*(x) to *Visited*(x)), and then shift the focus of attention to another segment, until all segments’ colors have been changed once.

Both location-based and direction-based *grouping* strategies rely on the same general principle: Grouping or “chunking” either locations or segments, in order to compensate memory limits. These strategies are useful for conditions with high number of targets (e.g., with n = 10 or higher value), in which the complete set or locations/segments cannot be memorized with their correct current status.

Table 1 Taxinomy of possible strategies in the MTSP task: location-based and direction-based strategies

Types of strategies		*Frames of reference*	*Relevant cues for visual inferences*	*Memory requirements*
Location-based	*Global*	*Allocentric* (i.e. indexed on environmental landmarks – e.g., screen edges)	(i) Vehicle location when a change of color occurs (target location). (ii) Segments’ angular velocity (inferences about relative distance).	Recall of all n target locations in the allocentric framework.
	Grouping	*Allocentric*	(i) Vehicle location when a change of color occurs. (ii) Segments’ angular velocity. (iii) Spatial proximity of subgroups of targets.	“Chunking” recall of a subset of the n target locations in the allocentric framework.
Direction-based	*Global*	*Deictic* (i.e. vehicle and gaze-fixation centered – relative to the current focus of attention)	(i) Segments’ changes of color. (ii) Spatio-temporal continuity of segments’ motion for visual tracking.	Recall of the current-status of each of all n segments in the deictic frame of reference.
	Grouping	*Deictic*	(i) Segments’ changes of color. (ii) Spatio-temporal continuity of segments’ motion for tracking (esp. when crossing over). (iii) Spatial proximity of subgroups of directional segments.	“Chunking” recall of the current-status of a subset of the n segments in the deictic frame of reference.

2.1.4 Available strategies in "allocentric" and "egocentric" conditions

In order to investigate which strategy was chosen by subjects, we introduced a new condition to the standard display, which we refer to as the "egocentric" condition. In the standard "*allocentric condition*" (see Figure 4), targets are fixed within the screen reference-system. The vehicle icon and arrow moves (according to the subject's commands) both in translation and rotation in the screen reference-system. In this condition, subjects may be able to use the location-based strategies for three reasons. Firstly, they can use the screen borders (and other stationary features of the visible environment) as the relevant allocentric landmarks for memorizing target location, since in this condition the target locations and inter-target spatial relations are fixed (stationary) in screen coordinates. Secondly, they can infer target positions with respect to the screen from the cues provided (directional segments motion and vehicle successive locations). Finally, they could construct in memory a spatial representation of target locations with respect to the screen frame of reference. However, location-based strategies may not be used in the "*egocentric condition*". In this condition, the vehicle icon and arrow remain fixed at the center of the screen (frontal direction upward). Only directional segments move according to subject's commands, as if subjects were directly moving the display beneath their apparently stationary vehicle. As a consequence, target locations on the screen are no longer fixed, even though target-target spatial relations are stable and coherent within an invisible rotating and translating frame of reference (which changes relative to the vehicle orientation and movements). Since the relevant frame of reference for locating target moves and remains invisible, it is unlikely that subjects use it in a location-based strategy. As a consequence, one would expect either direction-based strategies or the grouping location-based strategy to be chosen in the egocentric condition.

According to our experience of explaining the goal of the MTSP task to subjects, the most natural and ecological strategy is the global location-based strategy. If this were true, the difficulty of the MTSP task should be different in the two conditions. Because the egocentric condition may not allow recovery of target locations within an allocentric frame of reference, this condition should be difficult, if not impossible, if subjects were not spontaneously use a direction-based deictic strategy.

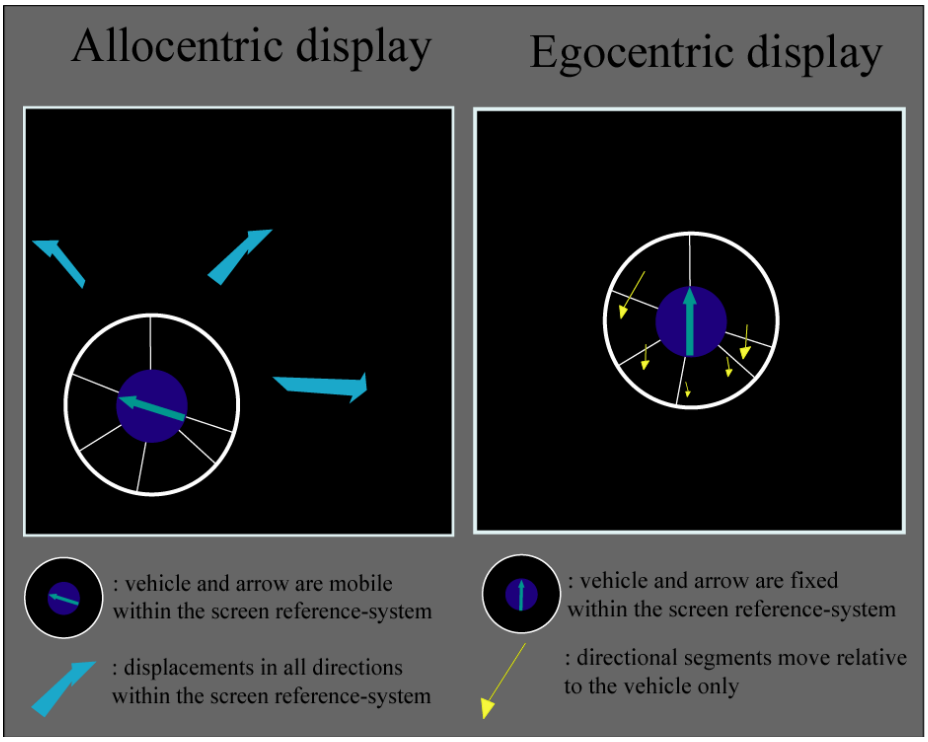


Figure 4 Schematic depiction of the "allocentric" and "egocentric" conditions of the MTSP task.

2.2 Method

Subjects controlled the movement of a vehicle displayed on a computer screen, and were told to visit each of a set of *n* targets. Only the target directions with respect to the controlled vehicle were displayed, so subjects had no *direct* way to evaluate target distances or locations. They were instructed to move the vehicle to the vicinity of each target in order to:

(1) To visit each target once and only once (highest priority constraint); and

(2) To try to minimize the travel distance (lower priority constraint).

This task was performed under 6 experimental conditions. There were three values of the number of targets (6, 8 or 10), and two types of displays (allocentric and egocentric). In the "*allocentric*" condition, targets had fixed positions with respect to the screen frame of reference, while the displayed vehicle changed position and orientation under subject control. In the "*egocentric*" condition, the vehicle position and orientation was fixed with respect to the screen, but target positions and directions with respect to the vehicle changed according to the displacements of the vehicle. Thus, in the egocentric condition, target locations, including their angular orientations relative to the screen, changed as the vehicle moved and rotated. We hypothesized that it would be very difficult to keep track of target locations in this moving frame of reference.

Subjects were 9 adult volunteers, 5 males and 4 females. All of them were either students or colleagues of the laboratory, and were naïve with respect to the hypotheses under examination. They were seated in front of a 17" computer screen at their usual working distance. They were given instructions prior to training with four simple examples (in the training trials the number of target was reduced to $n = 3$ and $n = 4$). They were familiarized with the vehicle controls using the computer keyboard. The Experiment was carried out in two sessions, with 3 blocks of 15 trials. During the first session, subjects performed the allocentric condition in three successive blocks, with 6, 8 and 10 targets. During the second session, subjects performed the egocentric condition, in three additional blocks with 6, 8, and 10 targets. The total duration of each block was about half an hour up to one hour and a half, but subjects could rest between trials and blocks. Subjects decided when each trial was finished, i.e. when *they* thought all targets had been reached. Within each block, target locations as well as the initial vehicle position and orientation were randomized. The average distances between closest locations on the screen (either targets or initial vehicle position) were about 197 pixels (for $n = 6$ targets), 167 pixels (for $n = 8$) and 144 pixels (for $n = 10$).

The visual display (resolution 1280*1024 pixels) was refreshed at 60 Hz. It consisted of a circular window (radius 64 pixels), in which the vehicle direction was indicated by an arrow, while target directions with respect to the vehicle were depicted by straight radial segments (see Figure 1). By pressing special keys in the keyboard, subjects could independently control the vehicle orientation and displacement. Vehicle displacement was only along the straight-ahead direction indicated by the arrow. Instantaneous vehicle position and orientation were recorded in data file for further processing.

Subjects' performance was evaluated by two criteria corresponding to the two goals of the task (which had different priorities). The first criterion was the *percentage of correct trials*. An individual trial was classified as correct if all targets had been visited once and only once, regardless of the travel time or distance. The second criterion, only applied to correctly performed tests, was the *normalized travel distance*, defined as the ratio of the actual travel distance to a normalization distance. The normalization distance was computed for each individual trial as the sum of the distance between initial location and the closest target, plus the distance between the first target to the closest remaining target, and so on to the last target. The normalization distance is slightly greater than the optimal solution of the traveling salesman problem, i.e. the shortest path starting from the initial point and passing through all targets. Percentage of correct trials and normalized travel distance were used to evaluate subject performance. We considered scores of 50% correct trials or better to indicate a capacity to solve the MTSP task. This threshold of 50 % is conservative, as random choice of targets would yield a percentage of correct trials equal to about 4% for 6 targets and less than 1% for 8 and 10 targets (that is, equal to 100 $(n-2)!/(n-1)^{n-2}$).

2.3 Results

Eight of the nine subjects were able to perform the task, even with 10 targets in the egocentric condition. An example of a trajectory in the 10 target condition is shown in figure Figure 5A. Only one subject systemically failed to perform the task in either of the two conditions. The trajectories produced by this subject (Figure 5B) were the convex hulls of the sets of targets. Thus, we excluded this subject.[16]

Percent correct trials (PCT) tends to decrease with the number of targets, although this effect was significant only in egocentric condition with 8 and 10 targets (t = 3.33, p = 0.012) and when egocentric and allocentric conditions were pooled between 6 and 10 targets (PCT_6 = 83.3% (SD = 15.2 %), PCT_{10} = 66.7 % (SD = 13.7 %), t = 2.65, p = 0.01). No statistical differences were found between egocentric and allocentric conditions, regardless of the number of targets. In fact, the difference always tended to be in favor of egocentric conditions (see table 2). That decrease of PCT between 8 and 10 targets is consistent with a constant probability of assignment error of about 3%,[17] and therefore, cannot be interpreted as an abrupt fall-of of performance reflecting subjects' limited resources (see also section 3).

Table 2 Percentage of correct trials (PCT) in each condition

	6 targets	*8 targets*	*10 targets*
PCT Allocentric	82.5 % (SD = 14.9 %)	72.5 % (SD = 14.4 %)	64 % (SD = 18.0 %)
PCT Egocentric	86.7 % (SD = 16.3 %)	84 % (SD = 5.96 %)	70 % (SD = 6.7 %)

No clear statistical difference between conditions was found for the normalized travel distance (NTD) score. Only a marginal increase of travel distance from allocentric to egocentric conditions was found, by pooling NTD for all target numbers ($NTD_{allocentric}$ = 1.26 (SD = 0.087), $NTD_{egocentric}$ = 1.33 (SD = 0.16), t = 1.83, p = 0.07). The analysis of trajectories showed that subjects changed their choice of the next target to visit from time to time (see Figure 5A for an example). This suggests that they were not able to accurately estimate or compare the distance to unvisited targets at the time they chose which target to visit next. The normalized traveled distance was clearly sub-optimal and was found to increase with the number of targets. The traveled distance was 30 % to 40 % higher than the performance predicted by the nearest neighbor heuristic, which itself provides about 10% distance above the optimal solution. As expected, these results differ from those of standard TSP experiments.[18] In contrast to PCT, NTD was higher in the egocentric condition compared with the allocentric condition. The

[16] This exclusion is justified by three reasons. First, her results was atypical (PCTallo 6 targets =26.7 %, 13.3% in PCTego 6 targets, and 0% in all other conditions). Second, it was impossible to compute NTD on all trials with her scores (NTD can be computed only for correct trials). Third, she apparently did not understood the nature of the task or did not understand why her strategy was not reliable for performing the task.

[17] A constant assignment error probability for each target give rise to a PCT roughly proportional to the target number – more precisely: P(correct trial) = $[1- \text{P(target assignment error)}]^n \approx 1- n\text{P(target assignment error)}$.

[18] MacGregor and co-workers (1996) have measured human performance in solving the standard traveling salesman problem for 10 and 20 points. They found that, on average, subjects solutions lie somewhere between the optimal and the nearest neighbor solutions (about 4 % above optimal length for 10 points). Clearly a number of experimental conditions may account for this difference, the most likely among them is that in MTSP, contrary to the standard TSP, the exact location of points and therefore the in-between distances are not directly given to subjects. Moreover, our data are computed from the actual travel distance including detours, while Euclidean distance between connected points are used in standard TSP.

allocentric display, which showed the vehicle displacement in display coordinates, seemed to improve the ability to evaluate target distances, but did not reduce the target assignment error.

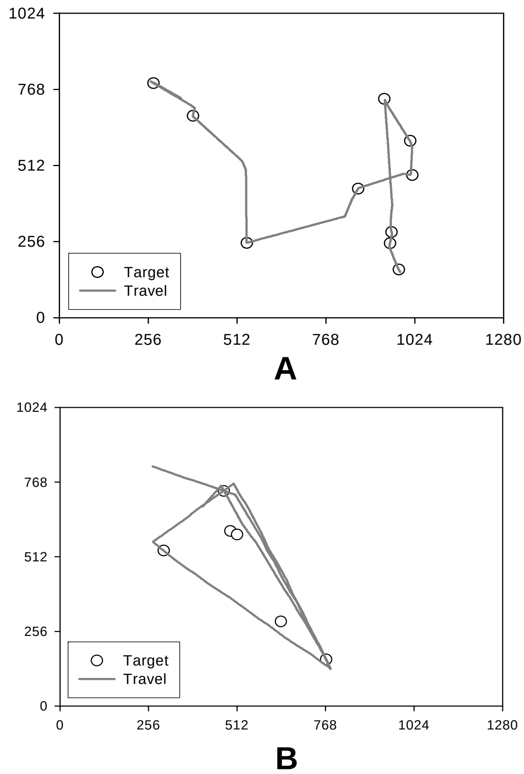


Figure 5. Examples of trajectories (solid lines) reconstructed from two different subjects. The location of targets are shown by small circles. X/Y coordinates are expressed in pixels. A: 10 targets in egocentric condition. B: 6 targets in allocentric condition. The subject in B fails to visit the three inner targets while visiting twice or more the outer ones.

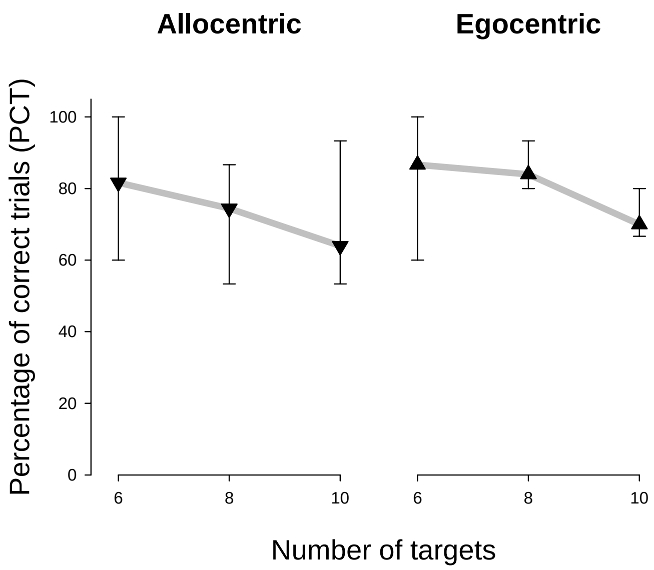


Figure 6. Percentage of correct trials (PCT), averaged for 8 subjects, with 6, 8 and 10 targets in both allocentric and egocentric conditions. Taking into account the large variability between subjects, there is no clear increase of mistakes in egocentric condition as compared to allocentric one. Error bars represent minimum and maximum performance for each condition.

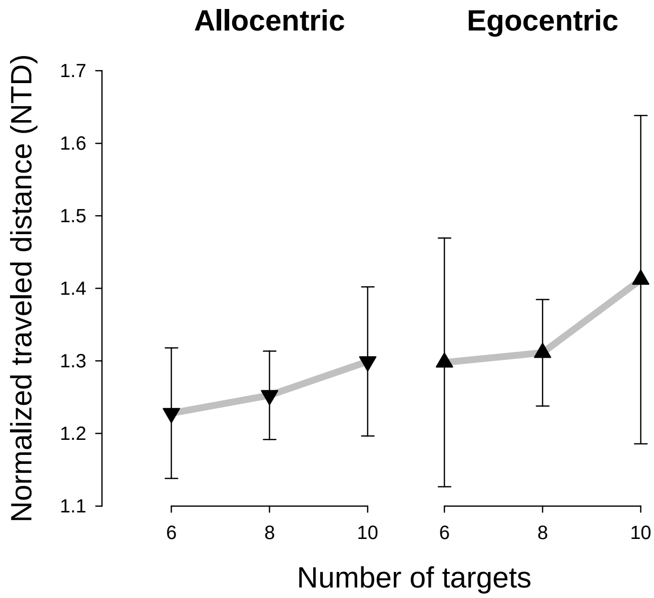


Figure 7 Normalized traveled distance (NTD), averaged for 8 subjects, with 6, 8 and 10 targets in both allocentric and egocentric conditions. There is a slight increase as a function of targets number. Traveled distance is also slightly increased in egocentric condition as compared to allocentric one (note also the increase of variability). Error bars represent standard deviation.

3 Discussion

3.1 Discrepancies with previous findings

We found that subjects were able to carry out the MTSP task correctly, visiting all targets in up to 70% of the trials, in both allocentric and egocentric conditions. This finding is surprising for several reasons. First, the result dispels our concern that the MTSP task with 8 and 10 targets might be beyond subjects' capacity. Our finding also contrasts with at least two kinds of results in the literature. For example, with Multiple Object Tracking (MOT) methodology (Pylyshyn & Storm, 1988), it was found that that people can track only about four or five target 'objects'.[19] Our finding suggests, therefore, a difference between tracking *n* independently-moving targets and keeping track of *n* static targets on the basis of their direction – as in the MTSP task. In addition, human short-term memory is known to have storage capacity limits, sometimes presented (controversially) as a "single, central capacity limit averaging about four chunks" (Cowan, 2000) and it has been suggested that visual working memory has a limit of about 4 items (Luck & Vogel, 1997). Our findings suggest a surprisingly higher limit: The number of 8 or 10 targets far exceeds the most common estimations of the limits of short-term memory, visual working memory or visual tracking. Secondly, our finding did not support our expectation that the global location-based strategy would be the most natural in the MSPT task (see section 2.1.4). Choice of this unique strategy would have led to poorer performance – the global location-based strategy requires a high memory load in order to retain locations of the 8-10 targets, as well as the current status predicates/tags associated with them.

3.2 The results explained by the use of different strategies

In order to account for the discrepancies between our findings and those of previous studies, we focus on the hypothesis that subjects used different strategies for performing the MTSP task (see Table 1) and that the use of segment strategies allowed succeed in the egocentric condition (with any number of target) and also in allocentric condition trials with large numbers of targets. Since direction-based strategies are deictic (see 2.1.2 and 2.1.3), this hypothesis emphasizes the use of deictic strategies as a crucial means for "overcoming" resource limitations (Ballard et al., 1997). If this hypothesis is true, then our findings provide, to the best of our knowledge, the first empirical evidence of the systematic use of a deictic strategy in a task that (only) at first glance has to be solved by location-based memory.

The global location-based strategy relies on memory of *all* target locations within a stationary allocentric frame of reference, and on the capacity to infer target locations on the basis of spatial cues. Hence, limitations of visuo-spatial working memory are its relevant constraints. Given the most common estimations of the limits of visuo-spatial working memory (Cowan, 2000; Logie, 1995; Luck & Vogel, 1997), it is unlikely that subjects could recall all target locations with 8 or 10 targets. This observation is consistent with explicit reports about

[19] This experimental paradigm was initially used by (Pylyshyn & Storm, 1988), and has now been developed in several experimental contexts (Scholl & Pylyshyn, 1999; Sears & Pylyshyn, 2000). Pylyshyn and colleagues hypothesized that the human visual system use visual indexes, or that humans could perceptually 'split' their selective attention (Scholl, 2001: 28) over multiple visual objects. In the paradigm presented by Pylyshyn & Storm (1988), participants visually track a specified subset of identical, randomly moving objects in a display. The members of the subset to be tracked (the targets) were identified by briefly flashing them, prior to the onset of movement. According to the FINST model (Pylyshyn, 1989), targets designated in this way are automatically indexed. During the tracking task, targets were indistinguishable from the nontarget distractors, which made the historical continuity of each target's motion the only available clue as to the targets' identity. Visual indexes provided a deictic link with targets which enabled them to be tracked. Participants tracked the target objects for 5 to 10 seconds, after which either a target or a distractor was indicated by superimposing a bright square over it. The participants' task was to determine whether this indicated object was a target or a distractor. The authors found that performance in this multiple object tracking task was particularly high for subsets of up to five elements: subjects could track simultaneously up to five target objects at an accuracy approaching 90%.

performing the MTSP task with 8 or 10 targets in allocentric and egocentric conditions, according to which it seems impossible to memorize all locations. However, one cannot exclude the possibility that subjects in the allocentric condition used the *gouping* location-based strategy in order to perform well with 8 or 10 targets.

The most interesting and surprising result remains however the performance on the egocentric condition, because location-based strategies are unlikely to be effective to explain the high performance in this condition. Several converging arguments support the view that only the *segment* strategies might explain performance in the egocentric condition. Firstly, visual inferences based on the *global location-based* strategy do not seem sufficient to explain performance for the egocentric condition, even with the smallest number of targets. The global location-based strategy is unlikely to be available in the egocentric condition since the location of targets in the screen frame of reference is constantly rotating and translating (sections 2.1.4 and 2.2). As a consequence, subjects cannot use the environmental stationary features (screen borders and other salient landmarks) as a frame of reference, because target locations are not fixed in relation to the screen and even move off the area of the screen. Finally, subjects in the egocentric condition reported about performing the MTSP task in the egocentric condition that it was impossible for them to figure out where the targets were, and described usually the choice of segment strategies.

While subjects in the egocentric condition could not have used the *global* location-based strategy, they may have used the *grouping* location-based strategy. For example subjects may have encoded the relative location of the closest targets as they moved through the space. In order to address this possibility, a small number of additional subjects were run on the experiment, with only 6 targets in the allocentric and egocentric conditions. They were instructed to visit each target once and only once, and to keep track of target locations. At the end of some trials, we debriefed these subjects concerning their memory for locations. If subjects used spatial memory in the egocentric condition, they might be able to recall, at least, the relative *distance* between each target and the vehicle, and therefore detect when targets moved off the screen borders (an event which happens frequently during most trials in the egocentric condition). We found, however, that *none* of the subjects noticed that this happened. In addition, every subject in the egocentric condition in this exploratory test reported that they were unable to recall where the targets had been – they had no idea where the targets were, except that they could see from the segments at the end of the trial which direction they were in. Thus, the most plausible explanation for how the task was carried out in the egocentric condition is that (and perhaps in the allocentric with 8 or 10 targets) subjects used direction-based deictic strategies (that is, visual tracking of directional segments). Given the limitation of visual working memory; the global location-based strategy might explain performance only for the smallest number of targets (n = 6) in allocentric condition.

3.3 The deictic direction-based strategies: applicability and cost

In contrast to the global location-based strategy, direction-based strategies may explain our findings. When using direction-based strategies, subjects do not compute target locations, but rather track local or proximal cues, i.e. subsets of directional segments. The primary focus of direction-based strategies is to track and focus on the directional segments, and so this strategy is less sensitive to visuo-spatial memory limitations. Nonetheless, the relevant constraint in this strategy is the capacity of visual tracking. As previously mentioned, human subjects can track up to four or five visual objects in parallel, and this constraint should apply when using segment strategies because directional segments in the MTSP task have similar characteristics to visual objects in classical (MOT) tracking experiments (i.e., they have identical features and the historical continuity of each segment's motion is the only available clue to represent or access its identity).

What differences between the MOT and MTSP paradigms may be the most relevant for explaining our findings with the MTSP task? First, part of the explanation could be that targets in MTSP do not follow arbitrary motion (in contrast to MOT), but change in accordance with the voluntary movement of the vehicle through a stationary configuration of targets in two dimensions. This coherence between the movements could facilitate tracking and reduce the errors current-status predicate ascription.

Second, in the MOT task subjects observe the motion of targets passively, whereas MTSP is an (active) subject-performed task in which the subject chooses movements in order to visit the next target. This might help memory performance since memory is improved in subject-performed tasks (Senkfor, Van Petten, & Kutas, 2002; Zimmer et al., 2001), even in the case of navigation in virtual environments (Brooks, Attree, Rose, Clifford, & Leadbetter, 1999). Also visual memory for object location benefits from active interactions with layout (Shelton & McNamara, 2001b).

Finally, in cases where the use of the global direction-based strategy does not circumvent memory limitations, grouping (or "chunking") subsets of segments may allow performance to exceed the visual memory limits – grouping or "chunking" strategies are often used to increase memory (Cowan, 2000: 87, 93), and to allow the enumeration of larger numbers of items (Klahr, 1973; Mandler & Shebo, 1982; Van Oeffelen & Vos, 1982). Moreover, effects of regularities in the point layout (esp. point clusters) have been found in performance in standard Traveling Salesman Problem (MacGregor et al., 1999; Ormerod & Chronicle, 1999). The addition of grouping strategies to segment tracking might allow subjects to make use of statistical clusters of segments to improve performance. Because the vehicle motion is under subjects' control it may be possible to switch between the chunks and the individual segments. For example, with 8 targets (x_1 to x_8) and 8 segments (s_1 to s_8), a subject might first group 4 adjacent segments on one side, or segments that happen to be close together into one chunk (set A, s_1 to s_4), and track the 4 others (set B, s_5 to s_8). Subjects might then switch to tracking the individuals in set A, treating set B (s_5 to s_8) as a group. By switching between groups and individual targets subjects might be able to use a time-sharing strategy. The operation of grouping allows subjects to draw inferences such as:

> "Given that all targets in this group g_1 have been visited targets; and that this segment belongs to g_1, therefore, s_1 and x_1 must have already been visited."

In this manner, the subject can recover the correct current-status predicate on any member of the group if he or she has correctly tracked the group and avoid swap between group members and other tracked targets. To avoid losing track of segments that cross over one another, subjects would have to switch between groups and individuals sufficiently quickly to anticipate that a pair of segments were about to cross and select those segments to track.

4 Concluding remarks

We have argued that models of deictic reference and indexical representations are a valuable alternative to constructivist models of perception. In the framework of research on deictic reference, we have presented an experimental study of the capacity to keep track of several targets in an informationally impoverished environment: the MTSP paradigm. We found that subjects were able to carry out the MTSP task with an accuracy approaching 70% of the trials, in both allocentric and egocentric conditions, with 6, 8 and 10 targets. Visual inferences based on global location-based strategy and grouping location-based strategy might explain performance for the smallest number of targets (n = 6), but does not explain performance for high number of targets (n = 8, 10) and for the egocentric condition. Visual inferences based on the direction-based deictic strategies might explain performance for high number of targets (n = 8, 10) in the large-*n allocentric condition* and in the *egocentric condition*, although it would require a chunking and time-sharing process that has not been modeled in detail. In sum, these results provide evidence for the systematic use of deictic strategies in a task that could only at first glance be solved via location-based memory.

Acknowledgements

We wish to thank the following persons and researchers for their considerable help with this paper: Roberto Casati, Ori Friedman, Jérôme Dokic, Jeslan Hopkins, Joëlle Proust, Mark Wexler, and two anonymous referees. This research was supported in part by "Action Incitative Cognitique" grant n°0693 from the French Ministry of Research.

5 References

Agre, P. E. (1997). *Computation and Human Experience*. Cambridge, MA: Cambridge University Press.

Agre, P. E., & Chapman, D. (1987). Pengi: An implementation of a theory of activity. *Proceedings of the American Association for Artificial Intelligence, 87*, 268-272.

Allwein, G., & Barwise, J. (Eds.). (1996). *Logical reasoning with diagrams*. New York: Oxford University Press.

Attneave, F., & Farrar, P. (1977). The visual world behind the head. *American Journal of Psychology, 90*(4), 549-563.

Attneave, F., & Pierce, C. R. (1978). Accuracy of extrapolating a pointer into perceived and imagined space. *American Journal of Psychology, 91*(3), 371-387.

Bach, K. (1987). *Though and Reference*. Oxford: Clarendon Press.

Baldwin, D. A. (1993). Infant contributions to the achievement of joint reference. In P. Bloom (Ed.), *Language Acquisition, Core Readings*. Cambridge, MA: MIT Press.

Ballard, D. H. (1991). Animate Vision. *Artificial Intelligence, 48*, 57-86.

Ballard, D. H. (1996). On the function of visual representation. In K. A. Akins (Ed.), *Problems in Perception. Proceedings, Simon Fraser Conference on Cognitive Science, February 1992*. Oxford: Oxford University Press.

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions: Biological Sciences, 337*(1281), 331-338.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*(4), 723-767.

Ballard, D. H., Hayhoe, M. M., Salgian, G., & Shinoda, H. (2000). Spatio-temporal organization of behavior. *Spatial Vision, 13*(2,3), 321-333.

Baron-Cohen, S. (1997). *Mindblindness*. Cambridge, MA: MIT Press.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature, 408*, 196-199.

Brooks, B. M., Attree, E. A., Rose, F. D., Clifford, B. R., & Leadbetter, A. G. (1999). The specificity of memory enhancement during interaction with a virtual environment. *Memory, 7*(1), 65-78.

Brooks, R. A. (1999). Intelligence without reason, *Cambrian Intelligence: The Early History of the New IA* (pp. 133-185). Cambridge, MA: MIT Press.

Bruner, J. S. (1983). *Child's Talk: Learning to Use the Language*. Oxford: Oxford University Press.

Burge, T. (1977). Belief *de re*. *Journal of Philosophy, 74*(6), 338-362.

Burgess, N., Jeffery, K. J., & O'Keefe, J. (Eds.). (1999). *The Hippocampal and Parietal Foundations of Spatial Cognition*. Oxford: Oxford University Press.

Butterworth, G., & Grover, L. (1990). Joint visual attention, manual pointing, and preverbal communication in human infancy. In M. Jeannerod (Ed.), *Attention and Performance XIII: Motor Representation and Control* (pp. 605-624). Hillsdale, NJ: Lawrence Erlbaum Associates.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: Clarendon Press.

Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch & J. L. Davis (Eds.), *Large Scale Neuronal Theories of the Brain* (pp. 23-60). Cambridge, MA: MIT Press.
Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences, 3*, 345-351.
Clark, A. (2000). *A Theory of Sentience*. Oxford: Clarendon Press.
Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.
Dennett, D. C. (1991). Multiple Drafts versus the Cartesian Theater, *Consciousness Explained* (pp. 101-138). Boston: Little, Brown and Company.
Dennett, D. C. (2001). Are we explaining consciousness yet? *Cognition, 79*, 221-237.
Desrochers, S., Morissette, P., & Ricard, M. (1995). Two perspectives on pointing in infancy. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development* (pp. 85-101). Hillsdale: Lawrence Erlbaum Associates.
Dokic, J. (1996). The dynamics of deictic thoughts. *Philosophical Studies, 2*(2), 179-204.
Dretske, F. (1981). The objects of perception, *Knowledge and the Flow of Information* (pp. 153-168). Cambridge, MA: MIT Press.
Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
Farrell, M. J., & Thomson, J. A. (1998). Automatic spatial updating during locomotion without vision. *The Quarterly Journal of Experimental Psychology, 51A*(3), 637-654.
Farrell, M. J., & Thomson, J. A. (1999). On-line updating of spatial information during locomotion without vision. *Journal of Motor Behavior, 31*(1), 39-53.
Findlay, J. M., & Gilchrist, I. D. (2001). Visual attention: The active vision perspective. In M. Jenkin & L. Harris (Eds.), *Vision and Attention* (pp. 83-103). New York: Springer-Verlag.
Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision: The Psychology of Looking and Seeing*. Oxford: Oxford University Press.
Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in adults and children. *Cognitive Science, 26*, 817-831.
Garnham, A. (1989). A unified theory of some spatial relational terms. *Cognition, 31*, 45-60.
Gattis, M., & Dupeyrat, C. (2000). Spatial Strategies in Reasoning. In W. Schaeken & G. De Vooght & A. Vandierendonck & G. D'Ydewalle (Eds.), *Deductive Reasoning and Strategies* (pp. 153-175). Mahwah: N.J.: LEA.
Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*, 516-522.
Graziano, M. S. A., Hu, X. T., & Gross, C. G. (1997). Coding the locations of objects in the dark. *Science, 277*, 239-241.
Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7*(1/2/3), 43-64.
Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research, 38*(1), 125-137.
Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*, 49-63.
Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498-504.
Hendriks-Jansen, H. (1996). *Catching Ourselves in the Act, Situated Activity, Interactive Emergence, Evolution, and Human Thought*. Cambridge, MA: MIT Press.
Hodgson, T. L., Bajwa, A., Owen, A. M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower of London task. *Journal of Cognitive Neuroscience, 12*(5), 894-907.
Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*, 575-577.

Humphreys, G. W. (1999). Neural representation of objects in space: A dual coding account. In G. W. Humphreys & J. Duncan & A. Treisman (Eds.), *Attention, Space and Action: Studies in Cognitive Neuroscience* (pp. 165-182). Oxford: Oxford University Press.
Hurford, J. R. (in press). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*.
Jeannerod, M. (1988). *The Neural and Behavioural Organization of Goal-directed Movements*. Oxford: Clarendon Press.
Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175-219.
Kaplan, D. (1989). Demonstratives. In J. Almog & J. Perry & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481-563). Oxford: Oxford University Press.
Karn, K. S., & Hayhoe, M. M. (2000). Memory representations guide targeting eye movements in a natural task. *Visual Cognition, 7*(6), 673-703.
Kirsh, D. (1995a). *Complementary strategies: why we use our hands when we think*. Paper presented at the Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ.
Kirsh, D. (1995b). The intelligent use of space. *Artificial Intelligence, 73*(1-2), 31-68.
Klahr, D. (1973). A production system for counting, subitizing and adding. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 527-546). New York: Academic Press.
Kosslyn, S. M. (1994). *Image and Brain, The Resolution of the Imagery Debate*. Cambridge, London: MIT Press.
Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions: Biological Sciences, 352*(1358).
Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559-3565.
Land, M. F., Mennie, N., & Rusted, J. (1999). The role of vision and eye movements in the control of activities of daily living. *Perception, 28*, 1311-1328.
Lespérance, Y., & Levesque, H. J. (1995). Indexical knowledge and robot action - a logical account. *Artificial Intelligence, 73*, 69-115.
Lewis, D. (1979). Attitudes *De Dicto* and *De Se*. *Philosophical Review, lxxxviii*, 513-543.
Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28*, 103-174.
Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom & M. A. Peterson & L. Nadel & M. F. Garrett (Eds.), *Language and Space* (pp. 493-529). Cambridge: MIT Press.
Logie, R. H. (1995). *Visuo-Spatial Working Memory*. Hove, Hillsdale: Lawrence Erlbaum Associates.
Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.
Lyons, J. (1977). Deixis, space and time, *Semantics, Volume 2* (pp. 636-724). Cambridge, MA: Cambridge University Press.
MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (1996). Human performance on the travelling salesman problem. *Perception & Psychophysics, 58*, 527-539.
MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (1999). Spatial and contextual factors in human performance on the travelling salesperson problem. *Perception, 28*, 1417-1427.
MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2000). A model of human performance on the travelling salesperson problem. *Memory & Cognition, 28*(7), 1183-1190.
Mandler, G., & Shebo, B. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*, 1-22.
Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.
McDowell, J. (1984). De Re Senses. *Philosophical Quarterly, 34*(136), 283-294.

McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In C. Fresksa & C. Brauer & C. Habel & S. Wender (Eds.), *Spatial Cognition III: Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Reasoning* (pp. 174-191). Berlin: Springer-Verlag.

Michie, D., Fleming, J. G., & Oldfield, J. V. (1968). A comparison of heuristic, interactive and unaided methods of solving a shortest-route problem. *Machine Intelligence, 2*, 245-255.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge, MA: Harvard University Press.

Milner, A. D. (1999). Neuropsychological studies of perception and visuomotor control. In G. W. Humphreys & J. Duncan & A. Treisman (Eds.), *Attention, Space and Action* (pp. 217-231). Oxford: Oxford University Press.

Milner, A. D., & Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.

Morris, R. G., Nunn, J. A., Abrahams, S., Feigenbaum, J. D., & Recce, M. (1999). The hippocampus and spatial memory in humans. In N. Burgess & K. J. Jeffery & J. O'Keefe (Eds.), *The Hippocampal and Parietal Foundations of Spatial Cognition* (pp. 259-289). Oxford: Oxford University Press.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*(3), 461-488.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(5), 939-1031.

Ormerod, T. C., & Chronicle, E. P. (1999). Global perceptual processing in problem solving: The case of the traveling salesperson. *Perception & Psychophysics, 61*(6), 1227-1238.

Palmer, S. E. (1999). *Vision Science, Photon to Phenomenology*. Cambridge, MA: MIT Press.

Perry, J. (2000). *The Problem of The Essential Indexical and Other Essays, Expanded Edition*. Stanford, CA: CSLI Publications.

Perry, J. (2001). *Reference and Reflexivity*. Stanford: CSLI Publications.

Posma, A., & De Haan, E. H. F. (1996). What was where? Memory for object locations. *The Quarterly Journal of Experimental Psychology, 49A*(1), 178-199.

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

Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences, 4*(5), 197-207.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127-158.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25*, 157-238.

Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independents targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179-197.

Recanati, F. (1993). *Direct Reference: From Language to Thought*. Oxford: Blackwell Publishers.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see : the need for attention to perceive change in scenes. *Psychological Science, 8*(5), 368-373.

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 15*(6), 1157-1165.

Russell, B. (1911). Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society, 11*, 108-128.

Russell, B. (1956). On the nature of acquaintance. In R. C. Marsh (Ed.), *Logic and Knowledge, Essays 1901-1950*. London: George Allen & Unwin.

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition, 80*, 1-46.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: clues to visual objecthood. *Cognitive Psychology, 38*(2), 259-290.

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processes. *Canadian Journal of Psychology, 54*, 1-14.

Segal, G. (1989). The return of the individual. *Mind, 98*, 39-57.

Senkfor, A., Van Petten, C., & Kutas, M. (2002). Episodic action memory for real objects: An ERP investigation with Perform, Watch, and Imagine Action encoding tasks versus a Non-Action encoding task. *Journal of Cognitive Neuroscience, 14*(3), 402-419.

Shelton, A. L., & McNamara, T. P. (2001a). Systems of spatial reference in human memory. *Cognitive Psychology, 43*, 274-310.

Shelton, A. L., & McNamara, T. P. (2001b). Visual memories from nonvisual experiences. *Psychological Science, 12*(4), 343-347.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences, 4*(4), 147-155.

Stenning, K. (2002). *Seeing Reason, Image and Language in Learning to Think*. Oxford: Oxford University Press.

Strawson, P. F. (1959). *Individuals, An Essay in Descriptive Metaphysics*. London: Methuen.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In T. Inui & J. L. McClelland (Eds.), *Attention and Performance XVI* (pp. 457-478). Cambridge, MA: MIT Press.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6*, 171-178.

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision, 3*(86-94).

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97-159.

Van Oeffelen, M., & Vos, P. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory and Cognition, 10*, 396-404.

Verblunsky, S. (1951). On the shortest path through a number of points. *Proceedings of the American Mathematical Society, 2*(6), 904-913.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24*, 295-340.

Zimmer, H. D., Cohen, R. L., Guynn, M. J., Engelkamp, J., Kormi-Nouri, R., & Foley, M. A. (2001). *Memory for Action: A Distinct Form of Episodic Memory?* Oxford: Oxford University Press.