Cognition 126 (2013) 181–197

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Object location memory: Integration and competition between multiple context objects but not between observers' body and context objects

Weimin Mou [*], Marcia L. Spetch

University of Alberta, Edmonton, Alberta, Canada T6G 2E9

ARTICLE INFO

Article history:
Received 3 January 2012
Revised 24 September 2012
Accepted 30 September 2012
Available online 9 November 2012

Keywords:
Spatial memory
Cue competition
Cue integration

ABSTRACT

Five experiments examined the integration and competition between body and context objects in locating an object. Participants briefly viewed a target object in a virtual environment and detected whether the target object was moved or not after a 10 s interval. Experiments 1 and 2 showed that performance when both the observer body and the context objects were consistent across study and test was not better than the optimal sum of performances when either one was the only consistent cue across study and test. In Experiments 3 and 4, in the competition conditions, both the body and the context objects were reference points at learning but only one stayed consistent during test. In the no competition conditions, only the body or the context objects were the primary reference points in learning and it stayed consistent in test. Detection performance did not differ between these conditions. Experiment 5 demonstrated the integration and competition between context objects as a reference point. Detection performance based on all four context objects was better than the optimal sum of the performance based on two close context objects and the performance based on two far context objects; detection performance based on two context objects was better when there were only these two context objects during learning than when there were four context objects during learning. These results suggest that body-object (body-target) and interobject (context-target) vectors are encoded independently and combined at test in an optimal way. Body-object and interobject vectors are not encoded in an integrated way and encoding of them does not compete. By contrast multiple interobject vectors are encoded in an integrated way in addition to the representations of individual interobject vectors and encoding close interobject vectors and encoding far interobject vectors interfere with each other.

1. Introduction

The memory of objects' locations is critical to our life. We need location memory to find our glasses when we wake up and find our way home after work. Location is specified with respect to a reference point (Klatzky, 1998; Levinson, 1996). For example, when we specify the location of a computer, we might say the computer is in front of us or north to the printer. In this example, the reference point is the body of the observer or the printer, another object.

Extensive research has been devoted to understanding the extent to which the observer's body and other objects are used as reference points in encoding the location of a target (e.g. Burgess, 2006; Burgess, Spiers, & Paleologou, 2004; Holmes & Sholl, 2005; Mou, McNamara, Rump, & Xiao, 2006; Mou, Xiao, & McNamara, 2008; Sargent, Dopkins, Philbeck, & Chichka, 2010; Sargent, Dopkins, Philbeck, & Modarres, 2008; Wang & Spelke, 2000, 2002; Xiao, Mou, & McNamara, 2009). Although the research has indicated that people might use both their body and other

---

[*] Corresponding author. Address: Department of Psychology, P217 Biological Sciences Bldg., University of Alberta, Edmonton, Alberta, Canada T6G 2E9.

E-mail address: wmou@ualberta.ca (W. Mou).

0010-0277/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2012.09.018

objects as a reference point to encode the target's location, it is not clear whether body-object spatial relations and interobject spatial relations are represented independently or dependently. The current project tackled this issue.

The question of whether different sources of spatial information, e.g. landmark and geometry, are encoded and represented independently or dependently has been the focus of substantial empirical and theoretical research within the domains of comparative cognition, developmental psychology, and neuroscience, mostly within the context of representing geometry or landmarks. These studies have indicated that different sources of spatial information sometimes compete with one another for control (i.e. show overshadowing or blocking effects; e.g. Cheng, 2008; Doeller & Burgess, 2008), and sometimes appear to be encoded independently of other information (i.e. geometry is often encoded independently of featural information; e.g. Cheng, 1986; Hermer & Spelke, 1996). In contrast to the extensive literature on the interaction between landmarks and geometry, the issue of whether body-object and interobject representations interact or are encoded independently has not been directly investigated.

In this project the relation between a target and a reference point is defined as a vector including the distance and direction (or bearing) between the target and the reference point (Klatzky, 1998). We refer to a representation using the reference point of an observer body as the body-object vector and a representation using the reference points of other objects as the interobject vector. The direction (or bearing) between the target and the reference point, whether the reference object is an observer body or another object, needs to be specified in terms of a reference direction (Klatzky, 1998; Levinson, 1996). The body axis defines a reference direction in the description that the computer is in front of us whereas the Earth's axis defines a reference direction in the description that the computer is north to the printer. In this study, we primarily focus on using reference points in organizing location memory whether the reference direction is the body axis or an axis independent of the observer (e.g. The Earth's axis).

Scene recognition, in particular position change detection, is an important paradigm to examine how objects' locations are represented in memory. In this paradigm, participants study a scene containing a target object and then in a subsequent test phase they are asked to indicate whether the target object has moved. Research using this paradigm has indicated that both body-object vector and interobject vectors might be represented and used in scene recognition.

It has been well documented that scene recognition is viewpoint dependent; that is, recognition performance is better when the test viewpoints are the same as the learning viewpoints than when the test viewpoints and the learning viewpoints differ (e.g. Diwadkar & McNamara, 1997; but see Friedman & Waller, 2008). This viewpoint dependency in scene recognition, however, does not distinguish between representations of interobject vectors and representations of body-object vectors because either type of representation could be viewpoint dependent. When the orientation of the test scene is not the same as the orientation of the study scene, equivalently the test viewpoint is not the same as the study viewpoint, people using a representation based on interobject bearings would need to align the orientation of the study scene in the scene memory with the orientation of the test scene and then check whether any interobject bearings in the test scene differ from those and in the scene memory. This alignment process leads to the viewpoint dependency pattern in scene recognition (Mou, Fan, McNamara, & Owen, 2008).

The use of body-object bearings would also result in viewpoint dependency when the viewpoint changes from study to test. In this case, in order to use the body-object bearings from the study viewpoint in the scene memory, people would need to calculate both the transformation information from their study position to their test position relative to the array of objects and the transformation information from their study orientation to their test orientation relative to the array of objects. If we assume that the difficulty in calculating the transformation information increases with the angular distance between the study viewpoint and the test viewpoint, then the viewpoint dependency results can be explained. Although both representations of interobject vectors and representations of body-object vectors can explain the viewpoint dependency result, using representations of body-object vectors requires not only the transformation of the viewing orientation but also the transformation of the viewing position whereas using representations of interobject vectors requires only the transformation of the viewing orientation. Hence if both interobject vectors and body-object vectors are represented, people might be more likely use the representations of interobject vectors in scene recognition when their study position and test position relative to the array are different. This conjecture will be used in designing experiments of the current project.

There is considerable evidence that both interobject vectors and body-object vectors can play a role in scene recognition (Burgess et al., 2004; Mou et al., 2008; Simons & Wang, 1998; Wang & Simons, 1999). Burgess et al. (2004) reported that scene recognition was better when the test scene was consistent with the orientation of a landmark, which indicated that there were interobject vector representations between the landmark and the array of objects. Burgess et al. also reported that scene recognition was better when the test scene was consistent with the representations of body-object vectors that were updated by participants' locomotion. Thus, Burgess et al. explicitly proposed that both interobject vectors and body-object vectors are represented and used in scene recognition.

This proposal is also supported by Mou et al. (2008). In Experiment 1 of their study, participants needed to judge whether a probed object had been moved with or without the movement of other objects (*context* objects). The results showed that position change detection of the probed object was more accurate when the context objects stayed stationary than when the context objects moved whether participants stayed at the study position or physically moved to a novel viewpoint. This finding indicated that the location of the probed object was represented with respect to the locations of the context objects. The results of Mou et al. (2008) also showed that position change detection was above chance even when the context objects changed and could not be used as a cue to locate the probed object; this detection occurred whether participants stayed at the study position or physically moved to a novel viewpoint. This implies that body-object vectors were used in scene recognition when there was no transformation between the learning position and test position and also when the transformation between the learning position and test position could be determined by the spatial updating process during locomotion.

In sum growing evidence indicates that both interobject and body-object vectors are represented in memory and used in scene recognition. However, to our knowledge no studies have directly examining how interobject and body-object vectors interact during memory organization and scene recognition. This project addresses this question. We hypothesize that interobject and body-object vectors are represented independently. We propose this hypothesis for two reasons. First in everyday life, when an observer moves in the environment, interobject vectors are usually stable within the environment whereas body-object vectors change. It might therefore be difficult to organize interobject vectors and body-object vectors in a single representation. Second the representation of interobject vectors might be a part of a cognitive map system that is the function of hippocampus whereas body-object vectors might be a part of a motion–action system that is the function of parietal lobe (Burgess, 2008; Byrne, Becker, & Burgess, 2007). Two testable predictions are derived from the hypothesis that interobject and body-object vectors are represented independently. Our first prediction is that performance based on both representations will be no better than the sum of the performances based on the individual representations. Our second prediction is that the two representations will not compete for common cognitive resources during encoding, and therefore neither source of information will overshadow the other.

Prediction 1: Body-object vectors and interobject vectors will not show super-additivity. If no integrated representation is formed in addition to the individual representations of interobject and body-object vectors, then the performance in localizing the target object when both interobject and body-object vectors are present and consistent across study and test should not be better than the sum of the performance based on either representation. This prediction can be written in formula 1. O refers to interobject vectors between the target and the other objects. B refers to body-object vector between the target object and the observer's body. Conditions are labeled to indicate which cues are available or informative during study and test, following the format of "study cues/test cues". Thus, OB/OB refers to a condition in which both the body-object vector and interobject vectors are expected to be encoded at learning and used at testing O/O refers to a condition in which only interobject vectors are expected to be encoded at learning and used at testing, and B/B refers to a condition in which only the body-object vector is expected to be encoded at learning and used at testing.

$$\mathrm{Performance}(\mathrm{OB/OB}) \leqslant \mathrm{Performance}(\mathrm{O/O}) + \mathrm{Performance}(\mathrm{B/B}) \qquad (1)$$

Alternatively, if interobject and body-object vectors are represented in an integrative way in addition to each individual representation, then performance in localizing the target object at test when both interobject and body-object vectors are consistent across study and test will be superior to the sum of the performances based on either representation. This super-additivity predication is written in the following formula:

$$\mathrm{Performance}(\mathrm{OB/OB}) > \mathrm{Performance}(\mathrm{O/O}) + \mathrm{Performance}(\mathrm{B/B}) \qquad (2)$$

Cheng, Shettleworth, Huttenlocher, and Rieser (2007) reviewed examples of using multiple individual cues in target localization. One famous example is from a task in which participants reproduced a dot location after viewing the dot within a circle. Participants showed a bias in responding toward the center of the quadrant in which the dot was presented, suggesting that they formed independent representations of categorical and metric information and summed them in determining the location of a dot (Huttenlocher, Hedges, & Duncan, 1991). It is hypothesized that people sum two independent cues in determining the location of the target in order to get a smaller variance of error in judgment of the location of the target. The minimum variance of error in judgment of the location of the target is obtained when participants optimally weight individual cues.

Our approach to testing whether the interobject vectors and body-object vectors are represented in an integrated way in addition to individual representations, is to determine whether the variance of error in locating targets when both interobject and body-object vectors are consistent across study and test ($\sigma^2_{\mathrm{OB/OB}}$) is smaller than the minimum $\sigma^2_{\mathrm{O/O+B/B}}$ ($\mathrm{MIN}\sigma^2_{\mathrm{O/O+B/B}}$). The minimum $\sigma^2_{\mathrm{O/O+B/B}}$ can be estimated according to formula 3. In formula 3, $\sigma^2_{\mathrm{O/O}}$ is the variance of error in locating targets when only interobject vectors are consistent across study and test; $\sigma^2_{\mathrm{B/B}}$ is the variance of error in locating targets when only the body-object vector is consistent across study and test.

$$\mathrm{MIN}\sigma^2_{\mathrm{O/O+B/B}} = \sigma^2_{\mathrm{B/B}}\sigma^2_{\mathrm{O/O}} \div (\sigma^2_{\mathrm{B/B}} + \sigma^2_{\mathrm{O/O}}) \qquad (3)$$

(see Appendix A for the proof)

Hence formula (1), which indicates that there is no integrative representation in addition to the individual representations of interobject and body-object vectors, is equivalent to

$$\sigma^2_{\mathrm{OB/OB}} \geqslant \sigma^2_{\mathrm{B/B}}\sigma^2_{\mathrm{O/O}} \div (\sigma^2_{\mathrm{B/B}} + \sigma^2_{\mathrm{O/O}}) \qquad (4)$$

Formula (2), which indicates that there is an integrative representation in addition to the individual representations of interobject and body-object vectors, is equivalent to

$$\sigma^2_{\mathrm{OB/OB}} < \sigma^2_{\mathrm{B/B}}\sigma^2_{\mathrm{O/O}} \div (\sigma^2_{\mathrm{B/B}} + \sigma^2_{\mathrm{O/O}}) \qquad (5)$$

The first purpose of this project was to distinguish between formula (1) and formula (2), using the equivalent formula (4) and formula (5).

Note that formula (5) should not be interpreted that at test people could combine two independent cues in a way better than the optimal (Bayesian) way, which usually indicates some flaws in experiment designs.[1] Instead, formula (5) describes a possibility that during encoding people may form a new integrative representation using both cues in addition to the representation using either cue. When both cues are reliable across learning and testing, three representations (one integrated representation and two individual representations using either cue) might still be combined in the Bayesian way. Hence super-additivity indicates that there is an additional representation formed during encoding rather than that people can outperform an ideal observer at test. Super-additivity is not new in experimental psychology. Gestalt phenomena are super-additive. For example, the phenomenon of apparent motion indicates that the whole is more than the sum of its parts (Wertheimer, 1912). Another example, which is related to location memory in the current project, is that when people perceive four dots forming a square, they might perceive the square as a whole in addition to individual pairs of dots. When only two of the four dots are presented, they only perceive the pair of the dots. Hence the square is the new representation in addition to the individual pairs of dots. The representation of the square, in addition to the representation of the individual pairs of dots, can be used to retrieve the location of an individual dot, producing super-additivity.

Formula (4) indicates that there is no integrated representation in addition to the individual representations of interobject and body-object vectors. There are two possibilities in using the individual representations at test. First people might combine the two individual representations in the optimal (Bayesian) way. In this case the left term equals the right term in the formula. Second people might only use the single most informative representations or they might combine them in non-optimal way. In this case the left term is larger than the right term in the formula indicating sub-additivity. How people combine two independent cues is still an open question (Cheng et al., 2007). For example, Nardini, Jones, Bedford, and Braddick (2008) reported that healthy human adults (see Fetsch, Turner, DeAngelis, & Angelaki, 2009 for the similar result for animals) did combine cues optimally whereas children failed to do so, showing sub-additivity.

More specifically the first purpose of this project was to test whether the performance in judgments of target location using body and other objects as cues follows super-additivity (formula (5)) or instead follows the Bayesian sum, or sub-additivity (formula (4)). Super-additivity indicates an additional integrated representation whereas Bayesian sum and sub-additivity do not indicate an additional integrated representation.

Prediction 2: Body-object vectors and interobject vectors will not compete with each other. If interobject and body-object vectors are represented independently, then the two representations might not compete for the common cognitive resource during encoding. Specifically, encoding interobject vectors is not influenced by the encoding of body-object vectors, and vice versa. This prediction is illustrated in Formulas (6) and (7). Performance (OB/B) refers to the performance when both interobject and body-object vectors are represented during study but only the body-object vector is consistent across study and test. Performance (B/B) refers to the performance when only body-object vectors are represented and constant across study and test. Performance (OB/O) refers to the performance when both interobject and body-object vectors are represented during study but only the interobject vectors are consistent across study and test. Performance (O/O) refers to the performance when only interobject vectors are represented and constant across study and test.

$$\mathrm{Performance(OB/B)} = \mathrm{Performance(B/B)} \tag{6}$$

$$\mathrm{Performance(OB/O)} = \mathrm{Performance(O/O)} \tag{7}$$

Alternatively, if interobject and body-object vectors are represented in an integrative way, then the two representations might compete for a common cognitive resource during encoding. In particular, encoding either interobject vectors or body-object vectors might be easier when encoding the other vectors is not required. This prediction is illustrated in formulas (8) and (9).

$$\mathrm{Performance(OB/B)} < \mathrm{Performance(B/B)} \tag{8}$$

$$\mathrm{Performance(OB/O)} < \mathrm{Performance(O/O)} \tag{9}$$

Cue competition in localization has been reported in animal and human spatial learning (Cheng, 2008; Doeller & Burgess, 2008). In Doeller and Burgess (2008), for example, human participants learned the locations of four objects with the presentation of a landmark alone or with the presentation of both the landmark and a boundary. They then located the objects with the presentation of the landmark alone. Participants who learned objects with only the presentation of the landmark were more accurate in locating the objects than those who learned objects with the presentation of both landmark and boundary. These results indicated that encoding the locations of the objects with respect to the boundary interfered with encoding the locations of the objects with respect to the landmark. The second purpose of the current project was to investigate whether interobject encodings and body-object encodings compete or interfere with each other by distinguishing between formula (6) and formula (8) and between formula (7) and formula (9).

Because we predict a null effect for both super-additivity and cue competition between interobject vectors and body-object vectors to support the independency between them in representation, our argument would be strengthened if we could demonstrate super-additivity and cue competition among multiple interobject vectors which we assume are represented in an integrative way. Multiple interobject vectors in an array of objects might be represented in an integrated way because people might represent the configuration of the array by perceptual grouping in addition to encoding individual interobject vectors as in the example of perceiving four dots forming a square discussed above. Therefore, the third purpose of

---

[1] We are grateful to an anonymous reviewer for the suggestion to add this discussion.

this project was to demonstrate super-additivity and cue competition between interobject vectors.

Experiments 1 and 2 tested whether interobject vectors and body-object vectors are represented in an integrative way in addition to each individual representation of interobject vectors and of body-object vectors by distinguishing between formula (4) and formula (5). Experiment 3 tested whether encoding interobject vectors interferes with encoding body-object vectors by distinguishing between formula (6) and formula (8). Experiment 4 tested whether encoding body-object vectors interferes with encoding interobject vectors by distinguishing between formula (7) and formula (9). Experiment 5 was designed to demonstrate super-additivity and cue competition between interobject vectors.

All experiments were implemented in a virtual environment to facilitate manipulation of the objects and spatial relations. The findings of this project will help us to further theorize how location memories are organized and used in scene recognition.

2. Experiment 1

The purpose of this experiment was to distinguish between formula (4) and formula (5) and to test whether interobject vectors and body-object vectors are represented in an integrated way in addition to each individual representation. In this experiment, participants saw an object probed by a red arrow on a circular grey background on the floor. After a short delay, participants standing at the same study position indicated whether this object (target) that was again probed by a red arrow had been moved (see Fig. 1). In one third of the trials (see OB/OB in Fig. 1), the target was presented with four context objects that stayed stationary for both study and test phases. As a result, the test array was consistent with representations of both the body-object vector and interobject vectors. Hence both the participants' body and the context objects at test could be used as a reference point to calculate the *correct* location of the target using the representations of body-object vector and interobject vectors. These trials were labeled as OB/OB, to indicate that both cues were consistent across learning and testing. The labels of the conditions are summarized in Table 1 for reference.

In another one third of the trials (B/B in Fig. 1), no context objects were presented in the study and test phases. In this way only the body-object vector was represented. Because participants' position remained the same across study and test and the study orientation and test orientation of the scene also stay the same, the test array was consistent with the representations of body-object vector. Hence participants' body could be used as a reference point to calculate the *correct* location of the target using the representations of body-object vector.

In the other one third of the trials (O/O in Fig. 1), the target was presented with four context objects and the tested position of the array was the same as in the other conditions. During the entire viewing time at study, the whole array was continuously translated on the floor from the to-be-tested position to a position one meter farther from the participant and then back to the to-be-tested position along the viewing direction. Otherwise the trials were identical to the trials of OB/OB. If we assume that interobject vectors were defined with respect to a reference direction parallel to the viewing direction, then a stable representation of interobject vectors could be formed with respect to the viewing direction even when the array was continuously translated along the viewing direction. Because the viewing direction did not change across study and test, interobject vectors were consistent across study and test in terms of the viewing direction. Hence context objects could be used as a stable reference point

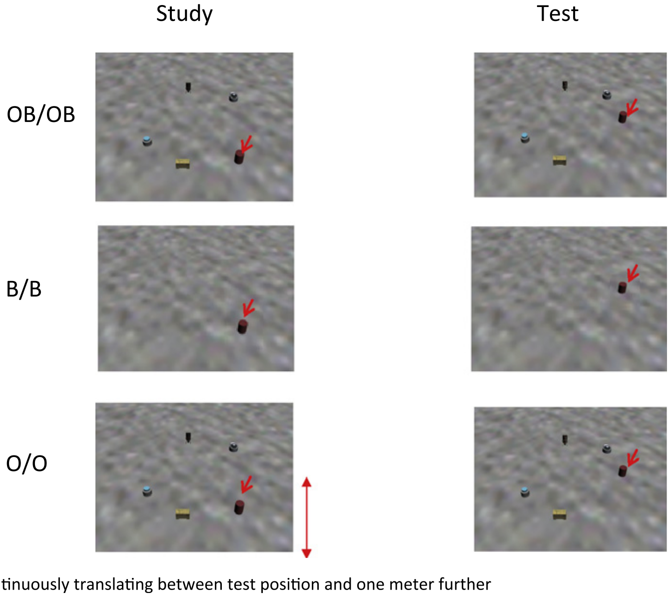


Fig. 1. The experiment conditions of Experiment 1.

Table 1
The labels of conditions used in experiments.

Condition	Available/informative vectors in study		Useful vectors in test	
	Body-object	Interobject	Body-object	Interobject
Experiments 1–5				
OB/OB	Yes	Yes	Yes	Yes
O/O	No	Yes	no	Yes
B/B	Yes	No	yes	No
BO/B	Yes	Yes	yes	No
OB/O_rotate	Yes	Yes	no	Yes
O/O_rotate	No	Yes	no	Yes
Experiment 5				
Interference				
Close_O (i.e. O/Close_O)	No	All	No	Close 2
Far_O (i.e. O/Far_O)	No	All	No	Far 2
No Interference				
Close_O (i.e. Close_O/O)	No	Close 2	No	Close 2
Far_O (i.e. Far_O/O)	No	Far 2	No	Far 2

to calculate the *correct* location of the target using the represented interobject vectors. In contrast, body-object vectors changed continuously at study discouraging participants from forming a reliable representation of body-object vector. We acknowledge that we cannot claim that there are no body-object vectors encoded even with this manipulation. It is safer to stipulate that participants may not develop a reliable representation of body-object vector and participants' body should less likely be used as a reference point to calculate the *correct* location of the target.

If the interobject vectors and the body-object vectors were represented independently and no additional integrated representation was formed, then detection of the movement of the target might not be better in the OB/OB trials than the sum of the performance in the B/B trials and the performance in the O/O trials, which was estimated according to the Bayesian principles (Cheng et al., 2007), as illustrated in formula (4). If the interobject and body-object vectors were represented in an integrated way in addition to the individual representations of the interobject vectors and of the body-object vector, then accuracy in detecting that the target had moved might be higher in the OB/OB trials than the sum of accuracy in the B/B trials and accuracy in the O/O trials as illustrated in formula (5).

2.1. Method

2.1.1. Participants

Thirty-six university students (18 men and 18 women) participated in this experiment as partial fulfillment of a requirement in an introductory psychology course.

2.1.2. Materials and Design

The experiment was conducted in a room of 4 by 4 m. The virtual environment with layouts and a circular grey background was displayed in the center of the real room in stereo with an nVisor SX60 head-mounted display (HMD, NVIS, Inc., Virginia). Participants' head motion was tracked with an InterSense IS-900 motion tracking system (InterSense, Inc., Massachusetts). The circular background with a diameter of 50 m was displayed on the floor of the real room to indicate the floor of the virtual environment. For each layout, the objects (wood, lock, battery, bottle, and candle, each about 5 cm in size) were placed on one or five of nine possible positions in an irregular array on the circular background. The distance between any two of the nine positions varied from 18 to 29 cm. The irregularity of the array ensured that no more than two objects were aligned with the observer throughout the experiment. There was a real chair, on which participants were seated, 1.5 m to the center of the room. A mouse was mounted on a real barstool, which was on the right side of the chair, and was used to submit participants' responses. Participants wore a blindfold before they donned the HMD. Hence all irrelevant environmental cues from the real room were removed. The earphone of the HMD was used to present instructions (e.g. to learn the layout, to close the eyes).

Sixty irregular configurations of five locations were created by selecting five of the nine possible locations. In each configuration, one of the five locations was selected randomly to be the location of the target object. On those trials when the target moved, the object was placed on one of the other four locations that were not selected to form the configuration. This new location of the object was usually the location closest to the original location.

The configurations were randomly divided into six groups of ten. Six sets of 60 trials were created by assigning the six groups of configurations to six combinations of target movement (moved or stationary) and cue conditions (OB/OB, B/B or O/O) with a Latin Square design. Consequently, in each set of trials, different sets of configurations were assigned to different combinations of target movement and cue conditions, and across the six sets of trials each set of configurations was assigned to every combination of target movement and cue condition once. Three men and three women were randomly assigned to one of the six sets of trials randomly so that across participants, any possible differences among the configurations were counter-balanced.

In each trial, participants saw an object with a red arrow pointing to it (Fig. 1) for 3 s and then made a judgment after a 10 s interval. In the OB/OB trials, the target was presented with four context objects at both study and test. In the B/B trials, the target was presented on its own at both study and test. In the O/O trials, the target was presented with four context objects during study and the whole array was continuously translating at 0.667 m/s (or 2 m for 3 s) from the test position to one meter farther from the participant and then was continuously translating back to the test position along the learning view. The target was presented with the context objects and the array was stationary at test.

The primary independent variable was the cue condition (OB/OB, B/B, or O/O) which was manipulated within participants. The dependent variable was the percentage of the correct judgments in deciding whether or not the target object changed positions. *d* Prime was also used to estimate the variance of error in localization of the target.

2.1.3. Procedure

Wearing a blindfold, participants walked into the testing room and sat on the viewing chair assisted by the experimenter. Participants donned the HMD and then removed the blindfold without peeking at the real room. Participants were instructed to look at the floor that was indicated by the grey circular background. Each trial was initiated by a key press of the experimenter and started with a verbal instruction via earphone ("please learn the location of the object."). After 3 s, the array disappeared and participants were instructed to stand up, close their eyes, and step in place ("please stand up, close your eyes, take four steps in place and sit down"). This was designed to prevent them from looking at the fixed point on the HMD corresponding to the target position during the interval even though the array of objects disappeared and the text of "close your eyes" appeared instead. Ten seconds after participants were instructed to stop viewing the layout, the test scene was presented and participants were instructed to determine whether the object designated by the red arrow had changed its position ("please make judgment of the location of the object"). The participant was instructed to respond as accurately as possible by clicking the mouse (left button for change and right button for no change); speedy response was discouraged ("please respond as accurately as possible, respond only after you are sure this is your decision"). After the response, the trial was ended. The 60 experimental trials were presented in a random order.

Before the 60 experimental trials, six extra trials (one in each of the six combinations of cue conditions and target movement) were used as practice to make sure that participants got used to the procedure.

2.2. Results and discussion

Mean percentage of correct judgment as a function of cue condition is plotted in Fig. 2. Percentage of correct judgments was computed for each participant and each cue condition, and analyzed in repeated measure analyses of variance (ANOVAs), with variables corresponding to cue condition.

The main effect of cue condition was significant, $F(2,70) = 27.66$, $p < .001$, $MSE = .01$, $\eta_p^2 = .44$. Planned comparisons showed that accuracy was significantly higher for the OB/OB trials than O/O trials and B/B trials, respectively, $ts(70) \geqslant 2.61$. Accuracy was also significantly higher for O/O trials than for B/B trials, $t(70) = 4.72$.

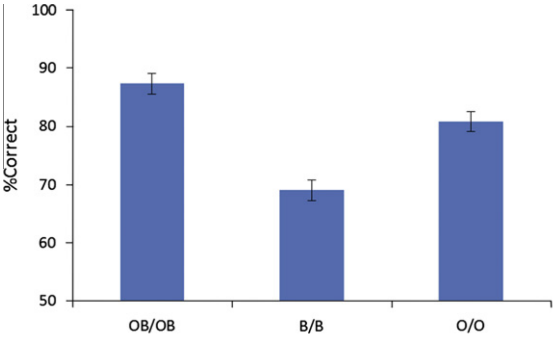


Fig. 2. Correct percentage in detecting position change as the function of cue condition in Experiment 1. (*Error bars are ±1 standard error of the mean.*)

More important, we also tested whether the interobject vectors and body-object vectors were represented independently and then combined in an optimal or sub-optimal way in locating the target's location, or instead whether the interobject vectors and body-object vectors were represented in an integrated way in addition to the individual representations. This was tested by distinguishing between formula (4) and formula (5).

We did not directly measure the variance of error ($\sigma^2_{\mathrm{B/B}}, \sigma^2_{\mathrm{O/O}}, \sigma^2_{\mathrm{OB/OB}}$) in locating targets in each condition. However, *d* prime has an inverse relation with the standard deviation (σ) according to:

$$d' = (\mathrm{AL_{new}} - \mathrm{AL_{old}})/\sigma \tag{10}$$

(see Appendix B for the proof).

$\mathrm{AL_{old}}$ refers to the actual original location of the target and $\mathrm{AL_{new}}$ refers to the actual new location to which the target was potentially moved; σ is the standard deviation of the error in remembering a location with respect to available cues.

Hence we tested whether *d* prime in the OB/OB condition ($d'_{\mathrm{OB/OB}}$) was comparable to or larger than the maximum *d* prime (MAX $d'_{\mathrm{B/B+O/O}}$) in using the optimal combination of the representations of body-object and interobject vectors according to:

$$\mathrm{MAX}\ d'_{\mathrm{B/B+O/O}} = \sqrt{d'^2_{\mathrm{B/B}} + d'^2_{\mathrm{O/O}}} \tag{11}$$

(see Appendix C for the proof)

Formula (4), which indicated that interobject vectors and body-object vectors are represented independently, is equivalent to the following:

$$d'_{\mathrm{OB/OB}} \leqslant \sqrt{d'^2_{\mathrm{B/B}} + d'^2_{\mathrm{O/O}}} \tag{12}$$

Formula (5), which indicated that interobject vectors and body-object vectors are represented in an integrative way in addition to the individual representations, is equivalent to the following:

$$d'_{\mathrm{OB/OB}} > \sqrt{d'^2_{\mathrm{B/B}} + d'^2_{\mathrm{O/O}}} \tag{13}$$

The mean *d* prime across participants in the OB/OB condition and the mean estimated maximum *d* prime across participants are plotted in Fig. 3. For comparison, the *d* primes in the O/O and B/B conditions are also plotted in Fig. 3. Paired t test showed that mean $d'_{\text{OB/OB}}$ (2.36) was not significantly different from the mean MAX $d'_{\text{B/B+O/O}}(2.40), t(35) = 0.26,\ p > .05$.

These results indicate that the performance in OB/OB condition is the Bayesian sum of the performance in O/O and the performance in B/B. There was no evidence that the interobject vectors and body-object vectors were integrated in a representation in addition to the individual representations of the interobject vector and of the body-object vector.

In the O/O condition, we tried to remove the influence of the representation of body-object vectors by translating the object array during learning. However, as the translating started and ended at the position of the test array, participants may still briefly represent body-object vectors that could be used during test. In Experiment 2, in the O/O condition, the object array translated continuously between the position 0.7 m farther than the position of the test array relative to the participant and the position 1.7 m farther than the position of the test array relative to participant such that there was no position overlap before the learning array and the test array. Hence even if there was a brief representation of the body-target vector during translation, it might not be readily used at test.

3. Experiment 2

3.1. Method

3.1.1. Participants

Twenty-four university students (12 men and 12 women) participated in this experiment as partial fulfillment of a requirement in an introductory psychology course. Two men and two women were randomly assigned to each of the six groups of trials in which the six groups of configurations were assigned to the six combinations of target movement (moved or stationary) and cue conditions (OB/OB, B/B or O/O) with a Latin Square design.

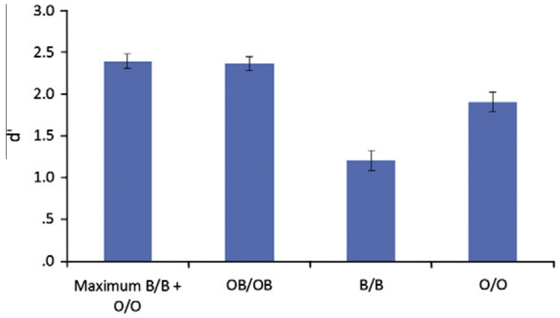


Fig. 3. *d* Prime as the function of cue condition in Experiment 1. (*Error bars are* ±1 *standard error of the mean.*) Maximum B/B + O/O corresponds to the estimation when optimally combining cues in the condition of O/O and in the condition of B/B for each participant, according to MAX $d'_{1+2} = \sqrt{d_1'^2 + d_2'^2}$.

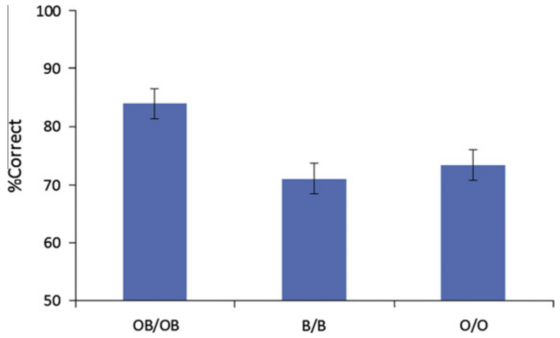


Fig. 4. Correct percentage in detecting position change as the function of cue condition in Experiment 2. (*Error bars are* ±1 *standard error of the mean.*)

3.1.2. Materials, design, and procedure

The materials, design, and procedure were the same as in Experiment 1 except that in the O/O condition, the learning object array started translation at the position 0.7 m farther than the test position relative to the participant, moved 1 m farther (i.e. 1.7 m from the testing position) and then moved back to the starting position. Hence there was no overlap in locations between the learning object array and the test object array.

3.2. Results and discussion

Mean percentage of correct judgment as a function of cue condition is plotted in Fig. 4. Percentage of correct judgments was computed for each participant and each cue condition, and analyzed in repeated measure analyses of variance (ANOVAs), with variables corresponding to cue condition.

The main effect of cue condition was significant, $F(2,48) = 7.08$, $p < .01$, $MSE = .016$, $\eta_p^2 = .24$. Planned comparisons showed that accuracy was significantly higher for the OB/OB trials than for the trials of O/O and B/B respectively, $ts(48) \geqslant 2.90$. The accuracy was not different for the trials of O/O and for the trials of B/B, $t(48) = .63$.

The mean *d* prime across participants in the OB/OB condition and the mean estimated maximum *d* prime across participants are plotted in Fig. 5. For comparison, the *d* primes in the O/O and B/B conditions are also plotted in Fig. 5. Paired t test showed that mean $d'_{\text{OB/OB}}$ (2.13) was not significantly different from the mean MAX $d'_{\text{B/B+O/O}}$ (2.21), $t(24) = 0.47$, $p > .05$.

These results indicate that the performance in OB/OB condition is the Bayesian sum of the performance in O/O and the performance in B/B. As in Experiment 1, there was no evidence that the interobject vectors and body-object vectors were integrated in a representation in addition to the individual representations of the interobject vector and of the body-object vector.

In Experiments 1 and 2, we used the null effect of the group comparison between $d_{(\text{OB/OB})}$ and the estimated $d_{(\text{maximum B/B + O/O})}$ to conclude participants combined the cues of the body and the other objects in the Bayesian way. To strengthen this conclusion, we plot individual data in Experiments 1–2 in Fig. 6 to examine the relations between the observed $d_{(\text{OB/OB})}$ and the estimated

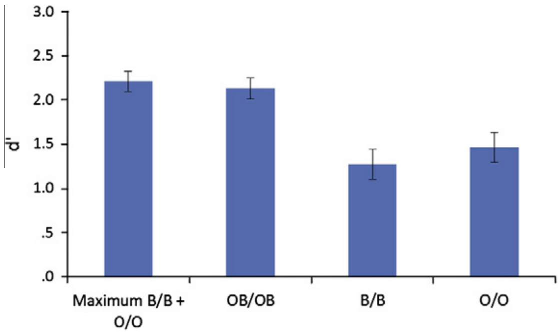


Fig. 5. *d* Prime as the function of cue condition in Experiment 2. (*Error bars are* ±1 *standard error of the mean*.) Maximum B/B + O/O corresponds to the estimation when optimally combining cues in the condition of O/O and in the condition of B/B for each participant, according to MAX $d'_{1+2} = \sqrt{d_1'^2 + d_2'^2}$.

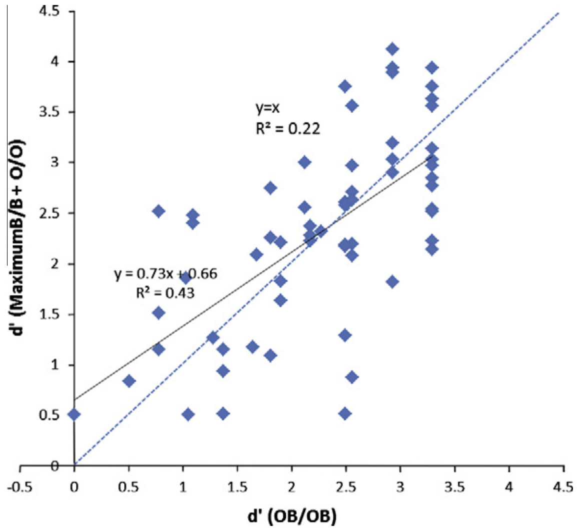


Fig. 6. The scatterplot shows the correlation between the observed d'(OB/OB) and the estimated d'(maximum B/B + O/O) in Experiments 1–2.

$d_{(\text{maximum B/B + O/O})}$.[2] The linear regression model ($R^2 = .43$) was significant, $F(1,58) = 44.02$, $p < .001$. Furthermore, roughly equal numbers of data points are located below (27 of 60) and above (33 of 60) the line of $y = x$, $\chi^2(1) = .60$, $p > .05$. In addition, $R^2 = .22$ when we fit the observed data with the line of $y = x$. These findings suggest that the estimated $d_{(\text{maximum B/B+O/O})}$ fits the observed $d_{(\text{OB/OB})}$ very well.

4. Experiment 3

The purpose of Experiment 3 was to investigate whether performance in scene recognition based on body-object vectors was better when only body-object vectors were represented than when both body-object vectors and interobject vectors were represented, distinguishing between formula (6) and formula (8). Formula (8) would support that encoding interobject vectors interferes with encoding body-object vectors whereas formula (6) would support that encoding interobject vectors does not interfere with encoding body-object vectors.

This experiment was identical to Experiment 1 except that the O/O trials in Experiment 1 were replaced with OB/B trials (see Fig. 7 and Table 1). In the OB/B trials, the target was presented with four context objects and the array was stationary so that both body and context objects were reference points in encoding the target's location. Only the target, however, was presented at test. Hence only the body is the valid cue in locating the target at test. If encoding interobject vectors interferes with encoding body-object vectors, then detection of position change should be better in the B condition, in which only encoding body-object vectors occurred, than in the OB/B condition, in which both encoding body-object and encoding interobject vectors occurred. The OB/OB trials were included to ensure that participants encoded both interobject and body-object vectors in the OB/B trials as participants could not determine whether the context objects would be presented at test when they saw the context objects at study.

4.1. Method

4.1.1. Participants

Twenty-four university students (12 men and 12 women) participated in this experiment as partial fulfillment of a requirement in an introductory psychology course. Two men and two women were randomly assigned to each of the six groups of trials in which the six groups of configurations were assigned to the six combinations of target movement and cue conditions with a Latin Square design.

4.1.2. Materials, design, and procedure

The materials, design, and procedure were the same as in Experiment 1 except that the trials of OB/B replaced the trials of O/O. The order of the sixty trials was randomized.

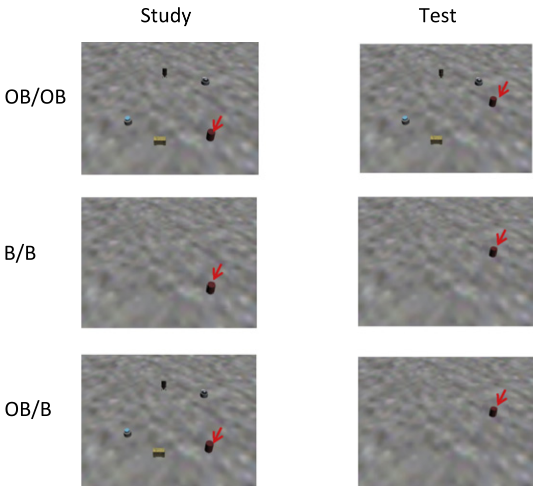


Fig. 7. The experiment conditions of Experiment 3.

---

[2] We are grateful to an anonymous reviewer for the suggestion to add this analysis.

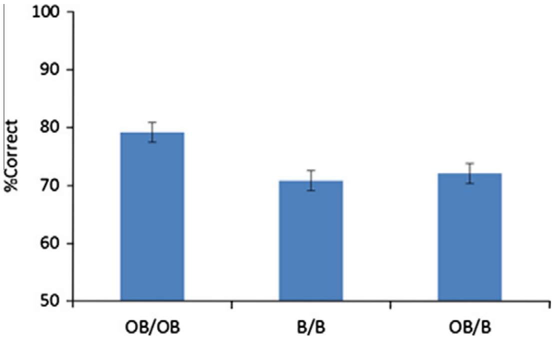


Fig. 8. Correct percentage in detecting position change as the function of cue condition in Experiment 3. (*Error bars are* ±1 *standard error of the mean.*).

4.2. Results and discussion

Mean percentage of correct judgment as a function of cue condition is plotted in Fig. 8. Percentage of correct judgments was computed for each participant and each cue condition, and analyzed in repeated measure analyses of variance (ANOVAs), with variables corresponding to cue condition.

The main effect of cue condition was significant, $F(2,46) = 7.08$, $p < .001$, $MSE = .007$, $\eta_p^2 = .24$. Planned comparisons showed that accuracy was significantly higher for the OB/OB trials than for the trials of B/B and OB/B respectively, $ts(46) \geqslant 2.97$. The difference between the latter two was not significant, $t(46) = 0.52$. Furthermore, accuracy for the OB/B trials was numerically close and even slightly higher than for the B/B trials (Fig. 8). Hence there is no evidence that accuracy for the B/B trials was better than that for the OB/B trials.

The similar accuracy on B/B and OB/B trials clearly indicates that encoding interobject vectors did not interfere with encoding body-object vectors.

5. Experiment 4

The purpose of Experiment 4 was to investigate whether performance in scene recognition based on interobject vectors was better when interobject vectors were represented but body-object vectors encoding was discouraged or impaired than when both body-object vectors and interobject vectors were represented, distinguishing between formula (7) and formula (9). Formula (9) would support that encoding body-object vectors interferes with encoding interobject vectors whereas formula (7) would support that encoding body-object vectors does not interfere with encoding interobject vectors.

The OB/OB trials in the previous experiments were used in this experiment. There are two new conditions of trials, OB/O_rotate and O/O_rotate (see Fig. 9 and Table 1). In both OB/O_rotate and O/O_rotate trials, there were context objects in both study and test phases. The test views were rotated 30° to the right of the study views so that context objects were the primary reliable cues at test because using interobject vectors only depended on the learning orientation but using body-object vectors depended on not only the learning orientation but also the learning location as discussed in the Introduction. We used O_rotate instead of O to indicate that the context objects were reliable cues but the test view was novel and different from that in Experiment 1. In the OB/O_rotate trials, the array was stationary at study so that contexts objects and body were both reference points at study. In the O/O_rotate trials, the array was translated during learning as in Experiment 1 discouraging participants from encoding reliable body-object vectors. Hence, encoding reliable body-object vectors were discouraged or impaired and context objects were the primary reference points at study. Although encoding body-object vectors might not be completely eliminated, a reduction in body-object encoding would still result in less interference with the interobject encoding if there is interference between these two encodings. Thus if encoding body-object vectors interferes with encoding interobject vectors, then detection of position change should be better in the O/O_rotate condition, in which encoding body-object vectors was reduced or impaired and encoding interobject vectors occurred primarily, than in the OB/O_rotate condition, in which encoding both interobject and body-object vectors occurred equivalently. The OB/OB trials were included to ensure that participants encoded both the interobject vectors and the body-object vector in the OB/O_rotate trials as participants could not predict whether the test view would be the same or different at test when they saw the stationary array at study.

5.1. Method

5.1.1. Participants

Twenty-four university students (12 men and 12 women) participated in this experiment as partial fulfillment of a requirement in an introductory psychology course. Two men and two women were randomly assigned to each of the six groups of trials in which the six groups of configurations were assigned to the six combinations of target movement and cue conditions with a Latin Square design.

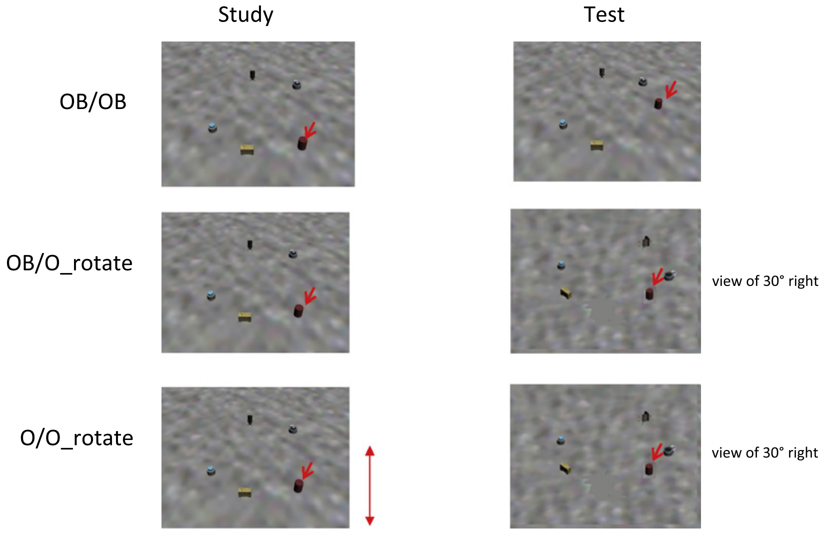
5.1.2. Materials, design, and procedure

The materials, design, and procedure were the same as in Experiment 1 except that the trials of OB/O_rotate and O/O_rotate replaced the trials of B/B and O/O.

5.2. Results and discussion

Mean percentage of correct judgment as a function of cue condition is plotted in Fig. 10. Percentage of correct judgments was computed for each participant and each cue condition, and analyzed in repeated measure analyses of variance (ANOVAs), with variables corresponding to cue condition.

The main effect of cue condition was significant, $F(2,46) = 15.16$, $p < .001$, $MSE = .01$, $\eta_p^2 = .40$. Planned comparisons showed that accuracy was significant higher for the OB/OB trials than for the trials of OB/O_rotate and O/O_rotate respectively, $ts(46) \geqslant 4.36$. The difference between the latter two was not significant, $t(46) = 0.73$. Furthermore, accuracy for the trials of OB/O_rotate and



Continuously translating between test position and one meter further

Fig. 9. The experiment conditions of Experiment 4.

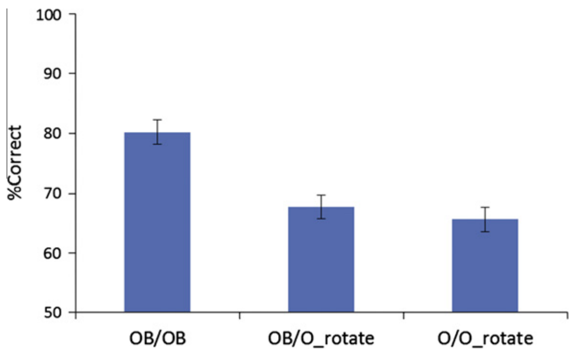


Fig. 10. Correct percentage in detecting position change as the function of cue condition in Experiment 4. (*Error bars are* ±1 *standard error of the mean.*)

O/O_rotate was numerically close and the former was even slightly higher than the latter (Fig. 10). Hence there is no evidence that accuracy for the trials of O/O_rotate was better than that for the trials of OB/O_rotate.

These results clearly indicate that encoding body-object vectors did not interfere with encoding interobject vectors.

6. Experiment 5

The purpose of this experiment was to demonstrate super-additivity and cue competition between interobject vectors. We assume that people represent the configuration of the object array in addition to individual interobject vectors. There are two predictions. First performance in detecting the target's position change using all context objects is better than the sum of the performances using either of two sub-sets of context objects. Second encodings of interobject vectors should interfere with each other.

In this experiment, half of the participants (see Fig. 11 and Table 1, interference condition) learned the target location with the presentations of four context objects. In the test, the target was presented with the presentations of all four context objects (O/O), only the two nearer objects (O/close_O) or only the two farther objects (O/far_O). According to the prediction of super-additive, the performance in the O/O condition should be better than the sum of the performance in the O/close_O condition and the performance in the O/far_O condition. According to the prediction of interference, encoding the vectors between the target and the two farther objects and encoding the vectors between the target and two nearer objects should interfered with each other. Hence we refer to this group of participants as the interference group. The other half of the participants (see Fig. 11 and Table 1, no-interference condition) also had the same O/O condition as in the interference group. Different from the O/close_O condition and O/far_O condition in the interference group, the participants in the no-interference group only learned the target location with the presentation of the two nearer objects in the close_O/O condition or with the presentation of the two farther objects in the far_O/O condition. Hence there should be no interference between encoding the vectors between the target and the two farther objects and encoding the vectors between the target and two nearer objects. We expected that the performance in far_O and close_O trials in the no-interference group should be better than the interference group.

In order to discourage or eliminate the encoding of body-object vectors, in all trials, the study array translated from 0.7 m, to 1.7 m, back to 0.7 m farther than the position of the test array relative to the participant as in the O condition of Experiment 2. Also, to ensure that both groups experienced a change in object number between study and test, the missing objects on the far_O/O and close_O/O trials were added to the testing display for the no-interference group. Because these objects were not present during study, they could not provide any usable information, but they resulted in a context change. As illustrated

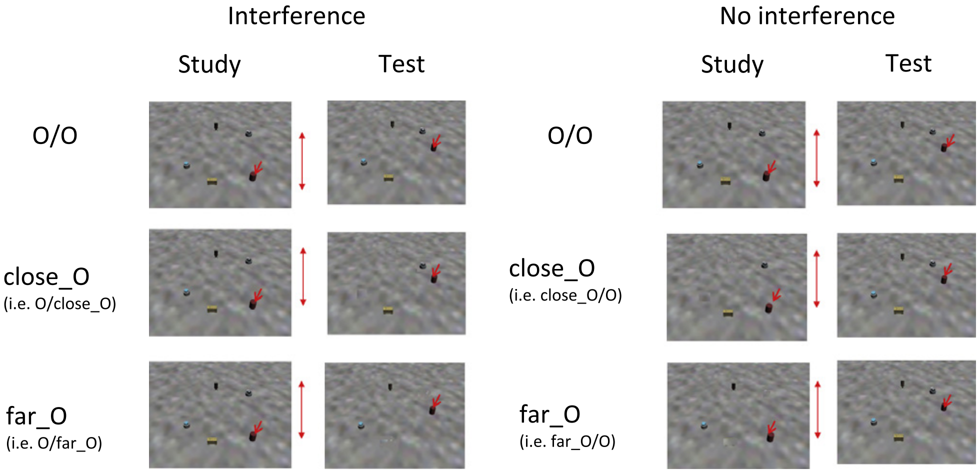


Fig. 11. The experiment conditions of Experiment 5.

in Fig. 11, for the no-interference group, the number of objects in the far_O/O and the close_O/O trials changed from three to five across study and test whereas for the interference group, the number of objects the O/far_O and the O/close_O trials changed from five to three across study and test. Hence both groups experienced a change in context from study to test. Therefore, superior performance in the no-interference group than in the interference group is unlikely to reflect the simple context change between study and test.

6.1. Method

6.1.1. Participants

Forty-eight university students (24 men and 24 women) participated in this experiment as partial fulfillment of a requirement in an introductory psychology course. Twenty-four (12 men and 12 women) was assigned to the interference group and the no-interference group respectively. For each group, two men and two women were randomly assigned to each of the six groups of trials in which the six groups of configurations were assigned to the six combinations of target movement and cue conditions with a Latin Square design.

6.1.2. Materials, design, and procedure

The materials, design, and procedure were the same as in Experiment 2 except that (1) there were interference and no-interference groups; (2) In each group, there were close_O (O/close_O for interference, close_O/O for no interference) and far_O (O/far_O for interference, far_O/O for no interference) trials replacing B/B and OB/OB trials.

6.2. Results and discussion

Mean percentage of correct judgment as a function of cue condition (within participants) and interference group (between participants) is plotted in Fig. 12. Percentage of correct judgments was computed for each participant and each cue condition, and analyzed in mixed-model analyses of variance (ANOVAs), with variables corresponding to cue condition and interference.

The main effect of cue condition was significant, $F(2,92) = 28.72$, $p < .001$, $MSE = .01$, $\eta_p^2 = .38$. The main effect of interference was significant, $F(1,46) = 8.39$, $p < .01$, $MSE = .048$, $\eta_p^2 = .15$. The interaction between cue condition and interference was significant, $F(2,92) = 6.33$, $p < .01$, $\eta_p^2 = .12$. The interaction was due to the different effects of interference between the O/O trials and the other trials (close_O and far_O), $ts(92) \geqslant 2.86$. The effect of interference did not differ between the close_O trials and the far_O trials, $t(92) = .40$.

There was no difference between the interference and no-interference groups for O/O trials, $F(1,46) = .22$, $p > .05$, $MSE = .024$, $\eta_p^2 = .005$, which is expected because the O/O trials were identical for the two groups. The simple effect of interference was significant for both the close_O trials, $F(1,46) = 10.06$, $p < .01$, $MSE = .023$, $\eta_p^2 = .18$, and the far_O trials, $F(1,46) = 10.06$, $p < .01$, $MSE = .023$, $\eta_p^2 = .18$. These results might indicate that encoding the target in terms of the close context objects and encoding the target in terms of the far context objects interfered with each other.

In the no-interference group, accuracy was not significantly different between the O/O trials and the close_O/O

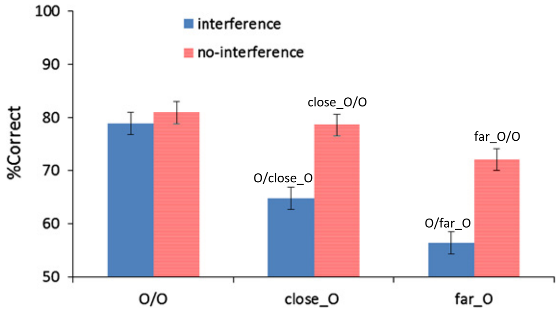


Fig. 12. Correct percentage in detecting position change as the function of cue condition in Experiment 5. (*Error bars are* ±1 *standard error of the mean.*)

trials, $t(92) = .78$. Accuracy for the O/O trials and the close_O/O trials was significantly higher than the far_O/O trials, $ts(92) \geqslant 2.27$. In the interference group, accuracy was significantly higher for the O/O condition than for the trials in the other two conditions, $ts(92) \geqslant 4.82$. Accuracy was significantly higher for the O/close_O trials than for the O/far_O trials $t(92) = 2.83$.

In addition to the inference account above, one may argue that there was an equally plausible explanation of the results. The superiority of the no-interference condition to the interference condition for the close_O trials might be due to the larger consistency of the *gestalt* across study and test in the no-interference condition compared to the interference condition. In particular, the *gestalt* present at encoding in the close_O/O is preserved at test, relative to the O/close_O condition. In close_O/O, the triangle formed by the close objects and the target object at study is still there at test. In O/close_O, the pentagram formed by the array at study is gone at test. This interpretation also makes sense in light of the fact that in the no-interference group, performance in the O/O and close_O/O conditions were the same and far_O/O was worse. In the O/O and close_O/O conditions the gestalt shapes present at encoding were there at test. But in far_O/O, the addition of the close objects at test breaks up the triangle formed by far and target objects, and thus disturbs the gestalt.[3] However, this explanation might not readily explain the same superiority of the no-interference condition for the far_O trials. According to this explanation, performance should be comparable in the far_O/O conditions and in the O/far_O condition as the *gestalt* was disrupted in test for both conditions. Hence we are more in favor of the interference account: encoding the target in terms of the close context objects and encoding the target in terms of the far context objects interfered with each other.

To demonstrate the super-additivity between interobjects vectors, we tested whether *d* prime in the O/O condition ($d'_{O/O}$) was larger than the maximum *d* prime in using the optimal combination of the interobject vectors involving the close context objects (close_O), and of the interobject vectors involving the far context objects (far_O).

As discussed above, encoding the target in terms of the close context objects and encoding the target in terms of the far context objects interfered with each other when all context objects were presented during learning. Note that all context objects were presented in the O/O trials even for the no-interference group (see Fig. 11). Encoding of the target in terms of the close context objects should be interfered with by encoding of the target in terms of the far context objects in O/O for the no-interference group. In contrast, no far context objects were presented during learning in close_O/O for the no-interference group so there was no interference from the far context object. Hence for the no-interference group, we can surmise that the representations of the vectors between the target and the close context objects were weaker in the O/O trials than in the close_O/O trials. Hence the contribution of the close context objects to detection of the correct location of the target in the O/O trials should be smaller than the contribution of the close context objects in the close_O/O trials. For the same reason, the contribution of the far context objects in the O trials should be smaller than the contribution of the far context objects in the far_O trials. Consequently we could not test super-additivity in the no-interference group but the individual representations would not be equally encoded across conditions. By contrast, in the interference group, the representations of the vectors between the close context objects and the target and between the far context objects and the target should be the same in the O/close_O and O/far_O conditions as in the O/O condition because participants saw all five objects in all trials and could not anticipate the test conditions during study. Hence we only used the interference group to test super-additivity.

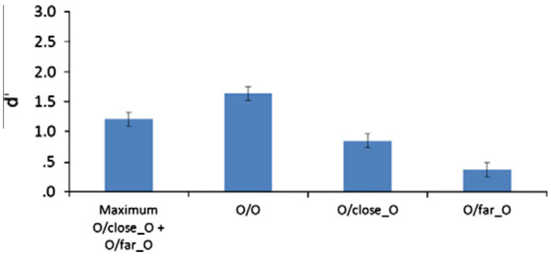


Fig. 13. *d* Prime as the function of cue condition in Experiment 5. (*Error bars are* ±1 *standard error of the mean.*) Maximum O/close_O + O/far_O corresponds to the estimation when optimally combining cues in the condition of O/close_O and the condition of O/far_O for each participant, according to MAX $d'_{1+2} = \sqrt{d'_1 + d'_2}$.

The mean *d* prime across participants in the O/O condition of the interference group and the mean estimated maximum *d* prime across participants in the interference group are plotted in Fig. 13. For comparison, the *d* primes in the O/close_O and O/far_O conditions of the interference group are also plotted in Fig. 13. Paired t test showed that mean $d'_{O/O}$ (1.63) was significantly higher than the mean MAX $d'_{O/close\_O+O/far\_O}$ (1.21), $t(24) = 7.04$, $p < .05$. Hence when participants saw four context objects and one target object during study, they formed some integrative representation in addition to the individual vectors between the context objects and the target object. Participants might group the five objects into a shape in addition to the individual vectors between the context objects and the target object.

The individual data of the observed $d_{(O/O)}$ and the estimated $d_{(\text{maximum O/close\_O+O/far\_O})}$ are plotted in Fig. 14. R^2 of the linear regression model was smaller than that in Experiments 1–2 (.19 vs. .43), but the linear regression was significant, $F(1,22) = 5.30$, $p < .05$. Furthermore, a majority of the data points (17 of 24) are located below the line of $y = x$, $\chi^2(1) = 4.17$, $p < .05$. These findings also suggest that overall the estimated $d_{(\text{maximum O/close\_O + O/far\_O})}$ was smaller than the observed $d_{(O/O)}$ consistent with the results of the t test. In addition, when we fit the observed data with the line of $y = x$, $R^2 = .01$, which is negligible and smaller than that estimated in Experiments 1–2 (.01 vs. .22). Hence the linear relations between the estimated

---

[3] We are grateful to a reviewer for the suggestion of discussing this explanation.

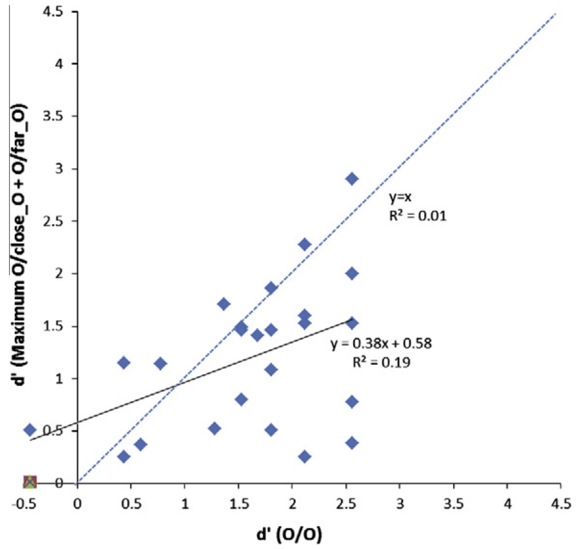


Fig. 14. The scatterplot shows the correlation between the observed d'(O/O) and the estimated d'(maximum O/close_O + O/far_O) in Experiment 5.

$d_{(\text{maximum O/close\_O+O/far\_O})}$ and observed $d_{(\text{O/O})}$ did not fit the line of $y = x$.

7. General discussion

The aims of this project were to investigate how interobject vectors and body-object vectors interact and how different interobject vectors interact in location memory organization and in scene recognition. Four important findings were obtained. First performance in position change detection based on context objects and participants' body can be perfectly estimated by the optimal sum of the performance based on only context objects or only the body according to the Bayesian principles (Cheng et al., 2007). Second performance based on only body or context objects was comparable whether participants encoded the target location relative to both cues or only relative to the valid cue. Third performance based on all four context objects was better than the optimal sum of the performance based on the two close context objects and the performance based on the two far context objects. Fourth performance based on the close or far context objects was better when no other context objects were learned than when other context object were learned.

These findings indicate that both interobject and body-object vectors are represented in spatial memory of a briefly viewed scene, consistent with the proposal by Burgess et al. (2004). The representation of the body-object vectors was evidenced by the superior performance when both interobject and body-object vectors were consistent than when only interobject vectors were consistent (Experiments 1 and 2). The representation of the interobject vectors was evidenced by the superior performance when both interobject and body-object vectors were consistent than when only body-object vectors were consistent (Experiments 1 and 2). Furthermore, the better performance when the close context objects were the consistent cue than when the far context objects were the consistent cue also strongly suggested that interobject vectors were represented, at least between the close context objects and the target.

The evidence of coexistence of interobject and body-object spatial representations challenges the theoretical position that people primarily use their own body to represent an object's location (e.g. Wang & Spelke, 2000, 2002). Wang and Spelke stipulated that people represent objects' locations with respect to their own body and that there are enduring spatial representations of geometric shapes, e.g. spatial representations among corners of a room, but not among objects. Clearly Experiments 1, 2 and 5 showed that there were interobject vectors in memory. The evidence of coexistence of interobject and body-object spatial representations equally challenges the theoretical position that people primarily represent interobject spatial relations (e.g. Mou, McNamara, Valiquette, & Rump, 2004). Mou et al. stipulated that there are both body-object and interobject spatial representations. However, they conjectured that body-object spatial representations might decay very quickly without perceptual support (Milner & Goodale, 1995). In Experiments 1 and 2, the interval, 10 s, was substantial and participants were asked to stand up, close their eyes, and step in place to remove the perceptual support. Hence the finding that body-object vectors facilitated localization of the target object indicated that there were enduring representation of body-object spatial relations (Ball, Smith, Ellison, & Schenk, 2010). Hence the findings of Experiments 1, 2 and Experiment 5, together with those in the previous studies (e.g. Burgess et al., 2004; Mou et al., 2008), might suggest that it is time to end the debate on whether there are interobject or body-object vector representations in memory (Burgess, 2006).

Therefore we think that a more important or appropriate research agenda on how body and other objects are used as reference objects in location memory organization is to study the way in which interobject vectors and body-object vectors interact in memory organization and scene recognition. The findings of the current study provided some preliminary understandings of this interaction. The first finding (Experiments 1 and 2) suggested that participants represented interobject vectors and body-object vectors independently and there was no integrated representation in addition to these individual representations. If there was an integrated representation, performance based on both body and context objects should be better than the optimal sum of the performance based on only context objects and the performance based on only the body. The first finding also indicated that participants combined the cue of the body and the cue of the other objects in an optimal way in scene recognition. Cue combination has been reported in the literature of spatial cognition (e.g. Huttenlocher et al., 1991; see Cheng et al., 2007 for a review). The finding of this study provided evidence that interobject and body-object vectors were also combined to increase the accuracy in locating an object's location and the combination is optimal. Cheng et al. (2007) used the term *integration* to refer to weighting independent cues in localization judgment. In this study, we distinguish between integration and combination. We use *integration* to refer to an integrated representation that is in addition to independent representations. We use cue *combination* to refer to weighting independent cues in judgment.

The second finding of this study showed that performance based on only body or context objects was comparable whether participants encoded the target location relative to both cues or only relative to the valid cue at test. In Experiment 3, in the OB/B condition, the interobject vectors should be encoded in addition to the body-object vector as Experiment 1 suggested. The interobject representations could not facilitate detection of the movement of the target object at test as all context objects were removed. In the B/B condition, participants could only encode the body-object vector as there were no context objects. The results showed comparable performance in the OB/B condition and in the B/B condition. In Experiment 4, in the OB/O_rotate condition, the body-object vector should be encoded in addition to the interobject vectors as Experiment 1 suggested. As discussed in the introduction, using representations of body-object vectors requires not only the transformation of the viewing orientation but also the transformation of the viewing position whereas using representations of interobject vectors requires only the transformation of the viewing orientation. Hence compared to the interobject vectors, the body-object vectors were less likely used in locating the target at test because the test view was different from the study view. In the O/O_rotate condition, participants should primarily encode interobject spatial relations because the array was translated along the viewing direction at study to discourage the encoding of body-object vectors. If encoding of targets in terms of body interfered with encoding of targets in terms of other objects, then this interference effect should be weaker in O/O_rotate than in the OB/O_rotate because of the reduced encoding of targets in terms of body in condition O/O_rotate. However, the results showed comparable performance in the OB/O_rotate condition and in the O/O_rotate condition.

The results of Experiments 3 and 4 strongly indicated that there was no competition between encoding the interobject vectors and encoding the body-object vector. If there was competition, encoding either vector should be easier when encoding of the other vector was eliminated or reduced than when the other vector was equivalently encoded. Because better encoding should lead to better retrieval, retrieval of either vector should be easier when the other vector was not encoded or encoded weakly than when the other vector was equivalently encoded. These predictions were clearly inconsistent with the results of Experiments 3 and 4. The lack of competition, or interference, between encoding the body-object vector and encoding the interobject vectors also indicated that encoding the body-object vector and encoding the interobject vectors are independent. Otherwise they would compete for the common resource.

In addition, this finding indicated that neither the observer body nor the set of context objects was more fundamental than the other in encoding an object's location (Cheng, 2008; Doeller & Burgess, 2008). Otherwise, the more fundamental cue might overshadow the other cue but not the other way around. Doeller and Burgess reported that a boundary overshadowed a landmark in encoding the location of a target but the landmark did not overshadow the boundary. The finding of no competition or overshadowing between context objects and observer body again undermines the debate about whether interobject or body-object vectors are primarily represented in spatial memory (Burgess, 2006).

The absence of both super-additivity and the competition between interobject vectors and body-object vectors supported the conclusion that they are represented independently. This conclusion is consistent with the proposal that interobject vectors might be a part of a cognitive map system that is based in the hippocampus and body-object vectors might be a part of a motion-action system that is based on the parietal lobe (Burgess, 2008).

The lack of an integrated representation between the body-object vector and the interobject vectors contrasts with the way in which multiple interobject vectors are represented. The third finding of the current study (Experiment 5) showed that participants performed better based on all four context objects than the optimal sum of the performance based on two close context objects and the performance based on two far context objects.

We speculated that participants mentally represented a shape formed by connecting the interobject vectors among adjacent objects together in addition to the individual interobject vectors. This shape-like representation provided some additional cue in detecting the location of the target when all context objects were presented. This shape-like representation might not be accessible or might be less accessible when two of the context objects were removed during test.

The lack of competition between interobject and body-object spatial vector also does not generalize to the encoding of multiple interobject vectors. The fourth finding of the current study (Experiment 5) showed that performance using close context objects or far context objects was reduced by the presence of the other context objects at study. Specifically, participants in the interference group, who had all objects present at study, performed worse with the close objects or the far objects alone at test than did participants in the non-interference group, who had only the near or far objects present in study. The addition of objects in test for the non-interference group helped to rule out that context change per se was the reason for the reduced performance. This finding indicated that there was competition between using the close and far context objects as a reference point of the target. The cue competition among multiple context objects might be attributed to a limited capacity of visual working memory (Cowan, 2001; Marois & Ivanoff, 2005). This speculation was also consistent with the evidence for an integrated representation of the four context objects in Experiment 5. When the number of vectors exceeds the capacity of the visual working memory, one efficient strategy may be to organize the context objects as a single item (e.g. a shape) in addition to the inaccurate individual interobject vectors.

Finally, we should note that in this project, we used the Bayesian principle introduced by Cheng et al. (2007) to estimate the sum of performance using different cues. This approach is strongly supported by the nice fit between the observed $d'_{(OB/OB)}$ and the estimated in Experiments 1–2 (see Fig. 6) We, however, have no intention to reject other mathematical or computation models of visual working memory (e.g. Johnson, Spencer, Luck, & Schöner, 2009). We do not think the data of the current project can distinguish between these mathematical models.

In conclusion, this project provided a systematic investigation on integration and competition of representing interobject and body-object vectors and of representing multiple interobject vectors. The results showed that there was neither integration nor competition of encoding interobject and body-object vectors. Instead interobject and body-object vectors were represented independently and combined in an optimal way in locating the target at test. By contrast there was integration and competition of encoding different interobject vectors.

Acknowledgments

Preparation of this paper and the research reported in it were supported in part by two separate grants from the Natural Sciences and Engineering Research Council of Canada to WM and to MLS. We are grateful to Laura Carlson, Timothy McNamara for comments on the interpretation of the data. We are also grateful to Taylor Murphy, Lauren Poon, Nicole Sieusahai, Sarah Tan for collecting data.

Appendix A

$$\text{Proofof MIN}\sigma^2_{\text{O/O+B/B}} = \sigma^2_{\text{B/B}} \times \sigma^2_{\text{O/O}} \div (\sigma^2_{\text{B/B}} + \sigma^2_{\text{O/O}}) \qquad (3)$$

According to Cheng et al. (2007), the optimal weights that lead to the minimum variance are:

$$W_{\text{O/O}} = \sigma^2_{\text{B/B}} \div (\sigma^2_{B/B} + \sigma^2_{\text{O/O}}), W_{\text{B/B}} = \sigma^2_{\text{O/O}} \div (\sigma^2_{\text{B/B}} + \sigma^2_{\text{O/O}})$$

$$\begin{aligned}\text{Hence MIN}\sigma^2_{\text{O/O+B/B}} &= W^2_{\text{O/O}} \times \sigma^2_{\text{O/O}} + W^2_{\text{B/B}} \times \sigma^2_{\text{B/B}} \\ &= \sigma^2_{\text{B/B}} \times \sigma^2_{\text{O/O}} \div (\sigma^2_{\text{B/B}} + \sigma^2_{\text{O/O}}).\end{aligned}$$

Appendix B

$$\text{Proofof } d' = (\text{AL}_{\text{new}} - \text{AL}_{\text{old}})/\sigma \qquad (10)$$

The represented location (RL) of an object is a random variable distributed normally with mean the actual location (AL) of the object, and variance. We write

$$\text{RL} \sim N(\text{AL}, \text{variance})$$

In the old–new recognition test, the represented location at study follows $\text{RL}_s \sim N(\text{AL}_s, \sigma^2)$, σ is the standard deviation of the error in remembering a location with respect to available cues.

The represented location at test is the actual location at test, AL_T, because the actual test position was always visible at test.

In the old–new recognition test, the judgment is based on difference (DIFF) between the represented location at study (RL_S) and the represented location at test (RL_T). We write

$$\text{DIFF} = \text{RLT} - \text{RLS}.$$

$$\text{Hence DIFF} \sim N(\text{AL}_T - \text{AL}_s, \sigma^2)$$

In the old–new recognition test, the half of the trials are the same trials with a pair of old–old locations across study and test and the other half of the trials are the different trials with a pair of old–new locations across study and test.

Hence for the same trials, $\text{DIFF}_{\text{same}} \sim N(\text{AL}_{\text{old}} - \text{AL}_{\text{old}}, \sigma^2)$ or $\text{DIFF}_{\text{same}} \sim N(0, \sigma^2)$ and for the different trials, $\text{DIFF}_{\text{different}} \sim N(\text{AL}_{\text{new}} - \text{AL}_{\text{old}}, \sigma^2)$.

d' measures the sensitivity of discrimination between the same and different trials or between $\text{DIFF}_{\text{same}}$ and $\text{DIFF}_{\text{different}}$.

d' is defined as the distance of the means of $\text{DIFF}_{\text{same}}$ and $\text{DIFF}_{\text{different}}$ normalized by the standard deviation of them. Hence

$$d' = (\text{AL}_{\text{new}} - \text{AL}_{\text{old}})/\sigma$$

Appendix C

$$\text{Proofof MAX } d'_{\text{O/O+B/B}} = \sqrt{d'^2_{\text{B/B}} + d'^2_{\text{O/O}}} \qquad (11)$$

Because $d' = (\text{AL}_{\text{new}} - \text{AL}_{\text{old}})/\sigma$, we replace σ with $1/d'$ in the formula of $\text{MIN}\sigma^2_{\text{O/O+B/B}} = \sigma^2_{\text{B/B}}\sigma^2_{\text{O/O}} \div (\sigma^2_{\text{B/B}} + \sigma^2_{\text{O/O}})$. MAX $d'^2_{\text{O/O+B/B}}$ is the inverse of $\text{MIN}\sigma^2_{\text{O/O+B/B}}$; $d'^2_{\text{B/B}}$ is the inverse of $\sigma^2_{\text{B/B}}$; $d'^2_{\text{O/O}}$ is the inverse of $\sigma^2_{\text{O/O}}$. We get the following:

$$\frac{1}{\text{MAX } d'^2_{\text{O/O+B/B}}} = \frac{\frac{1}{d'^2_{\text{O/O}}} * \frac{1}{d'^2_{\text{B/B}}}}{\left(\frac{1}{d'^2_{\text{O/O}}} + \frac{1}{d'^2_{\text{B/B}}}\right)} = \frac{1}{d'^2_{\text{B/B}} + d'^2_{\text{O/O}}}$$

$$\text{Hence MAX } d'_{\text{O/O+B/B}} = \sqrt{d'^2_{\text{B/B}} + d'^2_{\text{O/O}}}$$

References

Ball, K., Smith, D., Ellison, A., & Schenk, T. (2010). A body-centred frame of reference drives spatial priming in visual search. *Experimental Brain Research, 204*, 585–594.

Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences, 10*, 551–557.

Burgess, N. (2008). Spatial cognition and the brain. *Annals of the New York Academy of Sciences, 1124*, 77–97.

Burgess, N., Spiers, H. J., & Paleologou, E. (2004). Orientational manoeuvres in the dark: Dissociating allocentric and egocentric influences on spatial memory. *Cognition, 94*, 149–166.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review, 114*, 340–375.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition, 23*, 149–178.

Cheng, K. (2008). Whither geometry? Troubles of the geometric module. *Trends in Cognitive Sciences, 12*, 355–361.

Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin, 133*, 625–637.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–114.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*, 302–307.

Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 5909–5914.

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *Journal of Neuroscience, 29*, 15601–15612.

Friedman, A., & Waller, D. (2008). View combination in scene recognition. *Memory & Cognition, 36*, 467–478.

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition, 61*, 195–232.

Holmes, M. C., & Sholl, M. J. (2005). Allocentric coding of object-to-object relations in overlearned and novel environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1069–1087.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*, 352–376.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science, 20*, 568–577.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge* (pp. 1–17). Berlin: Springer-Verlag.

Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 109–169). MIT press.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences, 9*, 296–305.

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

Mou, W., Fan, Y., McNamara, T. P., & Owen, C. B. (2008). Intrinsic frames of reference and egocentric viewpoints in scene recognition. *Cognition, 106*, 750–769.

Mou, W., McNamara, T. P., Rump, B., & Xiao, C. (2006). Roles of egocentric and allocentric spatial representations in locomotion and reorientation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1274–1290.

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 142–157.

Mou, W., Xiao, C., & McNamara, T. P. (2008). Reference directions and reference objects in spatial memory of a briefly viewed layout. *Cognition, 108*, 136–154.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology, 18*, 689–693.

Sargent, J., Dopkins, S., Philbeck, J., & Chichka, D. (2010). Chunking in spatial memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 36*, 576–589.

Sargent, J., Dopkins, S., Philbeck, J., & Modarres, R. (2008). Spatial memory during progressive disorientation. *Journal of Experimental Psychology-Learning Memory and Cognition, 34*, 602–615.

Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychological Science, 9*, 315–320.

Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition, 70*, 191–210.

Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition, 77*, 215–250.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences, 6*, 376–382.

Wertheimer, M. (1912). Experimental studies on seeing of movement. *Zeitschrift Fur Psychologie Und Physiologie Der Sinnesorgane, 61*, 161–265.

Xiao, C., Mou, W., & McNamara, T. P. (2009). Use of self-to-object and object-to-object spatial relations in locomotion. *Journal of Experimental Psychology: Learning Memory and Cognition, 35*, 1137–1147.