

# Allocentric coding: spatial range and combination rules

Damien Camors, Christophe Jouffrais, Benoit Cottereau, Jean-Baptiste Durand

▶ **To cite this version:**

Damien Camors, Christophe Jouffrais, Benoit Cottereau, Jean-Baptiste Durand. Allocentric coding: spatial range and combination rules. Vision Research, 2015, 109, part A, pp.87-98. 10.1016/j.visres.2015.02.018 . hal-01387829

**HAL Id: hal-01387829**
**https://hal.science/hal-01387829**

Submitted on 26 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.




Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 15214

**To link to this article** : DOI:10.1016/j.visres.2015.02.018
URL : http://dx.doi.org/10.1016/j.visres.2015.02.018

**To cite this version** : Camors, Damien and Jouffrais, Christophe and Cottereau, Benoit and Durand, Jean-Baptiste *Allocentric coding: spatial range and combination rules.* (2015) Vision research, vol. 109 (n° A). pp. 87-98. ISSN 0042-6989

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

# Allocentric coding: Spatial range and combination rules

D. Camors [a,b,c,d], C. Jouffrais [c,d], B.R. Cottereau [a,b], J.B. Durand [a,b,*]

[a] Université de Toulouse, Centre de Recherche Cerveau et Cognition, Toulouse, France
[b] CNRS, CerCo, Toulouse, France
[c] Université de Toulouse, IRIT, Toulouse, France
[d] CNRS, IRIT, Toulouse, France

*Keywords:*
Space perception
Allocentric
Egocentric
Pointing
Cue combination

A B S T R A C T

When a visual target is presented with neighboring landmarks, its location can be determined both relative to the self (egocentric coding) and relative to these landmarks (allocentric coding). In the present study, we investigated (1) how allocentric coding depends on the distance between the targets and their surrounding landmarks (*i.e.* the spatial range) and (2) how allocentric and egocentric coding interact with each other across targets-landmarks distances (*i.e.* the combination rules). Subjects performed a memory-based pointing task toward previously gazed targets briefly superimposed (200 ms) on background images of cluttered city landscapes. A variable portion of the images was occluded in order to control the distance between the targets and the closest potential landmarks within those images. The pointing responses were performed after large saccades and the reappearance of the images at their initial location. However, in some trials, the images' elements were slightly shifted (±3°) in order to introduce a subliminal conflict between the allocentric and egocentric reference frames. The influence of allocentric coding in the pointing responses was found to decrease with increasing target-landmarks distances, although it remained significant even at the largest distances ($\geqslant 10°$). Interestingly, both the decreasing influence of allocentric coding and the concomitant increase in pointing responses variability were well captured by a Bayesian model in which the weighted combination of allocentric and egocentric cues is governed by a coupling prior.

## 1. Introduction

Being able to form, maintain and update representations of objects locations in surrounding space is mandatory for flexible and adaptive behaviors. Depending on the spatial task at hand and on the properties of the surrounding space, human subjects have been shown to rely on two non-mutually exclusive types of spatial representations: (1) egocentric representations, in which spatial locations are encoded with respect to the self and (2) allocentric representations, in which locations are encoded with respect to external landmarks (Burgess, 2006; Colby, 1998; Tatler & Land, 2011).

Converging lines of evidences indicate that subjects rely on egocentric, gaze-centered, representations when pointing/reaching toward the memorized locations of visual targets that were presented within neutral surrounds, *i.e.* in the absence of external landmarks (Cohen & Andersen, 2002; Lacquaniti & Caminiti, 1998; Medendorp et al., 2008; Thompson & Henriques, 2011). However, it is also been shown that when visual or cognitive landmarks are provided with the visual targets, these allocentric cues contributes to the mental representation of the targets location (Barry, Bloomberg, & Huebner, 1997; Carrozzo et al., 2002; Diedrichsen et al., 2004; Krigolson & Heath, 2004; Lemay, Bertram, & Stelmach, 2004a; Obhi & Goodale, 2005). Although some of the earliest studies have claimed that allocentric representations dominate and can even extinguish egocentric representations in the context of memory-based actions (Hay & Redon, 2006; Lemay, Bertram, & Stelmach, 2004a; Sheth & Shimojo, 2004), more recent studies rather support the idea of a weighted combination of these two types of spatial representation before action execution (Byrne, Cappadocia, & Crawford, 2010), with the weights being governed, at least partially, by the respective reliability of these two types of cues (Byrne & Crawford, 2010; Byrne & Henriques, 2013).

Cue reliability refers to the consistency of the representations derived from that cue alone, and is generally inferred from the reproducibility (the inverse of the variance) of the behavioral responses produced when only that cue is available (Battaglia, Jacobs, & Aslin, 2003; Ernst, 2006; Ernst & Banks, 2002; Girshick

* Corresponding author at: CNRS, CerCo, Toulouse, France.
*E-mail address:* jbdurand@cerco.ups-tlse.fr (J.B. Durand).

& Banks, 2009; Knill, 2007; Knill & Saunders, 2003; Körding et al., 2007; Landy et al., 1995; Young, Landy, & Maloney, 1993). Among the factors affecting the relative reliability (or weight) of allocentric versus egocentric cues, the importance of the temporal delay between targets/landmarks presentation and action execution has been repeatedly emphasized: a greater delay increases the weight allocated to allocentric cues, while decreasing the weight of egocentric cues (Carrozzo et al., 2002; Chen, Byrne, & Crawford, 2011; Hay & Redon, 2006; Obhi & Goodale, 2005; Sheth & Shimojo, 2004; but see Schütz, Henriques, & Fiehler, 2013, for an instance of constant allocentric/egocentric weights across time delays). In their majority, these observations are reminiscent to those reported for actions directed toward targets embedded in illusory visual contexts. Landmarks producing illusory biases in the targets' perceived size or position have been shown to exert a stronger influence on delayed actions than on those initiated during, or immediately after, the presentation of the illusory context (Bridgeman, Peery, & Anand, 1997; Gentilucci et al., 1996; Goodale, Westwood, & Milner, 2004; Hu & Goodale, 2000; Rossetti, 1998).

Other factors influencing the relative weight of allocentric and egocentric cues have been identified, such as the specificities in the task demand (Bridgeman, Peery, & Anand, 1997), the intrinsic stability (Byrne & Crawford, 2010) and predictability (Neely et al., 2008) of the allocentric landmarks, or the age of the subjects (Lemay, Bertram, & Stelmach, 2004b; Lemay & Proteau, 2003).

Surprisingly, another potentially important factor has received little interest: the spatial distance between the visual targets and their surrounding landmarks. Intuitively, it seems reasonable to assume that increasing the targets to landmarks distance should impair the precision to encode targets location relative to the surrounding landmarks. To our knowledge, only Krigolson and colleagues (2007) have addressed this question in a direct manner with a memory guided reaching task in which the landmarks occupied proximal (∼4°), medial (∼6°) or distal (∼8°) locations with respect to the visual targets during the memory encoding phase. They found that although proximal and medial landmarks had a beneficial effect regarding the reliability of the pointing responses (*i.e.* less dispersion in the pointing responses), the distal landmarks had no noticeable influence. This result suggests that allocentric coding is restricted to objects that are relatively close to each other (*i.e.* less than 8° apart in that particular experiment). The idea of a limited spatial range of allocentric coding is reinforced by studies (Diedrichsen et al., 2004; Schmidt, Werner, & Diedrichsen, 2003) showing that even when several landmarks are located within a few degrees apart from a target, only the closest one induces distortions in reaching end-points. An apparently related finding has been recently provided (Fiehler et al., 2014) with pictures of natural scenes in which both the targets and landmarks depicted real world objects. The proximal landmarks were found to influence the pointing responses while the distal ones did not. However, only the proximal landmarks were task-relevant in that previous study so it is unclear whether these proximal landmarks were more efficient because they were closer or because more attention was paid to them. A recent follow-up study from the same group supports this second hypothesis (Klinghammer et al., 2014).

It is important to note that the vast majority of the above-mentioned studies have relied on well controlled but highly artificial setups in which landmarks were isolated point-like objects, simple geometrical shapes or grid-like patterns. Even the last mentioned study (Fiehler et al., 2014) used relatively simple breakfast scenes, with a restricted number of visual objects on a table, and subjects could take as much time as needed to explore those scenes. Thus, the extent to which the accumulated knowledge can be generalized to more complex ecological situations remains largely unaddressed. For instance, does allocentric coding occur when landmarks are not provided in a neutral surround, or in relatively simple environments but must be extracted from short glimpses to cluttered visual scenes, such as those encountered when walking in crowded city streets? If allocentric coding does occur in more complex ecological conditions, is its spatial range more restricted than with artificial landmarks (Krigolson et al., 2007), and how does it combine with egocentric coding (Byrne & Crawford, 2010)? These questions on the spatial range and combination rules of allocentric coding are addressed in the present study.

We asked human subjects to perform a memory-based laser-pointing task toward previously gazed targets, which were briefly superimposed on background images of cluttered city landscapes. The presentation duration of both the targets and images was 200 ms, which is close to the average fixation time between two successive saccades during the exploration of natural visual scenes (Henderson, 2003; Martinez-Conde, Macknik, & Hubel, 2004; Rayner, 1998). Masks centered on the targets covered variable portions of the images in order to control the distance between the targets and the closest potential landmarks that are extracted from those images. After the memory encoding phase, subjects were required to perform a large saccade (25°) during total occlusion of the background images, which reappeared prior to the pointing response. The images reappeared at the same location but the content was, in some trials, slightly shifted (±3°) to the right or to the left side in order to introduce a subliminal conflict between the allocentric and egocentric reference frames. In the main experiment, subjects received the instruction to provide an allocentric-based judgment, *i.e.* to indicate where the fixation target was located within the picture, but they could rely on both allocentric and egocentric cues (either congruent or slightly incongruent) to perform the task. In additional control experiments, they were required to perform the same task in conditions where only the allocentric or egocentric cues were available. Such approach, in which cues are tested both separately and together, has already proven to be useful for inferring the cue combination rules in other contexts (Battaglia, Jacobs, & Aslin, 2003; Ernst, 2006; Ernst & Banks, 2002; Girshick & Banks, 2009; Knill, 2007; Knill & Saunders, 2003; Körding et al., 2007; Landy et al., 1995; Young, Landy, & Maloney, 1993).

The first main finding of the present study is that allocentric coding is functional in complex ecological conditions, *i.e.* when potential landmarks are seen briefly and must be extracted from cluttered visual scenes. This result still holds when the closest potential landmarks are 10° away from the target. The second main finding is that although subjects are required to produce allocentric-based spatial judgments, they rely on both allocentric and egocentric cues. Experimental results are well-captured by a Bayesian model in which the combination of allocentric and egocentric cues is governed both by their respective reliability and by a coupling prior (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2006).

## 2. Material and methods

### 2.1. Participants

Ten subjects (6 males and 4 females) performed the experiment. All of them were university students (24.4 ± 4.9 years old) and had normal or corrected to normal vision. They provided written informed consent before participating in the study and received a monetary reward (20 euros) at the end of the experiment. This study was approved by the local ethic committee (CLERIT, n° 2014-06-06-6) based notably on its compliance with the Helsinki Declaration and its newest (2013) amendments.

### 2.2. Stimuli

115 colored pictures under creative common license and depicting large and crowded city landscapes (*i.e.* markets, streets, pedestrian paths, etc...) were picked from the web site "flikr®". Pictures were selected for their high amount of salient objects and had a resolution of at least 2400 × 1200 pixels (24-bit). These initial pictures were cropped and resampled at 880 x 600 pixels (66° × 45°) in order to retain only a fraction containing a relatively homogeneous distribution of salient objects (and removing large portions of sky or other poorly informative regions). Another set of 115 pictures was generated by taking the mirror image of each of the cropped and resampled pictures from the original set, producing a total set of 230 pictures. Four exemplars of the pictures obtained after these manipulations are shown in Fig. 1A. Each of these 230 pictures served to generate 3 largely overlapping pictures of 800 × 600 pixels (60° × 45°), which differed from each other by a horizontal shift of 40 pixels (3°), as shown in Fig. 1B. The central-most picture was used as the reference while the rightward and leftward shifted versions (±3°) were used to produce a subliminal shift of the pictures inner elements in the incongruent conditions of the main experiment (*cf.* Design and procedure).

In the main experiment, the pictures were systematically presented in the same central portion of the screen in order to prevent subjects noticing the shifts. Thus, in the incongruent conditions, although the pictures inner elements were shifted horizontally, their borders were not, inducing a potential conflict between the inner landmarks and the vertical borders location. In order to minimize this conflict, the vertical borders of all the pictures were attenuated by partial transparency with the gray background through alpha compositing. The blending was governed by two vertically elongated Gaussian kernels centered on the pictures vertical borders ($\sigma$ = 10 pixels /0.75°).

### 2.3. Apparatus

Subjects were comfortably seated in a chair slightly inclined backward, with their head lying on the chair headrest and a laser pointer at hand. They were facing a very large and curved screen subtending 180° × 55° of visual angle at a viewing distance of 128 cm. The room was only illuminated by the screen reflection. Stimuli subtended 60° × 45° and were projected onto the screen by one of three video projectors (NEC NP1250) in order to occupy the central, left or right portion of the screen. The experiment was controlled by the Psychophysics Toolbox extensions version 3.0

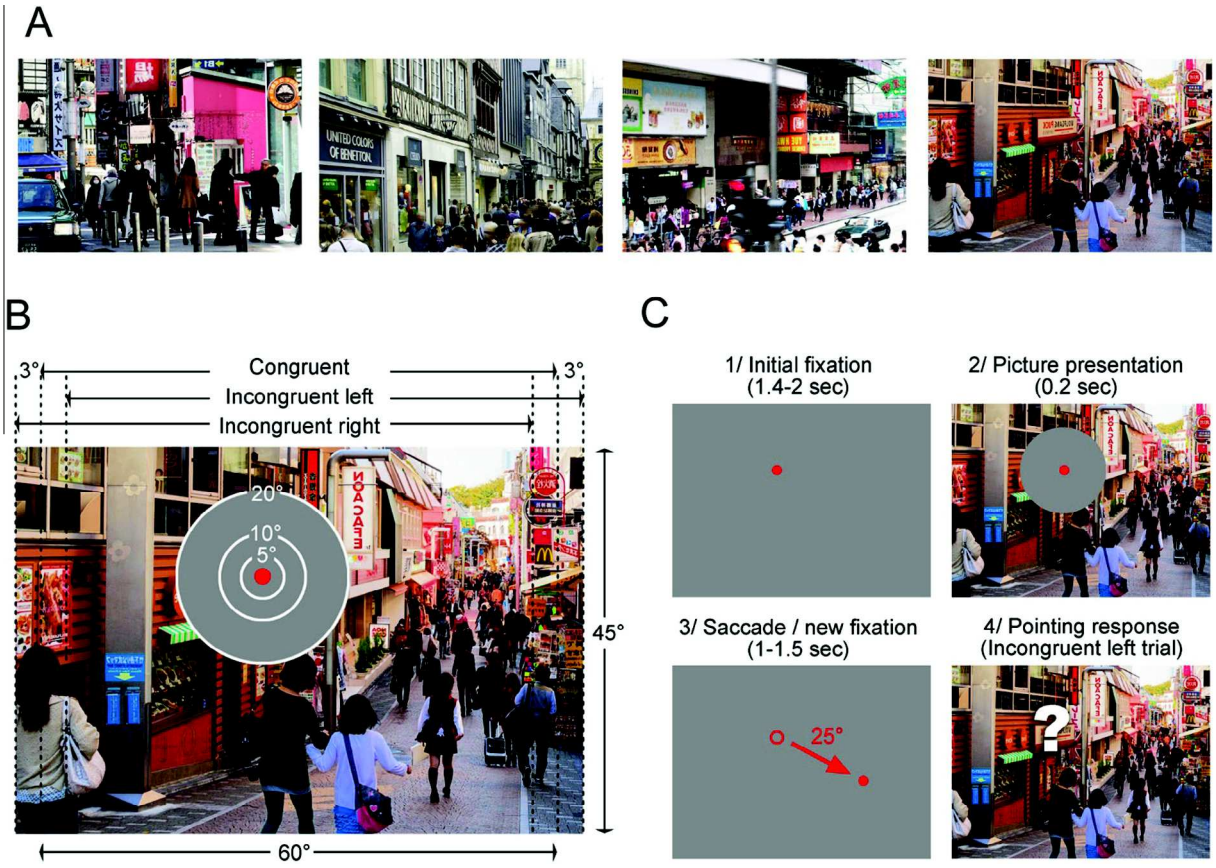



**Fig. 1.** (A) Exemplar images of cluttered city landscapes used in the present study. (B) Three images of 60° by 45° are created from each of the initial images by slightly shifting the view relative to the congruent image, both 3° left (incongruent left image) and 3° right (incongruent right image). The fixation target, whose location in the congruent image will be hold in memory for the pointing response is a red circle (0.5° diameter). Possible mask sizes surrounding the target, *i.e.* 0°, 5°, 10° or 20° diameter, are illustrated at the same scale. (C) Illustration of the cascade of events constituting a single trial in the main experiment. Subjects start by gazing the fixation target (whose location is to be memorized) after appearance on the screen for a random period between 1.4 and 2 s. Then, the background image and the mask are displayed for 200 ms. The fixation target, image and mask all disappear and a new fixation dot appears 25° away along a random direction. Subjects must saccade to this new location and keep their eyes on it for a random duration between 1 and 1.5 s. Finally, the image reappears on the screen, either the same version (congruent) or a slightly shifted incongruent, version (incongruent left in this particular example). Subjects have to point toward the initial target location within the picture.

(Brainard, 1997; Pelli, 1997) installed on Matlab® R2009 software, running on an Intel Core i5 based computer.

### 2.4. Design and procedure

#### 2.4.1. Main experiment with both allocentric (visual) and egocentric (oculomotor) spatial cues

Subjects were tested individually in a memory-based laser-pointing task, in presence of the experimenter. The sequence of events constituting a single trial is illustrated in Fig. 1C. At the beginning of each trial, the subjects were asked to fixate a small (0.5° diameter) red dot appearing on the screen. The position of this fixation dot was randomly selected within an imaginary rectangle (30° × 15°) centered on the middle of the screen. The fixation dot was displayed for a random time between 1.4 and 2 s. After that period, the fixation dot disappeared and immediately reappeared 25° away along a randomly selected direction. Subjects were instructed to execute a saccade toward this new location, and to maintain fixation until the disappearance of the fixation dot, which occurred after a random duration between 1 and 1.5 s. The subjects had then to indicate the initial location of the fixation dot by pointing on the screen with a laser pointer. They were asked to keep the pointer at this location until the experimenter, located far behind them, clicked with an optical mouse at the indicated location in order to record the pointing response. One of the 230 pictures was displayed on the screen during both the initial fixation and the pointing period. During the initial fixation, the picture was presented briefly, for 200 ms, just before the saccade occurs. A mask, whose diameter was randomly selected among four possible values (0°, 5°, 10° and 20°), surrounded the fixation dot, occluding a variable portion of the picture around the center of gaze (Fig. 1B). Because the mask was always centered on the target (the fixation dot), its size determined the target's distance from the closest potential landmarks within the pictures (0°, 2.5°, 5° and 10°) independently of possible fluctuations in fixation quality between trials. Importantly, even the biggest masks (20° diameter) covered only 12% of the total pictures area, insuring that dozens of potential landmarks were still available in those conditions (as can be appreciated in Fig. 1B). This point was important in order to insure that any decline in allocentric coding with increasing mask size is more likely caused by the targets-landmarks minimal distance rather than by a rarefaction of available landmarks. During the pointing period, the picture was always displayed without a mask and remained on the screen until the end of the trial. As the same picture was displayed during the initial fixation and the pointing period, subjects could rely on different sources of information for pointing: (a) *allocentric* (*visual*) *information,* specifying the location of the fixation dot within the picture, and (b) *egocentric* (*oculomotor*) *information,* specifying the location of the fixation dot with respect to the self and independently of the picture. Importantly, however, subjects received the instruction to indicate where the fixation target was located within the picture, which corresponds to an allocentric-based spatial judgment. To assess the respective contribution of allocentric and egocentric cues in that localization task, another crucial manipulation was introduced. Although all the pictures displayed during the initial fixation and pointing periods occupied exactly the same central portion of the screen, a perfect spatial overlap of the within-picture elements was preserved in only 1/3 of the trials (congruent conditions). In the remaining 2/3 of the trials, the picture displayed during the pointing period contained a slightly shifted view relative to that displayed initially (Fig. 1B), either 3° to the left (1/3) or 3° to the right (1/3), introducing a subliminal conflict between the locations specified by the allocentric and egocentric cues (incongruent conditions; *cf.* section "Stimuli"). Congruent and incongruent trials were randomly interleaved. In this main experiment, each subject completed 348 trials, divided into 4 blocks, for a total duration of about 1 h. At the end of the experiment, subjects were asked to report whether they had noticed some displacements of the pictures between the initial fixation and pointing periods. None of them reported having detected such displacements, which is consistent with previous studies on visual memory (Hollingworth, 2006) and change blindness (Henderson & Hollingworth, 2003).

#### 2.4.2. Control experiments with either allocentric or egocentric cues

In order to understand how allocentric and egocentric cues interact with each other, it is necessary to assess how each type of cue respectively contributes to the localization task. To that end, subjects underwent two additional control experiments in which they could only rely on the allocentric or egocentric cues. In the "*allocentric control*", the task was identical to that of the main experiment except that pictures were displayed either in the left or right portion of the screen during the pointing period. In such a case, the conflict between allocentric and egocentric cues was easily monitored (±60° shift), and subjects were instructed to estimate the location of the initial fixation dot with respect to the picture only. This "*allocentric control*" included 272 trials (68 trials per mask size), divided in 4 blocks, and lasted about 50 min. In the second control experiment, *i.e.* the "*egocentric control*", the task was also similar, to that of the first experiment except that pictures were not displayed during the initial fixation period. Random pictures were shown during the pointing period, but did not contain any relevant information for the task, forcing the subjects to rely exclusively on oculomotor cues. The "*egocentric control*" included 68 trials and lasted about 10 min.

### 2.5. Data analyses

Pointing responses, collected by the experimenter with an infrared mouse, were first converted from screen pixel coordinates to degrees of visual angle with respect to the initial fixation dot location (*i.e.* the location to memorize). We focused our analyses on the horizontal coordinates of the pointing responses since picture shifts occurred along that dimension in the incongruent trials of the main experiment.

#### 2.5.1. Outliers

Outlier trials were identified for each subject and each experimental condition of the main and control experiments by computing the 25th percentile (Q1), 75th percentile (Q3) and inter-quartile range (IQR) of the horizontal pointing response distributions. Pointing responses below Q1 − 1.5 × IQR or above Q3 + 1.5 × IQR were considered as outliers and discarded (representing 4.3% of the overall number of trials). All the analyses described below were also performed with a more permissive criterion (3 × IQR instead of 1.5 × IQR; leading to <1% of outlier trials) in order to insure that our conclusions were not critically affected by the exclusion criterion.

#### 2.5.2. Mean and standard deviation of pointing responses distributions

For each subject and each experimental condition, the mean and standard deviation of the pointing response distributions were computed. Mean pointing responses can reveal the existence of a systematic cognitive or motor bias when allocentric (visual) and egocentric (oculomotor) cues provide congruent spatial information, but also biases in favor of one or the other type of cue when they specify distinct locations (*incongruent* conditions). The standard deviations reflect the reliability of the localization/memorization processes, in terms of trial-to-trial reproducibility, and thus the noise associated to the allocentric and/or egocentric cues.

*2.5.3. Allocentric and egocentric weights*

The relative weights allocated to the allocentric ($\omega_{allocentric}$) and egocentric ($\omega_{egocentric}$) cues were determined by using the differences in mean pointing locations ($\Delta_{pointing}$) between the *incongruent left* and *incongruent right* conditions. Since these two conditions differed in terms of spatial locations specified by the allocentric cues ($\Delta_{allocentric} = 6°$), but not the egocentric cues ($\Delta_{egocentric} = 0°$), the relative weights were computed as follow:

$$\omega_{allocentric} = \Delta_{pointing}/\Delta_{allocentric} \quad (1)$$

$$\omega_{egocentric} = (\Delta_{allocentric} - \Delta_{pointing})/\Delta_{allocentric} = 1 - \omega_{allocentric}. \quad (2)$$

Thus, a difference in mean pointing responses matching the difference in locations specified by the allocentric cues ($\Delta_{pointing} = 6°$) would indicate that pointing responses are based solely on allocentric cues ($\omega_{allocentric} = 1$ and $\omega_{egocentric} = 0$). Conversely, no difference in mean pointing responses between these conditions ($\Delta_{pointing} = 0°$) would reveal that allocentric cues are not used for the task ($\omega_{allocentric} = 0$ and $\omega_{egocentric} = 1$). Intermediate differences in mean pointing responses (*e.g.* $\Delta_{pointing} = 3°$) would show that subjects rely on both allocentric and egocentric cues (*e.g.* $\omega_{allocentric} = 0.5$ and $\omega_{egocentric} = 0.5$).

## 2.6. Modeling

Besides studying the spatial range of allocentric coding in more ecological conditions, another aim of the present experiment was to gain some insight about the way allocentric and egocentric cues interact with each other. Since subjects received the instruction to provide an allocentric-based judgment, *i.e.* to indicate where the fixation targets were located within the pictures, at least three alternatives must be considered. First, the allocentric cues might be used alone and the egocentric cues simply ignored (no combination). Second, the egocentric and allocentric cues might be combined automatically for improving the localization process (full combination). A third alternative might be that although the allocentric and egocentric cues remain segregated, they nevertheless exert an influence on each other (partial combination). In order to consider these different alternatives, the cues combination was modeled using a Bayesian framework containing a "coupling prior" (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2006).

Like in most previous studies on cue combination, we assume that the sensory evidences derived from the allocentric and egocentric cues ($x_{allo}$ and $x_{ego}$) are independent of each other and adequately described by Gaussian functions whose means, $S_{allo}$ and $S_{ego}$, reflect the locations specified by these cues, and whose standard deviations, $\sigma_{allo}$ and $\sigma_{ego}$, indicate their respective sensory noise. These sensory evidences form the *likelihood* function (Fig. 2, upper row) which can be defined mathematically as follow:

$$p(x_{allo}|S_{allo})p(x_{ego}|S_{ego}) \propto N(S_{allo}; x_{allo}, \sigma_{allo})N(S_{ego}; x_{ego}, \sigma_{ego}),$$

where $N(i;\mu,\sigma)$ is the value of the normal distribution with mean $\mu$ and standard deviation $\sigma$ evaluated at $i$.

Following the proposal of Ernst and collaborators (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2006), the "*coupling prior*" reflects the previously acquired knowledge about the relevance of integrating the allocentric and egocentric cues (*i.e.* their probable correspondence). This coupling prior takes the form of a Gaussian function elongated along the identity diagonal ($S_{allo} = S_{ego}$) with a standard deviation $\sigma_{coupling}$ (Fig. 2, middle row):

$$p(S_{allo}, S_{ego}) \propto N(S_{ego}; S_{allo}, \sigma_{coupling}).$$

The combination strength of the allocentric and egocentric cues is thus solely determined by the parameter $\sigma_{coupling}$, and it can go from "full combination" when $\sigma_{coupling} \to 0$ (Fig. 2, leftmost column) to "no combination" when $\sigma_{coupling} \to \infty$ (rightmost column), passing through partial combination regimes for intermediate values (central column).

In this Bayesian framework, the final location estimate (revealed by the pointing responses) is obtained by reading out the maximum of a posterior distribution (Fig. 2, bottom row), which is formed by integrating both *likelihood* and *prior* functions according to the Bayes' rule:

$$p(S_{allo}, S_{ego}|x_{allo}, x_{ego}) \propto p(x_{allo}|S_{allo})p(x_{ego}|S_{ego})p(S_{allo}, S_{ego})$$

The inspection of Fig. 2 (bottom row) clearly shows that distinct pointing responses are expected depending on the combination regime dictated by the coupling prior. The next step is thus to formalize the model predictions in order to compare them to the empirical pointing response distributions.

For this formalization, we strictly followed that proposed by Körding et al. (2007; Supplementary material), with the idea of deriving the *allocentric* posterior distribution through integration along the egocentric dimension, as follows:

$$\begin{aligned} &p(S_{allo}|x_{allo}, x_{ego}) \propto p(x_{allo}|S_{allo}) \int p(x_{ego}|S_{ego})p(S_{allo}, S_{ego})dS_{ego} \\ &\propto N(S_{allo}; x_{allo}, \sigma_{allo}) \int N(S_{ego}; x_{ego}, \sigma_{ego})N(S_{ego}; S_{allo}, \sigma_{coupling})dS_{ego} \\ &\propto N(S_{allo}; x_{allo}, \sigma_{allo})N\left(S_{allo}; x_{ego}, \sqrt{\sigma_{ego}^2 + \sigma_{coupling}^2}\right) \\ &\propto N\left(S_{allo}; \frac{x_{allo}\sigma_{allo}^{-2} + x_{ego}(\sigma_{ego}^2 + \sigma_{coupling}^2)^{-1}}{\sigma_{allo}^{-2} + (\sigma_{ego}^2 + \sigma_{coupling}^2)^{-1}}, \frac{1}{\sqrt{\sigma_{allo}^{-2} + (\sigma_{ego}^2 + \sigma_{coupling}^2)^{-1}}}\right). \end{aligned}$$

Since the posterior distribution is a pure Gaussian function, it can be identified with the response distribution (Körding et al., 2007), leading to:

$$p(S_{allo}|x_{allo}, x_{ego}) \propto p(\hat{S}_{allo}|S_{allo}, S_{ego})$$

This allows deriving quantitative predictions for the means and standard deviations of the pointing responses distributions, which are given by the following formulas:

$$\bar{X}_{p(\hat{S}_{allo}|S_{allo}, S_{ego})} = \frac{S_{allo}\sigma_{allo}^{-2} + S_{ego}(\sigma_{ego}^2 + \sigma_{coupling}^2)^{-1}}{\sigma_{allo}^{-2} + (\sigma_{ego}^2 + \sigma_{coupling}^2)^{-1}} \quad (3)$$

$$\sigma_{p(\hat{S}_{allo}|S_{allo}, S_{ego})} = \frac{1}{\sqrt{\sigma_{allo}^{-2} + (\sigma_{ego}^2 + \sigma_{coupling}^2)^{-1}}}. \quad (4)$$

Since $S_{allo}$ and $S_{ego}$ are set experimentally and $\sigma_{allo}$ and $\sigma_{ego}$ are empirical measures (*i.e.* the standard deviations of the pointing responses in the allocentric and egocentric control experiments), it follows that the only free parameter in the present model is the coupling prior $\sigma_{coupling}$.

In the present study, Eq. (3) was used to generate predictions for the mean pointing responses in the incongruent left and incongruent right conditions. Their difference corresponds to $\Delta_{pointing}$, as illustrated in Fig. 2 (central panel of the bottom row), from which we computed the predicted allocentric weights based on Eq. (1). The value of the $\sigma_{coupling}$ parameter minimizing the difference between the model predictions and the measured allocentric weights over the 4 different targets-landmarks distances was determined by data fitting, using a least-square procedure. In a second step, those $\sigma_{coupling}$ values were used to generate predictions regarding the dispersions (standard deviations) of the pointing responses by using Eq. (4). These predicted dispersions were compared to those actually measured in the main experiment. Note that in this last procedure, predictions were derived from a fully constrained Bayesian model (no free parameter). For both the relative allocentric/egocentric weights and the responses dispersions, the model predictions were compared to those derived from two special cases of cue combination: (1) the full combination case,

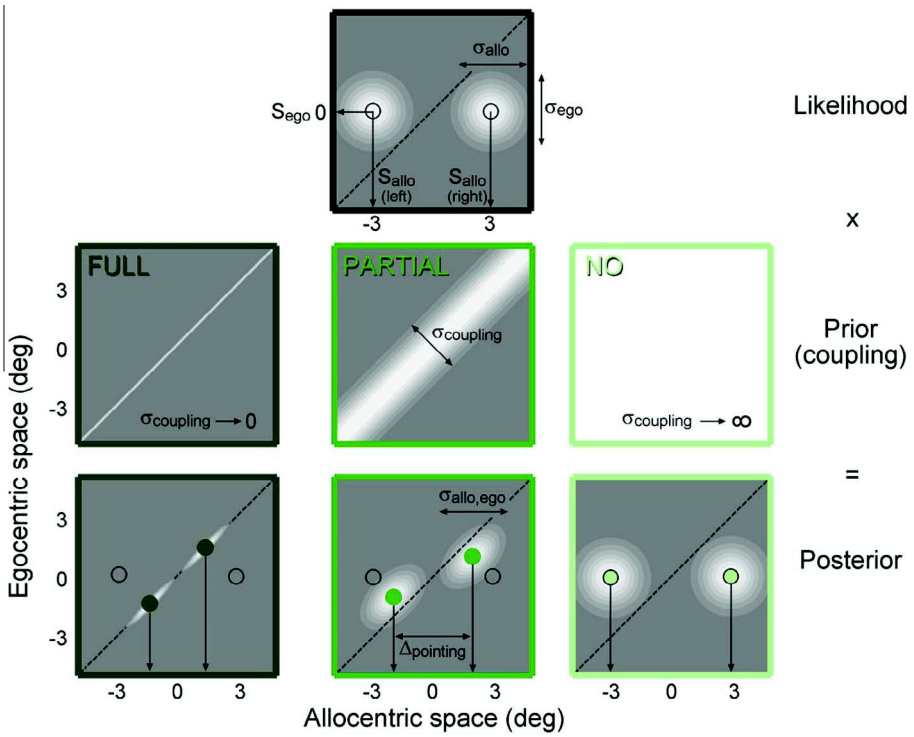



**Fig. 2.** Bayesian model of cue integration with a "coupling prior". In this model, the most probable pointing location corresponds to the maximum of the posterior distribution (bottom row) which is built by combining sensory (allocentric and egocentric) evidences (likelihood function; top row) with prior knowledge regarding their probable correspondence (coupling prior; middle row). The width of the coupling prior ($\sigma_{coupling}$) determines the strength of the coupling, going from full combination for $\sigma_{coupling}$ tending toward 0 (maximum likelihood estimate; leftmost column) to no combination for $\sigma_{coupling}$ tending toward infinity (single cue estimate; rightmost column).

or maximum likelihood estimate ($\sigma_{coupling} \to 0$), and (2) the no combination case, or single cue estimate ($\sigma_{coupling} \to \infty$).

### 2.7. Statistics

Parametric tests were used to assess the significance of our results, with a significance level set at $p < 0.05$ after Bonferroni correction for multiple comparisons. For the modeling, the goodness of fit was assessed by computing the coefficients of determination, $r^2 = 1 - SS_{residual}/SS_{total}$ ($SS_{residual}$: residual sum of squares between the observed and predicted values; $SS_{total}$: total sum of squares between the observed values and their overall mean). Coefficients of determination can be interpreted in terms of fractions of the total variance explained by a model, and are thus generally comprised between 0 and 1. Negative coefficients of determination can nevertheless occur when a model describes the data less well than a model assuming no difference between the observed values across conditions. All the analyses were performed with Matlab® R2009 software and the Statistical toolbox®.

## 3. Results

### 3.1. Allocentric coding as a function of targets-landmarks distance

The first objective of the present study was to assess whether allocentric coding can be used when landmarks have to be extracted from brief presentations of complex visual scenes, and how is this coding impacted by the distance between those landmarks and the visual targets. To that end, we used a memory-based pointing task (Fig. 1) in which subjects had to point toward the location of a previously gazed fixation target embedded, for a brief instant (200 ms), within images of complex visual scenes (city landscapes). Masks covered variable portions (0°, 5°, 10° or 20° in diameter) of the images surrounding the fixation targets in order to control the distance of the closest potential landmarks within those images (0°, 2.5°, 5° or 10°, respectively). Pointing responses were performed after a large saccade (25°) and the reappearance of the background images either with exactly the same spatial configuration (*congruent*) or with a 3° shift to the left (*incongruent left*) or to the right (*incongruent right*) of the inner images elements (and thus of the potential landmarks). Importantly, subjects debriefing after the main experiment confirmed that none of them perceived the shifts in the incongruent conditions.

Fig. 3A shows the pointing responses recorded in the main experiment for 3 representative subjects (lines) as a function of mask diameter (columns). Responses for the *congruent*, *incongruent left* and *incongruent right* conditions are given by the white circles, pale leftward triangles and dark rightward triangles, respectively. Individual responses and their overall means are signaled by small and large symbols respectively.

It can be seen that, with increasing distance between the targets and the closest potential landmarks (from left/yellow to right/red), the pointing responses for the *incongruent left* and *incongruent right* conditions tend to get closer to those measured in the *congruent* condition. The *congruent* and *incongruent* conditions differ regarding the locations specified by the allocentric cues (0° versus ±3°), not those specified by the egocentric cues (0° in both cases). Thus, the progressive convergence of pointing locations toward 0° indicates that the allocentric cues become progressively less influential (and the egocentric cues more influential) as the targets-landmarks distance increases.

Mean pointing responses for the 10 individual subjects and for the group are shown in Fig. 3B, with similar conventions as those used in Fig. 3A. The convergence of pointing responses toward 0° with increasing targets-landmarks distance is clearly observed

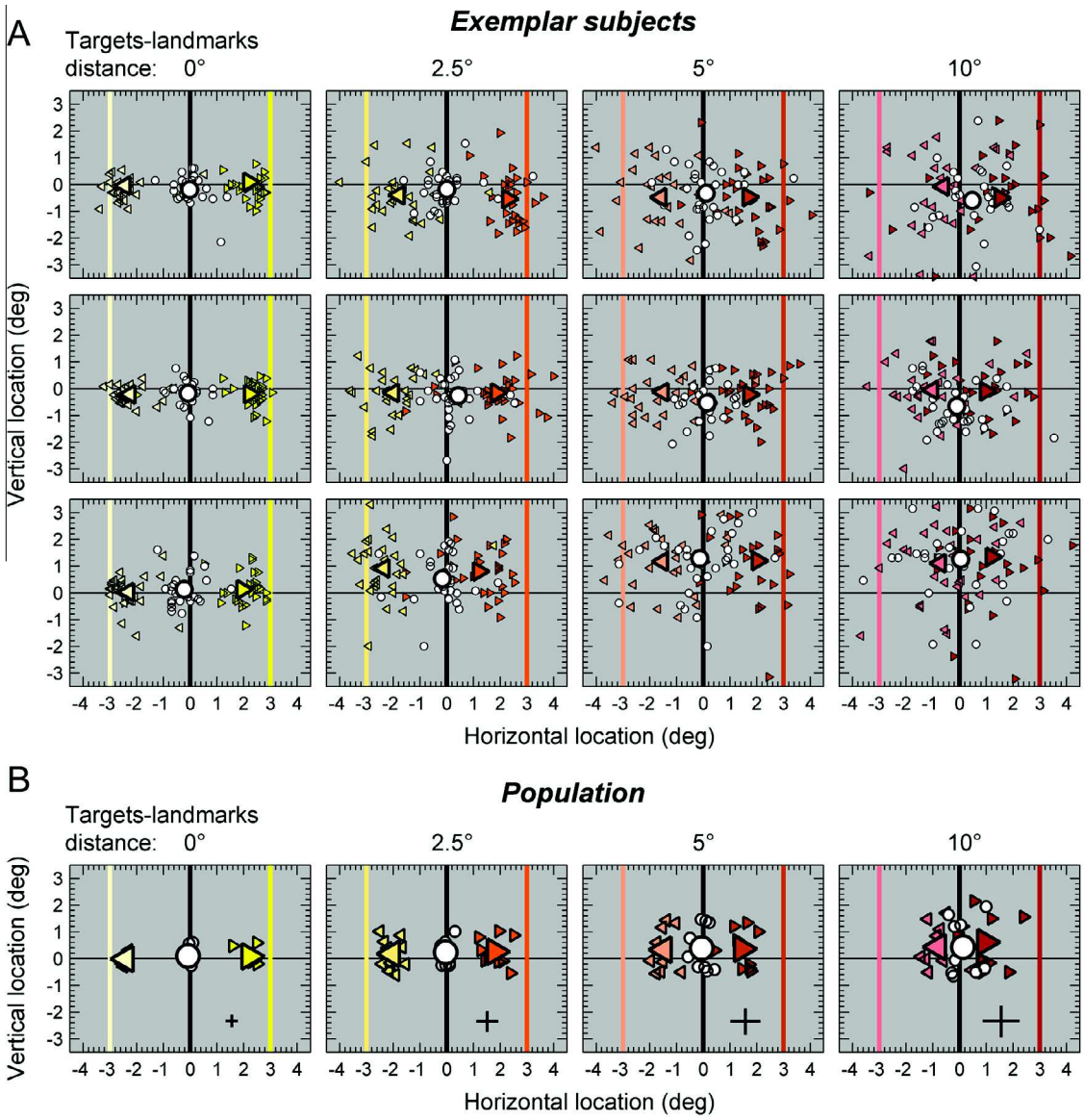



**Fig. 3.** (A) Pointing responses for 3 exemplar subjects (rows) as a function of the distance between the targets and the closest potential landmarks (columns). In each plot, the horizontal/vertical pointing directions are represented for the 3 conditions of the main experiment: congruent allocentric and egocentric cues (white circles), incongruent left: with allocentric cues shifted by −3° (pale leftward triangles) and incongruent right: with allocentric cues shifted by +3° (dark rightward triangles). The shifts are indicated by vertical lines with the same color code. Small symbols show individual responses and large symbols represent their mean. (B) Same representation as (A) for the mean responses of each of the 10 subjects (small symbols) and their overall mean (large symbols). Black crosses indicate the mean dispersion of the pointing responses along the horizontal and vertical dimensions.

for all subjects. Nevertheless, it can be seen that this convergence is only partial. Even with the largest targets-landmarks distances (5° and 10°), the distributions of mean pointing responses across subjects for the *incongruent left*, *congruent* and *incongruent right* conditions show only a weak overlap: −1.47° ± 0.35°, −0.06° ± 0.31°, 1.46° ± 0.45° respectively for a distance of 5° and −0.81° ± 0.54°, 0.12° ± 0.52°, 0.96° ± 0.74°, respectively for a distance of 10° (mean ± standard deviation). This point is important because it suggests that allocentric coding is effective over the whole range of targets-landmarks distances tested in the present study.

The extent to which allocentric cues influence the pointing responses was quantified by computing the relative weights of the allocentric and egocentric cues ($\omega_{allocentric}$ and $\omega_{egocentric}$), following the formulas (1) and (2) provided in the Section 2. Briefly, the two weights sum to 1 and the larger the relative weight allocated to one cue, the more the influence this cue exerts on the measured pointing responses. Concentrating on the allocentric weight, $\omega_{allocentric} = 1$ means that pointing responses fully follow the subliminal shifts (±3°) introduced in the background images during the incongruent conditions and, conversely, $\omega_{allocentric} = 0$ indicates that the mean pointing responses are not affected by the images shifts. Fig. 4A shows the relative allocentric weights as a function of mask size for the 10 individual subjects (profiles with circular symbols) as well as for the group (colored bar plot). The dark gray regions in the bar plot indicate the relative weight allocated to egocentric cues. It clearly appears that the weights of the allocentric cues decrease gradually as the distance of the closest potential landmarks increases: 077 ± 0.05, 0.64 ± 0.09, 0.48 ± 0.09 and 0.29 ± 0.14 for targets-landmarks distance of 0°, 2.5°, 5° and 10°, respectively (mean ± standard deviation). However, allocentric weights remain significantly higher than 0 over the whole range of tested targets-landmarks distances (*t*-test

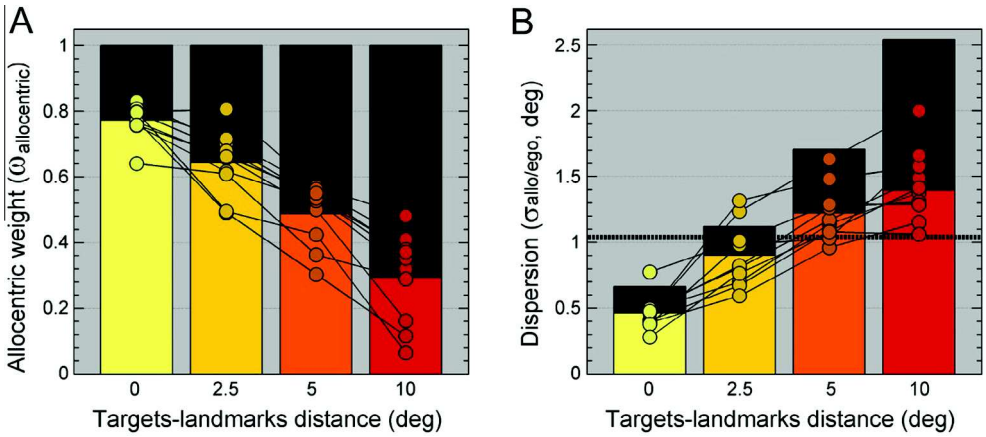



**Fig. 4.** (A) Relative weights of the allocentric cues as a function of the targets-landmarks distance, both for the population (colored bar plot) and for each individual subject (line plots with circular symbols). Related egocentric weights are shown in dark gray in the population bar plot. Allocentric weights are computed as the ratio of the distance between the mean horizontal pointing responses obtained in the incongruent left and incongruent right conditions ($\Delta_{pointing}$) against the actual distance in location specified by allocentric cues between these 2 conditions ($\Delta_{allocentric} = 6°$). (B) Dispersion (standard deviation) of the pointing responses in the main experiment ($\sigma_{allo/ego}$) as a function of the targets-landmarks distance, both for the population and the individual subject, with the same conventions than in (A). Dark bars indicate the related dispersions as a function of targets-landmarks distance in the allocentric only control experiment ($\sigma_{allo}$), while the horizontal dashed line indicates the dispersion measured in the egocentric only control experiment ($\sigma_{ego}$).

with Bonferroni corrected $p$ values; distance of 0°: $t_{9,1} = 46.9$, $p < 10^{-11}$; distance of 2.5°: $t_{9,1} = 21.1$, $p < 10^{-8}$; distance of 5°: $t_{9,1} = 16.68$, $p < 10^{-7}$; distance of 10°: $t_{9,1} = 6.82$, $p < 10^{-4}$), confirming that allocentric coding can occur after brief exposures (200 ms) to complex visual scenes, even when the closest potential landmarks are located 10° away from the visual targets.

Another important aspect of the data, easily observable in Fig. 3A, is that the dispersion of the pointing responses increases with increasing targets-landmarks distance. Dispersion profiles were obtained by averaging the standard deviations of the pointing responses obtained in the congruent, incongruent left and incongruent right conditions of the main experiment. These profiles of $\sigma_{allo/ego}$ are shown in Fig. 4B for the 10 individual subjects (profiles with circular symbols) and for the group (colored bar plot). The same procedure was repeated for the allocentric only control in order to obtain dispersion profiles of $\sigma_{allo}$ (dark gray bars in Fig. 4B). Mean (±standard deviation) for $\sigma_{allo/ego}$ and $\sigma_{allo}$ across the 4 targets-landmarks distances were, respectively, 0.46° ± 0.13°, 0.90° ± 0.26°, 1.22° ± 0.23°, 1.40° ± 0.27° and 0.66° ± 0.28°, 1.12° ± 0.25°, 1.71° ± 0.28°, 2.54° ± 0.47°. The dispersion measured in the egocentric only control, $\sigma_{ego}$, is also indicated in Fig. 4B by the horizontal dashed line (1.03° ± 0.22°). Thus the dispersion profile obtained in the main experiment lies systematically below that of the allocentric control (*i.e.* $\sigma_{allo/ego} < \sigma_{allo}$) but only half of it is also below the dispersion measured in the egocentric only control (*i.e.* $\sigma_{allo/ego} < \sigma_{ego}$ for distances of 0° and 2.5° but $\sigma_{allo/ego} > \sigma_{ego}$ for distances of 5° and 10°).

At that point, it is important to consider that the influence of the egocentric cues might be confounded by that of the allocentric cues that are not shifted in the incongruent conditions: the objects in the room, the borders of the screen and of the pictures. The debriefing with the subjects confirmed that the weak illumination of the room (only illuminated by the video-projector) and the wide area covered by the stimulation screen (180° × 55°) prevented the possibility of seeing distant objects or to rely on the screen borders. Thus, the pictures' vertical borders are the only external landmarks that could conflict with the horizontal shift of the within pictures landmarks in the incongruent conditions. To minimize their potential influence *a priori*, those borders were attenuated through alpha transparency (see Section 2). To assess their influence *a posteriori*, a control analysis was carried out by splitting the trials in 2 sets according to the targets distance from the closest pictures' vertical border. Trials in which the targets were the closest to the vertical borders (mean distance ± standard deviation: 17.5° ± 1.9°) and those that were the farthest (24.0° ± 2.5°) were indistinguishable in term of both allocentric weight profiles and dispersion profiles (Supplementary Fig. 1). This indicates that the remoteness of the closest borders (>15°), together with their transparency attenuation, prevented a significant contribution of these external landmarks in the pointing task.

These results indicate that in the present experimental conditions, the spatial representations derived from both allocentric and egocentric cues are more reliable than those derived from allocentric cues alone, but not necessarily more reliable than those built from egocentric cues alone. Although this last result might at first appear counterintuitive, the next section will show that a simple Bayesian model can capture all the aspects of the results described in this section.

### *3.2. Modeling the combination of allocentric and egocentric cues*

A second objective of the present study was to gain some insights regarding how allocentric and egocentric cues are combined during this memory-based pointing task. The fact that the pointing responses fall somewhere between the locations specified by allocentric and egocentric cues in the incongruent conditions indicates that subjects actually rely on spatial representations that combine both types of cues. Moreover, the decreasing influence of allocentric cues with increasing targets-landmarks distance is concomitant with a progressive loss of reliability of the allocentric cues (dark gray bars in Fig. 4B). This last observation suggests that the combination of allocentric and egocentric cues is at least partially driven by the respective reliability of these two types of cues.

Statistically-optimal cue combination (in term of reliability of the final estimate) implies that the weights allocated to the different cues are proportional to their respective reliability (more weight is attributed to the more reliable cues). This kind of optimal cue combination, known as maximum-likelihood estimate (MLE), has been evidenced in many instances of sensory cue combinations (Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003; Lovell, Bloj, & Harris, 2012), but recent studies suggests that it may not necessarily hold for the combination of egocentric and allocentric cues (Byrne & Crawford, 2010; Byrne & Henriques, 2013). In the present study, we adopted the Bayesian framework

proposed by Ernst and collaborators (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2006) to address this question.

This model described in the Section 2 and illustrated in Fig. 2, is attractive because it invokes a simple "coupling prior" whose value determines the combination regime of the allocentric and egocentric cues. In the case of a strong coupling prior ($\sigma_{coupling} \rightarrow 0$), the model boils down to the classical MLE model (full combination; Fig. 2, leftmost column). However, a weaker coupling prior allows the cues to remain partially segregated: one can rely dominantly on one of the cues, although the other will nevertheless exert an influence (partial combination; Fig. 2, central column). In the case of an increasingly weak coupling prior ($\sigma_{coupling} \rightarrow \infty$), the model then describes a complete segregation of the cues that exert no influence on each other (no combination; Fig. 2, rightmost column). Since subjects received the instruction to provide an allocentric-based judgment (*i.e.* to indicate where the fixation target was located within the picture), they might have relied on a statistically-optimal strategy (full combination), or dominantly on the allocentric cues with either an influence (partial combination) or no influence of the (non-dominant) egocentric cues.

Model fitting was achieved by statistical adjustment of this coupling prior ($\sigma_{\text{coupling}}$) in order to minimize the difference between the measured and predicted allocentric weight profiles across the four different mask sizes. Additionally, results of the statistical adjustment were compared to those obtained when forcing the model toward a full combination ($\sigma_{coupling} \rightarrow 0$) or a no combination regime ($\sigma_{coupling} \rightarrow \infty$). The same procedure was applied to the group and to each subject individually.

Fig. 5A shows the results of this procedure for the group of subjects. The allocentric weights profile ($\omega_{allocentric}$) as a function of targets-landmarks distance is given by square symbols. Importantly, neither the single cue estimate (no combination model; pale green curve) nor the maximum likelihood estimate (full combination model; dark green curve) can adequately account for the allocentric weights profile (coefficients of determination: $r^2_{\text{no combination}} < 0$; $r^2_{\text{full combination}} = 0.15$). However, introducing a single free parameter in the model, the *coupling prior* ($\sigma_{coupling}$), is sufficient to satisfactorily describe this profile (medium green curve). Data fitting leads to a $\sigma_{coupling}$ value of 1.16° for the group and the coefficient of determination indicates that the model account for 94% of the visual gain variance as a function of mask size ($r^2_{\text{partial combination}} = 0.94$). The same procedure was repeated for each of the 10 individual subjects with largely similar results ($r^2_{\text{full combination}} < 0$; $r^2_{\text{no combination}} < 0$; $r^2_{\text{partial combination}} = 0.75 \pm 0.23$; $\sigma_{coupling} = 1.18° \pm 0.58°$; mean ± standard deviation). Results for 3 representative subjects are shown in Fig. 5B (the same subjects as those shown in Fig. 3A). Their respective fitted values for $\sigma_{coupling}$ were 1.32°, 2.00° and 1.40°.

In a second step, the $\sigma_{coupling}$ values obtained by data fitting on the allocentric weights were used to generate predictions about the dispersion of the pointing responses when both allocentric and egocentric cues are available ($\sigma_{allo/ego}$). Note that this time, the model is fully constrained and does not contain any free parameter. The results of this procedure are shown in Fig. 5C for the group and Fig. 5D for the same individual subjects as in Fig. 5B. It can be seen that the *coupling priors* obtained by fitting the measured versus predicted allocentric weights allow describing fairly well the reliability of the pointing responses across the different targets-landmarks, both at the group level ($r^2_{\text{partial combination}} = 0.93$) and at the level of individual subjects ($r^2_{\text{partial combination}} = 0.65 \pm 0.26$; mean ± standard deviation). By contrast, neither the *full combination* model nor the *no combination* model can adequately explain these dispersion profiles (negative coefficient of determinations in both cases). Importantly, dispersion measures are highly sensitive to the criterion used for detecting and excluding outlier trials. The results described above were obtained after applying a 1.5 IQR rule (see Section 2). This rule is commonly used and resulted in excluding about 4% of the total number of trials. To insure that our conclusions did not critically depended on this criterion, all the analyses were repeated after applying a more permissive criterion (3 IQR) with which only 1% of the trials were excluded. The results, shown in Supplementary Fig. 2, indicate that although the estimated dispersions are higher with such criterion, both the dispersion profiles and the allocentric weight profiles keep very similar shapes and are still nicely captured by a Bayesian model with a coupling prior.

Altogether, the results on the relative allocentric weights and on the reliability of the combined estimates, strongly suggest that allocentric and egocentric cues are only *partially* combined for estimating the location of a previously gazed object within the surrounding visual scene.

## 4. Discussion

In the present study, we investigated the impact of the distance between gazed targets and the surrounding landmarks on the ability to form allocentric representations of the targets' locations (*i.e.* their locations with respect to the landmarks). To our knowledge, only one previous study (Krigolson et al., 2007) addressed a similar question, and it reported that allocentric coding might have a restricted spatial range: landmarks were found to exert a beneficial influence on the accuracy of the pointing responses when located up to 6° away from the targets, but this influence disappeared at a distance of 8°. One of our objectives was to evaluate whether this finding holds in more ecological conditions: when landmarks are not point-light objects presented for several seconds within an otherwise neutral background (Krigolson et al., 2007) but must instead be extracted from structured visual scenes within the average time of a stable fixation between two successive saccades, which is about 200 ms (Henderson, 2003; Martinez-Conde, Macknik, & Hubel, 2004; Rayner, 1998).

In line with the idea of offering more ecological conditions, we favored an immersive design with a wide screen and a large working space (180° × 55°). Therefore, we could not accurately monitor eye movements over such space. However, subjects confirmed during their debriefing that they followed the task instructions without any difficulty and the low inter-subject variability in our data advocates that they behaved essentially in the same way, which likely reflects those instructions. It is important to stress out that possible imprecision in fixation during some trials was not a critical issue here. First, trials randomization combined to the brief presentation of pictures make it unlikely that fixation quality differs across mask conditions. Second, in any given trial, the distance of the target from the closest potential landmarks depends only on the mask size, and not on the actual gaze position at the time of the picture presentation.

In the present study, subjects had to extract landmarks from cluttered city landscapes under strong temporal constraints. These experimental conditions are *a priori* less favorable to the expression of an allocentric coding than those employed by Krigolson and colleagues (2007), since subjects were exposed to isolated landmarks for several seconds. They are also less favorable than those employed by Fiehler and colleagues (2014), since subjects had as much time as they wanted to explore simple breakfast scenes. However, we found evidences for an allocentric coding in all our subjects, even when the closest potential landmarks were 10° away from the targets. Note that this 10° distance represents a minimal estimate of the allocentric spatial range since one cannot exclude that more distant landmarks within the pictures were also involved in the allocentric coding. Eye movements recording during the pointing period might help addressing this issue in a

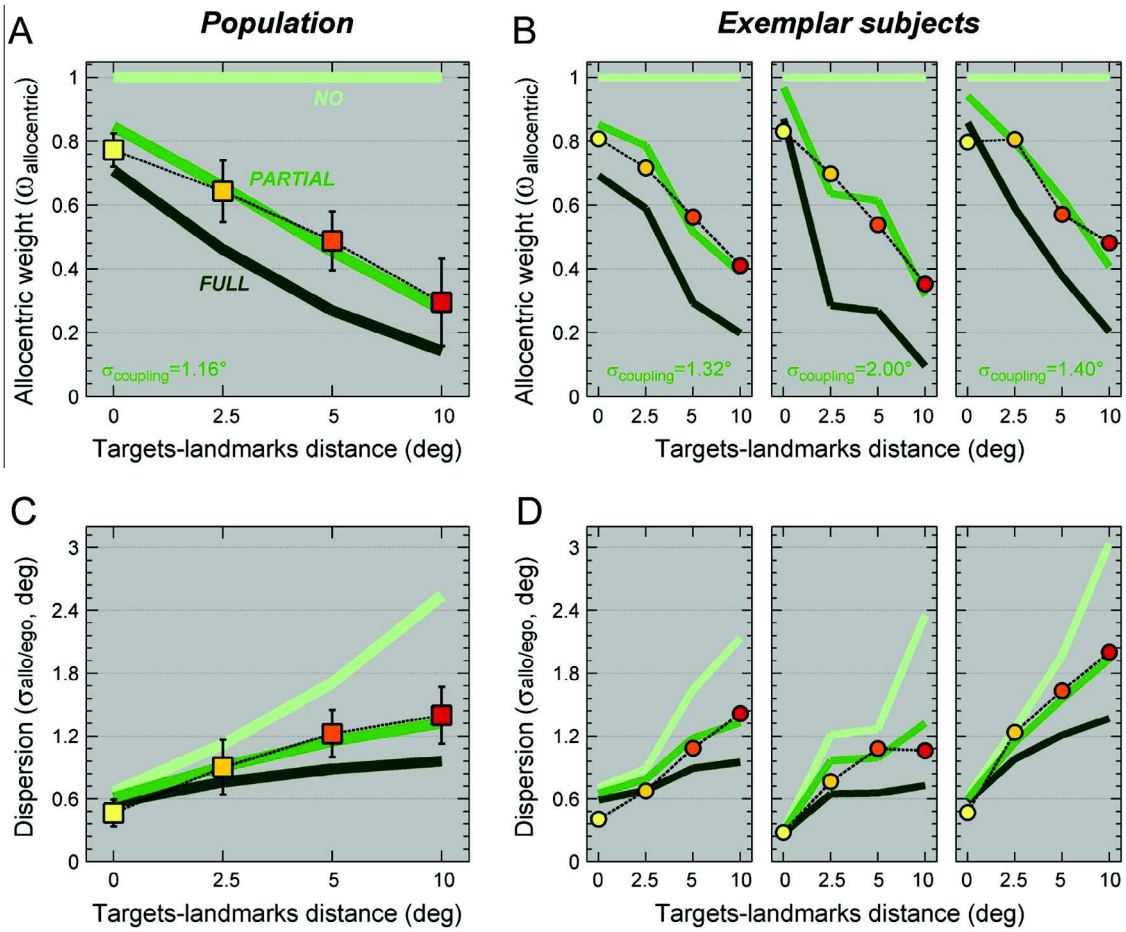



**Fig. 5.** (A) Mean allocentric weights profile (±standard deviation) across the 10 subjects (square symbols) and predictions derived from the Bayesian model for cases of full combination (FULL; dark green), no combination (NO; pale green) or partial combination (PARTIAL; medium green). Prediction for the case of partial combination was obtained by finding the value of the coupling prior ($\sigma_{coupling}$) that minimizes the residual sum of square between the model predictions and the experimental data. (B) Same as (A) for the 3 exemplar subjects shown in Fig. 3. (C) Mean dispersions profile (±standard deviation) across the 10 subjects and predictions derived from the Bayesian model for cases of full, no or partial combination (same color code). Here, the predictions for the partial combination model are derived from the value of $\sigma_{coupling}$ obtained by fitting the allocentric weights profile (fully constrained model). (D) Same as (C) for the 3 exemplar subjects.

future study. It may be tempting to interpret this wide spatial range, compared to that previously reported (Krigolson et al., 2007), as evidence that the visual system has developed special skills for working under the constraints that it faces in natural viewing conditions. Nevertheless, there is an alternative explanation, linked to a methodological difference between the two studies for measuring the influence of the allocentric cues in the pointing responses. Krigolson and colleagues looked at whether representations of the targets locations were more reliable when adding allocentric cues (leading to a reduction in pointing responses dispersion) while we studied both the reliability and the biases induced by introducing a subliminal conflict between the allocentric and egocentric cues. The tacit assumption behind the approach of Krigolson and colleagues is that adding visual landmarks (providing both allocentric and egocentric cues) should lead to more accurate spatial representations than in the absence of landmarks (*i.e.* from egocentric cues alone). However, in the present study, such an assumption would have led to conclude that allocentric cues are inefficient for targets-landmarks distances $\geqslant 5°$. As shown in Fig. 4B, the dispersion profile obtained with both allocentric and egocentric cues (colored bar plot) exceeds the dispersion measured with egocentric cues alone (horizontal dashed line) for targets-landmarks distances $\geqslant 5°$. Nevertheless, a simple inspection of the two rightmost graphs in Fig. 3C is sufficient to be convinced that allocentric cues actually bias the pointed locations for targets-landmarks distances of 5° and 10°. Thus, a simple way to reconcile both studies is to assume that the allocentric coding can occur over a wider spatial range than suggested by the dispersion of pointing responses, but that its combination with the egocentric coding does not necessarily lead to a combined spatial coding which is more reliable than that obtained with egocentric cues alone.

This point just meets the second aim of the present study, which was to gain a better understanding of the combination rules between the allocentric and egocentric cues. As defended in this work, a simple Bayesian model can explain the sub-optimal combination (in terms of reliability) of these two types of cues. The model that we adopted was proposed by Ernst and collaborators (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2006), and allows characterizing the influence of a non-dominant sensory estimate over another dominant sensory estimate. Here, the terms dominant and non-dominant characterize the sensory estimates that are directly and non-directly relevant for the task at hand. Indeed, in the present study, subjects were explicitly required to indicate the remembered location of the targets within the background images, *i.e.* to produce an allocentric-based judgment. Thus, in this case, the allocentric coding is considered dominant while the egocentric coding is non-dominant. The strength of this

model is that it contains only one free parameter, the coupling prior, which determines the strength of coupling between the sensory estimates, from full combination (maximum likelihood estimate) to no combination (single cue estimate), passing through partial combination (Fig. 2). We found that a partial fusion model fits nicely the profiles of allocentric weights across targets-landmarks distances, both at the group level and at the level of individual subjects, while neither maximum likelihood estimates nor single cue estimates could account for these profiles (Fig. 5). Importantly, the strength of coupling determined from the allocentric weight profiles allowed predicting with excellent accuracy the reliability of the estimated targets' locations, *i.e.* the profiles of pointing response dispersions. Note that these predictions for the dispersion were obtained from a fully constrained model, without any free parameter. Because of this partial combination regime with dominant allocentric cues, the reliability of the pointing responses with both allocentric and egocentric cues is less than what could be achieved from full fusion: it is statistically suboptimal. Importantly, this combination regime entails that the reliability is always better than that measured with allocentric cues alone, but not necessarily better than the reliability measured with egocentric cues alone. Thus, this model can reconcile *a priori* inconsistent data: adding visual landmarks can clearly bias the pointed locations while producing no sensible benefit in terms of reliability with respect to conditions in which there are no landmarks.

Note that the model on which we relied (Bresciani, Dammeier, & Ernst, 2006; Ernst, 2006) is not the only one in which the cue combination regime is governed by a prior (Körding et al., 2007; Roach, Heron, & McGraw, 2006; Shams, Ma, & Beierholm, 2005). Mathematically, these different models have been shown to be very close to each other (Körding et al., 2007) and the goal of the present study was not to determine which of them works better. Based on the few data points to fit to the models (*i.e.* the four targets-landmarks distances), we decided to avoid the risk of overfitting by using the more economical one (with only one free parameter), which offers the additional advantages of being linear and with posterior distributions that can be directly compared to the measured distributions of pointing responses (Körding et al., 2007). Further studies will be necessary to test and refine this model of allocentric/egocentric cues combination. Notably, an important prediction is that inverting the dominance of the cues, for instance by asking the subjects to localize the targets with respect to themselves, should modify the pattern of pointing responses according to the predictions that can be derived from Eqs. (3) and (4), simply by interchanging the "allo" and "ego" terms in these equations. To test whether mathematically more complex functions should be invoked to describe the coupling prior, it will also be necessary to vary the amount of incongruence between the egocentric and allocentric cues.

In a related study, Byrne and Crawford (2010) manipulated the stability of the visual landmarks by imposing either small or large dynamic vibrations to their spatial locations. Importantly, they found that such manipulation did not affect the reliability of the pointing responses based on allocentric cues alone, but it nevertheless impacted the weights allocated to these cues when egocentric cues were also available (less allocentric weight with large vibrations). Clearly, this result cannot be accounted for by a maximum-likelihood (full combination) model in which the weight of each cue is solely determined by its respective reliability. Such a model would predict constant weight for the allocentric cues (provided that the reliability of the egocentric cues is also constant). Thus, the authors proposed that an additional landmarks-stability heuristic contributes to the relative up- or down-weighting of the allocentric cues relative to the egocentric ones. Such a stability heuristic, learned through past experiences with the environment, might well be a factor shaping the coupling prior employed in the present study. According to this view, both the distance and the stability of the surrounding landmarks affect the weight allocated to the allocentric cues, but while the landmarks distance impacts directly the reliability of the allocentric cues, the stability has an indirect influence which can be adequately captured by a coupling prior. From a Bayesian perspective (Fig. 2), both the landmarks distance and landmarks stability impact the posterior distribution, but the former acts on the likelihood distribution while the latter operates through the prior distribution.

Altogether, the results of the present study indicate that visual landmarks allow allocentric coding of targets locations over an extended spatial range (at least 10° in diameter). Although allocentric and egocentric cues interact with each other, their combination cannot be explained by a classical maximum-likelihood model. A simple Bayesian model, with coupling prior, is shown to capture both the relative weights of allocentric and egocentric cues and the reliability of the spatial representations derived from these two types of cues.

## Acknowledgments

The authors would like to thank Florence Agbazahou for her help during the data collection. This work was supported by grants from the CNRS, the INSB-UNADEV and the ANR (ANR-13-JSV4-0007-01). D.C. is supported by a CDU from the doctoral school CLESCO (Université de Toulouse).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2015.02.018.

## References

Barry, S. R., Bloomberg, J. J., & Huebner, W. P. (1997). The effect of visual context on manual localization of remembered targets. *Neuroreport, 8*(2), 469–473.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 20*(7), 1391–1397.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision, 6*(5), 554–564. http://dx.doi.org/10.1167/6.5.2.

Bridgeman, B., Peery, S., & Anand, S. (1997). Interaction of cognitive and sensorimotor maps of visual space. *Perception and Psychophysics, 59*(3), 456–469.

Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences, 10*(12), 551–557. http://dx.doi.org/10.1016/j.tics.2006.10.005.

Byrne, P. A., Cappadocia, D. C., & Crawford, J. D. (2010). Interactions between gaze-centered and allocentric representations of reach target location in the presence of spatial updating. *Vision Research, 50*(24), 2661–2670. http://dx.doi.org/10.1016/j.visres.2010.08.038.

Byrne, P. A., & Crawford, J. D. (2010). Cue reliability and a landmark stability heuristic determine relative weighting between egocentric and allocentric visual information in memory-guided reach. *Journal of Neurophysiology, 103*(6), 3054–3069. http://dx.doi.org/10.1152/jn.01008.2009.

Byrne, P. A., & Henriques, D. Y. P. (2013). When more is less: Increasing allocentric visual information can switch visual-proprioceptive combination from an optimal to sub-optimal process. *Neuropsychologia, 51*(1), 26–37. http://dx.doi.org/10.1016/j.neuropsychologia.2012.10.008.

Carrozzo, M., Stratta, F., McIntyre, J., & Lacquaniti, F. (2002). Cognitive allocentric representations of visual space shape pointing errors. *Experimental Brain Research, 147*(4), 426–436. http://dx.doi.org/10.1007/s00221-002-1232-4.

Chen, Y., Byrne, P., & Crawford, J. D. (2011). Time course of allocentric decay, egocentric decay, and allocentric-to-egocentric conversion in memory-guided reach. *Neuropsychologia, 49*(1), 49–60. http://dx.doi.org/10.1016/j.neuropsychologia.2010.10.031.

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews. Neuroscience, 3*(7), 553–562. http://dx.doi.org/10.1038/nrn873.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron, 20*(1), 15–24. http://dx.doi.org/10.1016/S0896-6273(00)80429-8.

Diedrichsen, J., Werner, S., Schmidt, T., & Trommershäuser, J. (2004). Immediate spatial distortions of pointing movements induced by visual landmarks. *Perception and Psychophysics, 66*(1), 89–103.

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human Body Perception from the Inside Out*. New York, NY: Oxford University Press (Chapter 6, p. 105–131).

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433. http://dx.doi.org/10.1038/415429a.

Fiehler, K., Wolf, C., Klinghammer, M., & Blohm, G. (2014). Integration of egocentric and allocentric information during memory-guided reaching to images of a natural environment. *Frontiers in Human Neuroscience, 8*, 636. http://dx.doi.org/10.3389/fnhum.2014.00636.

Gentilucci, M., Chieffi, S., Deprati, E., Saetti, M. C., & Toni, I. (1996). Visual illusion and action. *Neuropsychologia, 34*(5), 369–376.

Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision, 9*(9). http://dx.doi.org/10.1167/9.9.8. 8.1–20.

Goodale, M. A., Westwood, D. A., & Milner, A. D. (2004). Two distinct modes of control for object-directed action. *Progress in Brain Research, 144*, 131–144.

Hay, L., & Redon, C. (2006). Response delay and spatial representation in pointing movements. *Neuroscience Letters, 408*(3), 194–198. http://dx.doi.org/10.1016/j.neulet.2006.08.080.

Henderson, J. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*(11), 498–504. http://dx.doi.org/10.1016/j.tics.2003.09.006.

Henderson, J. M., & Hollingworth, A. (2003). Global transsaccadic change blindness during scene perception. *Psychological Science, 14*(5), 493–497. http://dx.doi.org/10.1111/1467-9280.02459.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12), 967–992. doi:10:1167/4.12.1.

Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition, 14*(4–8), 781–807. http://dx.doi.org/10.1080/13506280500193818.

Hu, Y., & Goodale, M. A. (2000). Grasping after a delay shifts size-scaling from absolute to relative metrics. *Journal of Cognitive Neuroscience, 12*(5), 856–868.

Klinghammer, M., Wallner L., Blohm G. & Fiehler K. (2014). *Reaching for eggs and butter – Integrating spatial reference frames in natural scenes. Program No. 437.13. Neuroscience Meeting Planner*. Washington, DC: Society for Neuroscience, 2014. Online.

Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision, 7*(7). http://dx.doi.org/10.1167/7.7.5. 5.1–24.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43*(24), 2539–2558. http://dx.doi.org/10.1016/S0042-6989(03)00458-9.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One, 2*(9), e943. http://dx.doi.org/10.1371/journal.pone.0000943.

Krigolson, O., Clark, N., Heath, M., & Binsted, G. (2007). The proximity of visual landmarks impacts reaching performance. *Spatial Vision, 20*(4), 317–336. http://dx.doi.org/10.1163/156856807780919028.

Krigolson, O., & Heath, M. (2004). Background visual cues and memory-guided reaching. *Human Movement Science, 23*(6), 861–877. http://dx.doi.org/10.1016/j.humov.2004.10.011.

Lacquaniti, F., & Caminiti, R. (1998). Visuo-motor transformations for arm reaching. *The European Journal of Neuroscience, 10*(1), 195–203.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35*(3), 389–412.

Lemay, M., Bertram, C. P., & Stelmach, G. E. (2004a). Pointing to an allocentric and egocentric remembered target. *Motor Control, 8*(1), 16–32.

Lemay, M., Bertram, C. P., & Stelmach, G. E. (2004b). Pointing to an allocentric and egocentric remembered target in younger and older adults. *Experimental Aging Research, 30*(4), 391–406. http://dx.doi.org/10.1080/03610730490484443.

Lemay, M., & Proteau, L. (2003). Aging affects pointing to unseen targets encoded in an allocentric frame of reference. *Experimental Aging Research, 29*(3), 319–333. http://dx.doi.org/10.1080/03610730303723.

Lovell, P. G., Bloj, M., & Harris, J. M. (2012). Optimal integration of shading and binocular disparity for depth perception. *Journal of Vision, 12*(1). http://dx.doi.org/10.1167/12.1.1.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews. Neuroscience, 5*(3), 229–240. http://dx.doi.org/10.1038/nrn1348.

Medendorp, W. P., Beurze, S. M., Van Pelt, S., & Van Der Werf, J. (2008). Behavioral and cortical mechanisms for spatial coding and action planning. *Cortex, 44*(5), 587–597. http://dx.doi.org/10.1016/j.cortex.2007.06.001.

Neely, K. a., Tessmer, A., Binsted, G., & Heath, M. (2008). Goal-directed reaching: Movement strategies influence the weighting of allocentric and egocentric visual cues. *Experimental Brain Research, 186*(3), 375–384. http://dx.doi.org/10.1007/s00221-007-1238-z.

Obhi, S. S., & Goodale, M. a. (2005). The effects of landmarks on the performance of delayed and real-time pointing movements. *Experimental Brain Research, 167*(3), 335–344. http://dx.doi.org/10.1007/s00221-005-0055-5.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: A strategy for balancing the costs and benefits of audio-visual integration. *Proceedings Biological Sciences/The Royal Society, 273*(1598), 2159–2168. http://dx.doi.org/10.1098/rspb.2006.3578.

Rossetti, Y. (1998). Implicit short-lived motor representations of space in brain damaged and healthy subjects. *Consciousness and Cognition, 7*(3), 520–558. http://dx.doi.org/10.1006/ccog.1998.0370.

Schmidt, T., Werner, S., & Diedrichsen, J. (2003). Spatial distortions induced by multiple visual landmarks: How local distortions combine to produce complex distortion patterns. *Perception and Psychophysics, 65*(6), 861–873.

Schütz, I., Henriques, D. Y. P., & Fiehler, K. (2013). Gaze-centered spatial updating in delayed reaching even in the presence of landmarks. *Vision Research, 87*, 46–52. http://dx.doi.org/10.1016/j.visres.2013.06.001.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport, 16*(17), 1923–1927.

Sheth, B. R., & Shimojo, S. (2004). Extrinsic cues suppress the encoding of intrinsic cues. *Journal of Cognitive Neuroscience, 16*(2), 339–350. http://dx.doi.org/10.1162/089892904322984616.

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 366*(1564), 596–610. http://dx.doi.org/10.1098/rstb.2010.0188.

Thompson, A. A., & Henriques, D. Y. P. (2011). The coding and updating of visuospatial memory for goal-directed reaching and pointing. *Vision Research, 51*(8), 819–826. http://dx.doi.org/10.1016/j.visres.2011.01.006.

Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research, 33*(18), 2685–2689.