

# How attention simplifies mental representations for planning

Reviewed Preprint

v1 • September 9, 2025

Not revised

Jason da Silva Castanheira , Nicholas Shea, Stephen M Fleming

Department of Experimental Psychology, University College London, London, United Kingdom • Institute of Philosophy, School of Advanced Study, University of London, London, United Kingdom • Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, United Kingdom • Canadian Institute for Advanced Research (CIFAR), Brain, Mind and Consciousness Program, Toronto, Canada

 [https://en.wikipedia.org/wiki/Open\\_access](https://en.wikipedia.org/wiki/Open_access) Copyright information

## eLife Assessment

This **important** study utilizes behavioral data and computational modeling to show that spatial properties of visual attention affect human planning. The methodology and statistical analyses are **solid**, though the way attention is conceptualized and modeled could be refined. The findings of this study will interest cognitive scientists studying attention, perception, and decision-making.

<https://doi.org/10.7554/eLife.108034.1.sa3>

## Abstract

Human planning is efficient—it frugally deploys limited cognitive resources to accomplish difficult tasks—and flexible—adapting to novel problems and environments. Computational approaches suggest that people construct simplified mental representations of their environment, balancing the complexity of a task representation with its utility. These models imply a nested optimisation in which planning shapes perception, and perception shapes planning - but the perceptual and attentional mechanisms governing how this interaction unfolds remain unknown. Here, we harness virtual maze navigation to characterise how spatial attention controls which aspects of a task representation enter subjective awareness and are available for planning. We find that spatial proximity governs which aspects of a maze are available for planning, and that when task-relevant information follows natural (lateralised) contours of attention, people can more easily construct simplified and useful maze representations. This influence of attention varies considerably across individuals, explaining differences in people’s task representations and behaviour. Inspired by the ‘spotlight of attention’ analogy, we incorporate the effects of visuospatial attention into existing computational accounts of value-guided construal. Together, our work bridges computational perspectives on perception and decision-making to better understand how individuals represent their environments in aid of planning.

## Significance statement

Humans have an impressive ability to plan. Theoretical models in computer science propose that instead of using all the available information in a scene, a decision-maker should form a simplified mental representation of their environment over which they plan. However, little is known about how perceptual and attentional processes shape the planning process in humans. We find that people form simplified mental representations in line with the natural contours of spatial attention, whereby information limited to a visual hemifield is more readily available for planning, like a spotlight illuminating a part of the environment. We develop a novel computational model of the effects of attention on planning and characterise systematic variation between individuals in how they simplify their mental representations.

## Introduction

Humans have an impressive ability to plan. We are able to model the world, simulate potential outcomes, and select among possible courses of action. Take, for example, your first trip to London. You want to visit Buckingham Palace despite being jet-lagged. Looking at a map of the underground, you're overwhelmed with information but need to make a plan. How do you solve this problem? Even simple decisions like this involve many potential actions and outcomes, making it impossible to systematically evaluate every possible option, especially given limited cognitive resources<sup>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup>. Explaining how people plan efficiently and flexibly under these constraints is a long-standing challenge in human and machine intelligence<sup>5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup>.

Theories of human problem-solving conceptualize planning as a search through a ‘*decision tree*’ of all potential actions and their outcomes<sup>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup>. In our example, an individual may first list all possible tube stations within walking distance and then evaluate which action sequence will get them closer to their destination. Previous work proposes different algorithmic strategies for how an agent efficiently searches over a complex *decision tree*. These strategies include ignoring low-value actions (i.e., pruning)<sup>3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup>, limiting how far in the future one might search (i.e., depth)<sup>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup>, or relying on previously learnt strategies (i.e., habits)<sup>4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup>.

This previous work, however, largely assumes that a decision-maker has a fixed representation of the problem. When planning involves constructing and evaluating multiple multi-step trajectories within a *decision tree*, the computational burden increases with the complexity of the representation of the problem space. Consider planning in a two-dimensional spatial grid, for example. A fine-grained grid presents many choice points about which way to turn. A coarser-grained grid presents fewer choice points. Since the number of branches is a multiplicative function of the number of choice points, a simplified representation of the task space, if chosen appropriately, can have a profound effect on reducing the computational demands of planning.

One elegant approach to forming such a simplified representation is to adaptively select the granularity of information required to successfully complete the task<sup>18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup>, known as value-guided construal (VGC). Under VGC, a cognitively limited decision-maker selects a manageable subset of information over which to plan—i.e., a task representation—balancing utility and complexity<sup>18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100</sup>. In our example, the VGC algorithm predicts that an individual would plan over a few relevant tube lines rather than planning over all possible routes.

In previous work, Ho and colleagues discovered that people's awareness of, and memory for, obstacles in a maze varies in line with the predictions of a VGC model. The VGC model implies two nested optimisations - an outer loop of construal, and an inner loop that runs a plan conditional on a particular task representation. The VGC model is a normative model and remains agnostic as to the cognitive mechanisms controlling the construal. In particular, the perceptual and attentional mechanisms governing *how* information is selected to become part of a task representation remain unknown. Initiating such a nested computation plausibly rests on inductive biases - general principles that a perceptual system can apply to select task-relevant information, before refining it as part of the planning process<sup>19</sup>. Selective attention is proposed as one general mechanism by which the brain selects relevant information, either voluntarily (endogenous) or reflexively (exogenous)<sup>20–25</sup>.

Previous studies have demonstrated that attention guides the selection of particular features of the environment to support reinforcement learning<sup>26</sup>. However, it remains unknown whether and how attention shapes value-guided construal “on the fly” during planning. For instance, one possibility is that forming a simplified task representation is a “late” passive side-effect of the planning process - a tendency to focus on what we are thinking about. Alternatively, VGC may reflect an “early” selection of perceptual information, perhaps based on a rapid feedforward sweep of perceptual input<sup>27</sup>. These alternatives echo classic debates between early and late selection models of attention<sup>28,29</sup>, but now situated within the broader landscape of computational accounts of planning. More generally, despite the wealth of literature on attention, and pioneering efforts to incorporate attentional constraints into models of decision-making<sup>18,26</sup>, we lack a basic understanding of how attention influences planning.

To make progress on this question, we examined the role of visuospatial attention on how people construct simplified task representations across three experiments in human participants. We build on previous work using maze navigation to provide a rich readout of people's current task representations. We predicted that if visuospatial attention is guiding the formation of task representations, the construal process will be constrained by inductive biases characteristic of attentional selection. For instance, previous work has illustrated how attentional selection is biased by the spatial context in which information is presented: presenting distractors alongside task-relevant stimuli makes attentional selection more challenging<sup>30–32</sup>. Attention, in this case, spills over to the neighbouring stimuli. These findings align with the metaphorical attentional spotlight, which stipulates that the focus of visual attention can move around the visual field but is limited in spatial extent<sup>33,34</sup>. According to this model, individuals can, for example, orient their attention preferentially to a single hemifield—i.e., lateralizing—which is enabled by a hemispheric lateralization of alpha power over posterior cortex<sup>35–38</sup>.

We harness these classical signatures of attentional selection to characterise how attention shapes planning. First, we demonstrate “*attentional overspill*”: participants preferentially incorporate task-irrelevant information into their task representation when it is presented in spatial proximity to task-relevant information. Second, we observe that *attentional overspill* is reduced when task-relevant information is lateralised to a single hemifield, allowing participants to more effectively form optimal task representations. Finally, we extend the VGC model to incorporate visuospatial attention as a key psychological mechanism for constructing simplified task representations. Together, our findings furnish a computational account of how attention and perception guide simplified representations in the service of planning.

## Results

To examine the role of visuospatial attention in planning, we relied on a previously developed maze navigation paradigm in which participants solved 2-D mazes<sup>18</sup>, avoiding obstacles obstructing their path (**Figure 1a**, left panel). On every trial, participants reported their

awareness of specific obstacles (see Methods for details). The level of awareness attached to different obstacles provides a read-out of an individual's task representation while solving a particular maze. We first reanalyzed the data presented by Ho and colleagues (2022)<sup>18</sup> to examine the role of spatial attention in building task representations (datasets Ho 1 and 2). In a new experiment (dataset dSC 1), we designed novel mazes to test the effects of lateralization of attention in enabling efficient planning (see Methods & **Table S1**).

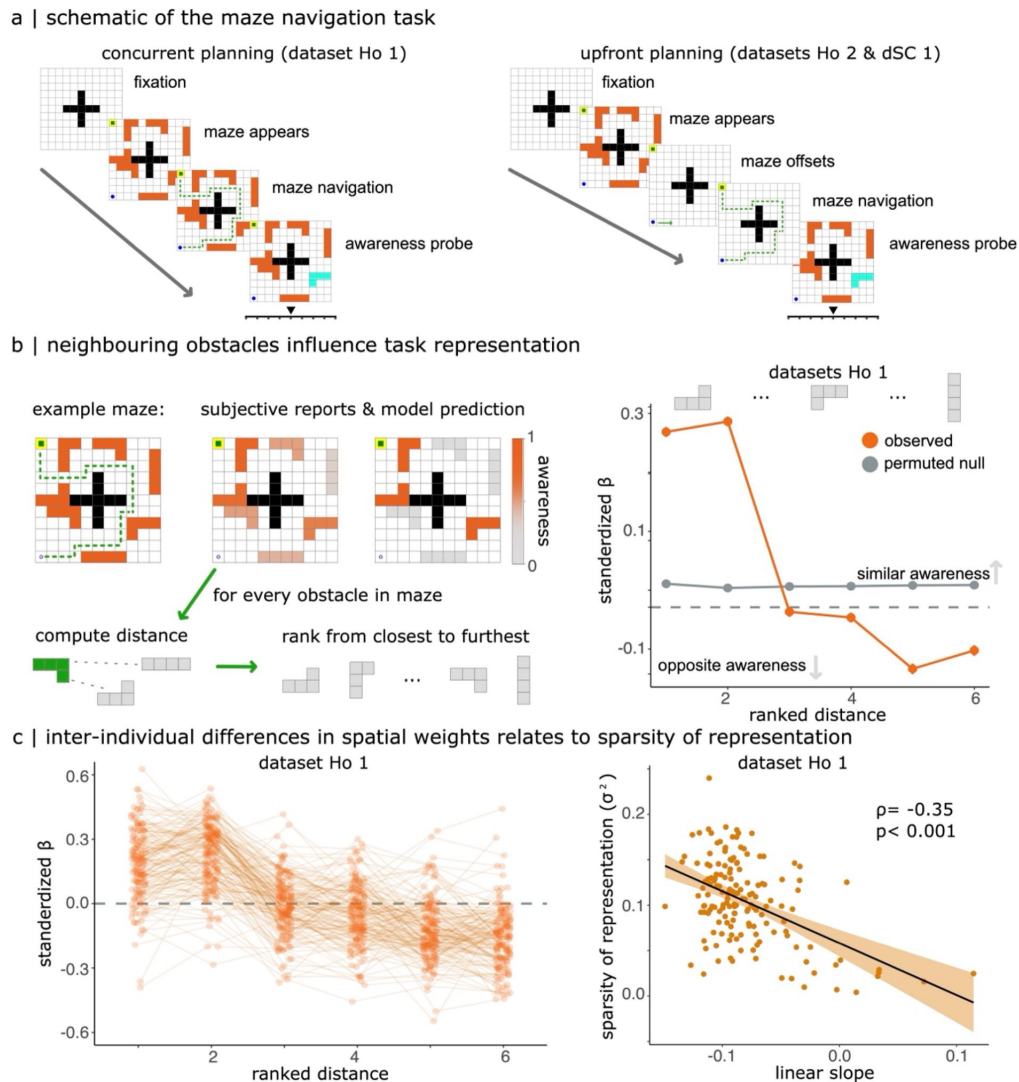
## A spotlight of attention influences task representations

We hypothesized that spatial attention would control which items are included in a task representation<sup>30–32</sup>. Specifically, we hypothesized that participants would deviate from the predictions of the VGC model and become distracted by task-irrelevant obstacles when they are presented in spatial proximity to task-relevant obstacles. To evaluate these predictions, we first computed the distance between a probe obstacle and every other obstacle in the maze. Second, we ranked the obstacles from the closest to the furthest from the probed item. Using the ranked obstacles, we trained a linear regression model to predict participants' awareness of the probed obstacle (in green) from their awareness of the remaining obstacles (in grey; **Figure 1b**).

Critically, we observed a significant effect of spatial context on task representations - an effect which is not predicted by the normative VGC model. Participants' awareness of a particular obstacle was positively predicted by the awareness of its close neighbours ( $\beta_1 = 0.26$ , SE = 0.01, 95% CI [0.25, 0.28];  $\beta_2 = 0.29$ , SE = 0.01, 95% CI [0.27, 0.30]), whereas awareness of its furthest neighbours negatively predicted participant reports ( $\beta_5 = -0.13$ , SE = 0.01, 95% CI [-0.15, -0.12];  $\beta_6 = -0.13$ , SE = 0.01, 95% CI [-0.15, -0.12]; see **Table S2**). In other words, the spatial context of an obstacle predicted whether it would be included in a simplified task representation – akin to a diffuse attentional spotlight which filters which aspects of the maze are available for planning. This effect remained significant for both task-relevant and task-irrelevant obstacles, and after controlling for the predictions of the VGC model (**Figure S7** & **Table S3**, respectively). We observed the same effect in a separate experiment where participants planned their route upfront before navigating the mazes (i.e., dataset Ho 2, see **Table S4** & **S5**). Finally, we replicated this pattern of results in our in-person experiment: closest neighbours positively predicted the awareness of an obstacle ( $\beta_1 = 0.19$ , SE = 0.007, 95% CI [0.18, 0.21]), whereas furthest neighbours negatively predicted participants' reports ( $\beta_3 = -0.10$ , SE = 0.01, 95% CI [-0.11, -0.08];  $\beta_4 = -0.26$ , SE = 0.007, 95% CI [-0.27, -0.25];  $\beta_5 = -0.29$ , SE = 0.007, 95% CI [-0.30, -0.27]; see **Table S6** & **S7** and **Figure S11**).

Next, we explored whether the influence of neighbouring obstacles on task representations varied across individuals. To do so, we fit the regression model described above to quantify each participant's attentional spillover, and quantified the linear slope of the resulting beta coefficients. Negative slopes indicate a significant effect of attentional spillover on task representation. The influence of attention varied considerably across participants: while on average, participants' task representations were influenced by attention (mean effect = -0.08; s.d. = 0.04), a subset of participants showed minimal influence of attention on their task representation (i.e., flat slopes; **Figure 1c**).

We hypothesized that participants with the largest attention effects (i.e., most negative slopes) would also show sparser task representation (i.e., a “spotlight of attention” which is focused only on a subset of obstacles). To test this, we computed the sparsity of participants' task representations by estimating the variance of their awareness reports, with higher variance indexing those participants who report being very aware of some obstacles and unaware of others. In line with our hypothesis, we observed that participants who were most influenced by neighbouring obstacles also showed sparser task representations (dataset Ho 1:  $\rho = -0.35$ ,  $p < 0.001$ ; dataset Ho 2:  $\rho = -0.49$ ,  $p < 0.001$ ; dataset dSC 1:  $\rho = -0.51$ ,  $p < 0.01$ ; see **Figure S11**). To address concerns of overfitting, we tested whether the spatial attention effects observed in a lateralized set



**Fig. 1**

### Spatial attention shapes task representations.

(a) Schematic of the maze navigation task. Participants fixated at the start of each trial, after which a maze was presented, which they were asked to navigate. Maze stimuli either remained on the screen during navigation (left panel; *concurrent planning experiments*) or were removed before navigation (right panel; *upfront planning experiments*). Once participants finished navigating the maze, they were asked to report their awareness of each obstacle in the maze.

(b) Left panel: schematic of the analysis pipeline. An example maze is shown where seven obstacles (plotted in orange) are presented on every trial according to pre-defined mazes. Participants report their awareness of every obstacle at the end of each trial (middle maze). The VGC model predicts which obstacles in a maze will likely be included in participants' task representation (right maze). We use participants' awareness reports to test the influence of neighbouring obstacles on the probe obstacle (presented in green). We compute the influence of neighbouring obstacles (in grey) on participants' awareness of the probed obstacle (in green). Right panel: Results of the ranked regression model for dataset Ho 1. We observed that obstacles closest to the probed item (rank 1 & 2) positively impact awareness reports. In contrast, obstacles furthest from the probed item negatively impact awareness reports (rank 5 & 6).

(b) Left panel: The effect of neighbouring obstacles on task representations varied across participants (each represented by a point). Right Panel: Inter-individual differences in the attentional effects correlate with the sparsity of participants' representations. Participants who showed the greatest influence of neighbouring obstacles (more negative slopes), showed the simplest representations (greatest variance in awareness reports).

of mazes generalized to task representations of non-lateralized mazes and vice versa (dataset dSC 1). We observed that inter-individual differences in spatial attention effects in one condition predicted the sparsity of task representations in the other ( $\rho = -0.48$ ,  $p < 0.01$ ;  $\rho = -0.42$ ,  $p < 0.05$ ).

## Attentional limits constrain the optimality of task representations

Prior psychological research indicates that attention can be efficiently allocated to a “hemifield” of visual space - with information being preferentially processed when presented in the attended hemifield<sup>39,41</sup>. Building on this work, we hypothesized that participants would select task-relevant information with greater ease - constructing task representations more closely aligned with the VGC model - when task-relevant information is spatially confined to a visual hemifield (i.e., presented unilaterally).

To test this hypothesis, we derived a measure of task-relevant lateralization inspired by the attention literature<sup>35,42,43</sup> (**Figure 2a**). Specifically, we separated maze stimuli across the vertical meridian and computed the ratio of task-relevant information presented on the left versus right side. For example, the maze shown in **Figure 2a** has twice the amount of task-relevant information presented in the left hemifield than in the right (lat. Index = 1/3). A lateralization index of 0.0 indicates that both hemifields contain equal amounts of task-relevant information (i.e., non-lateralized). We used this task-relevance lateralization index as a moderator in a hierarchical linear regression model to test whether participants’ awareness reports were better predicted by the original VGC model in mazes showing the greatest lateralization of task-relevant information.

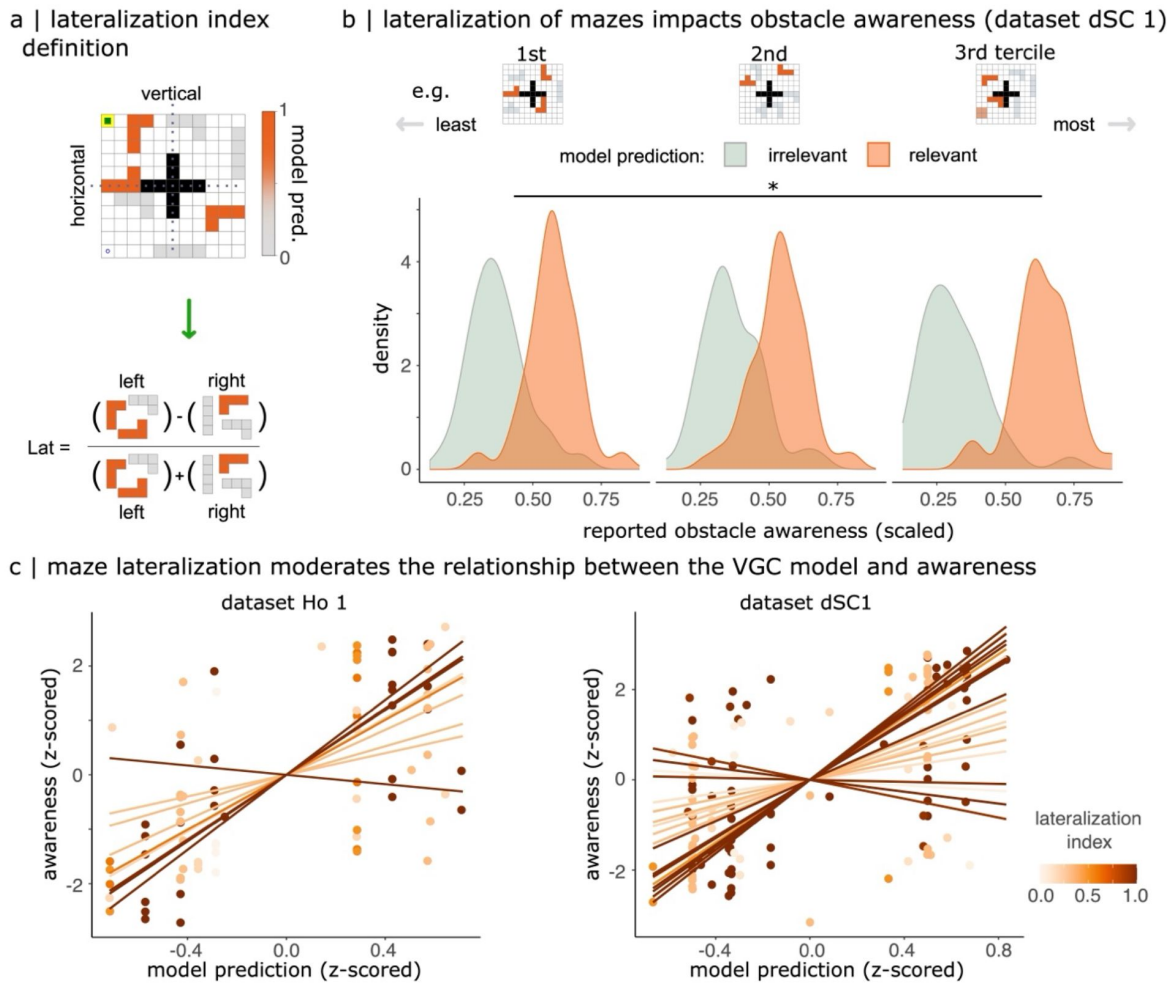
In line with our hypothesis, we observed a significant moderation effect whereby the greater the lateralization of task-relevant information across the vertical meridian, the better the original VGC model was at predicting participants’ awareness reports ( $\beta_{\text{interaction}} = 0.01$ ,  $\text{SE} = 2.65 \times 10^{-3}$ , 95% CI [0.01, 0.02],  $\rho_{\text{FDR}} < 0.001$ ; **Figure 2c** left panel & **Table S8**). We replicated these findings with the data collected in dataset Ho 2<sup>18</sup> ( $\rho_{\text{FDR}} < 0.01$ ; see **Table S9**). These results indicate that participants’ task representations are more closely aligned with the ideal observer (i.e., the original-VGC model) when task-relevant information is presented unilaterally.

In our new dataset (dSC 1), we designed novel maze stimuli to validate these lateralised effects of attention while addressing some limitations of previous experiments (see Methods). We again observed that lateralization of task-relevant information impacted participants’ awareness reports. Participants were less aware of task-irrelevant stimuli on trials where the lateralization of task-relevant information was larger (**Figure 2b**) and we replicated the moderation effect of information lateralization on the extent to which the original VGC model captured participants’ awareness reports ( $\beta_{\text{interaction}} = 0.01$ ,  $\text{SE} = 2.65 \times 10^{-3}$ , 95% CI [0.01, 0.02],  $p < 0.001$ ; **Figure 2c** & **Table S10**).

In contrast with our observations of consistent and strong attentional effects relative to the vertical meridian, effects relative to the horizontal meridian (superior vs. inferior) were inconsistent across experiments. Specifically, we observed a significant moderation effect in dataset Ho 2 ( $\alpha_{\text{interaction}} = 0.01$ ,  $\text{SE} = 2.85 \times 10^{-3}$ , 95% CI [0.00, 0.01],  $p^{\text{FDR}} < 0.05$ ; see **Table S9**), but not in dataset Ho 1, and the moderation effect was negative rather than positive in dataset dSC 1 ( $\beta_{\text{interaction}} = -0.01$ ,  $\text{SE} = 2.22 \times 10^{-3}$ , 95% CI [-0.01, 0.00],  $p < 0.05$ ). Both of these effects became insignificant after accounting for nuisance covariates (see **Table S14** & **S15**).

## Attentional spotlight model of task representations

Taken together, our results corroborate a critical role for visuospatial attention in constructing task representations. Notably, these filtering effects of attention on value-guided construal are not currently part of the original VGC framework proposed by Ho and colleagues. In what follows we explicitly incorporate the influence of a spotlight of attention into the original VGC model to formulate the spotlight-VGC model<sup>39,40</sup>.



**Fig. 2**

### Lateralization of task-relevant information affects task representations.

(a) For each maze, we computed a horizontal meridian and a vertical meridian lateralization index. This index reflects whether task-relevant information is lateralized to a hemifield. In the example plotted, there is more task-relevant information presented on the left than on the right of the maze, therefore this would correspond to a moderate level of vertical meridian (i.e., left vs right) lateralization.

(b) Density plots of the reported awareness of obstacles on the basis of whether the value-guided construal (VGC) model predicted them to be task-relevant ( $\geq 0.5$ ; in orange) or task-irrelevant ( $< 0.5$ ; in grey). Participants were more likely to be aware of obstacles predicted as task-relevant. We split maze stimuli based into terciles based on the degree to which task-relevant information was presented preferentially to one hemifield (x-axis). The leftmost plots are mazes where task-relevant information is presented on both hemifields. In contrast, the rightmost plot depicts mazes with the largest lateralization. We observed that the awareness reports of participants become increasingly aligned to the VGC model's predictions as lateralization increases.

(c) Scatter plot of the effect of maze lateralization on the relationship between the value-guided model and participants' awareness of obstacles. We observed a significant vertical meridian lateralization effect whereby participants' awareness reports were more strongly aligned with the VGC model's predictions when task-relevant information was presented unilaterally in all datasets.

To achieve this, we computed the predictions of the existing VGC model for each obstacle's task relevance in a given maze, and averaged these predictions within an attentional spotlight of 3 squares (**Figure 3a** & **S8**, see Methods for details). We depict the effects of this spatial spotlight in **Figure 3a**: task-irrelevant stimuli (plotted in grey; see middle left obstacle) neighbouring task-relevant obstacles (plotted in orange) become more task-relevant, whereas task-relevant information becomes less relevant when surrounded by task-irrelevant information (see bottom right orange obstacle). We hypothesized that this spotlight-VGC model would predict participants' reports better than the original VGC model, which does not account for spatial attention.

In line with this hypothesis, we observed that the spotlight-VGC model predicted participants' awareness reports better than the original VGC model in all three datasets (dataset Ho 1:  $\Delta\text{BIC} = 84.63$ ; Ho 2:  $\Delta\text{BIC} = 203.43$ ; dSC 1:  $\Delta\text{BIC} = 70.72$ ; see **Figure 3b** right panel). For dataset dSC 1, we observed a significant improvement in model fit for non-lateralized maze stimuli ( $\Delta\text{BIC} = 161.93$ ) but failed to find any improvement when maze stimuli were lateralized ( $\Delta\text{BIC} = -42.02$ ; see **Figure 3c**). These findings dovetail with the previously discussed moderation effects, and suggest that the spotlight-VGC model is particularly useful in improving our ability to explain human behaviour in situations when attentional filtering is more complex.

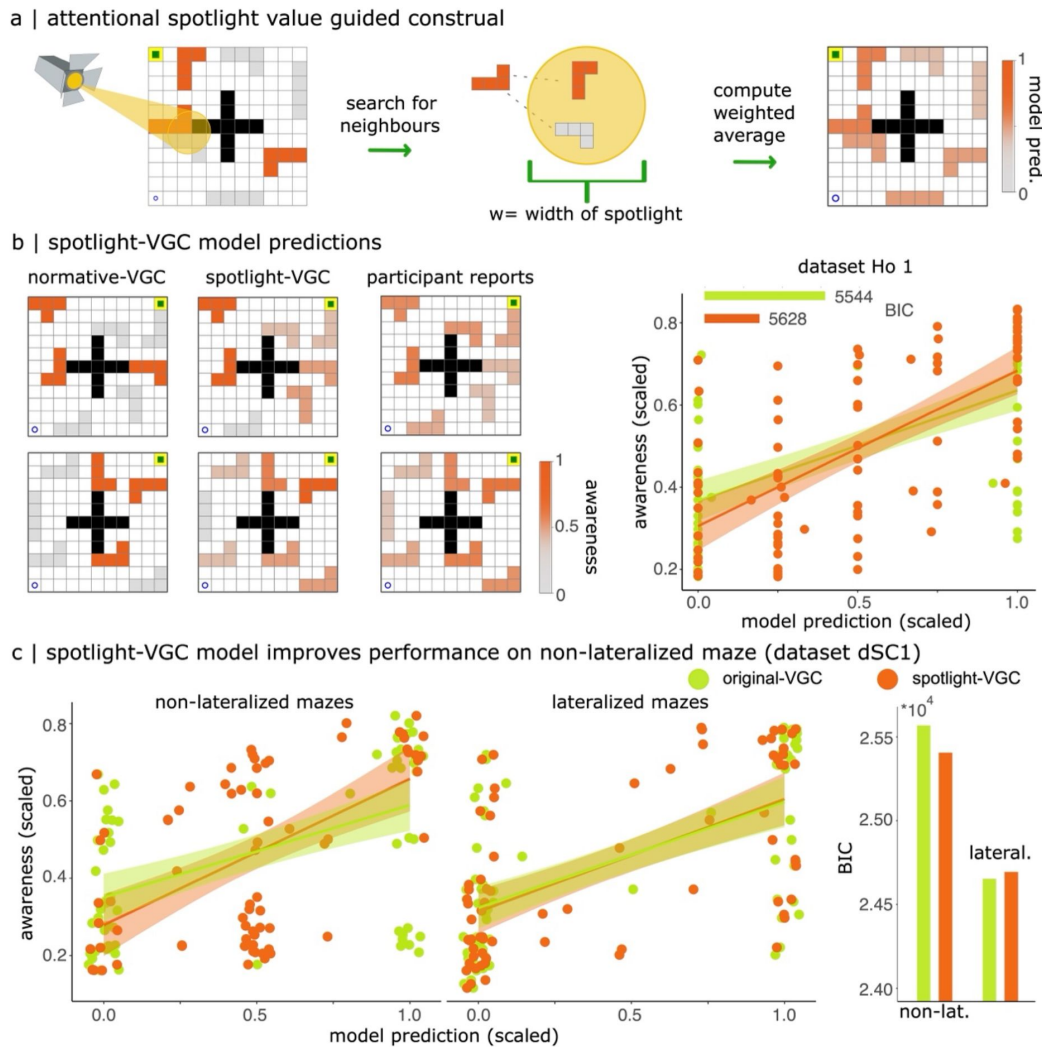
## Sensitivity analyses

We conducted a series of control analyses to verify the robustness of our experimental results. First, we verified that the spatial proximity effect (**Figure 1b**) was not driven solely by the spatial smoothness of participants' awareness reports by conducting null permutation tests (grey line, **Figure 1b**). For each maze stimulus, we permuted the rank of the neighbouring obstacles. We then fit the same linear model to assess the effect of spatial context on task representations. This procedure was repeated 1000 times to generate a null distribution of beta coefficients. The resulting null distribution showed no discernible effect of spatial context. Second, we used null permutation tests to verify that the improved fit of the spotlight-VGC model was not driven by greater spatial smoothness of the model predictions (see **Supplemental Figure 12**). Third, we assessed whether nuisance covariates could explain the moderation effects we observed. Specifically, we added the distance from the goal, starting location, center walls, and fixation as nuisance covariates in our hierarchical regression models. Maze lateralization remained a significant moderator of the relationship between the original VGC model and participants' awareness reports after controlling for these covariates (see **Table S11–13**). This was not the case, however, for lateralization effects along the horizontal meridian (see **Table S14–15**).

Finally, we sought to verify that the lateralization effects we observed were not driven by a change in eye movement patterns. For dataset dSC 1, we continuously tracked the position of participants' gaze. We explicitly instructed participants to maintain central fixation while planning (see Methods for details) and removed the obstacles from the screen after 6 seconds. This allowed us to verify that greater awareness of obstacles was not driven by longer fixation times. We confirmed that participants maintained central fixation on both lateralized and non-lateralized maze stimuli in most trials (see **Figure S13**). Excluding trials where participants exhibited excessive eye movements during planning, we continued to observe qualitatively similar lateralization effects (see **Figure S14** and **Table S16**).

## Discussion

Searching for a solution in a complex multi-step task is challenging. Recent computational work suggests that humans overcome this challenge by constructing simplified perceptual representations of their environment. In the present study, we reveal a role for visuospatial attention in constructing these simplified perceptual representations.



**Fig. 3**

### A VGC model augmented with an attentional spotlight model predicts participants' task representations.

(a) Schematic of the attentional spotlight model. Inspired by the spotlight of attention analogy, we recompute an obstacle's probability of being included in a task representation as the weighted average of its neighbours. We first search for all neighbours of obstacle<sub>*i*</sub> that are  $w$  squares away. We then compute  $P(\text{Obstacle}_i)$  as the weighted average of obstacle<sub>*i*</sub> and its neighbours. This generates more graded model predictions (far right panel).

(b) Left panel: Each row represents a different example maze stimulus. The left column depicts the original VGC model prediction  $P(\text{Obstacle}_i)$  for every obstacle in the example maze. The middle column shows the attentional-spotlight model prediction for every obstacle. Obstacles that were considered task-relevant (deep orange) in the original model become less important when surrounded by task-irrelevant information (grey obstacles). The right column shows the participants' average awareness of each obstacle in the example mazes. Right panel: Scatter plot of the linear relationship between participants' awareness reports of obstacles and model predictions (original-VGC in green and the spotlight-VGC model in orange) for dataset Ho 1. The latter fits participants' reports better than the original VGC model.

(c) Scatter plot of the linear relationship between participants' awareness reports of obstacles and model predictions (original VGC in green and the spotlight-VGC model in orange) for dataset dSC 1 separately for non-lateralized (left panel) and lateralized mazes (right panel). Although both models fit participants' awareness reports better for lateralized mazes, the advantage of the spotlight model over the original model (better model fit / lower BIC) was observed only in non-lateralized mazes.

## Participants' task-representations are informed by visuospatial attention

We provide several lines of evidence for the critical role of visuospatial attention in constructing task representations. First, we observed a significant effect of the spatial context in which information is presented. Participants were less likely to incorporate task-relevant information into their construal when it was surrounded by task-irrelevant information. These effects mirror perceptual crowding effects<sup>30,32</sup> which reveal that attention spills over to distractors presented alongside task-relevant stimuli when presented in close spatial proximity. Second, we observed that participants incorporated task-relevant information into their task representations more frequently when relevant obstacles were grouped together within the same hemifield (**Figure 2**). Participants' task representations in such settings were more closely aligned with an ideal observer model - suggesting that the natural contours of visuospatial attention interact with the capacity of observers to form efficient task representations.

We also observed significant inter-individual differences in attentional effects across participants (**Figure 1c**). While some participants were strongly influenced by the spatial context of neighbouring stimuli, others showed more limited evidence for an attentional effect (**Figure 1b**). Inter-individual differences in attention predicted the sparsity of participants' simplified representations: participants with larger attention effects exhibited sparser representations. Future research could explore how these individual differences constrain performance on other tasks that require planning and search in high-dimensional spaces.

## Incorporating attention into a model of value-guided construal

The VGC model articulates how an ideal decision-maker should represent an environment while balancing complexity and utility<sup>18</sup>. We developed an extension of this model that accounts for the effects of spatial attention on planning. Our model, inspired by the analogy of a spotlight of attention<sup>23,39</sup>, provides a better fit to participants' awareness reports than the original VGC model (**Figure 3**). This improved model fit was most evident for mazes where task-relevant information was presented to both hemifields (**Figure 3c**), suggesting the augmented model is helpful in explaining behaviour in contexts where attentional selection is more complex. These deviations from the original VGC model, therefore provide a useful benchmark to compare human performance and offer insights into natural constraints on human cognition. For instance, we demonstrate that spatial context biases whether information is to be included or excluded from a representation of the environment. These effects may reflect inductive biases in humans who have learned and evolved to select information from real-world environments where obstacles tend to be grouped together in visual scenes<sup>44,45</sup>. However, it is plausible that these inductive biases on value-guided construal may themselves be learnt, and vary according to other environmental demands and contexts which impose systematic regularities on useful task representations (e.g., attending preferentially to intersections when planning on the Tube). Future research can explore the flexibility of participants' task representations across environmental contexts, and ask how these inductive biases are acquired.

## Planning and Consciousness

Our experiments investigate the connection between planning and participants' reports of their awareness of features of the task environment. The results may therefore be relevant to understanding the functions of conscious experience. While an intimate connection between attention and consciousness is widely recognized<sup>46-49</sup>, there is less work explicitly considering the connection between planning and consciousness<sup>50,51</sup>. However, there are several reasons why the kind of planning at work in our experiments is likely to require the task to be represented consciously.

As we mentioned at the outset, simplified representations reduce the computational burden of planning in a branching multi-step task space. The same consideration suggests that planning should be based on conscious, rather than unconscious, representations<sup>52</sup>. Initial stages of perceptual processing can carry information about a range of different and incompatible possibilities at once, for example, a probability distribution across a range of possible orientations of a line. The probabilistic representation attaches some probability (or probability density) to many different possibilities. There is, of course, a certain burden in integrating and weighing probabilistic information of this kind, for which the brain is thought to deploy various solutions (e.g. approximate Bayesian inference). These initial stages of perceptual inference are typically thought to be unconscious. However, forward planning from multiple possibilities in a branching task space rapidly becomes intractable as the combinatorial possibilities explode. Consciousness, by contrast, provides a much sharper representation of the current state, from which planning can proceed forward<sup>53,54</sup>.

Given the computational cost of running through and comparing many potential multi-step action sequences, it makes sense to base that process on a reliable estimate of the current world state. While it is doubtless useful to produce some kinds of unlearned and habitual action very rapidly at the first hint of information, for example of the presence of a predator, with multi-step forward planning it makes sense to integrate information from more sensory modalities and across a longer timescale before then committing to using a representation as the basis for planning. This again suggests that conscious representations, which are known to integrate information across modalities and time<sup>55–59</sup>, are perfectly suited to the functional needs of this kind of planning task. Furthermore, planning depends on both facts and values. Potential actions are assessed based on the expected value of outcomes. The role of value was captured, in our studies, by an extension of the VGC (value-guided construal) model. Consciousness is thought to facilitate the integration of different sources of value<sup>60–62</sup>.

The task and associated computational model thus offer a flexible tool for characterising the computations by which conscious representations influence decisions and actions. Future work could tell us more about the way bottom-up attention-driven inputs and taskbased value jointly influence what information reaches conscious experience. This provides a novel ecologically-valid probe of the connections between attention, consciousness, and decision-making that does not require the explicit labelling of taskrelevant stimuli. Neural (e.g. M/EEG) data collected while participants plan could help understand the timescale and computational steps that lead to the formation of a conscious task representation. Modifications of the paradigm would also be suited to exposing the way non-consciously-presented cues do and do not influence the way participants plan.

## Methodological Considerations and Future Directions

The task we used requires participants to report their experience of each maze. Yet, retrospective reports are one step removed from perception and do not shed light on the time course of how perceptual representations are formed. In line with this limitation, the extended spotlight-VGC model accounts for visuospatial attention effects after a simplified representation has been computed (i.e., at the end of the construal process). This allowed us to evaluate the fit of the model to empirical data, but does not provide insight into how attention affects the construal process itself. Future research examining the neural dynamics of task representations will be critical to disentangling the time course of how attention and planning considerations interact in the formation of task representations. It will also be necessary to elaborate on how bottom-up and top-down aspects of attentional selection are combined to guide complex behaviours. Finally, open questions remain about the nature and form of visuospatial attention and how this intersects with planning. For instance, can multiple spatial locations be preferentially selected at once—i.e., are there multiple spotlights<sup>63–66</sup>? There is also discourse on how spatial attention may move from one location to another: are the intervening visual regions between attended locations similarly selected<sup>63,67–69</sup>? Our findings tentatively suggest that individuals are able to

attend to disparate spatial regions to form sparse task representations, yet there is substantial variability in how individuals orient their attention during the task. The present paradigm and computational modelling, in conjunction with carefully designed stimuli, may help resolve these outstanding questions.

While we observed clear lateralization effects along the vertical meridian (i.e., left vs right hemifield), effects along the horizontal meridian were less clear (see **Table S14** [↗](#)–**15** [↗](#)). A combination of factors may explain this finding. First and foremost, the maze stimuli were not designed to test horizontal-meridian lateralization effects, possibly leading to a lack of power to detect these effects. Second, prior research suggests that distractors produce a larger crowding effect when presented across the horizontal compared to the vertical meridian<sup>30</sup> [↗](#). The retinotopic organization of the cortex is believed to drive this effect: spatially adjacent stimuli can be retinotopically distant if presented on the opposite side of the vertical meridian, which is thought to facilitate distractor inhibition. A more detailed experiment explicitly designed to test these effects may clarify the interpretation of these null results.

## Conclusions

Complex daily decisions require a decision-maker to arbitrate over countless potential multi-step actions and their outcomes, making searching for a solution difficult. We shed light on how this is achieved by clarifying the role of visuospatial attention in forming simplified perceptual representations to aid in planning. We build on previous work on the effect of task relevance and develop a computational model which explicitly incorporates the role of attention in value-guided construal. Our model bridges the literature on perception, attention and computational models of planning to provide a more complete computational account of human cognition. We believe the results of this paper can inform future research on a comprehensive theory of human cognition and inspire novel biologically-informed intelligent algorithms.

## Methods

### Experimental Task

To test our hypotheses we relied on a previously established mazenavigation task where participants are asked to move a circle avatar from a starting location to a goal using the arrow keys<sup>18</sup> [↗](#). Each maze consisted of an 11x11 grid with blue obstacles (7 obstacles in datasets Ho 1 & 2, and 6 obstacles in dataset dSC 1), and black central walls arranged in the shape of a fixation cross. Each trial began with a fixation cross (center walls), after which participants were prompted to navigate to the goal. Experiments differed in terms of i) the mazes participants navigated, ii) whether the obstacles were presented before or during the execution of the plan, and iii) what the participant reported.

We reanalyzed the data of Ho and colleagues' experiments 1 and 2 for the present study. In experiment 1 (i.e., dataset Ho 1), participants were presented with the obstacles throughout the trial. At the end of each trial, participants were asked to rate "How aware of the highlighted obstacle were you at any point?" using a nine-point scale. In experiment 2 (i.e., dataset Ho 2), participants were similarly asked to rate their awareness of the various obstacles but were required to plan their solution before they began to solve the maze. See <sup>18</sup> [↗](#) for details concerning the experimental procedures.

We did not reanalyze the results of the fourth experiment by Ho and colleagues. In this experiment, participants were not presented with all the information (i.e., obstacles) at once to solve the maze. Instead, they discovered obstacles by hovering over them with a cursor.

To further test the effects of attention on task representations, we designed a novel set of maze stimuli. This consisted of 12 mazes with task-relevant obstacles lateralized to a hemifield (left or right) and 12 non-lateralized stimuli. Each maze consisted of six obstacles, three on each hemifield, none of which crossed the veridical meridian. This ensured that there were an equal number of obstacles for computing the lateralization index (see below). Maze stimuli of both sets were equated on several nuisance covariates (see Supplemental **Table S1**). For dataset dSC 1, participants solved each of these 24 mazes four times (i.e., all possible orientations of the maze).

The design of the in-person experiment (i.e., dataset dSC 1) closely followed the second experiment of Ho and colleagues<sup>18</sup>. On every trial, participants were presented with a maze stimulus for 6 seconds, over which they were required to plan. The maze stimulus was offset, and participants were required to solve the maze after a one-second delay. On every trial, participants reported on their task representations using a nine-point awareness scale.

## Participants

For datasets Ho 1 & 2, participants completed the task online on Prolific. In dataset Ho 1, 194 participants completed submissions, 161 of whom were included in the final sample after exclusions. In dataset Ho 2, 188 participants completed submissions, 162 of whom were included.

For dataset dSC 1, 35 participants (mean age = 23.14, SD = 5.35; 12 male) completed an in-person eye-tracking experiment (see 'Eye-tracking acquisition'). None of the participants were excluded from the data analysis. We excluded trials where participants' reaction times were longer than 20 seconds, or where participants deviated more than nine moves from the optimal path (which reflected 3SD above the mean).

## Ethics

All procedures were approved by the University College London ethics committee and adhered to the Declaration of Helsinki. Informed consent was obtained from each participant prior to each experiment.

## VGC model

We fit the previously described VGC model to our maze stimuli<sup>18</sup>. Briefly, this model computes the optimal simplified task representation such that it maximizes the utility of the representation while also minimizing the cognitive cost of keeping information in mind. This model assumes that a decision-maker combines a subset of cause-effect relationships to represent their environment in aid of planning. For every possible construal, the model computes the value of a representation:

$$\text{VOR}(c) = U(\pi_c) - C(c).$$

where  $U(\pi_c)$  is the utility of a construed plan  $\pi_c$ , and  $C(c)$  represents the cost of keeping that information in mind.

A task representation is selected according to a SoftMax decision rule. We then compute a marginalized probability for each obstacle being included within a construal,

$$P(\text{Obstacle}_i) = \sum ||\varphi_{\text{Obstacle}_i \in c}|| P(c),$$

where  $\phi_{\text{Obstacle}_i}$  is the cause-effect relationship for obstacle<sub>i</sub>,  $P(c)$  is the probability that the task representation is selected, and  $\|X\|$  is a statement which evaluates to 1 if  $X$  is true, and 0 when  $X$  is false. We use the values of  $P(\text{Obstacle}_j)$  for every obstacle in a maze to predict participants' awareness reports. See [18](#) for a detailed explanation of the computational model.

We focused our analyses on the *static* version of the VGC model (i.e., sVGC), whereby task representations are assumed to remain stable across planning. Our choice was informed by the design of the experiment where participants were required to plan over all obstacles at once.

## Spatial proximity effects

To examine how the spatial context of information influences participants' awareness reports, we ran a hierarchical linear regression model. First, for every obstacle in every maze, we rank-ordered all other obstacles based on spatial proximity. That is, the participant's awareness report of the closest item to obstacle<sub>i</sub> on the trial was used as a predictor of the participant's report of obstacle<sub>i</sub> in a hierarchical linear regression model. This yielded a regression model with 6 regression coefficients predicting participants' awareness reports based on spatial proximity:

$$\begin{aligned} \text{report}_{\text{obstacle } i} &= \beta_0 + \beta_1 * \text{report}_{\text{obstacle } 1} + \beta_2 * \text{report}_{\text{obstacle } 2} \dots + \beta_6 * \text{report}_{\text{obstacle } 6} \\ &+ (1 | \text{MazeID}) + (1 | \text{ParticipantID}) + \epsilon \end{aligned}$$

where  $(1 | \text{MazeID})$  and  $(1 | \text{ParticipantID})$  are random intercepts of each maze and participant, respectively, and  $\beta_1$  reflects the contribution of the closest obstacle to obstacle<sub>i</sub>. We interpret any significant effects in this model as the influence of neighbouring stimuli on participants' representations. We also fit the above hierarchical linear regression model for each participant separately. We report these individual beta coefficients in [Figure 1b](#).

To ensure that the above spatial proximity effects were not driven by the VGC model predictions, we regressed out the effects of VGC model predictions from participants' awareness reports, and used the residuals of the model as the dependent variable in a second regression where we similarly predicted the effects of neighbouring stimuli on representations.

We verified that these effects were not explained by the spatial smoothness of our data by conducting 1000 spatial null permutations. For every iteration, we permuted the mapping between each obstacle in a maze and their spatial location maintaining the number of neighbouring obstacles for every trial. We fit a hierarchical linear regression model using this permuted data and built a distribution of null beta coefficients to compare to our observed effects.

## The sparsity of task representations

We sought to test the relationship between i) interindividual differences in attention effects and ii) the sparsity of task representations. First, we estimated the magnitude of each person's attention effect by fitting a linear slope to the beta coefficients obtained (see Spatial proximity effects). A participant with a large negative slope, therefore, showed a larger effect of neighbouring obstacles on their representation. Second, we operationalized the sparsity of participants' simplified representation as the variance of their awareness reports. A participant with a sparse representation shows a high variance in their awareness of different obstacles in a given maze. Last, we tested the linear monotonic relationship between the sparsity of participants' representations and the attention effects using Spearman correlation.

## Lateralization index

To test the effects of lateralization of task-relevant stimuli on participants' awareness reports, we developed a lateralized index of task-relevance inspired by the alpha-power attention literature<sup>35,42,43</sup>. We divided each maze into a right and left hemifield and computed the ratio of task-relevant obstacles on both sides:

$$Lat.index = \frac{\sum sVGC_{left} - \sum sVGC_{right}}{\sum sVGC_{left} + \sum sVGC_{right}}$$

where sVGC is the model's prediction of each obstacle task-relevance for that maze. Note obstacles only with a majority of its blocks within a single hemifield were considered (3 or more squares). This yielded an index of task-relevance lateralization for each maze stimulus. We repeated the above procedure to obtain an index of task-relevance lateralization for the horizontal meridian (superior vs inferior hemifield).

We tested whether the lateralization index moderated the relationship between the value-guided model predictions (sVGC) and participants' awareness reports using a hierarchical linear regression model.

$$report = \beta_0 + \beta_1 * sVGC + \beta_2 * Lat + \beta_3 * Lat * VGC + (1 | Maze_{ID} + Participant_{ID})$$

where  $\beta_3$  represents the interaction between the VGC model predictions and the lateralization index.

## Spotlight-VGC model

Inspired by previous literature comparing visuospatial attention to a spotlight that moves across the visual field, we developed an extension of the VGC model to account for the effects of attentional selection in forming task representations.

To do this, we recomputed the  $P(Obstacle_i)$  as a weighted average of its neighbours. We computed the distance between every obstacle in the maze, and searched for obstacles with neighbours within 3 squares (Manhattan distance) away from obstacle<sub>i</sub>. The distance of 3 squares reflects the 'width' of the attentional spotlight, and was chosen based on the median distance between neighbouring obstacles from our ranked analyses (**Figure 1b** & **Figure S2**). Neighbouring obstacles that fell within the attention spotlight were averaged as follows:

$$P(Obstacle_i) = \frac{P(Obstacle_i) + mean(P(Obstacle_n))}{2}$$

where n is the number of obstacles that fall within the width of the attentional *spotlight* (i.e., neighbouring items). We repeat this procedure for all obstacles within each maze. If an obstacle did not have any neighbours, then the value of  $P(Obstacle_i)$  remained identical to the value of original VGC model.

We used the outputs of the attention spotlight model in a hierarchical linear regression to predict participants' awareness reports, where we included participant and maze random intercepts:

$$report = \beta_0 + \beta_1 * At.Sp.Model + (1 | Maze_{ID} + Participant_{ID})$$

All linear regression models were fit with the *lmer* package in R.

## Null permutations

To ensure that the improved model fit of the attentional spotlight model was not driven by the spatial smoothness of our data, we conducted a series of control analyses where we permuted the model predictions within mazes.

To do so, we re-assigned the  $P(\text{Obstacle}_i)$  of each obstacle in a given maze to a random item such that each obstacle was given a new model prediction. This permutation procedure maintains the distribution of  $P(\text{Obstacle}_i)$  across obstacles for each maze, while randomizing the location of task-relevant information. We repeated this procedure for each maze separately. We then used these random model predictions to predict participants' reports using the same hierarchical linear model described in [Spotlight-VGC model](#). We repeated this procedure 1000 times to generate a null distribution of beta coefficients. We compared the observed beta value for the spotlight-VGC model against this distribution. We note that averaging neighbouring obstacles before or after the permutation of the model predictions qualitatively yielded the same result.

## Eye-tracking acquisition

For dataset dSC1, participants completed the computer task while their eye-position and pupil size were monitored using an EyeLink 1000 Plus eye tracker at 1000Hz (SR Research, Osgoode, ON). Participants were seated comfortably in a dimly lit room in front of a 24-inch monitor set to the resolution of 1,920 x 1,080 pixels at 60 Hz. Participants were positioned 60 centimetres away from the screen and rested their heads on a mount. Stimuli were presented on MATLAB 2019a using psychtoolbox (3.0.16), synchronized with the eye tracker. Before the start of the experiment, participants completed a standard 5-point calibration procedure. Drift correction was applied after every block.

## Eye-tracking preprocessing & analysis

Eye-tracking data were preprocessed with the PuPL toolbox in MATLAB<sup>70</sup>. Impossible data points (i.e., gaze outside the screen's bounds) were removed, in addition to data 50ms before and 150 ms after eye blinks (identified by pupillometry noise<sup>71</sup>). Segments of missing gaze position, up to 400ms long, were interpolated using cubic splines. We analyzed eye position data between -1000 ms and 6000 ms around the presentation of the maze, which corresponds to the planning window and the one second prior to planning. To verify that participants did not move their eyes more frequently during planning for lateralized mazes, we computed the standard deviation of eye position along the X-axis for each trial. We compared the fluctuations across lateralized and non-lateralized trials with a two-sample t-test. To verify the robustness of our behavioural effects, we identified and removed from further analysis trials where participants' eye position exceeded two squares away from fixation.

## Data & code availability

All in-house code used for data analysis and visualization is available on GitHub <https://github.com/jasondsc/ConsciousDetour>. The reanalyzed data presented herein are available from <https://www.nature.com/articles/s41586-022-04743-9>. The data from experiment 3 are available from <https://osf.io/sa6vf/>.

## Acknowledgements

J.d.S.C. is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) postdoctoral fellowship program. S.M.F. is a CIFAR Fellow in the Brain, Mind and Consciousness Program and is supported by a Wellcome/Royal Society Sir Henry Dale Fellowship [206648/Z/17/Z] and UKRI under the UK government's Horizon Europe funding guarantee (selected as ERC Consolidator, grant number 101043666).

## Supplemental materials

### Spatial proximity effects

We explored inter-individual differences in spatial attention effects for the second experiment where participants were required to plan their route upfront.

Similar to the first experiment, obstacles closest to the probed item positively related to participants' awareness (1<sup>st</sup> rank:  $t(161) = 19.53$ ,  $p_{fdr} < 0.001$ ; 2<sup>nd</sup> rank:  $t(161) = 13.69$ ,  $p_{fdr} < 0.001$ ). In contrast, obstacles furthest from the probed item were negatively related to participants' awareness of the probed item (4<sup>th</sup> rank:  $t(161) = -5.36$ ,  $p_{fdr} < 0.001$ ; 5<sup>th</sup> rank:  $t(161) = -7.32$ ,  $p_{fdr} < 0.001$ ; 6<sup>th</sup> rank:  $t(156) = -10.16$ ,  $p_{fdr} < 0.001$ ; see **Figure S9**).

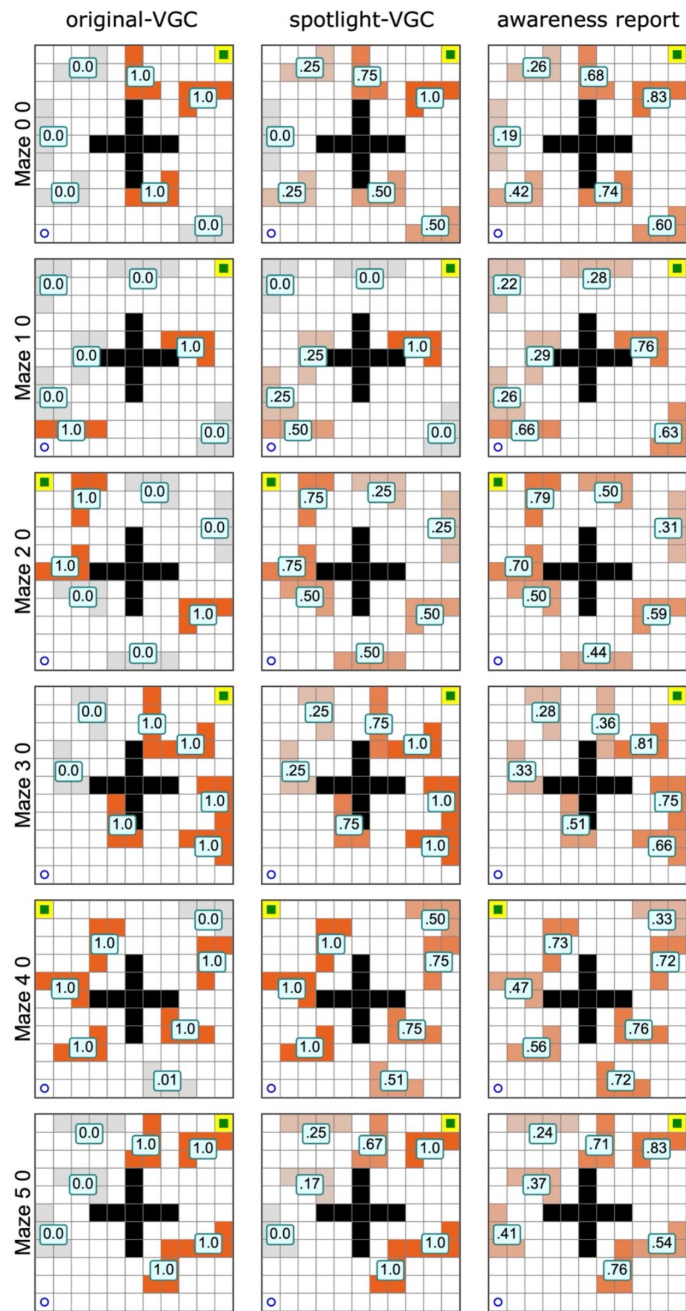
### Task-relevant lateralization effects

Given the observed spatial attention effects and previous attention literature<sup>39–41</sup> we hypothesized that participants would select task-relevant information with greater ease and therefore form a representation of the stimulus that is more aligned with the predictions of the VGC model when task-relevant obstacles are presented to a single hemifield. To test this, we computed a lateralization index per maze, and used this as a moderator in a hierarchical linear regression model. A summary of the moderation effects can be found in **Table S6** for both horizontal and vertical lateralization effects across both experiments.

### Attentional spotlight model

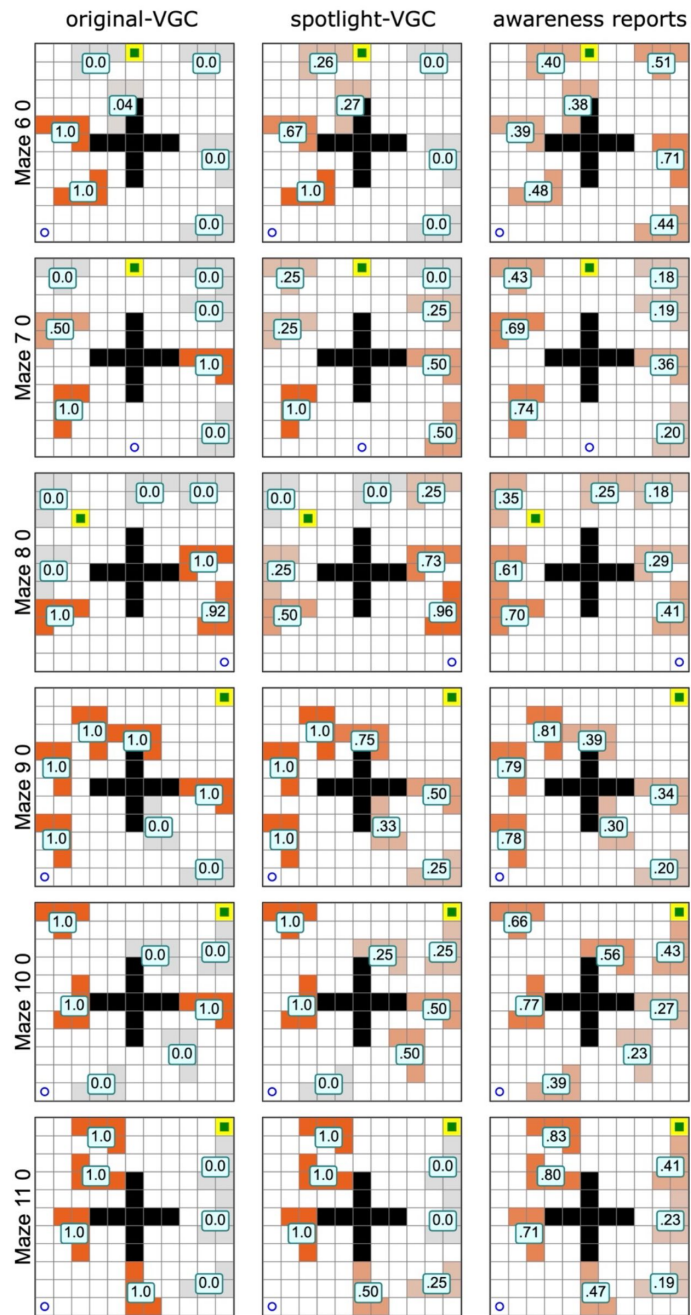
To ensure that our attentional spotlight model results were robust, we conducted control analyses where we ran spatial permutations of our model to ensure that the observed results were not simply due to the spatial smoothness.

### Sensitivity analyses



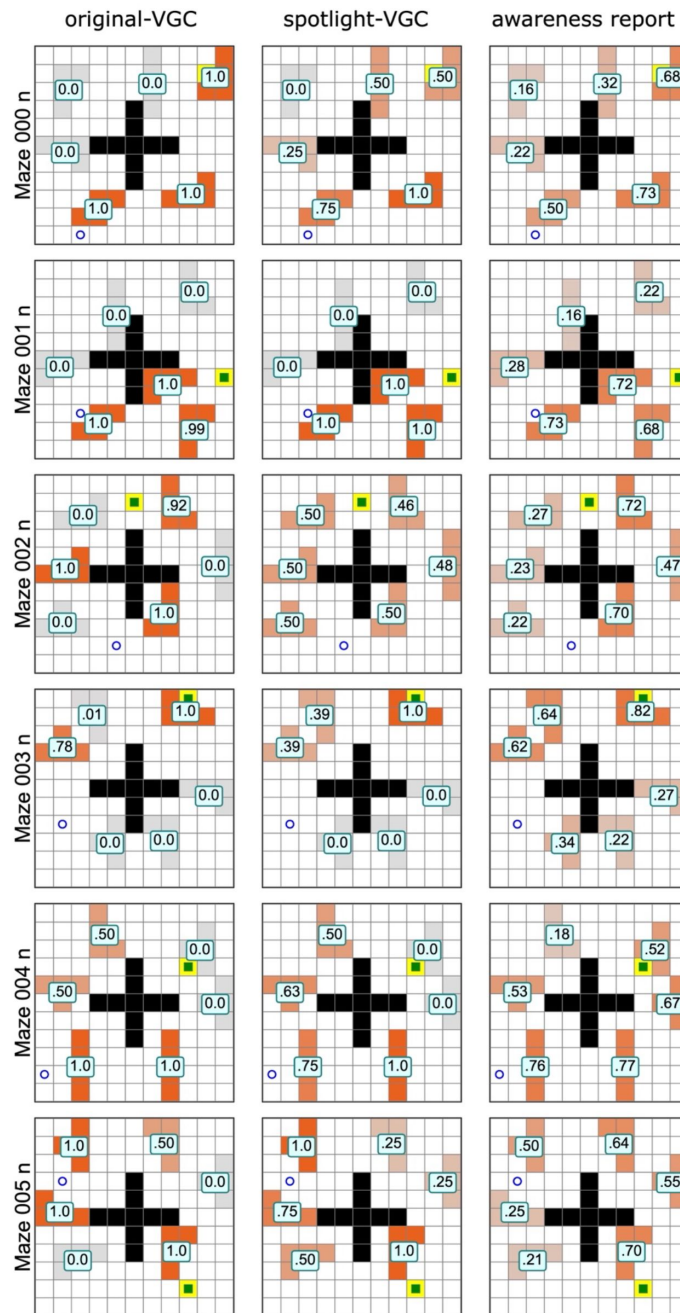
**Figure S1.**

**Model predictions and reported awareness on mazes 0 to 5 of reanalysed data.**



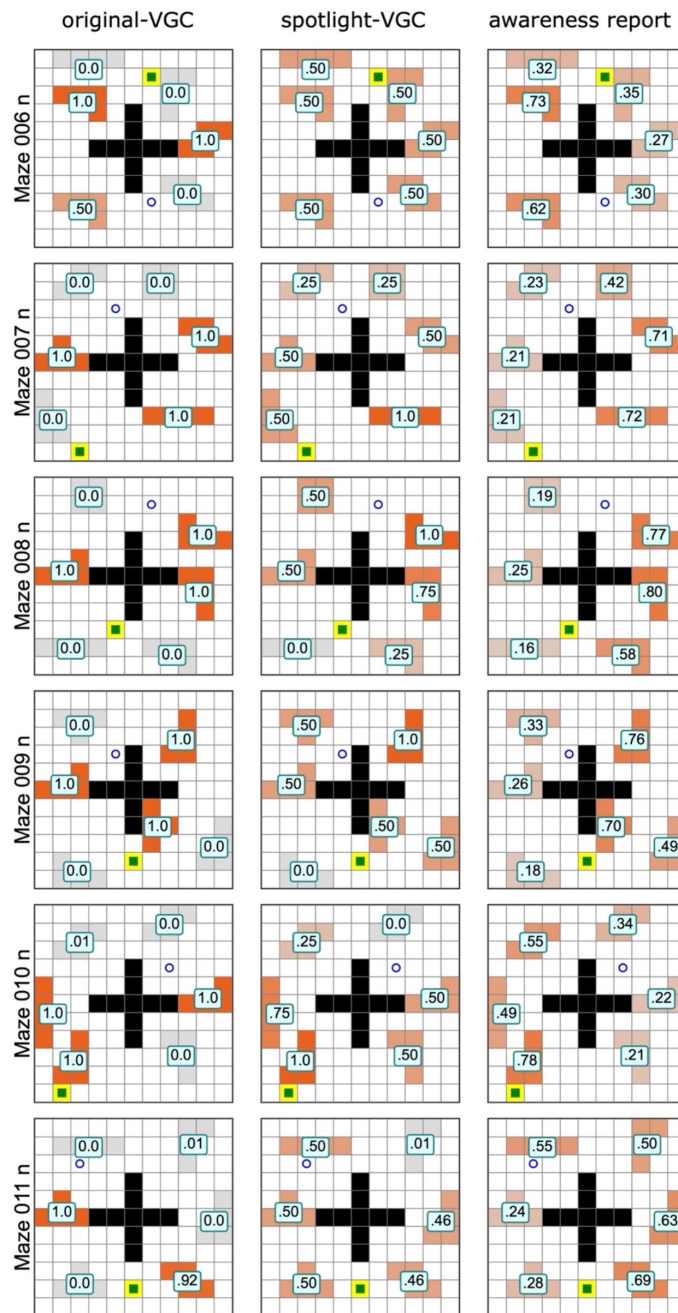
**Figure S2.**

**Model predictions and reported awareness on mazes 6 to 11 of reanalysed data.**



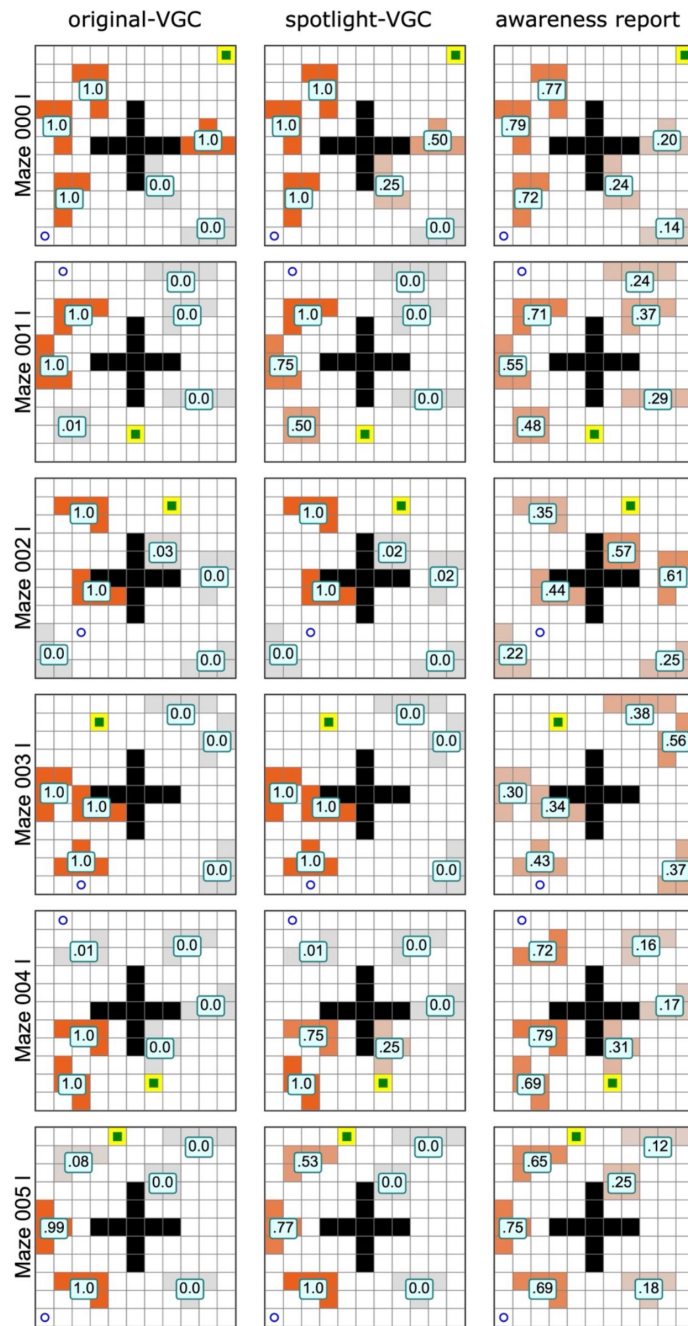
**Figure S3.**

Model predictions and reported awareness on non-lateralized mazes 0 to 5 of the new experiment.



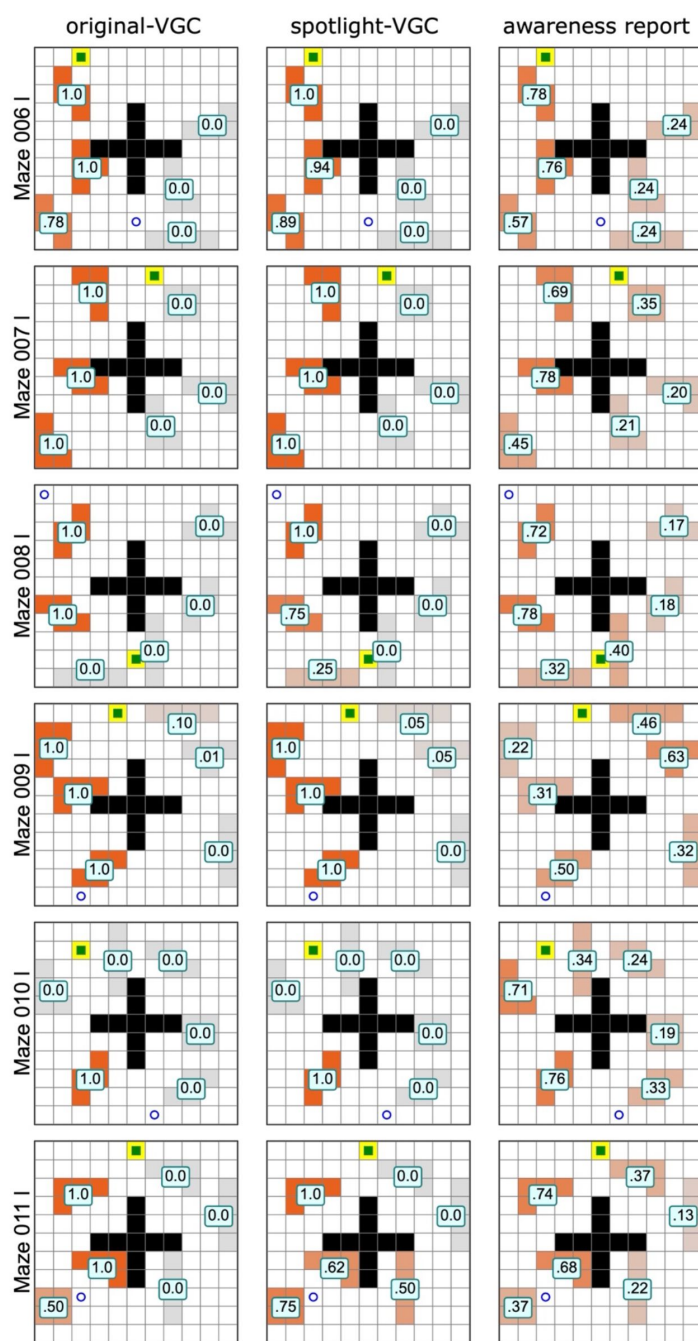
**Figure S4.**

**Model predictions and reported awareness on non-lateralized mazes 6 to 11 of the new experiment.**



**Figure S5.**

**Model predictions and reported awareness on lateralized mazes 0 to 5 of the new experiment.**



**Figure S6.**

**Model predictions and reported awareness on lateralized mazes 6 to 11 of the new experiment.**

|                    | <i>t</i> -statistic | Uncorrected <i>p</i> -value |
|--------------------|---------------------|-----------------------------|
| Optimal no. moves  | -0.53               | 0.65                        |
| Distance to goal   | 0.91                | 0.38                        |
| Distance to start  | 1.84                | 0.07                        |
| Distance to walls  | 0.99                | 0.33                        |
| Distance to center | 0.77                | 0.48                        |

**Table S1.**

**The lateralized and non-lateralized maze stimuli for experiment three did not differ on nuisance covariates.**

| Predictors                                           | Awareness of probe obstacle |               |                  |
|------------------------------------------------------|-----------------------------|---------------|------------------|
|                                                      | Estimates                   | CI            | p                |
| (Intercept)                                          | 0.00                        | -0.08 – 0.08  | 0.974            |
| Obstacle 1 (closest)                                 | 0.26                        | 0.25 – 0.28   | <b>&lt;0.001</b> |
| Obstacle 2                                           | 0.29                        | 0.27 – 0.30   | <b>&lt;0.001</b> |
| Obstacle 3                                           | 0.00                        | -0.01 – 0.02  | 0.699            |
| Obstacle 4                                           | -0.03                       | -0.04 – -0.01 | <b>0.001</b>     |
| Obstacle 5                                           | -0.13                       | -0.15 – -0.12 | <b>&lt;0.001</b> |
| Obstacle 6 (furthest)                                | -0.13                       | -0.15 – -0.12 | <b>&lt;0.001</b> |
| Random Effects                                       |                             |               |                  |
| $\sigma^2$                                           | 0.58                        |               |                  |
| T00 SubjectID                                        | 0.11                        |               |                  |
| T00 MazeID                                           | 0.01                        |               |                  |
| ICC                                                  | 0.17                        |               |                  |
| N MazeID                                             | 12                          |               |                  |
| N SubjectID                                          | 161                         |               |                  |
| Observations                                         | 13342                       |               |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.245 / 0.371               |               |                  |

**Table S2.**

**Hierarchical linear regression model predicting the awareness of an obstacle from its neighbours.**

Beta coefficients reflect the contribution of neighbouring obstacles to the awareness of the probe item. Beta coefficients reflect the rank order of the closest (1<sup>st</sup>) to furthest (6<sup>th</sup>) obstacle from the probed item.

**Table S3.**

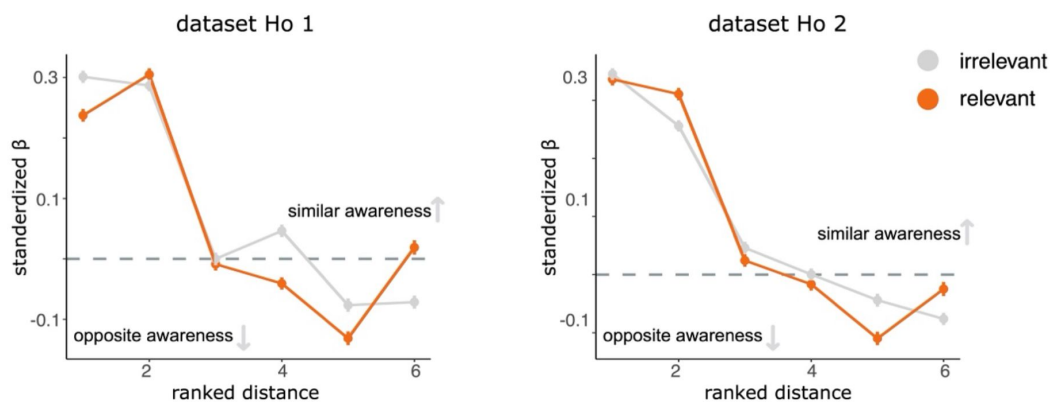
**Effect of neighbouring obstacles on awareness of probed item, after regressing the effect of the VGC model.**

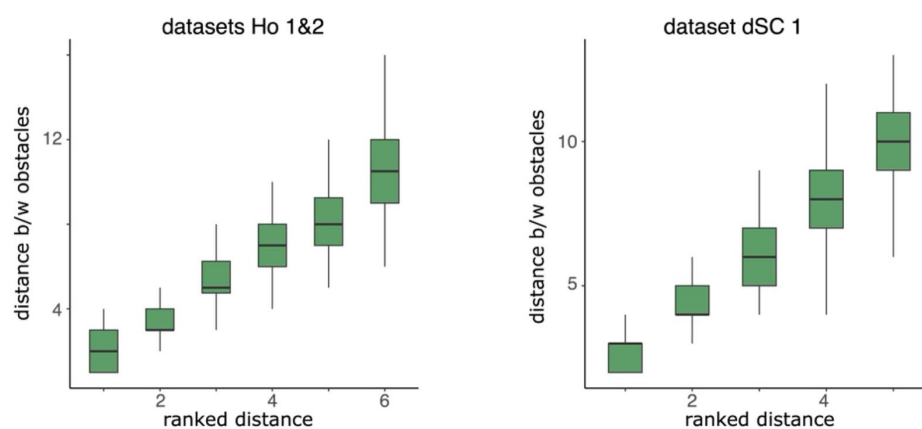
| Predictors                               | Awareness residuals |               |                  |
|------------------------------------------|---------------------|---------------|------------------|
|                                          | Estimates           | CI            | p                |
| (Intercept)                              | -0.00               | -0.02 – 0.02  | 1.000            |
| Obstacle 1 (closest)                     | 0.26                | 0.24 – 0.27   | <b>&lt;0.001</b> |
| Obstacle 2                               | 0.27                | 0.25 – 0.28   | <b>&lt;0.001</b> |
| Obstacle 3                               | -0.06               | -0.08 – -0.05 | <b>&lt;0.001</b> |
| Obstacle 4                               | -0.07               | -0.09 – -0.05 | <b>&lt;0.001</b> |
| Obstacle 5                               | -0.15               | -0.16 – -0.13 | <b>&lt;0.001</b> |
| Obstacle 6 (furthest)                    | -0.10               | -0.12 – -0.08 | <b>&lt;0.001</b> |
| Observations                             | 13342               |               |                  |
| R <sup>2</sup> / R <sup>2</sup> adjusted | 0.201 / 0.201       |               |                  |

**Fig. S7.**

**Spatial proximity predicts awareness for both task -relevant and -irrelevant obstacles.**

Standardized beta coefficients of the ranked regression model for task-relevant and -irrelevant obstacles separately across datasets Ho 1 & 2. We observed that obstacles closest to the probed item (rank 1 & 2) positively impact awareness reports, regardless of task-relevance. In contrast, obstacles furthest from the probed item negatively impact awareness reports (rank 5 & 6).





**Fig. S8.**

**Distribution of distances between obstacles for each rank in regression.**

Boxplots depicting the distance between obstacles. In datasets Ho 1 & 2, obstacles ranked 1<sup>st</sup> (closest) and 2<sup>nd</sup> in proximity were on average 2.14 (median= 2.0, sd= 0.91), and 3.60 squares away (median= 3.0, sd=1.30), respectively. These obstacles positively predict participants' awareness of items. In dataset dSC 1, obstacles ranked closest were on average 2.76 (median= 3.0, sd= 0.83) squares away. Based on these statistics, we expected an attentional spotlight of width 3 would provide the best fit to the data.

| Predictors                                           | Awareness of probe obstacle |               |                  |
|------------------------------------------------------|-----------------------------|---------------|------------------|
|                                                      | Estimates                   | CI            | p                |
| (Intercept)                                          | 0.00                        | -0.05 – 0.05  | 0.960            |
| Obstacle 1 (closest)                                 | 0.34                        | 0.33 – 0.36   | <b>&lt;0.001</b> |
| Obstacle 2                                           | 0.29                        | 0.27 – 0.30   | <b>&lt;0.001</b> |
| Obstacle 3                                           | 0.04                        | 0.03 – 0.06   | <b>&lt;0.001</b> |
| Obstacle 4                                           | -0.03                       | -0.05 – -0.02 | <b>&lt;0.001</b> |
| Obstacle 5                                           | -0.09                       | -0.11 – -0.07 | <b>&lt;0.001</b> |
| Obstacle 6 (furthest)                                | -0.12                       | -0.14 – -0.11 | <b>&lt;0.001</b> |
| <b>Random Effects</b>                                |                             |               |                  |
| $\sigma^2$                                           | 0.60                        |               |                  |
| T00 SubjectID                                        | 0.04                        |               |                  |
| T00 MazeID                                           | 0.00                        |               |                  |
| ICC                                                  | 0.07                        |               |                  |
| N MazeID                                             | 12                          |               |                  |
| N SubjectID                                          | 162                         |               |                  |
| Observations                                         | 13321                       |               |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.312 / 0.362               |               |                  |

**Table S4.**

**Hierarchical linear regression model predicting the awareness of an obstacle from its neighbours for dataset Ho 2.**

**Table S5.**

**Effect of neighbouring obstacles on awareness of probed item, after regressing the effect of the VGC model for dataset Ho 2.**

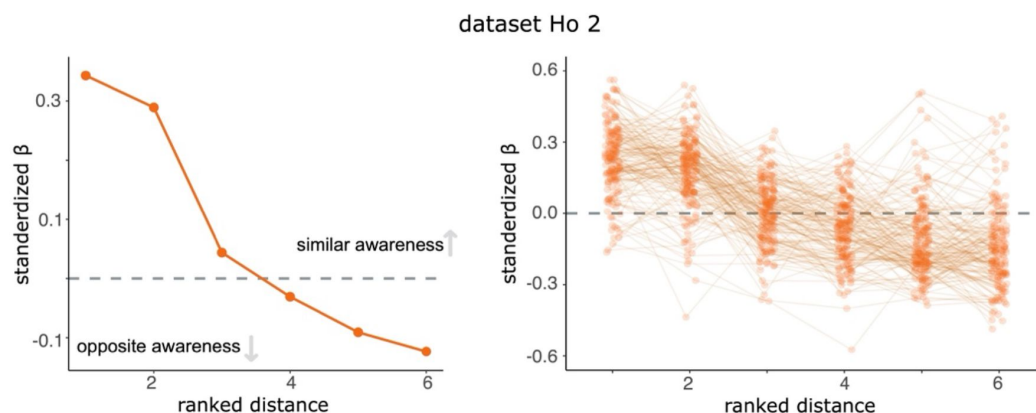
| Predictors                               | Awareness residuals |               |                  |
|------------------------------------------|---------------------|---------------|------------------|
|                                          | Estimates           | CI            | p                |
| (Intercept)                              | 0.00                | -0.01 – 0.01  | 1.000            |
| Obstacle 1 (closest)                     | 0.32                | 0.30 – 0.34   | <b>&lt;0.001</b> |
| Obstacle 2                               | 0.24                | 0.23 – 0.26   | <b>&lt;0.001</b> |
| Obstacle 3                               | -0.01               | -0.03 – 0.01  | 0.236            |
| Obstacle 4                               | -0.07               | -0.09 – -0.05 | <b>&lt;0.001</b> |
| Obstacle 5                               | -0.12               | -0.13 – -0.10 | <b>&lt;0.001</b> |
| Obstacle 6 (furthest)                    | -0.12               | -0.13 – -0.10 | <b>&lt;0.001</b> |
| Observations                             | 13321               |               |                  |
| R <sup>2</sup> / R <sup>2</sup> adjusted | 0.242 / 0.241       |               |                  |

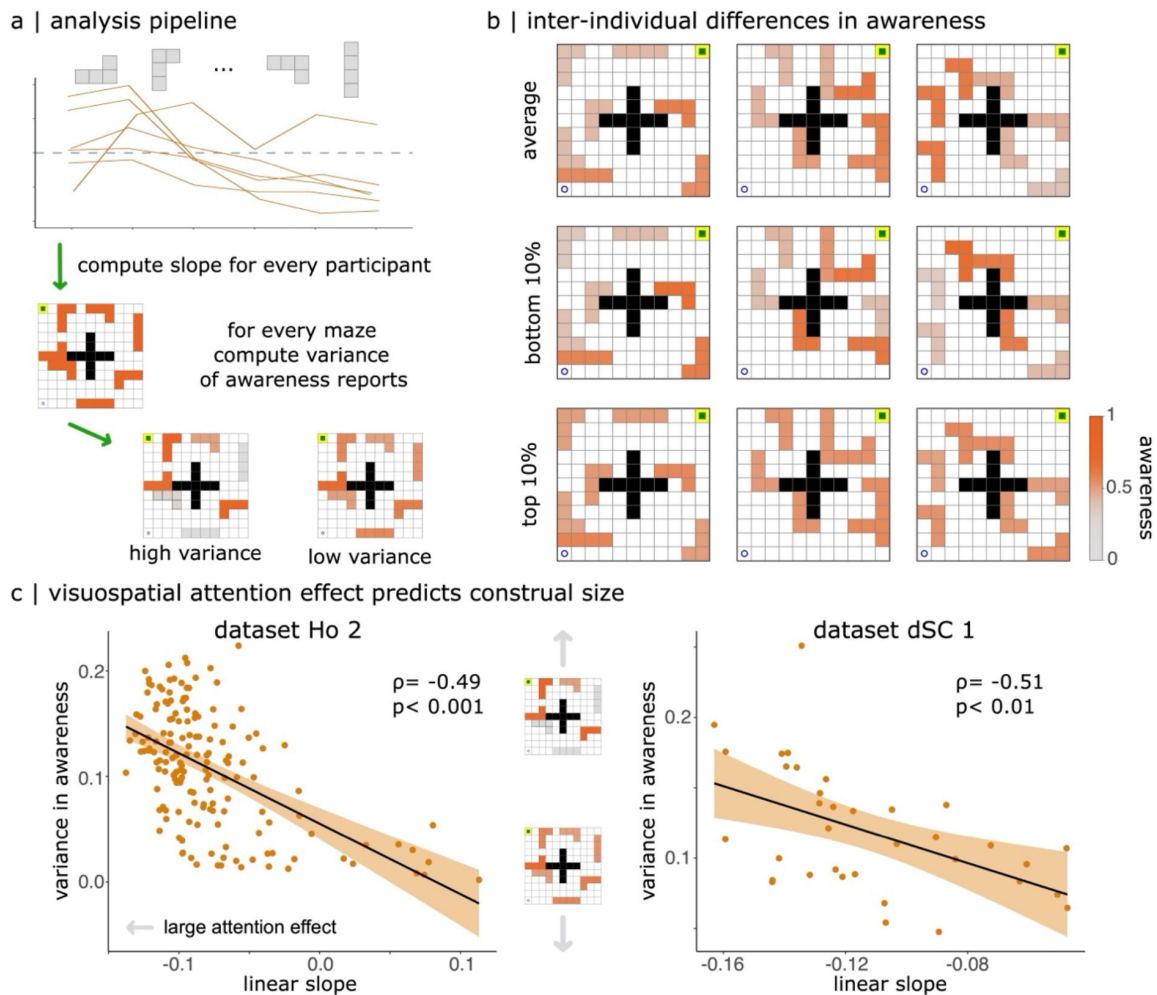
**Fig. S9.**

**Neighbouring obstacles predict inclusion/ exclusion from task representation.**

Right panel: Results of the ranked regression model for the upfront planning experiment (i.e., dataset Ho 2). Obstacles closest to the probed item (rank 1 & 2) positively impact awareness reports, whereas obstacles furthest from the probed item negatively impact awareness reports (rank 4, 5 & 6).

Left panel: Inter-individual differences in spatial attention effects in dataset Ho 2. Each participant is represented by a point and line.





**Fig. S10.**

### Inter-individual differences in simplified representations.

(a) Analysis pipeline to explore the relationship between spatial attention and task representations. First, we fit a linear model to beta coefficients obtained in [Figure 1b](#) left for each participant. This slope corresponds to a participant's attention effect, with more negative slopes indicating larger attention effects. We then computed the sparsity of participant's representations by computing the mean variance of their awareness reports.

(b) Examples of participants' task representations. The top row depicts the average awareness of each obstacle in three example mazes. The second row depicts the representations of participants with the most negative attention slopes. The bottom row plots the awareness effects of participants with the shallowest slopes (i.e., weakest spatial attention effects).

(c) Participants with stronger spatial attention effects had more sparse task representations. We replicated this effect in two independent samples across datasets Ho 2 and dSC 1.

| Predictors                             | Awareness of probe obstacle |               |                  |
|----------------------------------------|-----------------------------|---------------|------------------|
|                                        | Estimates                   | CI            | p                |
| (Intercept)                            | 0.00                        | -0.14 – 0.14  | 0.995            |
| Obstacle 1 (closest)                   | 0.19                        | 0.18 – 0.21   | <b>&lt;0.001</b> |
| Obstacle 2                             | 0.01                        | -0.00 – 0.02  | 0.115            |
| Obstacle 3                             | -0.10                       | -0.11 – -0.08 | <b>&lt;0.001</b> |
| Obstacle 4                             | -0.26                       | -0.27 – -0.25 | <b>&lt;0.001</b> |
| Obstacle 5 (furthest)                  | -0.29                       | -0.30 – -0.27 | <b>&lt;0.001</b> |
| <b>Random Effects</b>                  |                             |               |                  |
| $\sigma^2$                             | 0.65                        |               |                  |
| T00 SubjectID                          | 0.15                        |               |                  |
| T00 MazeID                             | 0.03                        |               |                  |
| ICC                                    | 0.21                        |               |                  |
| N MazeID                               | 24                          |               |                  |
| N SubjectID                            | 35                          |               |                  |
| Observations                           | 19140                       |               |                  |
| Marginal<br>Conditional R <sup>2</sup> | R <sup>2</sup> /            | 0.263 / 0.421 |                  |

**Table S6.**

**Hierarchical linear regression model predicting the awareness of an obstacle from its neighbours for dataset dSC 1.**

**Table S7.**

**Effect of neighbouring obstacles on awareness of probed item, after regressing the effect of the VGC model for dataset dSC 1.**

| Predictors                               | Awareness residuals |               |                  |
|------------------------------------------|---------------------|---------------|------------------|
|                                          | Estimates           | CI            | p                |
| (Intercept)                              | -0.00               | -0.01 – 0.01  | 1.000            |
| Obstacle 1 (closest)                     | 0.23                | 0.22 – 0.24   | <b>&lt;0.001</b> |
| Obstacle 2                               | 0.09                | 0.07 – 0.10   | <b>&lt;0.001</b> |
| Obstacle 3                               | -0.08               | -0.09 – -0.06 | <b>&lt;0.001</b> |
| Obstacle 4                               | -0.22               | -0.23 – -0.21 | <b>&lt;0.001</b> |
| Obstacle 5 (furthest)                    | -0.18               | -0.20 – -0.17 | <b>&lt;0.001</b> |
| Observations                             | 19140               |               |                  |
| R <sup>2</sup> / R <sup>2</sup> adjusted | 0.213 / 0.213       |               |                  |

**Fig. S11.**

### Spatial proximity predicts awareness for dataset dSC 1.

Standardized beta coefficients of ranked regression model for experiment 3. We observed that obstacles closest to the probed item (rank 1 & 2) positively impact awareness reports, regardless of the type of maze (i.e., lateralized vs non-lateralized; right panel). In contrast, obstacles furthest from the probed item negatively impact awareness reports (rank 5). We note that while the 2<sup>nd</sup> closest obstacle positively predicted the awareness reports, this effect was much weaker than the closest obstacle (i.e., rank 1).

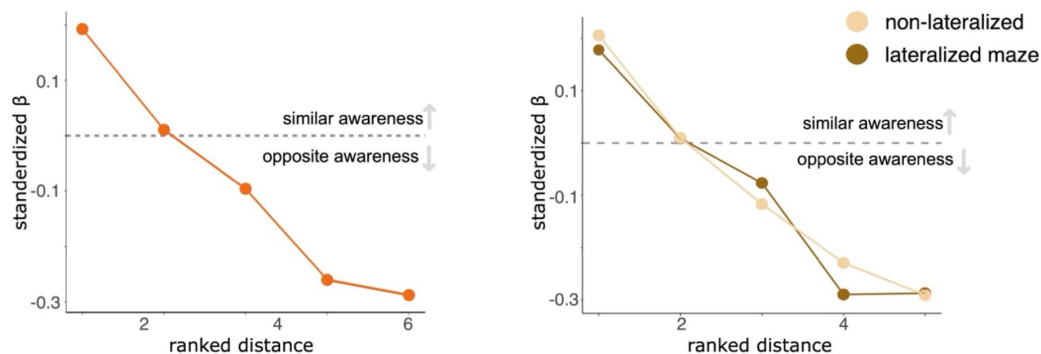


Table S8.

The degree to which task-relevant information is lateralized moderates the relationship between the VGC model and awareness (dataset Ho 1).

| Predictors                                           | Awareness     |              |                  |
|------------------------------------------------------|---------------|--------------|------------------|
|                                                      | Estimates     | CI           | p                |
| (Intercept)                                          | 0.50          | 0.47 – 0.54  | <b>&lt;0.001</b> |
| VGC model                                            | 0.13          | 0.13 – 0.14  | <b>&lt;0.001</b> |
| Lateralization index (vert)                          | 0.02          | -0.01 – 0.04 | 0.238            |
| VGC model * Lateralization index (vert)              | 0.01          | 0.01 – 0.02  | <b>&lt;0.001</b> |
| <b>Random Effects</b>                                |               |              |                  |
| $\sigma^2$                                           | 0.09          |              |                  |
| T00 SubjectID                                        | 0.03          |              |                  |
| T00 MazeID                                           | 0.00          |              |                  |
| ICC                                                  | 0.24          |              |                  |
| N MazeID                                             | 12            |              |                  |
| N SubjectID                                          | 161           |              |                  |
| Observations                                         | 13342         |              |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.138 / 0.348 |              |                  |

Table S9.

Effect of neighbouring obstacles on awareness of probed item, after regressing the effect of the VGC model.

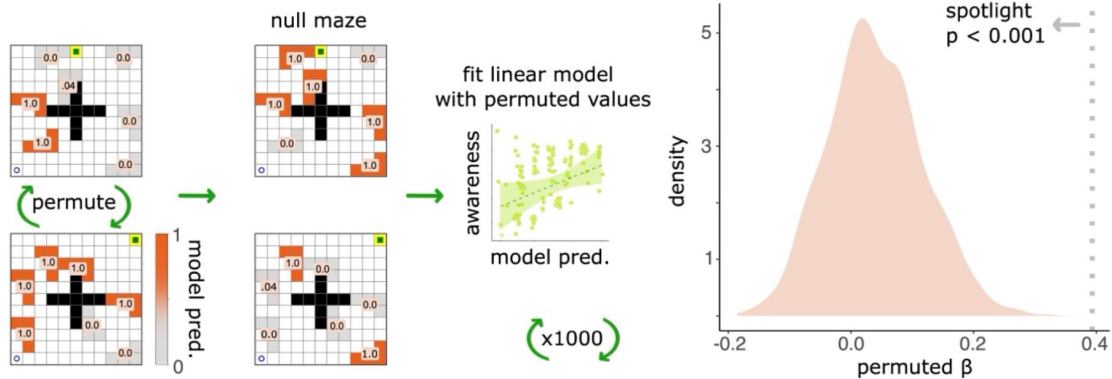
|              |            | Moderation Effect | CI            | p                | P FDR            | Marginal R <sup>2</sup> / Conditional R <sup>2</sup> |
|--------------|------------|-------------------|---------------|------------------|------------------|------------------------------------------------------|
| dataset Ho   | vertical   | 0.01              | 0.01 – 0.02   | <b>&lt;0.001</b> | <b>&lt;0.001</b> | 0.138 / 0.348                                        |
|              | horizontal | -0.00             | - 0.01 – 0.00 | 0.478            | 0.478            | 0.136 / 0.343                                        |
| dataset Ho 2 | vertical   | 0.01              | 0.00 – 0.02   | <b>0.001</b>     | <b>0.002</b>     | 0.092 / 0.246                                        |
|              | horizontal | 0.01              | 0.00 – 0.01   | <b>0.013</b>     | <b>0.017</b>     | 0.091 / 0.244                                        |

| Predictors                                           | Awareness     |              |                  |
|------------------------------------------------------|---------------|--------------|------------------|
|                                                      | Estimates     | CI           | p                |
| (Intercept)                                          | 0.45          | 0.41 – 0.48  | <b>&lt;0.001</b> |
| VGC model                                            | 0.13          | 0.12 – 0.13  | <b>&lt;0.001</b> |
| Lateralization index (vert)                          | -0.00         | -0.02 – 0.01 | 0.602            |
| VGC model * Lateralization index (vert)              | 0.01          | 0.01 – 0.02  | <b>&lt;0.001</b> |
| <b>Random Effects</b>                                |               |              |                  |
| $\sigma^2$                                           | 0.09          |              |                  |
| T00 SubjectID                                        | 0.01          |              |                  |
| T00 MazeID                                           | 0.00          |              |                  |
| ICC                                                  | 0.10          |              |                  |
| N MazeID                                             | 24            |              |                  |
| N SubjectID                                          | 35            |              |                  |
| Observations                                         | 19140         |              |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.137 / 0.222 |              |                  |

**Table S10.**

**The degree to which task-relevant information is lateralized moderates the relationship between the VGC model and awareness (dataset dSC 1).**

a | null permutations between mazes



b | null permutations of obstacles within a maze

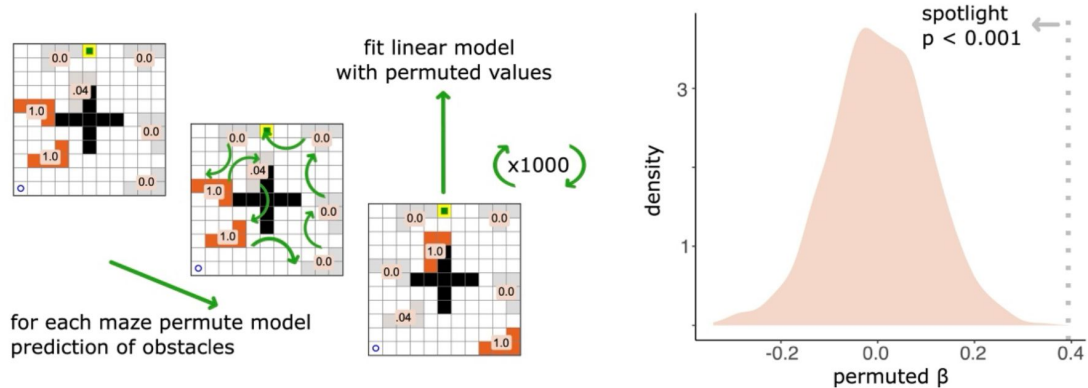
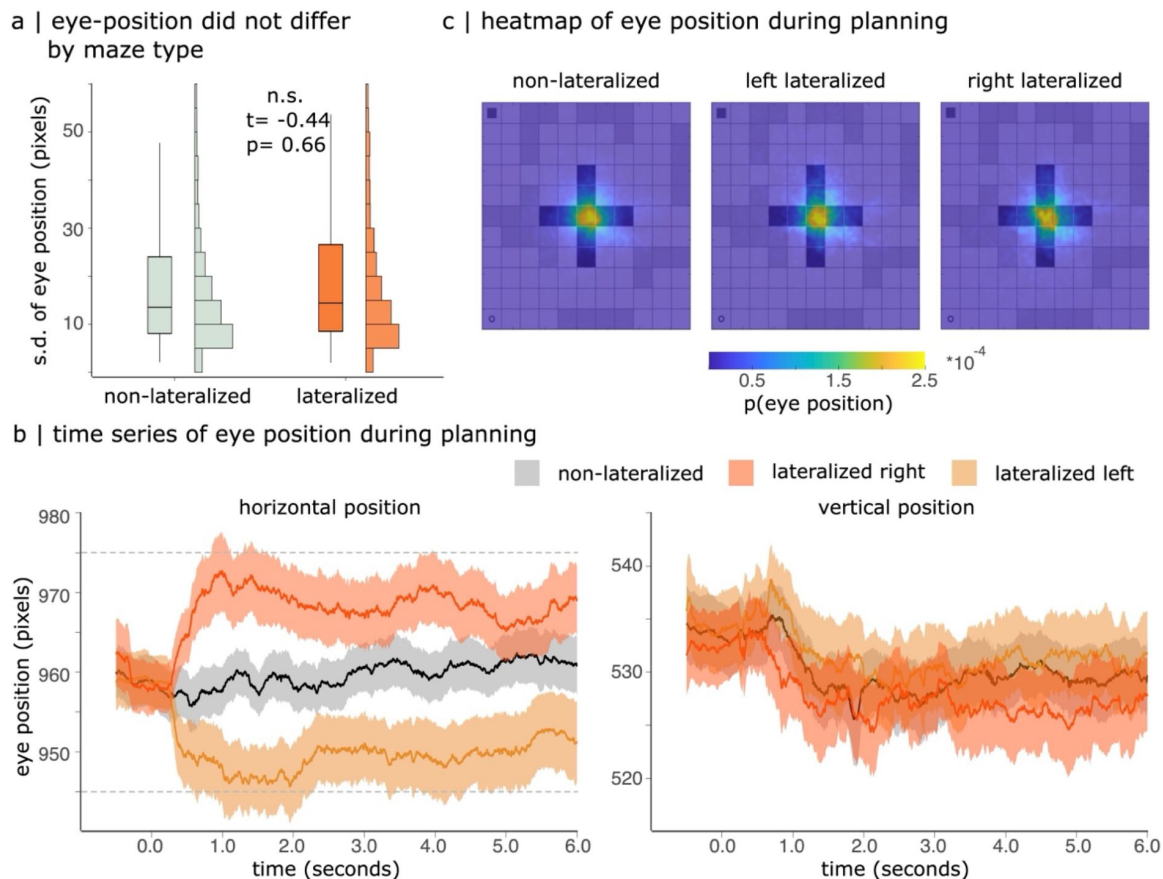


Fig. S12.

**Robustness of the attentional spotlight model to spatial autocorrelation.**

We assessed the robustness of the spotlight model to spatial autocorrelation by performing two sets of different spatial permutations. (a) In the first set of permutations, we permuted the predictions of the spotlight model across all mazes and used these null predictions in a hierarchical linear regression model. We repeated this procedure 1000 times to produce a null distribution of beta coefficients, depicted in the right panel. The observed spotlight model effect (dotted line) was significantly better than the spatial null permutations.

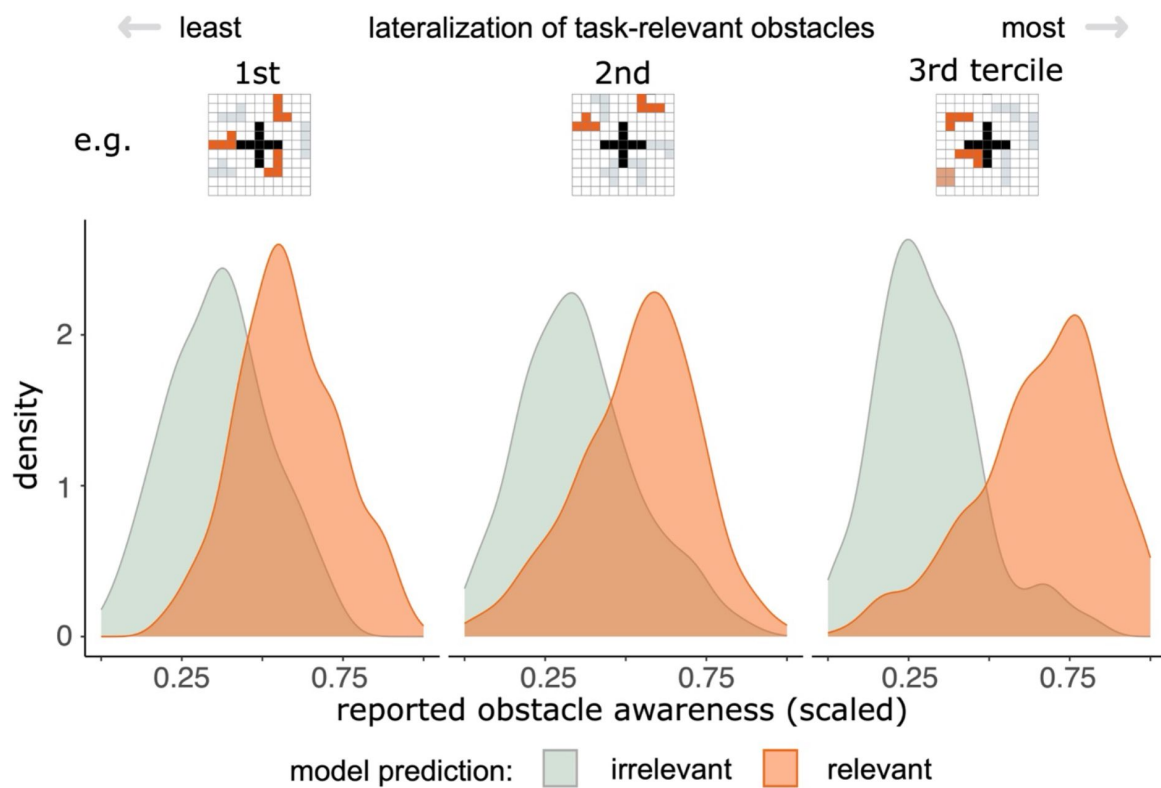
(b) We similarly permuted the spotlight model's predictions within each maze, assigning each obstacle a random prediction. We used these null predictions in a hierarchical regression model to predict participants' awareness reports. We repeated this procedure 1000 times to generate a null distribution, depicted in the right panel. The spotlight model predicted participants' awareness reports beyond the spatial autocorrelation of the data.



**Fig. S13.**

### Eye position during planning.

We verified that our behavioural effects were not driven by eye movements during the planning phase in dataset dSC 1. (a) The fluctuations of eye position, measured as the standard deviation of the eye-position time series, did not statistically differ between non-lateralized and lateralized maze stimuli. (b) Time series of the average position of participants' gaze during planning along the horizontal and vertical axis, left and right panels, respectively. Participants, on average, moved their eyes more toward the left when maze stimuli were lateralized to the left (yellow line) and toward the right when the maze stimuli were lateralized to the right (orange line). These eye movements, however, remained within the bounds of the central square (dotted grey lines). Eye gaze did not differ between maze stimuli along the y-axis. (c) Heat maps depicting where participants looked during the planning phase for non-lateralized, right-lateralized, and left-lateralized maze stimuli.



**Fig. S14.**

### Attention lateralization effects are robust to eye-gaze.

Density plots of the reported awareness of obstacles separated task-relevant ( $>0.5$ ; in orange) or task-irrelevant ( $<0.5$ ; in grey) obstacles as predicted by the VGC model for trials with minimal eye movements. Participants were more likely to be aware of task-relevant obstacles and unaware of irrelevant obstacles. This effect was moderated by the degree to which task-relevant information was presented preferentially to one hemifield (x-axis). From left to right, we plot the three tertiles of maze lateralization. Participants' awareness reports become increasingly aligned with the VGC model's predictions—i.e., the overlap between the two density plots decreases with lateralization.

| Predictors                                           | Awareness     |              |        |
|------------------------------------------------------|---------------|--------------|--------|
|                                                      | Estimates     | CI           | p      |
| (Intercept)                                          | 0.40          | 0.26 – 0.55  | <0.001 |
| goal distance                                        | 0.04          | 0.03 – 0.05  | <0.001 |
| start distance                                       | 0.03          | 0.02 – 0.03  | <0.001 |
| wall distance                                        | -0.02         | -0.08 – 0.03 | 0.354  |
| center distance                                      | 0.02          | -0.01 – 0.05 | 0.129  |
| VGC model                                            | 0.14          | 0.14 – 0.15  | <0.001 |
| Lateralization index (vert)                          | 0.02          | -0.01 – 0.04 | 0.196  |
| VGC model * Lateralization index (vert)              | 0.01          | 0.00 – 0.01  | 0.004  |
| <b>Random Effects</b>                                |               |              |        |
| $\sigma^2$                                           | 0.08          |              |        |
| T00 SubjectID                                        | 0.03          |              |        |
| T00 MazeID                                           | 0.00          |              |        |
| ICC                                                  | 0.24          |              |        |
| N MazeID                                             | 12            |              |        |
| N SubjectID                                          | 161           |              |        |
| Observations                                         | 13342         |              |        |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.148 / 0.356 |              |        |

**Table S11.**

**Robustness of lateralization moderation effect to nuisance covariates (dataset Ho 1).**

| Predictors                                           | Awareness     |              |                  |
|------------------------------------------------------|---------------|--------------|------------------|
|                                                      | Estimates     | CI           | p                |
| (Intercept)                                          | 0.43          | 0.27 – 0.59  | <b>&lt;0.001</b> |
| goal distance                                        | 0.06          | 0.05 – 0.07  | <b>&lt;0.001</b> |
| start distance                                       | 0.05          | 0.04 – 0.06  | <b>&lt;0.001</b> |
| wall distance                                        | 0.01          | -0.05 – 0.07 | 0.729            |
| center distance                                      | 0.01          | -0.02 – 0.04 | 0.427            |
| VGC model                                            | 0.13          | 0.13 – 0.14  | <b>&lt;0.001</b> |
| Lateralization index (vert)                          | 0.01          | -0.01 – 0.03 | 0.320            |
| VGC model * Lateralization index (vert)              | 0.01          | 0.00 – 0.01  | <b>0.013</b>     |
| <b>Random Effects</b>                                |               |              |                  |
| $\sigma^2$                                           | 0.10          |              |                  |
| T00 SubjectID                                        | 0.02          |              |                  |
| T00 MazeID                                           | 0.00          |              |                  |
| ICC                                                  | 0.17          |              |                  |
| N <sub>MazeID</sub>                                  | 12            |              |                  |
| N <sub>SubjectID</sub>                               | 162           |              |                  |
| Observations                                         | 13321         |              |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.113 / 0.267 |              |                  |

**Table S12.**

**Robustness of lateralization moderation effect to nuisance covariates (dataset Ho 2).**

| Predictors                                           | Awareness     |              |                  |
|------------------------------------------------------|---------------|--------------|------------------|
|                                                      | Estimates     | CI           | p                |
| (Intercept)                                          | 0.45          | 0.41 – 0.49  | <b>&lt;0.001</b> |
| goal distance                                        | -0.01         | -0.04 – 0.02 | 0.379            |
| start distance                                       | -0.02         | -0.04 – 0.01 | 0.258            |
| wall distance                                        | -0.03         | -0.12 – 0.06 | 0.545            |
| center distance                                      | 0.03          | -0.07 – 0.12 | 0.579            |
| VGC model                                            | 0.13          | 0.12 – 0.13  | <b>&lt;0.001</b> |
| Lateralization index (vert)                          | -0.00         | -0.02 – 0.02 | 0.772            |
| VGC model * Lateralization index (vert)              | 0.01          | 0.01 – 0.02  | <b>&lt;0.001</b> |
| <b>Random Effects</b>                                |               |              |                  |
| $\sigma^2$                                           | 0.09          |              |                  |
| T00 SubjectID                                        | 0.01          |              |                  |
| T00 MazeID                                           | 0.00          |              |                  |
| ICC                                                  | 0.10          |              |                  |
| N <sub>MazeID</sub>                                  | 24            |              |                  |
| N <sub>SubjectID</sub>                               | 35            |              |                  |
| Observations                                         | 19140         |              |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.136 / 0.222 |              |                  |

**Table S13.**

**Robustness of lateralization moderation effect to nuisance covariates (dataset dSC 1).**

| Predictors                                           | Awareness     |              |                  |
|------------------------------------------------------|---------------|--------------|------------------|
|                                                      | Estimates     | CI           | p                |
| (Intercept)                                          | 0.43          | 0.28 – 0.59  | <b>&lt;0.001</b> |
| goal distance                                        | 0.06          | 0.05 – 0.07  | <b>&lt;0.001</b> |
| start distance                                       | 0.05          | 0.04 – 0.06  | <b>&lt;0.001</b> |
| wall distance                                        | 0.01          | -0.05 – 0.07 | 0.714            |
| center distance                                      | 0.01          | -0.02 – 0.04 | 0.462            |
| VGC model                                            | 0.13          | 0.13 – 0.14  | <b>&lt;0.001</b> |
| Lateralization index (horz)                          | -0.01         | -0.03 – 0.01 | 0.205            |
| VGC model * Lateralization index (horz)              | 0.00          | -0.00 – 0.01 | 0.686            |
| Random Effects                                       |               |              |                  |
| $\sigma^2$                                           | 0.10          |              |                  |
| T00 SubjectID                                        | 0.02          |              |                  |
| T00 MazeID                                           | 0.00          |              |                  |
| ICC                                                  | 0.17          |              |                  |
| N <sub>grid</sub>                                    | 12            |              |                  |
| N <sub>SubjectID</sub>                               | 162           |              |                  |
| Observations                                         | 13321         |              |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.111 / 0.265 |              |                  |

**Table S14.**

**Horizontal lateralization moderation regression (dataset Ho 2).**

| Predictors                                           | Awareness     |              |                  |
|------------------------------------------------------|---------------|--------------|------------------|
|                                                      | Estimates     | CI           | p                |
| (Intercept)                                          | 0.45          | 0.41 – 0.48  | <b>&lt;0.001</b> |
| goal distance                                        | -0.01         | -0.04 – 0.02 | 0.635            |
| start distance                                       | -0.01         | -0.04 – 0.01 | 0.378            |
| wall distance                                        | -0.06         | -0.15 – 0.03 | 0.206            |
| center distance                                      | 0.06          | -0.03 – 0.15 | 0.215            |
| VGC model                                            | 0.13          | 0.12 – 0.13  | <b>&lt;0.001</b> |
| Lateralization index (horz)                          | 0.02          | -0.00 – 0.04 | 0.095            |
| VGC model * Lateralization index (horz)              | -0.00         | -0.01 – 0.00 | 0.349            |
| Random Effects                                       |               |              |                  |
| $\sigma^2$                                           | 0.09          |              |                  |
| T00 SubjectID                                        | 0.01          |              |                  |
| T00 MazeID                                           | 0.00          |              |                  |
| ICC                                                  | 0.10          |              |                  |
| N MazeID                                             | 24            |              |                  |
| N SubjectID                                          | 35            |              |                  |
| Observations                                         | 19140         |              |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.134 / 0.218 |              |                  |

**Table S15.**

**Horizontal lateralization moderation regression (dataset dSC 1).**

| Predictors                                           | Awareness     |              |                  |
|------------------------------------------------------|---------------|--------------|------------------|
|                                                      | Estimates     | CI           | p                |
| (Intercept)                                          | 0.45          | 0.41 – 0.48  | <b>&lt;0.001</b> |
| goal distance                                        | -0.01         | -0.04 – 0.02 | 0.360            |
| start distance                                       | -0.02         | -0.04 – 0.01 | 0.270            |
| wall distance                                        | -0.03         | -0.12 – 0.06 | 0.525            |
| center distance                                      | 0.03          | -0.06 – 0.12 | 0.570            |
| VGC model                                            | 0.13          | 0.13 – 0.13  | <b>&lt;0.001</b> |
| Lateralization index (vert)                          | -0.00         | -0.02 – 0.02 | 0.740            |
| VGC model * Lateralization index (vert)              | 0.01          | 0.01 – 0.02  | <b>&lt;0.001</b> |
| Random Effects                                       |               |              |                  |
| $\sigma^2$                                           | 0.09          |              |                  |
| T00 SubjectID                                        | 0.01          |              |                  |
| T00 MazeID                                           | 0.00          |              |                  |
| ICC                                                  | 0.10          |              |                  |
| N <sub>MazeID</sub>                                  | 24            |              |                  |
| N <sub>SubjectID</sub>                               | 35            |              |                  |
| Observations                                         | 16686         |              |                  |
| Marginal R <sup>2</sup> / Conditional R <sup>2</sup> | 0.139 / 0.223 |              |                  |

**Table S16.**

**Lateralization moderation effect on trials with minimal eye movements (dataset dSC 1).**

## Additional information

### Funding

**Natural Sciences and Engineering Research Council of Canada**

- Jason da Silva Castanheira

**Wellcome Trust**

<https://doi.org/10.35802/206648> 

- Stephen M Fleming

**Royal Society (206648/Z/17/Z)**

- Stephen M Fleming

**European Research Council (101043666)**

- Stephen M Fleming

## References

1. Callaway F., et al. (2022) **Rational use of cognitive resources in human planning** *Nat. Hum. Behav.* **6**:1112–1125 [Google Scholar](#)
2. Newell A., Simon H. (1956) **The logic theory machine-A complex information processing system** *IRE Trans Inf Theory* **2**:61–79 [Google Scholar](#)
3. Huys Q. J. M., et al. (2012) **Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees** *PLOS Comput Biol* **8**:e1002410 [Google Scholar](#)
4. Dezfouli A., Balleine B. W. (2013) **Actions, Action Sequences and Habits: Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized** *PLOS Comput Biol* **9**:e1003364 [Google Scholar](#)
5. Griffiths T. L., et al. (2019) **Doing more with less: meta-reasoning and meta-learning in humans and machines** *Curr. Opin. Behav. Sci.* **29**:24–30 [Google Scholar](#)
6. Daw N. D., Niv Y., Dayan P. (2005) **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control** *Nat. Neurosci.* **8**:1704–1711 [Google Scholar](#)
7. Saxe A., Nelli S., Summerfield C. (2021) **If deep learning is the answer, what is the question?** *Nat. Rev. Neurosci.* **22**:55–67 [Google Scholar](#)
8. Hassabis D., Kumaran D., Summerfield C., Botvinick M. (2017) **Neuroscience-Inspired Artificial Intelligence** *Neuron* **95**:245–258 [Google Scholar](#)
9. Quinlan J. R. (1986) **Induction of decision trees** *Mach Learn* **1**:81–106 [Google Scholar](#)
10. Breslow L. A., Aha D. W. (1997) **Simplifying decision trees: A survey** *Knowl Eng Rev* **12**:1–40 [Google Scholar](#)
11. Huys Q. J. M., et al. (2015) **Interplay of approximate planning strategies** *Proc. Natl. Acad. Sci.* **112**:3098–3103 [Google Scholar](#)
12. Mingers J. (1989) **An Empirical Comparison of Pruning Methods for Decision Tree Induction** *Mach Learn* **4**:227–243 [Google Scholar](#)
13. Knuth D. E., Moore R. W. (1975) **An analysis of alpha-beta pruning** *Artif Intell* **6**:293326 [Google Scholar](#)
14. Keramati M., Smittenaar P., Dolan R. J., Dayan P. (2016) **Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum** *Proc. Natl. Acad. Sci.* **113**:12868–12873 [Google Scholar](#)

15. Snider J., Lee D., Poizner H., Gepshtein S. (2015) **Prospective Optimization with Limited Resources** *PLOS Comput Biol* **11**:e1004501 [Google Scholar](#)
16. Korf R. E. (1985) **Depth-first iterative-deepening: An optimal admissible tree search** *Artif Intell* **27**:97–109 [Google Scholar](#)
17. Kool W., Cushman F. A., Gershman S. J. (2016) **When Does Model-Based Control Pay Off?** *PLOS Comput Biol* **12**:e1005090 [Google Scholar](#)
18. Ho M. K., et al. (2022) **People construct simplified mental representations to plan** *Nature* **606**:129–136 [Google Scholar](#)
19. Gershman S. (2021) **What Makes Us Smart: The Computational Logic of Human Cognition** Princeton, NJ, US: Princeton University Press :205 [Google Scholar](#)
20. Landry M., da Silva Castanheira J., Jerbi K. (2023) **Differential and Overlapping Effects between Exogenous and Endogenous Attention Shape Perceptual Facilitation during Visual Processing** *J. Cogn. Neurosci.* **35**:1279–1300 [Google Scholar](#)
21. Landry M., Da Silva Castanheira J., Sackur J., Raz A. (2021) **Investigating how the modularity of visuospatial attention shapes conscious perception using type I and type II signal detection theory** *J. Exp. Psychol. Hum. Percept. Perform.* **47**:402–422 [Google Scholar](#)
22. Chica A. B., Bartolomeo P., Lupianez J. (2013) **Two cognitive and neural systems for endogenous and exogenous spatial attention** *Behav. Brain Res.* **237**:107–123 [Google Scholar](#)
23. Carrasco M. (2011) **Visual attention: The past 25 years** *Vision Res.* **51**:1484–1525 [Google Scholar](#)
24. Carrasco M., Ling S., Read S. (2004) **Attention alters appearance** *Nat. Neurosci.* **7**:308–313 [Google Scholar](#)
25. Nobre K., Kastner S. (2014) **The Oxford Handbook of Attention** OUP Oxford [Google Scholar](#)
26. Niv Y. (2019) **Learning task-state representations** *Nat. Neurosci.* **22**:1544–1553 [Google Scholar](#)
27. Lamme V. A., Roelfsema P. R. (2000) **The distinct modes of vision offered by feedforward and recurrent processing** *Trends Neurosci.* **23**:571–579 [Google Scholar](#)
28. Broadbent D. E. (1958) **Perception and Communication** Elsevier Science & Technology [Google Scholar](#)
29. Nelson M. D., Crisostomo M., Khericha A., Russo F., Thorne G. L. (2012) **Classic Debates in Selective Attention: Early vs Late, Perceptual Load vs Dilution, Mean RT vs Measures of Capacity** *Perception* **41**:997–1000 [Google Scholar](#)
30. Liu T., Jiang Y., Sun X., He S. (2009) **Reduction of the Crowding Effect in Spatially Adjacent but Cortically Remote Visual Stimuli** *Curr. Biol.* **19**:127–132 [Google Scholar](#)
31. He S., Cavanagh P., Intriligator J. (1996) **Attentional resolution and the locus of visual awareness** *Nature* **383**:334–337 [Google Scholar](#)

32. Whitney D., Levi D. M. (2011) **Visual Crowding: a fundamental limit on conscious perception and object recognition** *Trends Cogn Sci* **15**:160–168 [Google Scholar](#)
33. Norman D. A. (1968) **Toward a theory of memory and attention** *Psychol Rev* **75**:522–536 [Google Scholar](#)
34. Posner M., Snyder C., Davidson B. (1980) **Attention and the detection of signals** *J. Exp. Psychol. Gen.* **109**:160–174 [Google Scholar](#)
35. Keefe J. M., Stormer V. S. (2021) **Lateralized alpha activity and slow potential shifts over visual cortex track the time course of both endogenous and exogenous orienting of attention** *NeuroImage* **225**:117495 [Google Scholar](#)
36. Bagherzadeh Y., Baldauf D., Pantazis D., Desimone R. (2020) **Alpha Synchrony and the Neurofeedback Control of Spatial Attention** *Neuron* **105**:577–587 [Google Scholar](#)
37. Landry M., da Silva Castanheira J., Raz A., Baillet S., Sackur J. (2024) **A lateralized alpha-band marker of the interference of exogenous attention over endogenous attention** *Cereb Cortex* **34**:bhad457 [Google Scholar](#)
38. Jensen O. (2024) **Distractor inhibition by alpha oscillations is controlled by an indirect mechanism governed by goal-relevant information** *Commun Psychol* **2**:1–11 [Google Scholar](#)
39. Posner M. I. (1980) **Orienting of attention** *Q J Exp Psychol* **32**:3–25 [Google Scholar](#)
40. Eriksen C. W., James J. D. (1986) **Visual attention within and around the field of focal attention: A zoom lens model** *Percept Psychophys* **40**:225–240 [Google Scholar](#)
41. James W. W. (1890) **The Principles of Psychology, Vol I** Henry Holt and Co <https://doi.org/10.1037/10538-000> | [Google Scholar](#)
42. Vollebregt M. A., et al. (2015) **Lateralized modulation of posterior alpha oscillations in children** *NeuroImage* **123**:245–252 [Google Scholar](#)
43. Ghafari T., Mazzetti C., Garner K., Gutteling T., Jensen O. (2024) **Modulation of alpha oscillations by attention is predicted by hemispheric asymmetry of subcortical regions** *eLife* **12**:RP91650 <https://doi.org/10.7554/eLife.91650> | [Google Scholar](#)
44. Peelen M. V., Kastner S. (2014) **Attention in the real world: toward understanding its neural basis** *Trends Cogn Sci* **18**:242–250 [Google Scholar](#)
45. Kaiser D., Quek G. L., Cichy R. M., Peelen M. V. (2019) **Object Vision in a Structured World** *Trends Cogn Sci* **23**:672–685 [Google Scholar](#)
46. Cohen M. A., Cavanagh P., Chun M. M., Nakayama K. (2012) **The attentional requirements of consciousness** *Trends Cogn Sci* **16**:411–417 [Google Scholar](#)
47. Tsuchiya N., Koch C. (2014) **On the Relationship Between Consciousness and Attention** In: Gazzaniga M. S., Mangun G. R., editors. *The Cognitive Neurosciences* The MIT Press <https://doi.org/10.7551/mitpress/9504.003.0092> | [Google Scholar](#)

48. Koch C., Tsuchiya N. (2007) **Attention and consciousness: two distinct brain processes** *Trends Cogn Sci* **11**:16–22 [Google Scholar](#)
49. Lamme V. A. F. (2003) **Why visual attention and awareness are different** *Trends Cogn Sci* **7**:12–18 [Google Scholar](#)
50. MacIver M. A., Finlay B. L. (2022) **The neuroecology of the water-to-land transition and the evolution of the vertebrate brain** *Philos Trans R Soc Lond B Biol Sci* **377**:20200523 [Google Scholar](#)
51. Fleming S., Michel M. (2025) **Sensory Horizons and the Functions of Conscious Vision** *Behav Brain Sci* **21**:1–53 [Google Scholar](#)
52. Shea N., Frith C. D. (2016) **Dual-process theories and consciousness: the case for ‘Type Zero’ cognition** *Neurosci Conscious* **2016**:niw005 [Google Scholar](#)
53. Stocker A. A., Simoncelli E. P. (2007) **A Bayesian Model of Conditioned Perception** *Adv Neural Inf Process Syst* **2007**:1409–1416 [Google Scholar](#)
54. Block N. (2018) **If perception is probabilistic, why does it not seem probabilistic?** *Philos Trans R Soc B Biol Sci* **373**:20170341 [Google Scholar](#)
55. Mudrik L., Faivre N., Koch C. (2014) **Information integration without awareness** *Trends Cogn Sci* **18**:488–496 [Google Scholar](#)
56. Deroy O., et al. (2016) **The Complex Interplay Between Multisensory Integration and Perceptual Awareness** *Multisensory Res.* **29**:585–606 [Google Scholar](#)
57. Bekinschtein T. A., et al. (2009) **Classical conditioning in the vegetative and minimally conscious state** *Nat. Neurosci.* **12**:1343–1349 [Google Scholar](#)
58. Strauss M., et al. (2015) **Disruption of hierarchical predictive coding during sleep** *Proc. Natl. Acad. Sci. U. S. A.* **112**:E1353–1362 [Google Scholar](#)
59. Herzog M. H., Drissi-Daoudi L., Doerig A. (2020) **All in Good Time: Long-Lasting Postdictive Effects Reveal Discrete Perception** *Trends Cogn Sci* **24**:826–837 [Google Scholar](#)
60. Dickinson A., Balleine B. (2010) **Hedonics: The cognitive-motivational interface** In: Kringelbach M. L., Berridge K. C., editors. *Pleasures of the brain* New York, NY, US: Oxford University Press pp. 74–84 <https://doi.org/10.1093/oso/9780195331028.003.0006> | [Google Scholar](#)
61. Braver T. S., et al. (2014) **Mechanisms of motivation-cognition interaction: challenges and opportunities** *Cogn Affect Behav Neurosci* **14**:443–472 [Google Scholar](#)
62. Dung L. (2022) **Assessing tests of animal consciousness** *Conscious Cogn* **105**:103410 [Google Scholar](#)
63. McMains S. A., Somers D. C. (2004) **Multiple Spotlights of Attentional Selection in Human Visual Cortex** *Neuron* **42**:677–686 [Google Scholar](#)

64. Pylyshyn Z. W., Storm R. W. (1988) **Tracking multiple independent targets: Evidence for a parallel tracking mechanism\*** *Spat Vis* 3:179–197 [Google Scholar](#)
65. Awh E., Pashler H. (2000) **Evidence for split attentional foci** *J. Exp. Psychol. Hum. Percept. Perform.* 26:834–846 [Google Scholar](#)
66. Shaw M. L., Shaw P. (1977) **Optimal allocation of cognitive resources to spatial locations** *J. Exp. Psychol. Hum. Percept. Perform.* 3:201–211 [Google Scholar](#)
67. McMains S. A., Somers D. C. (2005) **Processing Efficiency of Divided Spatial Attention Mechanisms in Human Visual Cortex** *J. Neurosci.* 25:9444–9448 [Google Scholar](#)
68. Kr C., Np B. (1999) **Visuospatial attention: beyond a spotlight model** *Psychon. Bull. Rev.* 6 [Google Scholar](#)
69. Dubois J., Hamker F. H., VanRullen R. (2009) **Attentional selection of noncontiguous locations: The spotlight is only transiently “split”** *J. Vis.* 9:3 [Google Scholar](#)
70. Kinley I., Levy Y. (2022) **PuPI: an open-source tool for processing pupillometry data** *Behav. Res. Methods* 54:2046–2069 [Google Scholar](#)
71. Hershman R., Henik A., Cohen N. (2018) **A novel blink detection method based on pupillometry noise** *Behav. Res. Methods* 50:107–114 [Google Scholar](#)

## Author information

### Jason da Silva Castanheira

Department of Experimental Psychology, University College London, London, United Kingdom

**For correspondence:** [j.castanheira@ucl.ac.uk](mailto:j.castanheira@ucl.ac.uk)

### Nicholas Shea

Institute of Philosophy, School of Advanced Study, University of London, London, United Kingdom

### Stephen M Fleming

Department of Experimental Psychology, University College London, London, United Kingdom, Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, United Kingdom, Canadian Institute for Advanced Research (CIFAR), Brain, Mind and Consciousness Program, Toronto, Canada

## Editors

Reviewing Editor

### Nathan Faivre

Centre National pour la Recherche Scientifique et Technique (CNRS), Grenoble, France

Senior Editor

**Michael Frank**

Brown University, Providence, United States of America

**Reviewer #1 (Public review):**

**Summary:** This study investigated how visuospatial attention influences the way people build simplified mental representations to support planning and decision-making. Using computational modeling and virtual maze navigation, the authors examined whether spatial proximity and the spatial arrangement of obstacles determine which elements are included in participants' internal models of a task. The study developed and tested an extension of the value-guided construal (VGC) model that incorporates features of spatial attention for selecting simpler task mental representation.

**Strengths:**

- (1) **Original Perspective:** The study introduces an explicit attentional component to established models of planning, offering an approach that bridges perception, attention, and decision-making.
- (2) **Methodological Approach:** The combination of computational modeling, behavioral data, and eye-tracking provides converging measures to assess the relationship between attention and planning representations.
- (3) **Cross-validated data:** The study relies on the analysis of three separate datasets, two already published and an additional novel one. This allows for cross-validation of the findings and enhances the robustness of the evidence.
- (4) **Focus on Individual Differences:** Reports of how individual variability in attentional "spillover" correlates with the sparsity of task representations and spatial proximity add depth to the analysis.

**Weaknesses:**

- (1) **Clarity of the VGC model and behavioral task:** The exposition of the VGC model lacks sufficient detail for non-expert readers. It is not clear how this model infers which maze obstacles are relevant or irrelevant for planning, nor how the maze tasks specifically operationalize "planning" versus other cognitive processes.

The method for classifying obstacles as relevant or irrelevant to the task and connecting metacognitive awareness (i.e., participants' reports of noticing obstacles) to attentional capture is not well justified. The rationale for why awareness serves as a valid attention proxy, as opposed to behavioral or neurophysiological markers, should be clearer.

- (2) **Attention framework:** The account of attention is largely limited to the "spotlight" model. When solving a maze, participants trace the correct trail, following it mentally with their overt or covert attention. In this perspective, relevant concepts are also rooted in attention literature pertaining to object-based attention using tasks like curve tracing (e.g., Pooresmaeili & Roelfsema, 2014) and to mental maze solving (e.g., Wong & Scholl, 2024), which may be highly relevant and add nuance to the current work. This view of attention may be more pertinent to the task than models of simultaneously tracking multiple objects cited here. Prior work (notably from the Roelfsema group) indicates that attentional engagement in curve-tracing tasks may be a continuous, bottom-up process that progressively spreads along a trajectory, in time and space, rather than a "spotlight" that simply travels along the path. The spread of attention depends on the spatial proximity to distractors - a point that could also be pertinent to the findings here.

Moreover, the tracing of a "solution" trail in a maze may be spontaneous and not only a top-down voluntary operation (Wong & Scholl, 2024), a finding that requires a more careful framing of the link to conscious perception discussed in the manuscript.

Conceptualizing attention as a spatial spotlight may therefore oversimplify its role in navigation and planning. Perhaps the observed attentional modulation reflects a perceptual stage of building the trail in the maze rather than a filter for a later representation for more efficient decision making and planning. A fuller discussion of whether the current model and data can distinguish between these frameworks would benefit readers.

(3) Lateralization of attention: The analysis considers whether relevant information is distributed bilaterally or unilaterally across the visual display, but does not sufficiently address evidence for attentional asymmetries across the left and right visual fields due to hemispheric specialization (e.g., Bartolomeo & Seidel Malkinson, 2019). Whether effects differ for left versus right hemifield arrangements is not made explicit in the presented findings.

(4) Individual differences: Individual differences in attentional modulation are a strength of the work, but similar analyses exploring individual variation in lateralization effects could provide further insight, and the lack of such analyses may mask important effects.

(5) Distinction between overt and covert attention: The current report at times equates eye movement patterns with the locus of attention. However, attention can be covertly shifted without corresponding gaze changes (see, for example, Pooremaeili & Roelfsema, 2014).

The implications for interpreting the relationship between eye movement, memory, and attention in this setting are not fully addressed. The potential dynamics of attention along a maze trajectory and their impact on lateralization analysis would benefit from further clarification.

#### Appraisal of Aims and Results:

The study sets out to determine how spatial attention shapes the construction of task representations in planning contexts. The authors provide evidence that spatial proximity and arrangement influence which environmental features are incorporated into internal models used for navigation, and that accounting for these effects improves model predictions. There is clear documentation of individual variation, with some participants showing greater attentional spillover and more sparse awareness profiles.

However, some conceptual and methodological aspects would be clearer with greater engagement with the broader literature on attention dynamics, a more explicit justification of operational choices, and more targeted lateralization analyses.

<https://doi.org/10.7554/eLife.108034.1.sa2>

#### Reviewer #2 (Public review):

##### Summary:

Castanheira et al. investigate the role of spatial attention for planning during three maze navigation experiments (one new experiment and two existing datasets). Effective planning in complex situations requires the construction of simplified representations of the task at hand. The authors find that these mental representations (as assessed by conscious awareness) of a given stimulus are influenced by (spatially) surrounding stimuli. Individual participants varied in the degree to which attention influenced their task representations, and this attentional effect correlated with the sparsity of representations (as measured by the range of awareness reports across all stimuli). Spatially grouping task-relevant information

on either the left or right side of the maze led to mental representations more similar to optimal representations predicted by the value-guided construal (VGC) model - a normative model describing a theoretical approach to simplifying complex task information. Finally, the authors propose an update to this model, incorporating an attentional spotlight component; the revised descriptive model predicts empirical task representations better than the original (normative) VGC model.

#### Strengths:

The novelty of this study lies in the proposal and investigation of a cognitive mechanism through which a normative model like value-guided construal can enable human planning. After proposing attention as this mechanism, the authors make concrete hypotheses about mismatches between the VGC predictions and real human behavior, which are experimentally validated. Thus, not only does this study describe a possible mechanism for simplification of task information for planning, but the authors also propose a descriptive model, revising VGC to incorporate this attentional component.

A strength of this paper is the variety of investigative approaches: analysis of existing data, novel experiment, and a computational approach to predict experimental findings from a theoretical model. Analyzing pre-existing datasets increases the size of the participant cohort and strengthens the authors' conclusions. Meanwhile, comparing the predictions of the existing normative model and the authors' own refined model is a clever approach to substantiate their claims. In addition, the authors describe several crucial controls, which are key to the interpretability of their results. In particular, the eye tracking results were critical.

In summary, this paper constitutes an important step toward a more complete understanding of the human ability to plan.

#### Weaknesses:

(1) There is a critical conceptual gap in the study and its interpretation, mainly due to the reliance on a self-report metric of awareness (rather than an objective measure of behavioral performance).

a. Awareness is tested by a 9-point self-report scale. It is currently unclear why awareness of task-irrelevant obstacles in this task would necessarily compromise optimal planning. There is no indication of whether self-reported awareness affects performance (e.g., navigation path distance, time to complete the maze, number of errors). Such behavioral evidence of planning would be more compelling.

b. Relatedly, it would have been more convincing to have an objective measure of awareness, for instance, how the presence or absence of a "task-irrelevant" obstacle affects performance (e.g., change navigation path distance or time to complete the maze), or whether participants can accurately recall the location of obstacles.

c. Consequently, I'm not sure that we can conclude that the spatial context does impact participants' ability to plan spatial navigation or to "incorporate task-relevant information into their construal". We know that the spatial context affects subjective (self-reported) awareness, but the authors do not present evidence that spatial context affects behavioral performance.

d. Another concern that may complicate interpretation is the following: Figure 3c shows improved VGC model predictions (steeper slope) for mazes with greater lateralization. However, there are notable outliers in these plots, where a high lateralization index does not correspond to good model performance. There is currently no discussion/explanation of these cases.

(2) I noticed an issue with clarity regarding task-relevance. It is currently not fully clear which obstacles are "task irrelevant". Also, the term is used inconsistently, sometimes conflating with "awareness". For example, in the "Attentional spotlight model of task representations" section, the authors state that "task-relevant information becomes less relevant when surrounded by task-irrelevant information". But they really mean that participants become less aware of those task-relevant obstacles. I assume task-relevance is an objective characteristic related to maze organization, not to a participant's construal. Indeed, the following paragraph provides evidence of model predictions of awareness.

(3) The behavioral paradigm has some distinct disadvantages, and the validity of the task is not backed up by behavioral data.

- a. I understand the need for central fixation, but it also makes the task less naturalistic.
- b. The task with its top-down grid view does not seem to mimic real human navigation. Though this grid may be similar to mental maps we form for navigation, the sensory stimuli corresponding to possible paths and to spatial context during real-life navigation are very different.
- c. Behavioral performance is not reported, so it is unknown whether participants are able to properly complete the task. The task seems pretty difficult to navigate, especially when the obstacles disappear, and in combination with the central fixation.
- d. There is no discussion of whether/how this navigation task generalizes to other forms of planning.

<https://doi.org/10.7554/eLife.108034.1.sa1>

### Reviewer #3 (Public review):

#### Summary:

The authors build on a recent computational model of planning, the "value-guided construal" framework by Ho et al. (2022), which proposes that people plan by constructing simple models of a task, such as by attending to a subset of obstacles in a maze. They analyze both published experimental data and new experimental data from a task in which participants report attention to objects in mazes. The authors find that attention to objects is affected by spatial proximity to other objects (i.e., attentional overspill) as well as whether relevant objects are lateralized to the same hemifield. To account for these results, the authors propose a "spotlight-VGC" model, in which, after calculating attention scores based on the original VGC model, attention to objects is enhanced based on distance. They find that this model better explains participant responses when objects are lateralized to different hemifields. These results demonstrate complex interactions between filtering of task-relevant information and more classical signatures of attentional selection.

#### Strengths:

- (1) The paper builds on existing modeling work in a novel manner and integrates classic results on attention into the computational framework.
- (2) The authors report new and extensive analyses of existing data that shed light on additional sources of systematic variability in responses related to attentional spillover effects
- (3) They collect new data using new stimuli in the original paradigm that directly test predictions related to the lateralization of task-relevant information, including eye tracking

data that allows them to control for possible confounds.

(4) The extended model (spotlight-VGC) provides a formal account of these new results.

Weaknesses:

(1) The spotlight-VGC model has a free parameter - the "width" of the attentional spotlight. This seems to have been fixed to be 3 squares. It would be good if the authors could describe a more principled procedure for selecting the width so that others can use the model in other contexts.

(2) Have the authors considered other ways in which factors such as attentional spillover and lateralization could be incorporated into the model? The spotlight-VGC model, as presented, involves first computing VGC predictions and only afterwards computing spillover. This seems psychologically implausible, since it supposes that the "optimal" representation is first formed and then it gets corrupted. Is there a way to integrate these biases directly into the VGC framework, perhaps as a prior on construals? The authors gesture towards this when they talk about "inductive biases", but this is not formalized.

(3) Can the authors rule out that the lateralization effects are the result of memory biases since the main measure used is a self-report of attention?

<https://doi.org/10.7554/eLife.108034.1.sa0>