Spatial coding for action across spatial scales

Katja Fiehler[1, 2†] and Harun Karimpur[1, 2]

[1]Experimental Psychology, Justus Liebig University Giessen, Giessen, Germany

[2]Center for Mind, Brain and Behavior, University of Marburg and Justus Liebig University Giessen, Germany

†e-mail: katja.fiehler@psychol.uni-giessen.de

Published in:

Fiehler, K., & Karimpur, H. (2023). Spatial coding for action across spatial scales. *Nature Reviews Psychology, 2*(2), 72-84.

Abstract

Humans perform goal-directed actions such as reaching for a light switch or grasping a coffee mug, thousands of times a day. Behind the scenes of these seemingly simple actions, the brain performs sophisticated calculations to locate the target object of the action and correctly guide the hand towards it. In this Review, we discuss how the brain establishes spatial representations used for visually-guided actions. In addition to reviewing simple tasks and paradigms, we discuss spatial coding in complex and naturalistic environments. We highlight the importance of high-level cognitive factors, such as memory, task constraints and object semantics, which influence the use of spatial representations for action. To move the field forward, we suggest that future research should integrate across different scales of action spaces from small-scale finger movements to large-scale navigation. Doing so would enable identification of general mechanisms that underlie spatial coding across different actions and spaces.

Introduction

In everyday life, humans continuously interact with their environment by executing goal-directed actions. Grasping a cup, flipping a light switch or sliding one's finger across a mobile phone to answer a call are seemingly simple actions that can be performed with hardly any apparent effort. However, multiple spatial transformations are required to achieve the intended action goal. A spatial representation of the target needs to be established, spatially updated every time the body moves and finally transformed into specific commands that drive the movements of the respective body part. Spatial coding for action refers to how spatial representations for goal-directed actions are established and used.

To achieve a desired motor output (for example, grasping a cup), a spatial representation of the target object (cup) and the effector (hand) needs to be established for later use in calculating the movement path that brings the hand to the cup. Effectors are body parts used to execute an action, such as the hand for picking up a mug or the foot for kicking a ball. Different spatial representations[1,2] are involved in goal-directed actions. Given the two dimensional (2D) nature of the visual representations and the three-dimensional (3D) nature of the motor signals, visual information must undergo sensorimotor transformation into effector-specific movement commands[3,4]. Accounts of how visual target information is used to guide goal-directed actions make different assumptions about what is represented (for example, visual target vs. action outcome) and to which degree sensorimotor transformations are needed[5–7]. Despite conceptual differences in these accounts, sensorimotor transformations have been demonstrated on the behavioral, neural and computational level and for different effectors, such as arm[8,9] and eye movements[3,10]. Spatial coding for action relies on the computation of the spatial location of objects to which we want to guide our effectors to successfully interact with the environment.

Previous findings on spatial coding for action have primarily relied upon highly controlled laboratory experiments. In these paradigms, experimenters presented simple and abstract visual stimuli, such as dots, lines or LED lights that participants had to point to or reach for (FIG. 1)[11,12]. These experiments have provided valuable results, but they lack key facets of everyday behaviour (such as sequential movements) and the complexity of natural environments which are multisensory by nature. Studies of spatial coding for grasping provide greater sensory and motor complexity by presenting real 3D target objects and involving multi-joint arm and hand movements. Furthermore, grasping 3D objects requires spatial coding of visible and invisible (hidden by the object) grasp locations [13] and the continuous processing of haptic feedback. Positional information of vision and touch is flexibly integrated depending on the current sensory feedback[14], leading to improved grasping performance[15]. When visual and haptic information do not match the scaling of the grip is

3

generally reduced and the natural range of individual differences in grip scaling largely restricted[16]. Even if consistent haptic feedback is provided, visual information of the real 3D object as compared to a virtual 3D object is required for natural grasping[16]. To fully understand spatial coding for action, researchers need to consider the multisensory 3D nature of real-world environments.
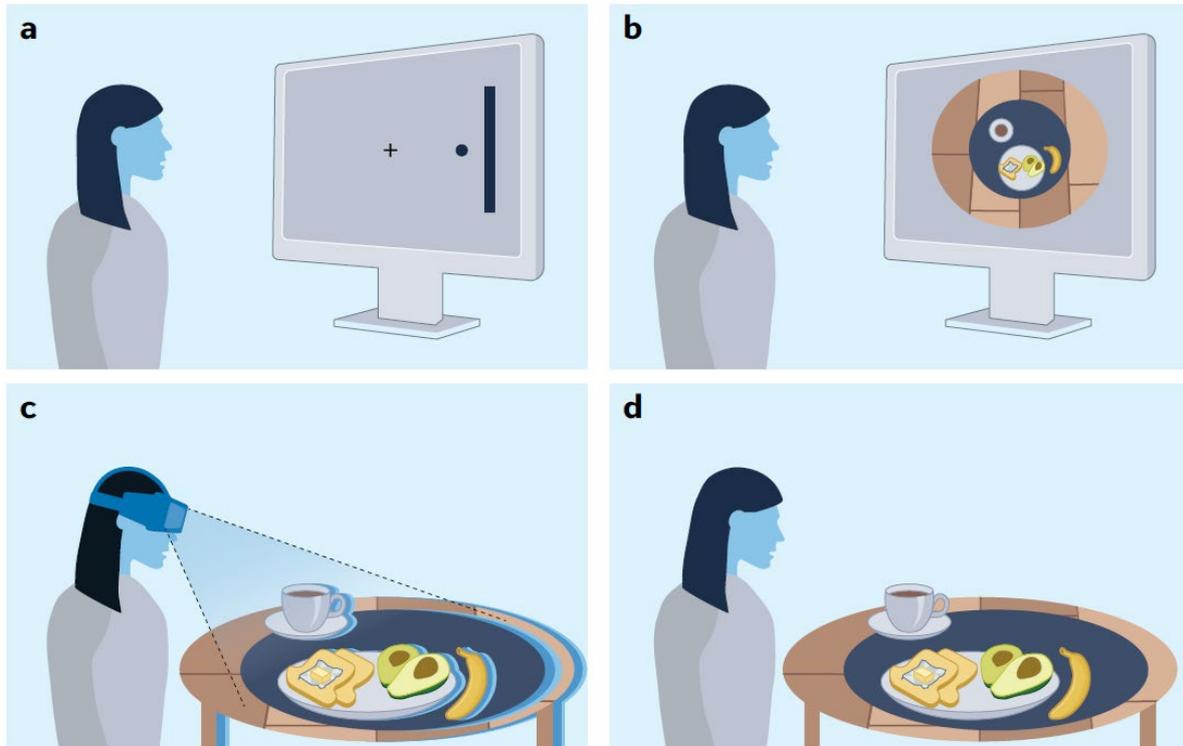
Figure 1: Example experimental tasks used to study spatial coding for action. These include the use of abstract stimuli (a) and naturalistic images (b) presented on a monitor, to virtual environments presented in virtual reality (c), to real environments (d).

In the real world, actions differ in multiple features, including their timing (executed immediately or after a short delay), the actor's goals, and target context. These complexities lead to influences of high-level cognitive factors such as memory, task goals, and semantic processing on action. These three factors have clear implications for spatial information processing, and each represent one aspect of the perception-action cycle by shedding light on mechanisms related to the agent, the action itself or the action target. Another key focus for research is considering actions across spatial scales (from small spaces within reach to large spaces explored by walking), which will provide a greater understanding of ecologically-valid actions.

In this Review, we provide an overview of how spatial representations of single or multiple action targets are established. Thus, we focus on the early stages of spatial coding for action; spatial updating of visual target representations and later stages of sensorimotor transformation and movement execution are beyond the scope of this review (for more information[17,3]). We consider influences on how spatial representations are established, highlighting the role of high-level

4

cognitive factors. Next, we review the use of spatial representations across different scales from small-scale to large-scale movements. Finally, we suggest future directions for combining findings across different scales of action spaces and different concepts to identify general mechanisms that underlie spatial coding in natural human behaviour.

Spatial coding schemes

Establishing and maintaining spatial target representations are required for movement planning. Using a 2D visual image on the retina, the brain computes spatial representations that can be used to guide actions in the 3D world. This transformation from vision to action unfolds gradually in space and time[18]. In the brain, the movement plan is represented in the posterior parietal cortex (PPC) and dorsal premotor cortex (dPM) and then transferred to the primary motor cortex[3,17,19]. The primary motor cortex generates a motor command which is sent downstream to the spinal cord to activate the relevant muscles and finally move the effector to the desired location[20].

We focus on the processing stages related to movement planning, in particular the establishment of spatial target representations for arm movements and the brain areas crucially involved in this process, PPC and dPM. The importance of spatial representations becomes clear when brain regions within the frontoparietal network are damaged, such as after a stroke. For example, patients with optic ataxia with lesions in the PPC show impaired visuomotor behaviour, such as difficulties in reaching to targets in their visual periphery compared to targets in central vision[21]. This highlights the central role of the PPC in establishing visuospatial representations and transforming them into motor commands. In the following, we review the spatial coding schemes used in action coding, their neural implementation as well as their impact on behaviour. We discuss the distinction between visual and motor goal representations, how coding changes when multiple action targets are present, as well as the use of egocentric and allocentric reference frames.

Visual and motor goals

In everyday actions, the action target one sees (the visual goal) and where they would like to move their hand (the motor goal) are typically aligned. One sees a mug, and reaches their hand to it. Thus, it is difficult to experimentally examine whether the two goals share common spatial representations. However, the pro-movement and anti-movement paradigm is well-suited to combine identical visual inputs with different motor outputs, dissociating the visual and motor goal[22]. Participants are presented with a visual target followed by a cue that instructs them whether the movement should be directed towards the target (pro-reach) or towards the mirror opposite location (anti-reach) (FIG. 2). In the anti-reach condition, participants infer the reach goal from the position of the visual target, requiring a spatial transformation[23]. Due to the spatial selectivity of neurons in the frontoparietal network, the presence of lateralized brain activation can inform

whether the visual or motor goal information is represented[24,25]. Electrophysiological studies in non-human primates and functional neuroimaging studies in humans have shown that activity in the PPC can be dissociated from the visual goal[26,27]. The PPC sometimes shows transient activation associated with the visual goal but ends up aligned with the movement direction[28].
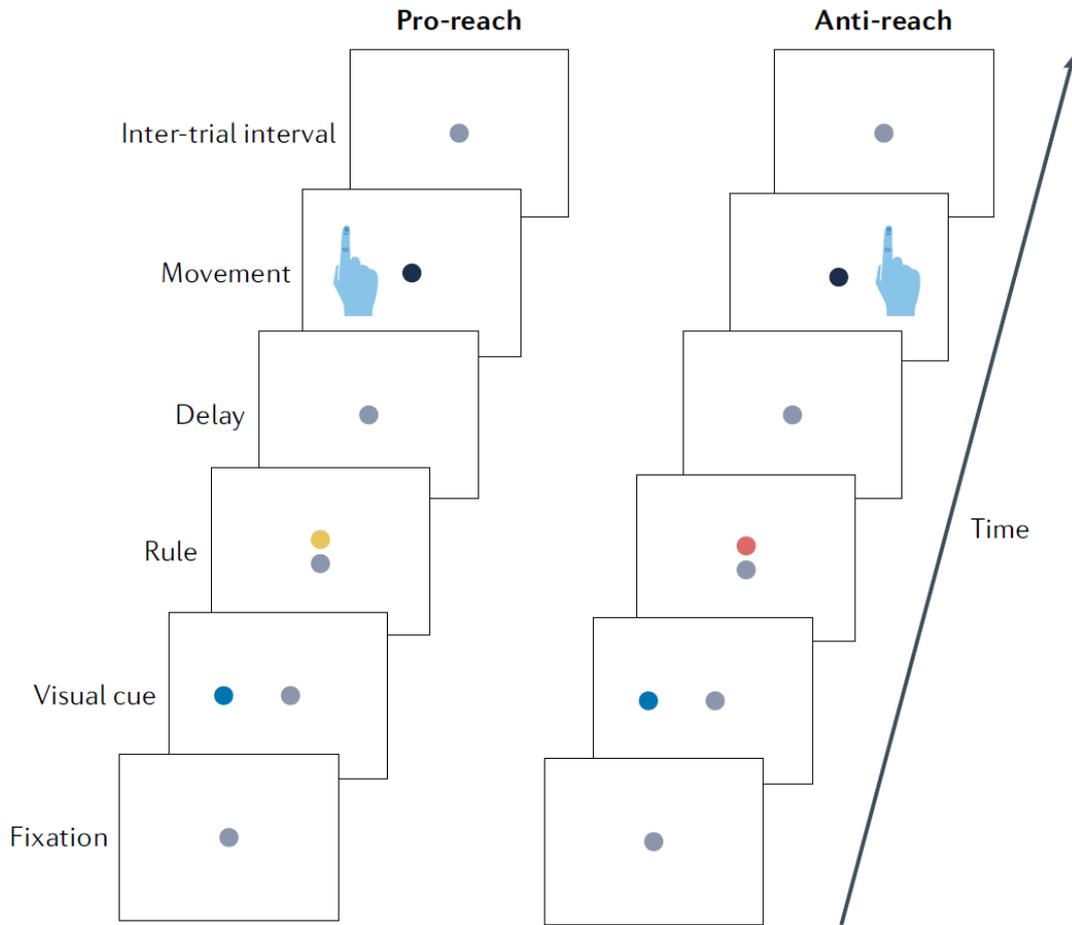
Figure 2: Example trial structure of a pro-reach and anti-reach task. Participants fixate at the center of the screen throughout the trial (gray circle). A visual cue is presented (blue circle), followed by a rule cue indicating whether participants should reach to the location of the visual cue (yellow circle, pro-reach) or to its mirror opposite location (red circle, anti-reach). After a short delay, they perform the reach within a specified time window. The next trial starts after an inter-trial interval (ITI).

Directionally-selective activity aligned to the motor goal was also found when participants wear prism goggles that reverse the spatial contingency between the viewed target and the direction of the movement, requiring a movement to the right to reach a target located to the left[29]. These findings suggest that the PPC does not represent the visual target, rather it translates this information into the motor goal. Similar spatial coding patterns were found in the dPM, which is tightly connected to the PPC[27], but also to the primary motor cortex[30] and the spinal cord[31], reflecting a stronger motor-related character of the PMd compared to the PPC. Both PPC and dPM encode the motor goal and can maintain this spatial representation throughout a delay of several seconds[32,33].

145 Additionally, both represent the effector selected for the movement (for instance, left versus right

146 arm). As their activation is stronger when both the motor goal and the effector are specified, PPC

147 and dPM seem to be optimally suited to transform visuospatial representations into motor

148 commands[26,34].

149 ***Multiple movement targets***

150   In simple experimental settings, only a single action target might be presented. But in

151 everyday behaviour, humans usually select a target from a number of possible alternatives under

152 consideration. Evidence for the active consideration of multiple action targets comes from reaching

153 experiments in which participants are shown multiple potential action targets and the final target is

154 only specified after the start of the movement. The resulting hand movement direction corresponds

155 to the average movement towards each of the targets, which reflects the number and position of all

156 potential motor goals[35]. As soon as the motor goal is specified, the movement trajectory is adjusted

157 on the fly to successfully end up at the target. A similar averaging strategy was found for wrist

158 orientation, which is influenced by all potential target orientations[36]. However, spatial averaging is

159 only present in rapid movements under tight time-constraints[37]. When sufficient time for movement

160 planning is available, humans tend to pre-select one target and only modify their movement path

161 online if correction is required[37]. Therefore, changes in the aiming target on-the-fly is more frequent

162 under moderate time constraints[38].

163   The aforementioned results leave open whether multiple competing movement plans are

164 built in parallel or whether a single averaged movement plan is specified. To distinguish between

165 these options, an experiment was run in which an obstacle was placed close to one of two targets,

166 which produced a change of the initial movement if the obstacle was present, despite no change in

167 the target locations[39]. This result was inconsistent with a simple averaging of the visuospatial target

168 locations, but instead corresponded to the averaged direction of multiple movement plans.

169 Accordingly, the researchers suggested that the brain simultaneously prepares multiple fully

170 specified reaching movements to all potential targets, encoding motor goals rather than visual

171 goals[40]. Theses behavioural results are in line with electrophysiological and functional neuroimaging

172 studies in humans[41,42] and non-human primates[43,44] that show activity in the frontoparietal network

173 that represents parallel competing motor goals prior to action selection.

174   Preparing multiple movements can be advantageous because the movement generated to

175 the selected target can borrow components of the non-selected movement to an alternative target,

176 resulting in shorter times to start the movement (reaction times) and shorter times to complete the

177 movement (movement time). This co-optimization of motor plans is assumed to be highly automatic

178 and to occur largely outside of conscious awareness[45]. However, results from experimental studies

179 and computational modelling in the past two years challenge the spatial averaging account. A new

180 proposal is that humans instead generate a single motor plan that optimizes task performance

181 rather than prepare for all potential movement goals in parallel and then average the corresponding

182 motor plans[46]. The associated mechanistic framework for this optimization process can improve the

183 understanding of motor goal representation and motor planning in certain and uncertain

184 environments and stimulate future research.

185 ***Egocentric reference frames***

186 Different spatial reference frames can be used to define the location of an action target. A

187 reference frame is a rigid body in which coordinate axes are embedded. Two broad classes of

188 reference frames have been proposed: egocentric and allocentric, which differ in the point of origin

189 of their coordinate axes[1,2,47] (FIG. 3). In egocentric reference frames, target locations are represented

190 relative to the observer and anchored to a specific body part. Egocentric reference frames are

191 termed according to the body part that serves as the point of origin, for instance, head-centered,

192 hand-centered or body-centered reference frames. For movement control, the stable insertion point

193 of a set of muscles is generally chosen as the egocentric reference frame, such as the torso for head

194 movements[48].

195

196 **Figure 3:** Reference frames coding the location of a target object (mug) for grasping. In egocentric reference frames (ego),

197 targets are encoded relative to the observer, such as the mug relative to the direction of gaze (red, dashed arrow). In

198 allocentric reference frames (allo), targets are encoded relative to other objects or landmarks in the environment, such as

199 the mug relative to the corner of the desk (yellow, solid arrow).

200

201 Visual targets for eye and arm movements are primarily represented in a gaze-centered

202 egocentric reference frame, which represents the location of the movement target relative to gaze

203 direction[17,49]. In a later processing stage, these gaze-centered spatial target representations are

204 transformed into an effector-specific motor frame (such as limb-centered for reaching) that can be

8

205 read out by the motor system to generate a joint-centered muscle-based motor command[3,8]. Thus,

206 the visual target representation is transformed into coordinates suitable for producing the proper

207 muscle contraction to guide the effector to the target.

208 Behavioral evidence for gaze-centered spatial coding and updating of action targets was

209 provided by investigating directional pointing errors toward remembered visual targets [11]. Taking

210 advantage of the fact that pointing movements show a directional bias when the target is viewed in

211 the visual periphery[50], in this paradigm participants fixated a briefly flashed target before performing

212 an eye movement to place the location of the no-longer-visible target into the visual periphery.

213 When participants pointed to the remembered target location, their pointing movements were

214 biased in the same direction as pointing movements to targets viewed in the visual periphery. This

215 result suggests that participants established a gaze-centered visual target representation that was

216 updated into the visual periphery when they moved their eyes, and this updated target

217 representation was then used to calculate the movement vector, leading to directional reaching

218 errors. This directional bias in reaching is one of the key behavioural measures of gaze-centered

219 spatial coding and updating. Similar results have been found for grasping[51] and reaching visual

220 targets in near (reachable) and far (beyond reach) space[52]. For proprioceptive, tactile and auditory

221 targets, positional judgements systematically vary with gaze direction. When participants reached to

222 their unseen own index finger, they reach too far to the right when gaze was directed to the left of

223 the target hand and vice versa[53]. Such gaze-dependent directional errors also occurred when the

224 location of a tone had be adjusted so that it was perceived in the median plane of the head[54]. This

225 suggests that gaze-centered reference frames are applied across different actions and sensory target

226 modalities[55].

227 Neural evidence for gaze-centered spatial coding of action targets has been found in the PPC

228 of human[56,57] and non-human[12] primates that codes and updates reach and grasp targets relative to

229 gaze direction and lesions in the human PPC can disturb gaze-centered spatial updating[58,59]. Activity

230 in the human PPC reflects the location of the action target relative to gaze position, for example a

231 visual reach target viewed to the left of gaze fixation leads to a stronger right than left hemisphere

232 PPC activity[56]. Due to the gaze-centered spatial representation of action targets, when a lateral eye

233 movement brings the visual target to the opposite side of gaze fixation, information must be

234 exchanged across the PPC hemispheres.

235 Although gaze-centered reference frames play a primary role in spatial coding of visual

236 action targets[60], body-centered[61] and head-centered[62,63] representations contribute as well. One

237 critical factor determining the preferential use of reference frames is the sensory target modality.

238 Visual targets reach us via our eyes in a gaze-centered reference frame, tones reach us via our ears

that are fixed on the head in a head-centered reference frame and proprioception and touch via our body in a body-centered reference frame. Accordingly, body-centered reference frames contribute to goal-directed reaching especially in situations when the reach target is unseen and needs to be derived from proprioceptive information[61]. Likewise, head-centered target representations are involved in eye movements directed to auditory targets[62]. There is not a single frame of reference to represent action targets, instead spatial coding for action is of a hybrid nature in which multiple reference frames are expressed[4,55]. Such a combination of multiple egocentric reference frames facilities the use of the available sensory information and allows for their flexible use depending on the sensory target modality.

Allocentric reference frames

The second class of reference frames are allocentric reference frames, which represent target locations with respect to other objects (also called landmarks) or features in the environment (FIG. 3). Therefore, allocentric reference frames are also referred to as object-centered or world-centered reference frames. In contrast to egocentric reference frames, the point of origin lies outside the observer and therefore does not change if the observer changes position.

The use of allocentric information has been shown to improve movement performance[64–66]. In memory-guided movements, allocentric representations are more advantageous than egocentric representations because they are more spatially invariant to changes in their point of origin and therefore more stable over time. However, these representations can vary with the observer's viewpoint. Allocentric information can also misguide movements as reflected in systematic reaching errors when the position of a landmark changes[67,68].

Lesion[69] and functional neuroimaging studies in humans[70,71] suggest that inferior occipital-temporal brain areas encode action targets in an allocentric reference frame relative to visual landmarks. However, the functional and anatomical segregation of egocentric and allocentric encoding is challenged by electrophysiological results in monkeys showing that the same neurons in PPC and dPM that encode egocentric target information can also encode object-centered allocentric target information depending on dynamically changing task demands[72].

There is converging evidence that spatial coding for action relies on both classes of reference frames: When egocentric and allocentric information are available, reaching performance can be best explained by a combination of both types of information rather than either source alone[73–75]. One framework that could explain the combination of egocentric and allocentric representations is optimal Bayesian integration, in which multiple sources of noisy sensory information are combined in a statistically optimal fashion based on their variability. Information with high variability (less reliable) is assigned a low weight and information with low variability (highly reliable) is assigned a

high weight, leading to a combined representation that is closer to the source with the lower variability[76–79]. Bayesian integration leads to less variability in the combined representation compared to the variability of the single information source[76,80,81].

This framework was originally proposed for the combination of information from multiple sensory modalities, but is also a powerful approach for the integration of spatial representations across sensory modalities[82]. In one experiment, the reliability of egocentric and allocentric information was manipulated and reach behaviour measured to determine the use of the two sources of information. To change the reliability of egocentric information, participants performed large or small gaze shifts after encoding the target (before reaching) to introduce more or less variability in the gaze-centered representation of the target[67]. Reliability of the allocentric target information was varied by adding large or small jumps to the spatial position of four visual landmarks surrounding the target. This manipulation also influenced the perceived stability of the landmarks. Consistent with the Bayesian framework, egocentric and allocentric target information were combined in a statistically optimal fashion based on their variabilities in reaching movements. Landmark stability also influenced the weighting of allocentric information. However, the role of landmark stability can vary by task; changes in the task (perception or action) can alter the weighting of spatial information and even learned unstable landmarks (such as a moving person) can become reliable[83]. Similarly, more variable shifts of landmarks[84] or larger distances between target and landmarks[85] result in an overall weaker contribution of allocentric information.

A promising extension of the optimal Bayesian integration approach is causal Bayesian integration[86]. In this framework, information is also weighted by the probability that two sources of information share a common cause. In cases of a high probability of a common cause (such as two stimuli moving in the same direction), they are weighted more strongly than if distinct causes are probable[87]. The power of this framework in explaining spatial coding for real-time and memory-guided movements needs to be tested in future work.

In summary, movement planning relies on spatial representations of the motor goal, which are implemented in the PPC and dPM. When actors are presented with more than one movement option, multiple competing motor goals are represented in parallel before implementing one of them. Such parallel specification might enhance rapid, effective movements in dynamically changing environments and under uncertainty. A competing point of view suggests that the brain generates a single motor plan that optimizes task performance. Motor goals are represented in multiple spatial reference frames before they are converted into a spatial frame suitable for read out by the motor system. When both egocentric and allocentric information are available, they are combined and movement performance is usually improved. The extent to which both sources of information

307 contribute to spatial coding for action depends on various high-level cognitive factors that we

308 elaborate upon next.

Cognitive factors in spatial coding

310   Moving beyond simple and abstract stimuli, spatial coding for complex naturalistic actions

311 cannot be fully understood by investigating only low-level perceptual factors, such as those innate to

312 space (such as proximity)[66,88] or the object (such as saliency)[89,90]. High-level cognitive factors play a

313 crucial role in understanding spatial coding for action. Here we review three essential factors related

314 to the agent (short-term memory), the action itself (task constraints) and the action target (object

315 semantics).

316 ***Memory effects***

317   Goal-directed actions are usually performed on objects in view. Such movements are

318 characterized by continuous visual feedback about the target, surrounding objects, their location

319 and the moving effector (real-time movements). Real-time movements towards visual targets are

320 typically highly accurate and precise[65] and rely on brain structures associated with the dorsal action

321 pathway that projects into the PPC[91]. Actions can also be successfully directed to objects after a

322 short temporal delay using memory representations of the previously viewed object (memory-

323 guided movements). Memory-guided movements are less accurate and more variable than real-time

324 movements[92,93] and these effects increase with longer memory delays due to decay of visual

325 information in memory[94–96]. In contrast to real-time movements, memory-guided movements are

326 processed along both the dorsal action and ventral perception pathways[97–99], with the latter

327 projecting to the inferior temporal cortex that is interconnected with memory structures in the

328 medial temporal lobe[91].

329   Egocentric and allocentric coding are primarily each involved in real-time and memory-

330 guided movements, respectively[100,101]. However, this distinction is not perfectly clear-cut[102,103]. For

331 example, egocentric (gaze-centered) coding was found in memory-guided reaching with delays of up

332 to twelve seconds[104,105], and allocentric coding was shown in real-time reaching[65]. Additionally,

333 dynamic allocentric information such as applied background motion influences both memory-

334 guided[106] and real-time movements[68,107,108].

335   Real-time movements likely rely on implicit online corrections based upon updated

336 allocentric information[109]. This influence was tested in an object-shift paradigm. In this paradigm,

337 participants reach to a visual target while all other objects presented in a scene (allocentric cues) are

338 unnoticeably shifted to the left or right during an eye blink (induced by brief air puff)[110]. These shifts

339 occurred at different times with respect to the movement onset. According to the real-time control

340 of action hypothesis, real-time movements require visual information about the action target at the

time of movement onset and rely on visuomotor mechanisms. In the other two situations where visual target information is provided directly prior to movement onset (memory-guided movements) or with some temporal delay before movement onset (delayed memory-guided movements), actions rely on perceptual mechanisms[101]. The results of the object shift paradigm showed systematic reaching errors in the direction of the objects' shift (allocentric effect). These errors were found irrespective of the type of movement, confirming the use of allocentric information in real-time, memory-guided, and delayed memory-guided movements. However, the allocentric influence was stronger in memory-guided movements (irrespective of delay) compared to real-time movements. This finding is in line with evidence that egocentric representations degrade over temporal delays whereas allocentric representations remain relatively stable[111,112]. Consequently, allocentric cues improve the accuracy and precision of memory-guided movements[64,113,114]. On the other hand, in real-time movements visual feedback of the target and the hand provides highly reliable egocentric information. Therefore, egocentric reference frames dominate in guiding ongoing actions even if reliance upon allocentric cues would be advantageous, for example when movement planning is easier when representing an action target relative to a cursor on the screen (allocentric) than to the own hand (egocentric)[115,116].These findings suggest that real-time and memory-guided actions rely on the same spatial reference frames whereas memory demands alter the contribution of egocentric and allocentric representations.

Task effects

Real-world tasks often have well-defined goals (such as making a cup of tea), which make specific demands on the spatial coding system for action. The effects of a given task on spatial coding for action can be tested by instructing participants about the task's goals or subgoals and the steps required to reach these goals. For example, participants could be asked to find a kitchen tool, grasp it and hand it over to another person with the requirement to rotate the tool in a way so that the other person can grasp the handle. Such task instructions are often classified as top-down factors[117,118], in contrast to bottom-up factors such as physical salience (such as a bright and colorful cup) or object history (such as using the same cup every day).

Tasks help to structure behaviour and selection of task-relevant objects (for instance, tea bag and cup) from task-irrelevant objects (such as a coffee filter or pan). Selecting some objects at the expense of others to guide immediate or subsequent behaviour is essential given the plethora of information available in naturalistic scenes and humans' limited information processing capacities[119]. Attentional selection has been conceptualized as a priority map that regulates the sensory input by enhancing the perceptual representation of the attended objects and attenuating noise[120–122]. Perceptual performance (accuracy and speed in detection, discrimination and location tasks) at the

375 attended location is improved[123] which facilitates subsequent motor behaviour. Thus, attended

376 locations are associated with improved accuracy, precision, and shorter latency in eye movements

377 over unattended locations[124,125]. Mechanisms of attentional selection can gain access to internal

378 representations of different spatial coding schemes[126,127]. Visuospatial attention is maintained in

379 egocentric, gaze-centered coordinates and updated with each eye movement[127]. Evidence for a

380 simultaneous access of multiple allocentric reference frames comes from studies examining

381 inhibition of return, which refers to the inhibition of processing of objects that had recently been

382 attended [128]. If a recently attended object moves to a new location the inhibition moves with the

383 object, supporting an object-centered reference frame. In addition, inhibition occurs at the previous

384 object location independent of the object movement, supporting a location-centered reference

385 frame[129]. These findings are confirmed by patients with spatial neglect (an attention deficit toward

386 the side of space opposite to brain damage) who are worse at detecting information on the

387 contralesional side in both object-centered and location-centered reference frames[126].

388 Task effects on the use of allocentric reference frames have been demonstrated using

389 complex, naturalistic 3D-rendered scenes presented on a monitor. Such scenes allow for multiple

390 relevant or irrelevant objects to be displayed depending on the task (FIG. 4). When participants were

391 asked to point to the remembered location of one object, reaching behaviour was clearly affected by

392 location shifts of the other surrounding objects relevant for the given task[130]. For example, in a task

393 asking participants to 'reach to the location of the missing table object,' reaching behaviour was

394 impacted by shifts of the surrounding table objects but hardly impacted even by extensive shifts of

395 objects in the immediate surround, and vice versa. These findings suggest that task goals determine

396 the selection of allocentric cues for the spatial coding of action targets By directing attention to the

397 relevant areas in the scenes. Such attentional selection was supported by eye fixations (a well-

398 established marker for overt attention[131]) directed to the task-relevant region of the scene. This

399 evidence strengthens previous findings of more and longer fixations on task-relevant objects in real-

400 world and naturalistic scenes[132,117,133]. Due to the prioritization of retaining task-relevant objects in

401 visual short-term memory[134–137], the selected allocentric cues are likely to be effective in both real-

402 time and memory-guided movements.

403 Effects of task instruction on allocentric coding and eye fixations were also found when

404 participants explicitly learned which objects were task-relevant[84] or were informed about the reach

405 target (in contrast to other tasks where they had search for the missing target and then reach to its

406 remembered location[130])[138]. Incidental learning of the statistical regularities of the environment can

407 be just as powerful as having explicit knowledge about task-relevant objects in directing spatial

408 attention. Locations where task-relevant information occurs with a high probability are generally

14

prioritized over low-probability locations[139]. Incidental learning of spatial information took place relative to the viewer's perspective in an egocentric reference frame rather than relative to the environment despite the presence of multiple landmarks. A working hypothesis derived from results on visual search suggests that attention driven by explicit and implicit knowledge is associated with distinct spatial reference frames, with the former based on an allocentric reference frame that explicitly selects relevant locations over others, and the latter based on an egocentric reference frame that modulates how attention is moved through space[139]. This idea provides an interesting future perspective for spatial coding in goal-directed actions. How the use of action-based reference frames is influenced by explicit and implicit knowledge and their relation to attentional selection have not been considered so far.
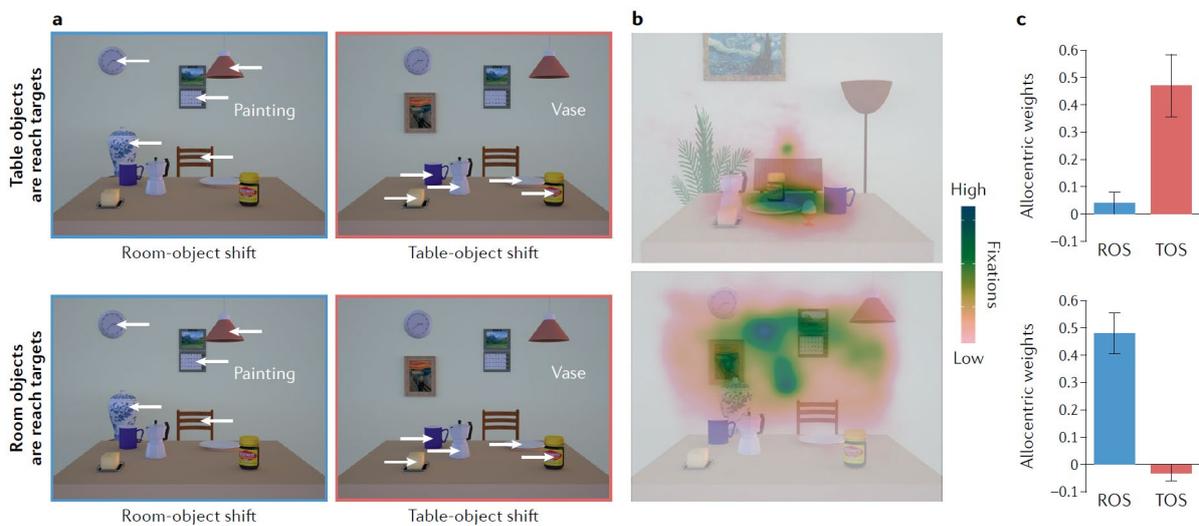
Figure 4: Stimuli, eye movement patterns, and use of reference frames in an object shift task[130].(a) Participants were presented with scenes containing multiple objects on a table and in the room that could be shifted leftward or rightward (as indicated by the white arrows). Participants were instructed to reach either to one of the table objects or to one of the room objects, rendering those objects 'task relevant'. (b) More fixations were observed in the area of the scene containing task-relevant objects. (c) Reach endpoints deviated in the direction of the object shifts, but only when task-relevant objects were shifted. This pattern is reflected in the allocentric weights that define the ratio of the lateral reach endpoint deviations to the average lateral displacement of the objects in the scene. Allocentric weights were substantially increased for table-object shifts (TOS) and not significantly increased for room-object shifts (ROS) when table-objects were potential reach targets (upper row), and vice versa for room-objects (lower row).

Semantic effects

Scene and object semantics play a prominent role in attention by guiding the eyes to where certain content is likely to occur[140,141]. Scene semantics refers to the meanings of and relationships between scenes and objects[142]. Objects can occur in semantically congruent (a kettle in the kitchen) or incongruent (a kettle in the bathroom) locations. This high-level cognitive factor can explain where humans look within a scene substantially better than low-level factors, such as salience.

15

Accordingly, salient features within a scene have been found to be hardly attended whereas semantic features are[143]. Importantly, an attentional advantage for semantic features rapidly develops and biases gaze toward semantically rich regions of a scene[144,145]. Such an attentional shift toward scene semantics seems to happen involuntarily, even while performing visual search tasks that are independent of semantics[146–148].

Apart from attentional shifts, semantic labelling of objects can also impact human kinematics. For example, humans adjust their grip aperture according to the semantic label ('small' versus 'large') during movement planning but not at later stages of movement control[149]. A similar result is that including labels on weights impacts judgements of how heavy an objects would be before lifting it but not the actually perceived heaviness or the force used to lift the weights[150]. Given this involuntary integration of semantic information during movement planning and the guidance of attention by scene and object semantics, semantic information has the power to influence the spatial representations for goal-directed actions.

The establishment of spatial representations might be facilitated by semantics through object grouping. According to the grouping hypothesis[151], humans represent scene configurations as a virtual polygon with the vertices connected to the locations of individual objects (such as a hexagon connecting all kitchen utensils visible in a scene). Objects are grouped according to their semantic similarity. This hypothesis was supported by findings showing that humans rely on contextual relationships, such as familiarity, functional relationships, or physical plausibility, to form spatial configurations[152]. The influence of object semantics on spatial representations for action was examined in a virtual reality (VR) experiment that presented objects belonging to two different semantic categories (man-made and natural objects)[153]. A computational approach using representational similarity analysis[154] was applied to identify object categories. To do so, participants dragged and dropped 49 pre-selected objects within an arena based on similarity. This resulted in a metric, high-dimensional feature space, where nearer distances between objects correspond to objects of the same category. Object semantics had a strong influence on reaching behaviour. When objects (natural objects: banana and pear) that belonged to the same category as the target object (natural objects: apple) were shifted, reaching movements were influenced more than twice as much by the objects' shift as when the reach target belonged to a different category (man-made objects: puncher). These findings show that object semantics indeed facilitate allocentric coding, likely due to involuntary shifts of attention based on object semantics.

In sum, high-level cognitive factors related to the agent, the action itself, and the action target play a decisive role in spatial coding for action. First, temporal movement characteristics can change the contribution of egocentric and allocentric information. Because allocentric reference

frames are more invariant to changes in their point of origin, they can provide more stable representations and are therefore more effective for memory-guided movements than egocentric reference frames when visual feedback of the target is not present. By contrast, fast guidance of online movements mainly relies on egocentric reference frames. Second, the task defines the movement goals and subgoals and therefore the information that is relevant to accomplish these goals. Task-relevant information is selected and then integrated into allocentric target representations, whereas task-irrelevant information is widely neglected. Third, object semantics can facilitate the formation of spatial configurations for action. Objects that belong to the same semantic category are more effective allocentric cues than objects of different semantic categories, indicating a direct influence of object semantics on spatial coding of action targets. We encourage the scientific community to consider this trichotomy in future research, aiming to integrate separate research fields on working memory, attention and scene perception in the context of spatial coding for action.

Actions across spatial scales

Decades of experiments on computer monitors using small-scale movements, such as pointing or reaching, support a deep understanding of spatial coding for action targets. These screen-based experiments were conducted with the implicit assumption that the findings on a 2D plane would generalize to the 3D world in which humans typically act. However, there is a crucial difference between the two spaces: The 2D pictorial space is a space an observer can observe —such as the space depicted in a painting—whereas the 3D visual space is the space an observer is part of and interacts in[155–158]. This distinction is made across vision science, philosophy, and the history of art[159–161]. In pictorial space, an observer's location is ill-defined because they do not have an own location in that space. Therefore, moving in front of a picture does not change the vantage point from which the picture has been taken or painted.

Because egocentric information is less reliable in pictorial space, humans should prefer allocentric information for space perception and action. This prediction was tested in a VR experiment that allowed participants to execute memory-guided reaching movements either to 3D objects or to 2D objects depicted on a computer monitor located within the virtual space[162]. Allocentric coding was found in both visual and pictorial space, with a higher contribution in visual space. These results were replicated after controlling for object size and presentation variability in depth[162]. This surprising result might be due to the prototypical observer location (in front of a monitor), which might have increased the reliance on egocentric information in pictorial space. If allocentric coding is crucial for human-object interactions, then the reliance on allocentric

information could be enhanced in an environment that allows humans to actually perform actions (visual space) in contrast to pictorial spaces in which human-object interactions are limited.

Another distinction within types of spaces is a well-established binary between peripersonal and extrapersonal action spaces. This distinction stems in part from pioneering work on monkey neurophysiology that demonstrated neuronal populations that are sensitive to 3D action targets close to the monkey's body[163–165]. The boundary of peripersonal space can be determined using physical distance ('within arm's reach'), or using the space in which behavioural responses are modulated as a proxy. For example, reaction times to tactile stimulation are faster the closer task-irrelevant auditory stimuli are presented to the hand, face and trunk. Importantly, the size and location of peripersonal space differ depending on the stimulated body part, which is reflected in reaction times that continuously increase with larger distances[166]. Moreover, peripersonal space can be extended in several ways, such as with a tool[167–169] or a virtual avatar hand[170]. For example, when using a tool such as a hand brush the interaction space is increases and effects that previously only occurred within arm's reach do now extend into the new enlarged peripersonal space[167–169]. Thus, the simple binary of within and outside arm's reach does not do justice to the complexity of the findings regarding peripersonal space.

A related theoretical approach is the action field theory of peripersonal space [171]. This theory describes two main characteristics of peripersonal spaces. First, they are graded (rather than binary), and the activity of multimodal neurons or reaction time advantages gradually decrease with increasing distance. Second, the size of peripersonal spaces varies as a function of the relevance of an action to avoid or make contact. For example, when standing, responding to a tactile stimulus is faster the closer a looming auditory stimulus is presented, but when participants walk, the reaction time advantage occurs for auditory stimuli a meter further away, effectively extending the peripersonal space[172]. From an action field theory perspective, the peripersonal space was extended because walking in the direction of the sound rather than being stationary increases the chance of an early impact with the stimulus and is therefore highly relevant for avoiding contact. The action field theory is supported by several similar findings confirming the gradual nature of reaction times and neural activity patterns as a function of distance[171]. Whether this theory will stand the test of time remains to be seen but it has raised an interesting debate by questioning the functional dichotomy between peripersonal and extrapersonal action spaces[150,151].

Studies on peripersonal space or visual and pictorial space are often carried out in reach distance and overlook spaces beyond these boundaries. A more comprehensive classification of space distinguishes between four types of spaces that differ in scale (FIG. 5)[173]. The smallest spaces, figural spaces, are smaller than the observer's body. These spaces are split into two-dimensional

pictorial surfaces (a journal cover) and three-dimensional object surfaces (a journal). This division is reminiscent of the distinction between pictorial and visual spaces, but pictorial and visual spaces are not bound by scale. Vista spaces are larger than the observer's body, but apprehensible from a single vantage point (such as an office). Environmental spaces are also larger than the observer's body (such as an office building), but require considerable locomotion to be comprehended. Finally, geographical spaces (such as a country or city) can only be grasped by using a map, which effectively reduces the geographical space to a figural space.
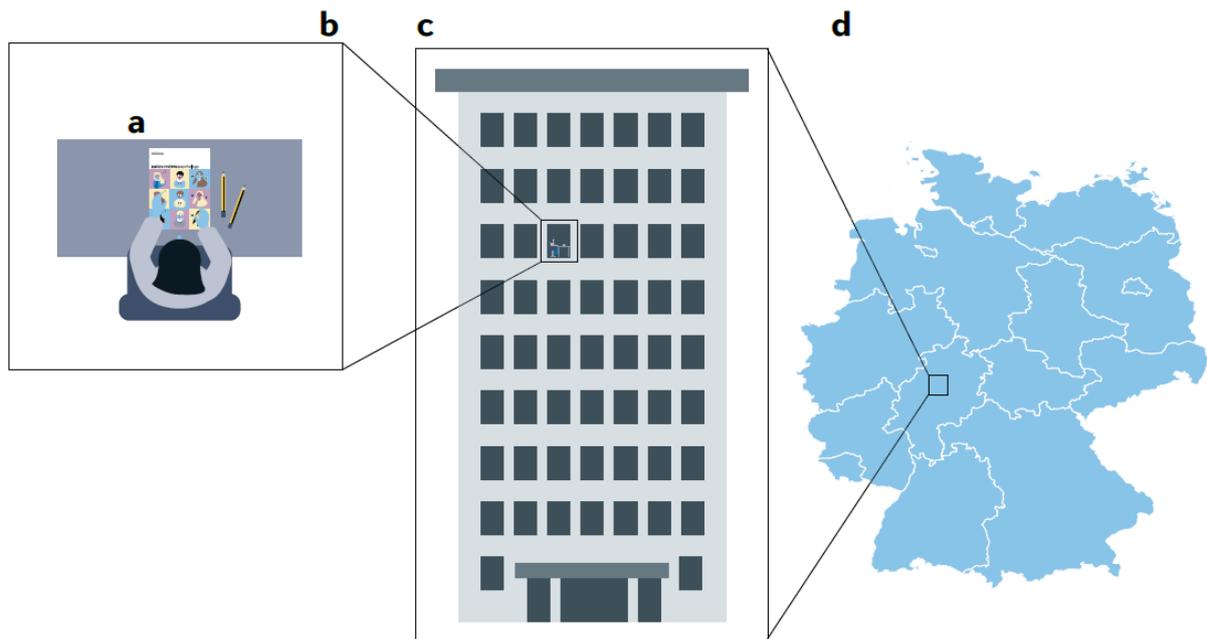
Figure 5: Classification of different types of spaces[173]. Four spaces that differ in their spatial scale: (a) Figural spaces are smaller than the observer's body (a journal). (b) Vista spaces are larger than the observer's body (an office room). (c) Environmental spaces are also larger than the observer's body, but require considerable locomotion to be comprehended (an office building). (d) Geographical spaces are too large to view and can only be grasped by using a map (the country where the office building is located).

Researchers rarely investigate how findings on spatial coding of action targets generalize across these different spatial scales. Typically, a research lab might be solely focused on human grasping (figural spaces) or navigation (vista or environmental spaces). However, the evidence for allocentric coding of reach goals in figural spaces can be generalized to larger spaces to some extent. In a virtual walk-and-place task[83], participants had to encode the landing location of a ball thrown onto a soccer field. When they walked to place a ball on the memorized location, they were biased by subtle shifts of the midfield line and the thrower. These results are comparable to previous studies on small-scale reaching to scenes presented on a monitor[130,138,174] and in virtual reality[139,175]. Importantly, when participants intercepted the ball with their foot during encoding, shifts of the midfield line no longer influenced performance and the thrower became the sole crucial allocentric

cue impacting action. These findings are further supported by sophisticated neuroscience methods that allow recordings of neural activity from head-unrestrained[176–178] and freely moving monkeys[179]. For example, walk-and-reach targets were found to be encoded in the same frontoparietal network as targets within a monkey's immediate reach space[179]. However, in humans self-movement and the perceived availability of items in the environment can change the contribution of egocentric and allocentric reference frames[180].

A series of VR experiments tested how interconnected vista spaces (such as multiple rooms) relate to global environmental spaces. When participants were asked to memorize object locations while walking along interconnected vista spaces, their pointing accuracy for cued objects was higher when their body was aligned within the reference frames of individual vista spaces compared to a reference frame of a global environmental space[181]. Similarly, memorized object locations in an environmental space were affected by the order in which participants learned the objects (earlier retrieval of object locations that were learned earlier), as well as their traveled distance (better performance for closer targets). In vista spaces, even when walking trajectory and successive presentation were controlled, this result was not found[182]. Thus, humans encode pairwise connected vista spaces and their respective reference frames rather than subsuming vista spaces in a global reference frame of the environmental space.

Psychological space is more than just a shallow construct. Humans act and interact in multiple spaces, and somehow connect these spaces and accordingly adjust their spatial coding strategies. It is important for research to move beyond classical stationary tasks and compare larger psychological spaces to uncover behavioural and neural commonalities and differences.

Summary and future directions

In everyday life, humans perform actions in a multitude of psychological spaces: They act on 2D planes (such as touchscreens) and in 3D virtual and real-world environments, they reach toward targets that are placed within or out of reach, and they navigate toward targets far away on a map. All these actions require the establishment and maintenance of spatial representations of the motor goal that can be implemented into a motor plan. Different classes of spatial reference frames are used to represent the action target with respect to the observer (egocentric) or the environment (allocentric). The contributions of egocentric and allocentric information to spatial coding is determined by a combination of low-level perceptual factors and high-level cognitive factors including memory, task constraints and object semantics. The high-level cognitive factors we reviewed here are likely far from exhaustive. Future research should try to identify further determinants of spatial coding of action targets (such as social factors during cooperative or

20

competitive actions) and scrutinize their specific contributions and interactions depending on the psychological space.

To fully understand the underlying mechanisms of spatial coding for action, research needs to actively address the richness of perception and action by exploring natural behaviour in complex environments. Well-controlled laboratory experiments need to be enriched, for example by presenting naturalistic scenes or using realistic 3D scenarios. Extending the spatial scale and investigating the richness of human experiences in complex environments might not have been feasible in the past without diminishing control over experimental manipulations. However, advances in computer graphics, markerless motion tracking and extended reality (Box 1), together with the computational architecture to process massive amounts of data in real-time, make this approach feasible. Similar strategies in other fields of research have resulted in novel findings. For example, established theories like the feature-integration theory of attention[183] do not fully account for visual attention in naturalistic scenes[184,185]. This example helps highlight the need to move research closer to the richness of 3D real-world scenarios.

The success of future research might also depend on whether scientists adapt an 'enactive' approach. This approach describes the effort to investigate how human cognition can facilitate actions[186]. For example, allocentric coding is not a mechanism that functions in isolation. Instead, it varies in utility depending on the scene context and the task at hand. To understand natural human behaviour, all these factors need to be considered and spatial coding needs to be investigated along different temporal scales and in different contexts.

Further research is also required to investigate how psychological spaces are biologically embedded. For example, the classic two-streams model of visual perception and action[91] has been linked to egocentric and allocentric reference frames associated with the dorsal action pathway and the ventral perception pathway, respectively[187,69]. This broad anatomical distinction seems to be too simplistic[70,71,188]. However, the neural correlates of egocentric and allocentric coding for action are still not clearly defined. Additionally, research in the field of sensorimotor control needs to be extended to targets beyond the sensory horizon and their neural underpinnings need to be compared, especially with respect to different scales of space[189].

Finally, the field needs to investigate similarities and differences of spatial coding across psychological spaces: Pictorial spaces are not equal to visual spaces. Furthermore, the binary view of peripersonal spaces might be replaced with the concept of a graded action field[171]. Future experiments might allow further understanding of how figural spaces are related to vista and environmental spaces, as well as whether and how the plethora of research findings that have been generated in figural space generalize to large-scale spaces. This would open up the opportunity to

628   bridge the gap and to stimulate a dialogue between the different research communities that focus

629   on space perception for action and spatial cognition.

630

Box 1: Extended reality to study action space

Studying human behaviour in the real world is the gold standard to investigate perception and action. However, there is limited experimental control in the real world. Extended reality (XR) technologies (an umbrella term for augmented, mixed, and virtual reality) can provide powerful tools to investigate spatial coding for actions in naturalistic 3D scenarios, while allowing experimenters to control stimulus presentation and experimental conditions. For example, augmented reality allows the enhancement of real-world environments by computer-generated content, displayed via specialized glasses. Mixed reality extends augmented reality in that physical and computer-generated content both co-exists and physically interacts with each other.

Given that head-mounted virtual reality (VR) displays have now become reasonably priced and go through fast development cycles with improved displays being released nearly every year, current research is increasingly focused on VR. This technology enables researchers to create manipulations that would not be possible in the real world (such as violating the laws of physics), which can help to isolate factors contributing to particular action behaviours. High quality VR displays now provide a very realistic stereoscopic image at refresh rates high enough for imperceptible spatial lag introduced by head movements. Experimenters creating virtual environments have full flexibility to generate scenes that range from simplified to realistic or introduce changes that would not be possible in reality (such as in room structure or lighting). Real environments can be rebuilt exactly as they are in VR, to provide the opportunity to examine factors such as the influence of prior knowledge and scene memory on spatial perception and action. In combination with eye and body movement tracking systems, actions of different effector systems and varying complexity can be examined across different spatial scales. VR-compatible auditory and tactile stimulation devices allow for the investigation of spatial coding in multiple sensory modalities that can be experimentally aligned or misaligned with the visual presentation, to test the limits of multisensory integration.

However, some limitations of VR technology can compromise a realistic experience. One of the most prominent current limitations is the inability to create realistic virtual touch, which is essential for human-object interaction. Additionally, the field of view remains restricted relative to real world environments, which leads to more head movements when exploring a virtual scene. Simulator sickness (such as nausea) can occur, especially when the movements of the participant's body and the environment are not perfectly aligned and therefore create a conflict between vestibular and visual information. Realistic-looking and kinematically accurate avatars are still a subject of current research. Despite these and other limitations, XR technology has now been leveraged in many fields of psychological research and future XR solutions can serve as useful tools to investigate spatial coding for action across different spatial scales.

23

References

1   Colby, C. L. Action-oriented spatial reference frames in cortex. *Neuron*. **20**(1), 15–24 (1998).

2   Klatzky, R. L. in *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (eds Freska, C., Habel, C. & Wender K. F.) 1–17 (Springer, 1998).

3   Crawford, J. D., Henriques, D. Y. P., Medendorp, W. P. Three-dimensional transformations for goal-directed action. *Annu. Rev. Neurosci*. **34**, 309–331 (2011).

4   McGuire, L. M. M., Sabes, P. N. Sensory transformations and the use of multiple reference frames for reach planning. *Nat. Neurosci*. **12**(8), 1056–1061 (2009).

5   O'Regan, J. K., Noë, A. A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci*. **24**(5), 939-973 (2001).

6   Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W. The Theory of Event Coding (TEC): a framework for perception and action planning. *Behav. Brain Sci*. **24**(5), 849-878 (2001).

7   Rizzolatti, G., Craighero, L. The mirror-neuron system. *Annu. Rev. Neurosci*. **27**, 169–192 (2004).

8   Flanders, M., Tillery, S. I. H., Soechting, J. F. Early stages in a sensorimotor transformation. *Behav. Brain Sci*. **15**(2), 309–320 (1992).

9   Blohm, G., Crawford, J. D. Computations for geometrically accurate visually guided reaching in 3-D space. *J. Vis*. **7**(5), 4 (2007).

10   Crawford, J. D., Guitton, D. Visual-motor transformations required for accurate and kinematically correct saccades. *J. Neurophysiol*. **78**(3), 1447–1467 (1997).

11   Henriques, D. Y. P., Klier, E. M., Smith, M. A., Lowy, D., Crawford, J. D. Gaze-centered remapping of remembered visual space in an open-loop pointing task. *J. Neurosci.* **18**(4), 1583–1594 (1998).

12   Batista, A. P., Buneo, C. A., Snyder, L. H., Andersen, R. A. Reach plans in eye-centered coordinates. *Science*. **285**(5425), 257–260 (1999).

13   Voudouris, D., Smeets, J. B. J., Fiehler, K., Brenner, E. Gaze when reaching to grasp a glass. *J. Vis*. **18**(8), 16 (2018).

14   van Beers, R. J., van Mierlo, C. M., Smeets, J. B. J., Brenner, E. Reweighting visual cues by touch. *J. Vis*. **11**(10) (2011).

15   Camponogara, I., Volcic, R. Integration of haptics and vision in human multisensory grasping. *Cortex*. **135**, 173–185 (2021).

16   Cuijpers, R. H., Brenner, E., Smeets, J. B. J. Consistent haptic feedback is required but it is not enough for natural reaching to virtual cylinders. *Hum. Mov. Sci*. **27**(6), 857–872 (2008).

17   Medendorp, W. P. Spatial constancy mechanisms in motor control. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. **366**(1564), 476–491 (2011).

18   Blohm, G., Alikhanian, H., Gaetz, W., Goltz, H. C., DeSouza, J. F. X., Cheyne, D. O., Crawford, J. D. Neuromagnetic signatures of the spatiotemporal transformation for manual pointing. *Neuroimage*. **197**, 306–319 (2019).

19   Medendorp, W. P., Beurze, S. M., van Pelt, S., van der Werf, J. Behavioral and cortical mechanisms for spatial coding and action planning. *Cortex*. **44**(5), 587–597 (2008).

705   20   Culham, J. C., Valyear, K. F. Human parietal cortex in action. *Curr. Opin. Neurobiol*. **16**(2), 205–
706   212 (2006).

707   21   Perenin, M. T., Vighetto, A. Optic ataxia: a specific disruption in visuomotor mechanisms. I.
708   Different aspects of the deficit in reaching for objects. *Brain*. **111 (Pt 3)**, 643–674 (1988).

709   22   Hallett, P. E. Primary and secondary saccades to goals defined by instructions. *Vision Res*.
710   **18**(10), 1279–1296 (1978).

711   23   Gail, A., Andersen, R. A. Neural dynamics in monkey parietal reach region reflect context-
712   specific sensorimotor transformations. *J. Neurosci*. **26**(37), 9376–9384 (2006).

713   24   DeSouza, J. F. X., Dukelow, S. P., Gati, J. S., Menon, R. S., Andersen, R. A., Vilis, T. Eye position
714   signal modulates a human parietal pointing region during memory-guided movements. *J.*
715   *Neurosci.* **20**(15), 5835–5840 (2000).

716   25   Medendorp, W. P., Goltz, H. C., Crawford, J. D., Vilis, T. Integration of target and effector
717   information in human posterior parietal cortex for the planning of action. *J. Neurophysiol*.
718   **93**(2), 954–962 (2005).

719   26   Gail, A., Klaes, C., Westendorff, S. Implementation of spatial transformation rules for goal-
720   directed reaching via gain modulation in monkey parietal and premotor cortex. *J. Neurosci.*
721   **29**(30), 9490–9499 (2009).

722   27   Gertz, H., Lingnau, A., Fiehler, K. Decoding movement goals from the fronto-parietal reach
723   network. *Front. Hum. Neurosci*. **11**, 84 (2017).

724   28   Westendorff, S., Klaes, C., Gail, A. The cortical timeline for deciding on reach motor goals. *J.*
725   *Neurosci.* **30**(15), 5426–5436 (2010).

726   29   Fernandez-Ruiz, J., Goltz, H. C., DeSouza, J. F. X., Vilis, T., Crawford, J. D. Human parietal "reach
727   region" primarily encodes intrinsic visual direction, not extrinsic movement direction, in a visual
728   motor dissociation task. *Cereb. Cortex*. **17**(10), 2283–2292 (2007).

729   30   Dum, R. P., Strick, P. L. Frontal lobe inputs to the digit representations of the motor areas on
730   the lateral surface of the hemisphere. *J. Neurosci.* **25**(6), 1375–1386 (2005).

731   31   He, S. Q., Dum, R. P., Strick, P. L. Topographic organization of corticospinal projections from the
732   frontal lobe: motor areas on the lateral surface of the hemisphere. *J. Neurosci.* **13**(3), 952–980
733   (1993).

734   32   Connolly, J. D., Goodale, M. A., DeSouza, J. F., Menon, R. S., Vilis, T. A comparison of
735   frontoparietal fMRI activation during anti-saccades and anti-pointing. *J. Neurophysiol*. **84**(3),
736   1645–1655 (2000).

737   33   Gertz, H., Fiehler, K. Human posterior parietal cortex encodes the movement goal in a pro-
738   /anti-reach task. *J. Neurophysiol*. **114**(1), 170–183 (2015).

739   34   Beurze, S. M., Lange, F. P. de, Toni, I., Medendorp, W. P. Integration of target and effector
740   information in the human brain during reach planning. *J. Neurophysiol*. **97**(1), 188–199 (2007).

741   35   Chapman, C. S., Gallivan, J. P., Wood, D. K., Milne, J. L., Culham, J. C., Goodale, M. A. Reaching
742   for the unknown: multiple target encoding and real-time decision-making in a rapid reach task.
743   *Cognition*. **116**(2), 168–176 (2010).

744 36 Stewart, B. M., Baugh, L. A., Gallivan, J. P., Flanagan, J. R. Simultaneous encoding of the
745 direction and orientation of potential targets during reach planning: evidence of multiple
746 competing reach plans. *J. Neurophysiol*. **110**(4), 807–816 (2013).

747 37 Hesse, C., Kangur, K., Hunt, A. R. Decision making in slow and rapid reaching: Sacrificing success
748 to minimize effort. *Cognition*. **205**, 104426 (2020).

749 38 Onagawa, R., Kudo, K. Sensorimotor strategy selection under time constraints in the presence
750 of two motor targets with different values. *Sci. Rep*. **11**(1), 22207 (2021).

751 39 Stewart, B. M., Gallivan, J. P., Baugh, L. A., Flanagan, J. R. Motor, not visual, encoding of
752 potential reach targets. *Curr. Biol*. **24**(19), R953-R954 (2014).

753 40 Gallivan, J. P., Stewart, B. M., Baugh, L. A., Wolpert, D. M., Flanagan, J. R. Rapid Automatic
754 Motor Encoding of Competing Reach Options. *Cell Rep*. **18**(7), 1619–1626 (2017).

755 41 Gallivan, J. P., Logan, L., Wolpert, D. M., Flanagan, J. R. Parallel specification of competing
756 sensorimotor control policies for alternative action options. *Nat. Neurosci*. **19**(2), 320–326
757 (2016).

758 42 Praamstra, P., Kourtis, D., Nazarpour, K. Simultaneous preparation of multiple potential
759 movements: opposing effects of spatial proximity mediated by premotor and parietal cortex. *J.
760 Neurophysiol*. **102**(4), 2084–2095 (2009).

761 43 Cisek, P. Cortical mechanisms of action selection: the affordance competition hypothesis.
762 *Philos. Trans. R. So.c Lond. B Biol. Sci*. **362**(1485), 1585–1599 (2007).

763 44 Cisek, P., Kalaska, J. F. Neural correlates of reaching decisions in dorsal premotor cortex:
764 specification of multiple direction choices and final selection of action. *Neuron*. **45**(5), 801–814
765 (2005).

766 45 Gallivan, J. P., Barton, K. S., Chapman, C. S., Wolpert, D. M., Flanagan, J. R. Action plan co-
767 optimization reveals the parallel encoding of competing reach movements. *Nat. Commun*. **6**,
768 7428 (2015).

769 46 Alhussein, L., Smith, M. A. Motor planning under uncertainty. *Elife*. **10**, e67019 (2021).

770 47 Battaglia-Mayer, A., Caminiti, R., Lacquaniti, F., Zago, M. Multiple levels of representation of
771 reaching in the parieto-frontal network. *Cereb. Cortex*. **13**(10), 1009–1022 (2003).

772 48 Soechting, J. F., Flanders, M. Moving in three-dimensional space: frames of reference, vectors,
773 and coordinate systems. *Annu. Rev. Neurosci*. **15**, 167–191 (1992).

774 49 Thompson, A. A., Henriques, D. Y. P. The coding and updating of visuospatial memory for goal-
775 directed reaching and pointing. *Vision Res*. **51**(8), 819–826 (2011).

776 50 Bock, O. Contribution of retinal versus extraretinal signals towards visual localization in goal-
777 directed movements. *Exp. Brain Res*. **64**(3), 476–482 (1986).

778 51 Selen, L. P. J., Medendorp, W. P. Saccadic updating of object orientation for grasping
779 movements. *Vision Res*. **51**(8), 898–907 (2011).

780 52 Medendorp, W. P., Crawford, J. D. Visuospatial updating of reaching targets in near and far
781 space. *Neuroreport*. **13**(5), 633–636 (2002).

782 53 Mueller, S., Fiehler, K. Effector movement triggers gaze-dependent spatial coding of tactile and
783 proprioceptive-tactile reach targets. *Neuropsychologia*. **62**, 184–193 (2014).

784 54 Lewald, J., Ehrenstein, W. H. The effect of eye position on auditory lateralization. *Exp. Brain Res*.
785 **108**(3), 473–485 (1996).

786 55 Pouget, A., Deneve, S., Duhamel, J.-R. A computational perspective on the neural basis of
787 multisensory spatial representations. *Nat. Rev. Neurosci*. **3**(9), 741–747 (2002).

788 56 Medendorp, W. P., Goltz, H. C., Vilis, T., Crawford, J. D. Gaze-centered updating of visual space
789 in human parietal cortex. *J. Neurosci.* **23**(15), 6209–6214 (2003).

790 57 Leoné, F. T. M., Monaco, S., Henriques, D. Y. P., Toni, I., Medendorp, W. P. Flexible reference
791 frames for grasp planning in human parietofrontal cortex. *eNeuro*. **2**(3) (2015).

792 58 Dijkerman, H. C., McIntosh, R. D., Anema, H. A., Haan, E. H. F. de, Kappelle, L. J., Milner, A. D.
793 Reaching errors in optic ataxia are linked to eye position rather than head or body position.
794 *Neuropsychologia*. **44**(13), 2766–2773 (2006).

795 59 Khan, A. Z., Pisella, L., Rossetti, Y., Vighetto, A., Crawford, J. D. Impairment of gaze-centered
796 updating of reach targets in bilateral parietal-occipital damaged patients. *Cereb. Cortex*. **15**(10),
797 1547–1560 (2005).

798 60 Ambrosini, E., Ciavarro, M., Pelle, G., Perrucci, M. G., Galati, G., Fattori, P., Galletti, C.,
799 Committeri, G. Behavioral investigation on the frames of reference involved in visuomotor
800 transformations during peripheral arm reaching. *PLoS One*. **7**(12), e51856 (2012).

801 61 Bernier, P.-M., Grafton, S. T. Human posterior parietal cortex flexibly determines reference
802 frames for reaching based on sensory context. *Neuron*. **68**(4), 776–788 (2010).

803 62 Mullette-Gillman, O. A., Cohen, Y. E., Groh, J. M. Eye-centered, head-centered, and complex
804 coding of visual and auditory targets in the intraparietal sulcus. *J. Neurophysiol*. **94**(4), 2331–
805 2352 (2005).

806 63 Mullette-Gillman, O. A., Cohen, Y. E., Groh, J. M. Motor-related signals in the intraparietal
807 cortex encode locations in a hybrid, rather than eye-centered reference frame. *Cereb. Cortex*.
808 **19**(8), 1761–1775 (2009).

809 64 Obhi, S. S., Goodale, M. A. The effects of landmarks on the performance of delayed and real-
810 time pointing movements. *Exp. Brain Res*. **167**(3), 335–344 (2005).

811 65 Krigolson, O., Heath, M. Background visual cues and memory-guided reaching. *Hum. Mov. Sci*.
812 **23**(6), 861–877 (2004).

813 66 Krigolson, O., Clark, N., Heath, M., Binsted, G. The proximity of visual landmarks impacts
814 reaching performance. *Spat. Vis*. **20**(4), 317–336 (2007).

815 67 Byrne, P. A., Crawford, J. D. Cue reliability and a landmark stability heuristic determine relative
816 weighting between egocentric and allocentric visual information in memory-guided reach. *J.
817 Neurophysiol*. **103**(6), 3054–3069 (2010).

818 68 Taghizadeh, B., Gail, A. Spatial task context makes short-latency reaches prone to induced
819 Roelofs illusion. *Front. Hum. Neurosci*. **8**, 673 (2014).

820 69 Schenk, T. An allocentric rather than perceptual deficit in patient D.F. *Nat. Neurosci*. **9**(11),
821 1369–1370 (2006).

822 70 Chen, Y., Monaco, S., Byrne, P., Yan, X., Henriques, D. Y. P., Crawford, J. D. Allocentric versus
823 egocentric representation of remembered reach targets in human cortex. *J. Neurosci.* **34**(37),
824 12515–12526 (2014).

71 Chen, Y., Monaco, S., Crawford, J. D. Neural substrates for allocentric-to-egocentric conversion of remembered reach targets in humans. *Eur. J. Neurosci*. **47**(8), 901–917 (2018).

72 Taghizadeh, B., Fortmann, O., Gail, A. Position- and scale-invariant object-centered spatial selectivity in monkey frontoparietal cortex dynamically adapts to task demand. *bioRxiv* (2022). https://doi.org/10.1101/2022.01.26.477941

73 Carrozzo, M., Stratta, F., McIntyre, J., Lacquaniti, F. Cognitive allocentric representations of visual space shape pointing errors. *Exp. Brain Res*. **147**(4), 426–436 (2002).

74 Neely, K. A., Tessmer, A., Binsted, G., Heath, M. Goal-directed reaching: movement strategies influence the weighting of allocentric and egocentric visual cues. *Exp. Brain Res*. **186**(3), 375–384 (2008).

75 Thompson, A. A., Henriques, D. Y. P. Locations of serial reach targets are coded in multiple reference frames. *Vision Res*. **50**(24), 2651–2660 (2010).

76 Ernst, M. O., Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. **415**(6870), 429–433 (2002).

77 Knill, D. C. Robust cue integration: a Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *J. Vis*. **7**(7), 5.1-24 (2007).

78 Körding, K. P., Wolpert, D. M. Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci*. **10**(7), 319–326 (2006).

79 Vaziri, S., Diedrichsen, J., Shadmehr, R. Why does the brain predict sensory consequences of oculomotor commands? Optimal integration of the predicted and the actual sensory feedback. *J. Neurosci*. **26**(16), 4188–4197 (2006).

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

81 Knill, D. C., Richards, W. *Perception as Bayesian Inference* (Cambridge University Press, 1996).

82 Tagliabue, M., McIntyre, J. A modular theory of multisensory integration for motor control. *Front. Comput. Neurosci*. **8**, 1 (2014).

83 Karimpur, H., Kurz, J., Fiehler, K. The role of perception and action on the use of allocentric information in a large-scale virtual environment. *Exp. Brain Res*. **238**(9), 1813–1826 (2020).

84 Klinghammer, M., Blohm, G., Fiehler, K. Scene Configuration and Object Reliability Affect the Use of Allocentric Information for Memory-Guided Reaching. *Front. Neurosci*. **11**, 204 (2017).

85 Camors, D., Jouffrais, C., Cottereau, B. R., Durand, J. B. Allocentric coding: spatial range and combination rules. *Vision Res*. **109**(Pt A), 87–98 (2015).

86 Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., Shams, L. Causal inference in multisensory perception. *PLoS One*. **2**(9), e943 (2007).

87 Sato, Y., Toyoizumi, T., Aihara, K. Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput*. **19**(12), 3335–3355 (2007).

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

864 89 Neggers, S. F. W., Schölvinck, M. L., van der Lubbe, R. H. J., Postma, A. Quantifying the
865   interactions between allo- and egocentric representations of space. *Acta Psychol. (Amst)*.
866   **118**(1-2), 25–45 (2005).

867 90 Ruotolo, F., van der Ham, I. J. M., Iachini, T., Postma, A. The relationship between allocentric
868   and egocentric frames of reference and categorical and coordinate spatial information
869   processing. *Q. J. Exp. Psychol*. **64**(6), 1138–1156 (2011).

870 91 Goodale, M. A., Milner, A. Separate visual pathways for perception and action. *Trends Neurosci*.
871   **15**(1), 20–25 (1992).

872 92 Heath, M., Westwood, D. A. Can a visual representation support the online control of memory-
873   dependent reaching? Evident from a variable spatial mapping paradigm. *Motor Control*. **7**(4),
874   346–361 (2003).

875 93 Westwood, D. A., Heath, M., Roy, E. A. No evidence for accurate visuomotor memory:
876   systematic and variable error in memory-guided reaching. *J Mot Behav*. **35**(2), 127–133 (2003).

877 94 Hesse, C., Franz, V. H. Memory mechanisms in grasping. *Neuropsychologia*. **47**(6), 1532–1545
878   (2009).

879 95 Hesse, C., Franz, V. H. Grasping remembered objects: exponential decay of the visual memory.
880   *Vision Res*. **50**(24), 2642–2650 (2010).

881 96 Franz, V. H., Hesse, C., Kollath, S. Visual illusions, delayed grasping, and memory: no shift from
882   dorsal to ventral control. *Neuropsychologia*. **47**(6), 1518–1531 (2009).

883 97 Fiehler, K., Bannert, M. M., Bischoff, M., Blecker, C., Stark, R., Vaitl, D., Franz, V. H., Rösler, F.
884   Working memory maintenance of grasp-target information in the human posterior parietal
885   cortex. *Neuroimage*. **54**(3), 2401–2411 (2011).

886 98 Freud, E., Plaut, D. C., Behrmann, M. 'What' is happening in the dorsal visual pathway. *Trends
887   Cogn Sci*. **20**(10), 773–784 (2016).

888 99 Himmelbach, M., Nau, M., Zündorf, I., Erb, M., Perenin, M.-T., Karnath, H.-O. Brain activation
889   during immediate and delayed reaching in optic ataxia. *Neuropsychologia*. **47**(6), 1508–1517
890   (2009).

891 100 Goodale, M. Frames of reference for perception and action in the human visual system.
892   *Neurosci. Biobehav. Rev*. **22**(2), 161–172 (1998).

893 101 Westwood, D. A., Goodale, M. A. Perceptual illusion and the real-time control of action. *Spat.
894   Vis*. **16**(3-4), 243–254 (2003).

895 102 Schenk, T., McIntosh, R. D. Do we have independent visual streams for perception and action?
896   *Cogn. Neurosci*. **1**(1), 52–62 (2010).

897 103 Schenk, T., Hesse, C. Do we have distinct systems for immediate and delayed actions? A
898   selective review on the role of visual memory in action. *Cortex*. **98**, 228–248 (2018).

899 104 Schütz, I., Henriques, D. Y. P., Fiehler, K. Gaze-centered spatial updating in delayed reaching
900   even in the presence of landmarks. *Vision Res*. **87**, 46–52 (2013).

901 105 Schütz, I., Henriques, D. Y. P., Fiehler, K. No effect of delay on the spatial representation of
902   serial reach targets. *Exp. Brain Res*. **233**(4), 1225–1235 (2015).

903 106 Bridgemen, B., Kirch, M., Sperling, A. Segregation of cognitive and motor aspects of visual
904   function using induced motion. *Percept. Psychophys*. **29**(4), 336–342 (1981).

107   Brenner, E., Smeets, J. B. Fast Responses of the human hand to changes in target position. *J. Mot. Behav*. **29**(4), 297–310 (1997).

108   Grave, D. D. J. de, Brenner, E., Smeets, J. B. J. Illusions as a tool to study the coding of pointing movements. *Exp. Brain Res*. **155**(1), 56–62 (2004).

109   Gomi, H. Implicit online corrections of reaching movements. *Curr. Opin. Neurobiol*. **18**(6), 558–564 (2008).

110   Lu, Z., Fiehler, K. Spatial updating of allocentric landmark information in real-time and memory-guided reaching. *Cortex*. **125**, 203–214 (2020).

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

112   Chen, Y., Byrne, P., Crawford, J. D. Time course of allocentric decay, egocentric decay, and allocentric-to-egocentric conversion in memory-guided reach. *Neuropsychologia*. **49**(1), 49–60 (2011).

113   Hay, L., Redon, C. Response delay and spatial representation in pointing movements. *Neurosci. Lett*. **408**(3), 194–198 (2006).

114   Sheth, B. R., Shimojo, S. Extrinsic cues suppress the encoding of intrinsic cues. *J. Cogn. Neurosci*. **16**(2), 339–350 (2004).

115   Crowe, E. M., Bossard, M., Karimpur, H., Rushton, S. K., Fiehler, K., Brenner, E. Further evidence that people rely on egocentric information to guide a cursor to a visible target. *Perception*. **50**(10), 904–907 (2021).

116   Crowe, E. M., Bossard, M., Brenner, E. Can ongoing movements be guided by allocentric visual information when the target is visible? *J. Vis*. **21**(1), 6 (2021).

117   Land, M. F., Hayhoe, M. In what ways do eye movements contribute to everyday activities? *Vision Res*. **41**(25-26), 3559–3565 (2001).

118   Rothkopf, C. A., Ballard, D. H., Hayhoe, M. M. Task and context determine where you look. *J Vis*. **7**(14), 16.1-20 (2007).

119   Broadbent, D. E. *Perception and Communication* (Pergamon Press, 1958).

120   Bisley, J. W., Goldberg, M. E. Attention, intention, and priority in the parietal lobe. *Annu. Rev. Neurosci*. **33**, 1–21 (2010).

121   Desimone, R., Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci*. **18**, 193–222 (1995).

122   Fecteau, J. H., Munoz, D. P. Salience, relevance, and firing: a priority map for target selection. *Trends Cogn. Sci*. **10**(8), 382–390 (2006).

123   Carrasco, M. Visual attention: the past 25 years. *Vision Res*. **51**(13), 1484–1525 (2011).

124   Moehler, T., Fiehler, K. Effects of spatial congruency on saccade and visual discrimination performance in a dual-task paradigm. *Vision Res*. **105**, 100–111 (2014).

125   Moehler, T., Fiehler, K. The influence of spatial congruency and movement preparation time on saccade curvature in simultaneous and sequential dual-tasks. *Vision Res*. **116**, 25–35 (2015).

126   Behrmann, M., Tipper, S. P. Attention accesses multiple reference frames: evidence from visual neglect. *J. Exp. Psychol. Hum. Percept. Perform*. **25**(1), 83–101 (1999).

945 127 Chun, M. M., Golomb, J. D., Turk-Browne, N. B. A taxonomy of external and internal attention. *Annu. Rev. Psychol*. **62**, 73–101 (2011).

947 128 Abrams, R. A., Dobkin, R. S. Inhibition of return: effects of attentional cuing on eye movement latencies. *J. Exp. Psychol. Hum. Percept. Perform*. **20**(3), 467–477 (1994).

949 129 Tipper, S. P., Weaver, B., Jerreat, L. M., Burak, A. L. Object-based and environment-based inhibition of return of visual attention. *J. Exp. Psychol. Hum.Percept. Perform*. **20**(3), 478–499 (1994).

952 130 Klinghammer, M., Blohm, G., Fiehler, K. Contextual factors determine the use of allocentric information for reaching in a naturalistic scene. *J Vis*. **15**(13), 24 (2015).

954 131 Posner, M. I. Orienting of attention. *Q. J. Exp. Psychol*. **32**(1), 3–25 (1980).

955 132 Ballard, D. H., Hayhoe, M. M. Modelling the role of task in the control of gaze. *Vis. Cogn*. **17**(6-7), 1185–1204 (2009).

957 133 Mills, M., Hollingworth, A., van der Stigchel, S., Hoffman, L., Dodd, M. D. Examining the influence of task set on eye movements and fixations. *J. Vis*. **11**(8), 17 (2011).

959 134 Golomb, J. D., Pulido, V. Z., Albrecht, A. R., Chun, M. M., Mazer, J. A. Robustness of the retinotopic attentional trace after eye movements. *J. Vis*. **10**(3), 19 (2010).

961 135 Jonikaitis, D., Moore, T. The interdependence of attention, working memory and gaze control: behavior and neural circuitry. *Curr. Opin. Psychol*. **29**, 126–134 (2019).

963 136 Maxcey-Richard, A. M., Hollingworth, A. The strategic retention of task-relevant objects in visual working memory. *J. Exp. Psychol. Learn. Mem. Cogn*. **39**(3), 760–772 (2013).

965 137 Schneider, W. X., Einhäuser, W., Horstmann, G. Attentional selection in visual perception, memory and action: a quest for cross-domain integration. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. **368**(1628), 20130053 (2013).

968 138 Lu, Z., Klinghammer, M., Fiehler, K. The role of gaze and prior knowledge on allocentric coding of reach targets. *J. Vis*. **18**(4), 22 (2018).

970 139 Jiang, Y. V., Swallow, K. M., Sun, L. Egocentric coding of space for incidentally learned attention: effects of scene context and task instructions. *J. Exp. Psychol. Learn. Mem. Cogn*. **40**(1), 233–250 (2014).

973 140 Henderson, J. M. Human gaze control during real-world scene perception. *Trends Cogn. Sci*. **7**(11), 498–504 (2003).

975 141 Henderson, J. M. Gaze Control as Prediction. *Trends Cogn. Sci*. **21**(1), 15–23 (2017).

976 142 Võ, M. L.-H., Wolfe, J. M. Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol. Sci*. **24**(9), 1816–1823 (2013).

978 143 Hayes, T. R., Henderson, J. M. Scene semantics involuntarily guide attention during visual search. *Psychon. Bull. Rev*. **26**(5), 1683–1689 (2019).

980 144 Henderson, J. M., Hayes, T. R. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nat. Hum. Behav*. **1**(10), 743–747 (2017).

982 145 Oliva, A., Torralba, A. Chapter 2 Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res*. 155, 23–36 (2006).

146 Cornelissen, T. H. W., Võ, M. L.-H. Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Atten. Percept. Psychophys*. **79**(1), 154–168 (2017).

147 Malcolm, G. L., Rattinger, M., Shomstein, S. Intrusive effects of semantic information on visual selective attention. *Atten. Percept. Psychophys*. **78**(7), 2066–2078 (2016).

148 Peacock, C. E., Hayes, T. R., Henderson, J. M. Meaning guides attention during scene viewing, even when it is irrelevant. *Atten. Percept. Psychophys*. **81**(1), 20–34 (2019).

149 Glover, S., Dixon, P. Semantics affect the planning but not control of grasping. *Exp. Brain Res*. **146**(3), 383–387 (2002).

150 Naylor, C. E., Power, T. J., Buckingham, G. Examining Whether Semantic Cues Can Affect Felt Heaviness When Lifting Novel Objects. *J. Cogn*. **3**(1), 3 (2020).

151 Yantis, S. Multielement visual tracking: Attention and perceptual organization. *Cogn. Psychol*. **24**(3), 295–340 (1992).

152 Hock, H. S., Gordon, G. P., Whitehurst, R. Contextual relations: The influence of familiarity, physical plausibility, and belongingness. *Percept. Psychophys*. **16**(1), 4–8 (1974).

153 Karimpur, H., Morgenstern, Y., Fiehler, K. Facilitation of allocentric coding by virtue of object-semantics. *Sci. Rep*. **9**(1), 6263 (2019).

154 Kriegeskorte, N., Mur, M. Inverse MDS: Inferring Dissimilarity Structure from Multiple Item Arrangements. *Front. Psychol*. **3**, 245 (2012).

155 Goldstein, E. B. Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *J. Exp. Psychol. Hum. Percept. Perform*. **13**(2), 256–266 (1987).

156 Koenderink, J.J., van Doorn, A. in *Looking into pictures: An interdisciplinary approach to pictorial space* (eds Hecht, H., Schwartz, R. & Atherton, M.) 239-299 (MIT Press, 2003).

157 Koenderink, J., van Doorn, A. The structure of visual spaces. *J. Math. Imaging Vis*. **31**(2-3), 171–187 (2008).

158 Vishwanath, D., Girshick, A. R., Banks, M. S. Why pictures look right when viewed from the wrong place. *Nat. Neurosci*. **8**(10), 1401–1410 (2005).

159 Heidegger, M. *Being and time: A translation of Sein und Zeit* (Suny press, 1996).

160 Wollheim, R. On pictorial representation. *J. Aesthet. Art Crit*. **56**(3), 217-226 (1998).

161 Troje, N. F. Reality check. *Perception*. **48**(11), 1033–1038 (2019).

162 Karimpur, H., Eftekharifar, S., Troje, N. F., Fiehler, K. Spatial coding for memory-guided reaching in visual and pictorial spaces. *J. Vis*. **20**(4), 1 (2020).

163 Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H., Acuna, C. Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *J. Neurophysiol*. **38**(4), 871–908 (1975).

164 Rizzolatti, G., Scandolara, C., Matelli, M., Gentilucci, M. Afferent properties of periarcuate neurons in macaque monkeys. I. Somatosensory responses. *Behav. Brain Res*. **2**(2), 125–146 (1981).

165 Rizzolatti, G., Fadiga, L., Fogassi, L., Gallese, V. The space around us. *Science*. **277**(5323), 190–191 (1997).

166 Serino, A., Noel, J.-P., Galli, G., Canzoneri, E., Marmaroli, P., Lissek, H., Blanke, O. Body part-centered and full body-centered peripersonal space representations. *Sci. Rep.* **5**, 18603 (2015).

167 Iriki, A., Tanaka, M., Iwamura, Y. Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport.* **7**(14), 2325–2330 (1996).

168 Farnè, A., Serino, A., Làdavas, E. Dynamic size-change of peri-hand space following Tool-use: determinants and spatial characteristics revealed through cross-modal extinction. *Cortex.* **43**(3), 436–443 (2007).

169 Longo, M. R., Lourenco, S. F. On the nature of near space: effects of tool use and the transition to far space. *Neuropsychologia.* **44**(6), 977–981 (2006).

170 Mine, D., Yokosawa, K. Disconnected hand avatar can be integrated into the peripersonal space. *Exp. Brain Res.* **239**(1), 237–244 (2021).

171 Bufacchi, R. J., Iannetti, G. D. An action field theory of peripersonal space. *Trends Cogn. Sci.* **22**(12), 1076–1090 (2018).

172 Noel, J.-P., Grivaz, P., Marmaroli, P., Lissek, H., Blanke, O., Serino, A. Full body action remapping of peripersonal space: the case of walking. *Neuropsychologia.* **70**, 375–384 (2015).

173 Montello, D. R. in *Spatial information theory: A theoretical basis for GIS.* (eds Frank, A. U. & Campari, I.) 312-321 (Proceedings of COSIT'93, 1993).

174 Fiehler, K., Wolf, C., Klinghammer, M., Blohm, G. Integration of egocentric and allocentric information during memory-guided reaching to images of a natural environment. *Front. Hum. Neurosci.* **8**, 636 (2014).

175 Klinghammer, M., Schütz, I., Blohm, G., Fiehler, K. Allocentric information is used for memory-guided reaching in depth: A virtual reality study. *Vision Res.* **129**, 13–24 (2016).

176 Sadeh, M., Sajad, A., Wang, H., Yan, X., Crawford, J. D. The influence of a memory delay on spatial coding in the superior colliculus: is visual always visual and motor always motor? *Front. Neural Circuits.* **12**, 74 (2018).

177 Sadeh, M., Sajad, A., Wang, H., Yan, X., Crawford, J. D. Spatial transformations between superior colliculus visual and motor response fields during head-unrestrained gaze shifts. *Eur. J. Neurosci.* **42**(11), 2934–2951 (2015).

178 Sajad, A., Sadeh, M., Keith, G. P., Yan, X., Wang, H., Crawford, J. D. Visual-motor transformations Within frontal eye fields during head-unrestrained gaze shifts in the monkey. *Cereb. Cortex.* **25**(10), 3932–3952 (2015).

179 Berger, M., Agha, N. S., Gail, A. Wireless recording from unrestrained monkeys reveals motor goal encoding beyond immediate reach in frontoparietal cortex. *Elife.* **9**, e51322 (2020).

180 Draschkow, D., Nobre, A. C., van Ede, F. Multiple spatial frames for immersive working memory. *Nat. Hum. Behav.* **6**(4), 536-544 (2022).

181 Meilinger, T., Riecke, B. E., Bülthoff, H. H. Local and global reference frames for environmental spaces. *Q. J. Exp. Psychol.* **67**(3), 542–569 (2014).

182 Meilinger, T., Strickrodt, M., Bülthoff, H. H. Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation. *Cognition.* **155**, 77–95 (2016).

183 Treisman, A. M., Gelade, G. A feature-integration theory of attention. *Cogn. Psychol*. **12**(1), 97–136 (1980).

184 van Gomple, R. P. G. *Eye movements* (Elsevier, 2007).

185 Wolfe, J. M., Võ, M. L.-H., Evans, K. K., Greene, M. R. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn. Sci*. **15**(2), 77–84 (2011).

186 Engel, A. K., Maye, A., Kurthen, M., König, P. Where's the action? The pragmatic turn in cognitive science. *Trends Cogn. Sci*. **17**(5), 202–209 (2013).

187 Adam, J. J., Bovend'Eerdt, T. J. H., Schuhmann, T., Sack, A. T. Allocentric coding in ventral and dorsal routes during real-time reaching: Evidence from imaging-guided multi-site brain stimulation. *Behav. Brain Res*. **300**, 143–149 (2016).

188 Chen, Y., Crawford, J. D. Allocentric representations for target memory and reaching in human cortex. *Ann. N. Y. Acad. Sci*. **1464**(1), 142–155 (2020).

189 Wolbers, T., Wiener, J. M. Challenges for identifying the neural mechanisms that support spatial navigation: the impact of spatial scale. *Front. Hum. Neurosci*. **8**, 571 (2014).