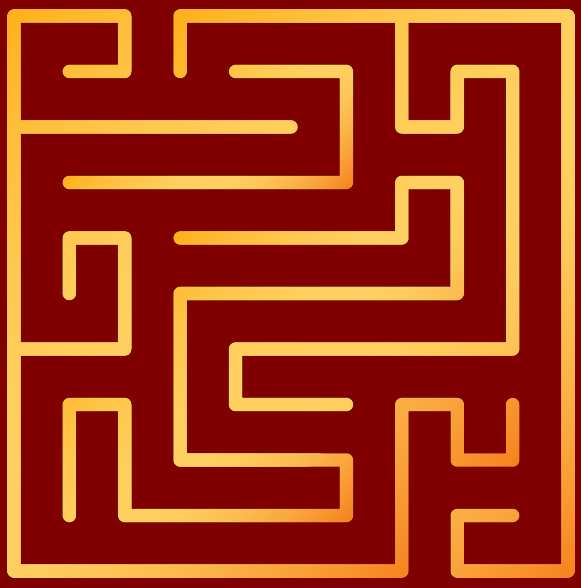
On the Relation between Representation and Embodiment in Spatial Cognition



Nicolas Kuske

On the Relation between Representation and Embodiment in Spatial Cognition

Dissertation zur Erlangung des Grades
"Doktor der Naturwissenschaft"
im Fachbereich Humanwissenschaften der Universität Osnabrück

vorgelegt von

Nicolas Kuske

Osnabrück, August 2020

Supervisor:

Prof. Dr. Peter König
Osnabrück University, Osnabrück, Germany

Additional Reviewers:

Dr. Tobias Meilinger
Eberhard Karls University of Tübingen,Tübingen, Germany

Prof. Dr. Fred Hamker
Chemnitz University of Technology, Chemnitz, Germany

Abstract

The claim that neuronal activity represents properties of the world outside of the brain plays a fundamental role in most of cognitive science. Enacted embodiment is the name of a theoretical framework which purports cognition to be rooted in bodily action. To investigate the relationship between representation and embodiment, the work presented here focuses on reasoning processes involving spatial relations among objects which are not part of the cognizing agent's body. These environmental spatial relations are called allocentric.

After defining the term "representation," a model of human cognition is developed which allows us to experimentally distinguish representation from computation and action in the brain. It is further argued that the relation of neural activity in areas classically considered sensory and motor is a fundamental organizational principle of the brain. Importantly, the structure of the relation also depends on the embodiment of the agent.

Finally, a study is presented in which participants explored a virtual reality (VR) city in different embodiment conditions and on multiple days. After each exploration participants completed tasks asking for different allocentric spatial relations. Performance in the spatial tasks interacts with both exploration time and embodiment condition. The findings indicate allocentric spatial representations to be structured by bodily action. Remaining variance can be explained through individual differences in spatial aptitude. Embodiment cannot account for the individual levels of ability. In conclusion, bodily action only partially structures the cognitive processes which represent spatial relations among objects in the agent's environment.

Table of Contents

Prologue

Bodily action is one of the primary means by which we explore and shape the space around us. In this dissertation I try to shed some light on the question whether bodily action is also one of the primary means by which we explore and shape the space inside of us. In particular, the structures which play an indispensable role in enabling spatial cognition.

The dissertation is divided into three main chapters. It begins with evaluating the role of the term “representation” in cognitive science. After developing my own definition of a cognitive representation I move on to show that the commonly used definition leads to an unacceptable inflation of the terminology. Then, following a simple model about the brain, I develop several properties applicable in experiments to differentiate representation from other cognitive processes in the brain.

However, since the brain is a recurrent network, the notion of representation, especially in the sense of flow of information about environmental properties, again proves controversial. Notwithstanding, using the tools developed earlier in the paper, a detailed analysis of the remaining modularity in the brain allows me to update the simple model while leaving some of its core structure intact. The chapter ends with a summary of my account of representation in the brain as well as a short discussion of remaining issues.

The second chapter introduces the notion of embodiment in cognitive science. It focuses on the claim that the brain has no inherent hierarchy. I discuss in detail the alternative view of sensorimotor relations being the major organizational principle. The chapter continues with an overview of the take on representation from proponents of embodiment, which I relate to my own position. Finally, I discuss limitations and questions raised by critics of the embodied framework.

The third chapter starts by introducing spatial cognition as the prime example to investigate the relationship between embodiment and representation. In particular, cognition involving allocentric spatial relations, or spatial relations among objects

which are not part of the agent's body. That means environmental spatial relations. The chapter continues with a detailed introduction to spatial cognition with a focus on allocentric spatial knowledge. I further discuss a pilot study our lab conducted on the relation of embodiment and allocentric knowledge.

The main part of the chapter begins with an introduction of the study I took part in during my PhD. In my doctoral project, we set up a virtual reality (VR) city, which participants explored in one of three embodiment conditions: in VR, studying a map of the city, or exploring the city in VR while wearing a belt that indicates north. In each of the three experiments, participants explored the city on multiple days. We found that accuracy when judging alignment of housefront and north rises with consecutive experimental sessions, while judging relative house orientations remained constant. Exploration with belt lowered accuracy when aligning houses to north. The findings suggest a difference in allocentric encoding of house orientation and cardinal direction due to different actions associated with the two spatial concepts. Furthermore, pointing from house to house was more accurate after VR- than after map exploration, relative to the other two tasks. Thus, all three embodiment conditions led to different spatial representations.

We also found behavioral variables to inform differences in spatial aptitude and strategy across all three experiments. Related variance in task performance does not interact with the above stated results and, thus, provides evidence for their stability. Nevertheless, it is not clear how embodiment can account for the large individual differences in spatial aptitude. In conclusion, bodily action structures not all, but a part of the allocentric representations underlying spatial cognition.

My dissertation closes with an epilogue. Here, I bring together the different chapters once again. The epilogue provides a more general discussion and conclusion concerning the relation of representation, embodiment, and spatial cognition.

Chapter I. Representation

Talk of representation is ubiquitous in cognitive science. This is especially true when investigating the role of brain processes. Here, the terminology encountered most often is that of neural code (Perkel & Bullock, 1968; deCharms & Zador, 2000; Brette, 2018). Both terms, representation and code, can be considered synonymous.

Romain Brette (2018) argues, however, that to speak of a neural code is to use a metaphor deriving from the use of symbols in language. The metaphor becomes apparent, for example, when considering basic examples in coding theory. There, a code is a set of codewords being mapped onto another set of words following some probability distribution (see, for example, Roth, 2006). Admittedly, metaphors are important cognitive tools that shape our thinking (Lakoff & Johnson, 1980). Nevertheless, a metaphor can only accurately match a part of what it describes. A correct and comprehensive description of a state of affairs is not a metaphor.

For this and reasons I will go into later, Brette as well as other researchers believe that cognitive science should get rid of the false friend representation, at least when talking about basic cognition (Chemero, 2009; Engel, Maye, Kurthen, & König, 2013; Hutto & Myin, 2017). Schlicht and Starzak (2019) caution against such a "radical" move. They point out that the opponents of representation still have to come up with a convincing concept shouldering the same explanatory burden as the supposed metaphor of neural code seems to do. In fact, one may ask if it is really a metaphor to speak of a neural representation. Claude Shannon already allowed for real valued analog signals in his seminal work on message passing in 1948. Consequently, while coding theory is applicable to finite sequences of symbols, it can also be applied to continuous real valued sequences (Roth, 2006). A fact which seems to seamlessly transfer to the idea that continuous change of neuronal structure and charge can represent processes in the environment.

To be precise, a neuron's representative power need not be limited to objects, or relations among objects in its environment. Given a skilled surgeon and a curved microscope, a patient might be able to observe her own neurons in her own visual cortex, for example.

If we assume the visual cortex to take part in the representation of that neuron, then a neuron can take part in creating its own representation. This is not a paradox because the light reflecting off the neuron at time t will have to travel through the patients eyes before it can be represented. Therefore, the same neuron at time t + 1 will simply represent a past version of itself. In conclusion, self-representation is possible if that which is represented is a past version of that which represents. That way we can still distinguish representation from identity. In conclusion, a neuron can represent objects or relations among objects in its environment, as well as its own past state.

To find the mapping between both the representing set of neuronal activity R and the represented set of environmental and/or neuronal processes E, the experimenter needs to resort to the following logic: She prepares the experimental condition E (or not E). Then she measures the value, or time series of values, of one or more properties of the elements in R. One example of such a property would be the frequency of action potentials, also known as firing rate. She then goes on to compare the value(s) of R measured in condition E to the value(s) of R in another condition, not E. The condition E can be defined by instruction, (intracranial) stimulus, or lesion, or even by another participant, e.g. comparing another brain's response to the same stimulus. In all and only the cases where the experimenter can identify a statistically significant difference between the two measurements can she clearly map the properties' values to their respective conditions. Indeed, only if such a relationship is found is it generally accepted to speak of a neural code or representation.

Admittedly, however, this general acceptance rests on a hidden assumption that may be unwarranted. The experimenter has shown that she can leave the experimental room (trusting she is not needed for the procedure) and by looking at the neural measurement alone can deduce the experimental condition. Thus, for the experimenter, a certain measurement represents a certain condition. The measurement of R reduces the experimenter's uncertainty if she were to decide between E and not E. She has not shown that R informs other parts of the participant's brain about E. That is, the existence of a correlation, or more general mutual information, between experimental condition and neural activity

might not be sufficient to call this a representation *for* the brain of the participant.

A representation seems to get its name *because* it informs someone about something. We need at least three components. Something needs to be represented, the environmental or past-self condition E. The represented is also known as the content of the representation. Another thing needs to represent, or code or inform. As introduced above, I abbreviate the representation as R. Philosophers also refer to that which represents as the vehicle (Dennett 1991, Millikan 1991, 1995). Finally, there is the interpreter, receiver or consumer (in the same order Peirce, 1906; Shannon, 1948; Millikan, 1995). That is the system (the someone) for which the representation actually means something. I will often refer to the act of interpreting the representation as computation and will therefore abbreviate the system which receives the representation as C.

I define a representation R for a system C as a structure which informs the system about a property of another, the represented, structure E. Two structures are different if they exist at a different place, at a different time or both. Here, to inform means that R influences C to act overtly, that means through bodily movement, on E. Alternatively, R can also inform C if it leads to covert, i.e. mental action, on E.

What it means to "act on" something seems relatively unproblematic to define for bodily behavior. There, the experimenter can compare movement trajectories or more simply overt choices, like when the participant has to decide between one of two options. If the participant can solve a task *while* her brain exhibits R values that the experimenter has already linked to E, and the participant cannot solve this task without R, then it can be said that the participant represents E. Admittedly, since significance is a statistical term, this assessment would not always be correct in the concrete situation. Nevertheless, in most of these cases, R is representing E for the participant.

So far, the dissociation between third-person decodability and decodability for the participant has only been hypothesized. That such cases exist, however, becomes apparent for example in the case of distractor-induced blindness. In the paradigm, a participant has to press a button if a certain movement pattern occurs after a

cue appears. The experimenters found that, if the pattern occurred shortly before the cue, participants would overlook the next pattern after the cue. EEG measurements, however, did not reveal a difference in the neural activity in the visual cortex in either condition (Niedeggen, Michael, & Hesselmann, 2012). While EEG is a somewhat coarse-grained method the result at least corroborates the hypothesis. Even when the participant missed the movement pattern due to distraction, the pattern, that is the represented E, was in principle still decodable from the visual cortex, just not so for the participant. In conclusion, there are situations in which neural activity in someone's brain is only decodable for a third person.

Thus, while task performance allows the experimenter to distinguish between third- and first-person information through deliberate action, such a measure is not available to identify covert or mental action. Consider, for example, cognitive processes involved during reasoning. Even some opponents of the idea that basic cognition involves representations admit that high-level processes can make use of code (Hutto & Myin, 2017). But how do we differentiate parts of the brain that represent from those that act covertly on those representations? After all, representations are neural activity as well.

A simple but influential model of the brain, known as the sandwich model (Hurley, 1998), actually distinguishes among three parts. There are those parts that represent, those that do computations on these representations, and those that transform the result of the computation into bodily behavior. Note that the model involves an arrow of time. The sandwich gets pierced starting in the representational layer. The arrow then reaches the meat of the computation and ends in the action layer. Hence, the neural system which does the computations does the first mental act involving representation.

If we assume the sandwich model to be correct, then, for an experimenter to talk about representation in intracranial processes she needs to separate four forms of neural activity. First is the neural activity that only carries mutual information for the experimenter. I will refer to that activity as O. Second is the activity R that represents for the participant. Third is the neural system C that interprets these representations through computations. Since

C is part of the participant, it remains correct to speak of R as a representation for the participant as well. Fourth and final is the transformation into bodily action. I will refer to the region which transforms the output from C into bodily action as A.

In the remainder of this chapter, I will identify a set of properties which can help to distinguish representative activity R in the sandwich model from O, C and A. I will then show that, due to recurrent processing in the brain, the sandwich model cannot be held up in its current form. Nevertheless, keeping the properties of representation developed along the way, we will see that an alternative model can bring together both representation and recurrence.

I.1. Oh Representation, Where Art Thou?

Neural activity O holds information about experimental condition E which is exclusively available to the experimenter and not usable for the participant. However, O is only a subset of the total neuronal activity which carries information about E for the experimenter. It might well be that the information R about E for the participant, the activity C which interprets R, and the action transformation A using the result of the C's interpretation, might all be decodable for the experimenter as well. Consequently, analysis of the sum total of third-person-decodable neural activity is a good starting point to identify measurable properties in the brain that allow us to identify representation for the brain.

A major issue for the simplest possible notion of representation in the brain, namely the third-person decodability, is that decoding location depends on decoding method. The ability to decode information about E from a certain location in the brain increases with precision in measurement and modelling technology. There is no question that we can train a (non-linear) classifier to distinguish animal from inanimate object stimuli given we provide it with a neural activity vector from the retinotopic striate cortex. Within the first twenty years of the development of functional magnetic resonance imaging, Kriegeskorte et al. (2008) were able to show information about this difference to be decodable from the inferotemporal cortex in humans as well. In 2019, Kietzmann et al. investigated the temporal dynamics of representation and found that categorical difference also becomes decodable from the

middle temporal cortex (Kietzmann, Spoerer, Sörensen, Cichy, Hauk, & Kriegeskorte, 2019).

Given the ongoing improvement in measurement precision and modeling techniques there seems to be, in principle, no reason why we should not be able to eventually decipher this information from other cortical areas as well. The premotor cortex, for example, can carry information about sensory stimulus as well as (perceived) motor action (Etzel, Gazzola, & Keysers, 2008; Schwartz, Moran, & Reina, 2004). It also has been shown that attentional and decision processes are decodable both from the frontal cortex and areas classified as sensory (Armstrong, Fitzgerald, & Moore, 2006; Wilming, Murphy, Meyniel, & Donner, 2020). In contrast, Kapoor et al. have shown the decodability of the visual percept in a no-report binocular rivalry paradigm in the prefrontal cortex of the macaque monkey (Kapoor et al., 2020). Furthermore, sensory cortices show multimodal properties as becomes apparent, for example, when visual information can be decoded from the auditory cortex (Kayser, Petkov, & Logothetis, 2008). Note that all these examples present experimental evidence for representational qualities of the cerebral cortex. We did not even touch on subcortical structures like the basal ganglia or thalamic nuclei, which are also much harder to obtain data from, especially in humans. The inflation of the number of representative brain processes when one uses third-person decodability as sufficient criterion is illustrated by the following quote from Michael Anderson (2010, p.2): “An empirical review of 1,469 subtraction-based fMRI experiments in eleven task domains reveals that a typical cortical region is activated by tasks in nine different domains.” He continues (ibid.), “… one gets the same pattern of results even when dividing the cortex into nearly 1,000 small regions.” In conclusion, given sophisticated experimental design, sufficient data and intelligence of decoding method, representation seems to be everywhere.

I.2. Lesion to the Rescue

One method to find out about a component part’s function in a working system is to observe that part. Another method is taking the part out of the system and observing what the system cannot do anymore. While double dissociation studies can shed light on

mutual dependence of a system's parts (Parkin, 1996; Teuber, 1955), single dissociation is sufficient to show a dependence of a certain function on a certain part. Consequently, lesion studies allow us to differentiate neural activity O, decodable only for the researcher, from neural activity that has a function for the participant.

Upcoming bodily movement can be decoded from frontal, parietal and motor areas (Gallivan, McLean, Valyear, Pettypiece, & Culham, 2011). However, lesions to the frontal lobe, for example, usually affect a host of cognitive functions (Grattan, Bloomer, Archambault, & Eslinger, 1994; Stuss, Floden, Alexander, Levine, & Katz, 2001). Lesions in the classical motor areas, on the other hand, will almost exclusively affect the ability to activate well defined muscle groups (Shelton & Reding, 2001), at least in the short term. We will return in more detail to the relationship between motor cortex and action transformation A in the next section.

Patients with lesions in sensory cortices still have control over their bodily actions. Their issues are mostly related to quick goal-directed action, which becomes much more cumbersome without the proper sense. Importantly, however, goal directed action generally does not become impossible. Indeed, apart from becoming slower at solving certain tasks, it is remarkably hard to define what function the cognitive system actually looses when the lesion only affects sensory modalities. Helen Keller makes a prominent example as the first deaf-blind person to earn a Bachelor of Arts degree. Her describtion of colors is suprisingly elaborate. Yellow, for example, is "like the sun. It means life and is rich in promise" (Keller, 1929).

The redundancy of sensory modalities can be partly understood since many tasks in everyday life involve abstract concepts, that means, concepts which are invariant with respect to the modal input. If an experimenter can decode from a brain region during a task involving such an abstract concept, she refers to these regions typically as supramodal representations. Numbers, for example, which one might regard as a hallmark of abstraction, show differential activation compared to letters and colors across modalities in the horizontal intraparietal sulcus (Eger, Sterzer, Russ, Giraud, & Kleinschmidt, 2003). Even emotional content is an

abstract feature of a stimulus since it has been shown to be linearly decodable from the posterior superior temporal cortex, irrespective of auditory or visual stimulus modality (Sievers, Parkinson, Kohler, Hughes, Fogelson, & Wheatley, 2018). Similarly, the famous place cells, the activity of which allow the experimenter to decode the position of (human) animals (Ekstrom et al., 2003; O'Keefe & Dostrovsky, 1971) in well-known environments, also form in early-blind rats (Save, Cressant, Thinus-Blanc, & Poucet, 1998).

Lesions in the above mentioned areas severely affect navigation ability, emotion recognition, and processing of numerical magnitudes, respectively (in the order of effect: Pearce, Roberts, & Good, 1998; Meletti et al., 2003; Kadosh et al., 2007). Therefore, supramodel representations seem to become decodable for the experimenter only when they also play a functional role for the participant.

Independence of modality, however, is not a necessary property for a brain region to be third-person decodable and take part in a first-person function. The well-known face area in the temporal cortex, for example, is necessary for the ability to recognize people's faces without haptic input (Kanwisher, McDermott, & Chun, 1997). Activity in the region also allows the experimenter to decode if there was a face in the participants visual field or not. Indeed, this type of a more detailled task description lets us decipher the function of early sensory cortices, as well. Vision, for example, allows us to pick out an object of a certain color, given it is situated at a certain angle relative to our retina, without having to rely on other sensory modalities. Area V8 in humans is a functional focal point of this ability (Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998; Nunn et al. 2002). Again, activity in the region can also provide third-person readable code for color.

The lesion approach of representation dissociation presented here, has become a more flexible tool over the last decades. Electrical current induced in the brain through transcranial magnetic stimulation (TMS) allows for non-invasive, temporary lesions (Hallet, 2000; Jacquet & Avenanti, 2015; Müri & Nyffeler, 2008). We will discuss an example of TMS in more detail below.

To conclude, lesion is a helpful tool to identify neural activity which represents something for the experimenter but not for the participant, that is neural activity O. We found our first property on the way to find neurons R that represent for the participant.

Being able to identify a certain brain region as necessary for a certain function related to that which is represented (E), does not make this brain region a representation for the participant, however. In the sandwich model of brain function we still have to distinguish between R, the region C which interprets R, and the region that acts A.

I.3. Choose Your Model

For simplicity, let us assume in the remainder of this section that the lesion filter of functionality worked perfectly. All the mutual information about E the experimenter is left with either stems from R, C or the action region(s) A of the participant. Our task would then reduce to mapping the remaining third-person decodable neural activitiy on R, C and A, respectively. In this section, I will introduce two more properties that can help us brave the representational chaos. Both properties, or conditions, were proposed by Peter König (2017). First, he argued that in order to differentiate coding neural activity R from neural activity C that interprets R through computation in the participant's brain, the experimenter should consider the complexity of the decoding model she uses. Only if a linear model using parameters from the condition E is the best fit for the neural activity under investigation can we call that activity a representation of E. Concretely, only models of the following form are allowed:

$$\mathrm{r} = Ew + b \tag{1}$$

Here r is a vector made up of all N measured values in the purported region of representation. In the statistical-analysis literature this vector is usually referred to as the dependent variable (see for example Cohen, J, Cohen, P, West, & Aiken, 2003). The parameters describing the experimental condition E make up the matrix E. The number of rows of that matrix is again equal to the number of measurements of neural activity. The number of columns of E, however, can range from one to N. This number is often referred to as the number of independent variables. It cannot exceed N. That is because w is a coefficient weight vector with the

same number of elements as number of columns in E. We can only find unique optimal values for these weights if the number of column "base" vectors is below or equal to the dimensionality of r. Otherwise they would lose their supposed mutual indepedence. The last term b is a constant vector of size N again. If the measurement is time dependent while the experimental condition E stays constant, then the weight vector and the constant term may be modeled as depending on time as well.

The reasoning behind this restriction on the complexity of decoding method is that if, in order to explain a measurement, we need to assume an interaction of a parameter describing condition E with itself, i.e. $E_{ij}E_{ij}$ or with another parameter $E_{ij}E_{kl}$ then this should be considered a computation on E. An interaction is not "simply" a representation of E. Note that the Taylor expansion of a non-linear differentiable function $f(Ew)$ contains interaction terms of the above form as well. Hence, similar reasoning excludes these from being included in our linear model. In conclusion, the property of linearity seems to resolve, in part, the dependence of decoding location on decoding method by allowing the experimenter to sort measurements into representation and computation.

What about action? Recall that lesions in the classical motor areas have very specific effects. They mostly impair overt behavior, i.e. bodily action. There is, however, another reason why the motor cortex has been dubbed as such, namely that it allows the experimenter to decode upcoming bodily movement using linear models. Single neurons in the motor cortex show activity dependent on movement direction. Thus, a linear sum of these movement "vectors" weighted with their respective (normalized) activity predicts the direction of arm movement, for example (Georgopoulos, Schwartz, & Kettner, 1986).

The relationship is, in principle, revertable. Given an experiment in which the participant has simple movement instructions, using a linear model, the researcher could roughly predict the activity of each of these motor neurons. Does that mean the condition E, movement instruction, is represented (for the participant) in the motor cortex? No, because one condition required to speak of a representation R has not been met. The arrow of time in the sandwich model has been violated. Recall that representation

comes before computation, which again comes before action. Primary motor cortex efferents almost exclusively project through the internal capsule to neurons in the spinal cord (Shelton & Reding, 2001), which again project to muscle groups. Consequently, the primary motor cortex does not represent action *for* the brain. The primary motor cortex can be identified as a part of the region A which transforms computation into action.

As mentioned in Section I.1., however, the role of the region known as premotor cortex is not so clearly discernible. Andrew Schwartz and colleagues had a monkey move a joystick tracing the outline of an ellipse (Schwartz, Moran, & Reina, 2004). The monkey could not see the actual movement but instead saw a video screen depicting a computer-generated representation (CGR) of the joystick's movement in real time. The experimenters, then, gradually added an inward force at certain points during the monkey's movement such that the real trajectory changed into a circle. At the same time the eccentricity of the CGR movement remained unchanged. The ape moved his arm in a circle but saw an ellipse. Using the linear model of the population vector the experimenters could decode the circle from the motor cortex and the ellipse from the premotor cortex. Furthermore, lesion studies provide evidence that the premotor cortex does not only correlate with, but also plays a functional role for perception (Candidi , Urgesi, Ionta, & Salvatore & Aglioti, 2008; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007). Similarly, mental-imagery experiments have found above-baseline activity in the premotor cortex when participants were imagining bodily movements but not (or debated) in the primary motor cortex (Dechent, Merboldt, & Frahm, 2004). In conclusion, the premotor cortex is not part of the bodily action region A.

Indeed, the findings question the compatibility of the sandwich model and linearity of representation. At least, if we consider perception and mental imagery, both, to play *only* the role of a representation. We have established that the primary motor cortex is for action. If the premotor cortex takes part in perception and, hence, representation, there seems to be no computational layer in between representation and action. The clash of definition and reality could be circumvented if perception and mental imagery would involve representation and computation. Otherwise, either the sandwich model or linearity of representation is wrong.

Apart from this potential shortcoming, linear decodability as criterion to identify representation has two more issues. First, sophisticated models are not the only reason why we find third-person-decodable activity all over the brain. Improvement of measurement techniques plays an important part in this inflation of representation, as well. Second, in terms of variance explained, a model generally gets worse with improving measurement precision. We have to agree on what precision in space and time suffices to distinguish a linear from a non-linear explanation. Fortunately, however, once we agree on measurement resolution, we can solve the first issue.

The solution to the problem of unexplained variance is simple. It is the second property proposed by König to dissociate representation R from interpreting computation C and action transformation A. The linear model needs to explain all the variance in the signal. In order to find out what a neural signal represents, the factors in the model need to add up, explaining all the signal's variance except of the noise inherent to the finite measurement precision. If any variance is left, we either have not found all the neuron represents or the neuron is not only involved in representing. The latter explanation would not fit with the layered sandwich approach. To give an example of the former, the variance in firing rate recorded from the auditory cortex has been shown to be best explained by including both the properties of the sound source and the visual stimuli (Kayser et al., 2008).

As noted above, the imprecision of the measurement method defines the resolution of the necessary model. Thus, König's proposal could lead to an interesting consequence. With better methods, events long past or far-future plans might have to be added as representational factors into the model in order to identify the remaining signal. A similar case is made by Stanley Klein considering the concept of memory (2015). He argues that, following the standard definition, "memory […] can be seen in virtually every mental state we are capable of having" (ibid., in abstract) and, hence, the term becomes void of meaning. Instead, only memories which are actually perceived as past events should be considered memories. I believe, however, that the criterion of linear decodability prevents too extreme extension of

representation into the past, and in our more general case, also the future.

I will conclude this section with a case study involving the well known representational (dis-)similarity analysis. The method has been developed over the last decade and has since seen a surge of interest (Devereux, Clarke, Marouchos, & Tyler, 2013; Kriegeskorte, Mur, & Bandettini, 2008; Marchette, Vass, Ryan, & Epstein, 2014; Popal, Wang, & Olson, 2019). The case study uses the method in a particular context which fits almost, but not quite, to our definition of a linear model. The study is of particular interest because, by making use of the noise-limit approach, it reveals that we have to update our sandwich model.

The study was conducted by Kietzmann et al. (2019, see above). In their experiment, they presented visual stimuli to the participant, for example, an elephant or a pineapple. Recall that Kietzmann was able to decode the animal-vs.-object category from the temporal cortex. A rough explanation of Kietzmann's method is that he adapted the weights w of a linear model until the sum of experimental condition property vectors E matched a vector of cortical response patterns r (see Equation 1). In more detail, however, both property and response pattern vectors were already simple functions of property and response patterns.

The model was a linear combination of low-level stimuli properties like spatial frequency or orientation and high-level properties like real-world size or category. For each level of stimuli properties, Kietzmann et al. created a square matrix, which we can identify as one column vector in E. All matrices were symmetrical along the diagonal. Each matrix element had a value from zero to one depending on how similar the specific property of the object in the row (pineapple) was to the corresponding column-object's (elephant) property. For example, one high-level-property matrix had only a square of ones in the upper left corner because it modeled the property of object category. In the left part of the column and the upper part of the row were only animals. The rest of the matrix contained zeros. The low-level-property matrix of spatial frequency, on the other hand, contained high values for those elements that had similar spatial-frequency content, independent of their category or other properties. Each element of

one similarity-of-property matrix (vector in E) was, thus, already a simple function of two properties of experimental condition E.

Likewise, the response matrix or vector r was obtained by comparing the similarity of the neuronal response to two different stimuli. Namely, by comparing the one in the row (pineapple) to the corresponding one in the column (elephant). The comparison was the normalized scalar product of the MEG response vectors of the respective stimuli.

A scalar product of two neuronal response vectors to two different stimuli can be considered a sum of interactions. The result is not revertible. For these reasons, I claim that representational similarity analysis does not necessarily compare representations. It can only give us an estimate of how similar neural activity among different conditions E is. A priori, we do not know if we are observing representation R, interpretation C, or action A.

Nevertheless, activity-similarity analysis remains an important tool even in our context. That is because, if R would result from the linear combination of properties from E, then the vector r of scalarproducts comparing those representations would result from linear combinations of scalar products comparing property vectors of E. In other words, if there were are linear model from E to R, the model of scalar products should fit as well. Consequently, representational similarity analysis allows us to exclude the possibility of a linear model explaining neuronal activity and, therefore, to exclude the possibility of the activity being a representation.

Kietzmann found that the weight of the high-level-property-similarity matrix or vector was negligible in the early visual cortex. Furthermore, low-level similarity could be ignored in the temporal cortex. Following the noise-limit approach, we, thus, can conclude that the property of spatial frequency, for example, is not represented in the temporal cortex. At least, if we accept the MEG measurement precision to be acceptable for such a claim. Given that the arrow of time in the sandwich model first pierces the sensory cortices, it would further follow that either the early visual cortex or the temporal cortex represent and compute. Some regions in the brain use the representations in other regions to

compute new representations. In conclusion, Kietzmann's findings call for an update of the sandwich model.

I.4. Circular Information Flow

The simple sandwich model would assume one layer of representation R, one layer which makes up the system C that interprets R and the action layer A. We have identified the primary motor cortex as part of the action layer. We assume a chain of causality leading to action that starts at the sensory cortex. Then, as our analysis of Kietzmann's experiment revealed, the three-layer model is not sufficient. The reason is that representations of high-level properties are not found in early sensory cortices.

One solution to this problem would be to assume a multilayer sandwich model. The simplest such model would consist of one low-level-property R layer, followed by one C layer interpreting that representation and sending the result to a high-level R layer, which is again followed by another C layer and then action A. Such a five-layer model could be easily adapted to a seven-layer model and so on. The model has the issue, however, that it seems redundant to have an extra representational layer after each computation, especially since the representation is simply the result of the computation. A computation takes time. Therefore, we could consider a layer to be an C layer while it is performing a computation on a lower-level representation and an R layer (high-level) when its computation is done. Layers simply transform from C to R and back if the lower-level input changes. Only the lowest or earliest layer would have to be assumed to always be representing.

Importantly, this switchable, multilayer sandwich model would preserve the causal arrow. Representation leads to computation and eventually to (functional) bodily action. Such a model is also known as the feedforward model.

The brain, however, is not a multilayer, feedforward network. It is well known, for example, that the visual cortex contains a significant number of active feedback pathways (Felleman, Van Essen, 1991; Timo Van Kerkoerle et al., 2014). To quote Markov et al. (2013) on the macaque visual cortex: "FF [feedforward] projecting neurons dominate over short distances, and FF and FB [feedback] are about equal over long distances; hence, on average

nearly 80% of neurons projecting less than 10 mm participate in an FF pathway, while on average 60% of neurons engaged in projections longer than 10 mm are in an FB pathway." Indeed, the effect of probably comparably rich recurrent wiring in the human visual cortex also becomes apparent in the study by Kietzmann.

The reason the experimenters used MEG was to observe the development of neural activity with high temporal resolution. Participants were presented with each stimulus for 500ms. Thus, similarity of neural response for each combination of two stimuli (pineapple and elephant) could be computed for each millisecond. For the model nothing changed except that the coefficient weight vector w (see Equation 1) became time dependent.

Kietzmann found that the factor of category, that is animal versus object, emerges in the inferotemporal (IT) region at 140ms after stimulus onset and vanishes at 200ms. The category then emerges at about 250ms in the middle temporal (V4t) region while also returning to IT, albeit considerably weaker. In the time window from about 250 to 350ms after stimulus onset, the animal-vs.-object category was decodable for the experimenter from both V4t and IT. Note that, from a feedforward perspective, information about a change in stimulus reaches V4t before it reaches IT.

Furthermore, causality analysis revealed that the only time in which IT did not significantly influence activity in V4t was during the first 30ms after the wave of activity change due to stimulus change had reached V4t. During the largest part of the 100ms-long time interval in which the category was decodable from both regions, changes in IT predicted changes in V4t and the other way around. Significance of category feature and causality relation over time are depicted in Figure 1.

Kietzmann went on to compare these experimental findings to feedforward and recurrent neural network models. Dynamics of the feedforward-only model fit dynamics of early visual cortices. Only the recurrent model, however, featuring both feedforward and feedback connections, showed activity patterns matching experimental data in early as well as in late-processing regions.

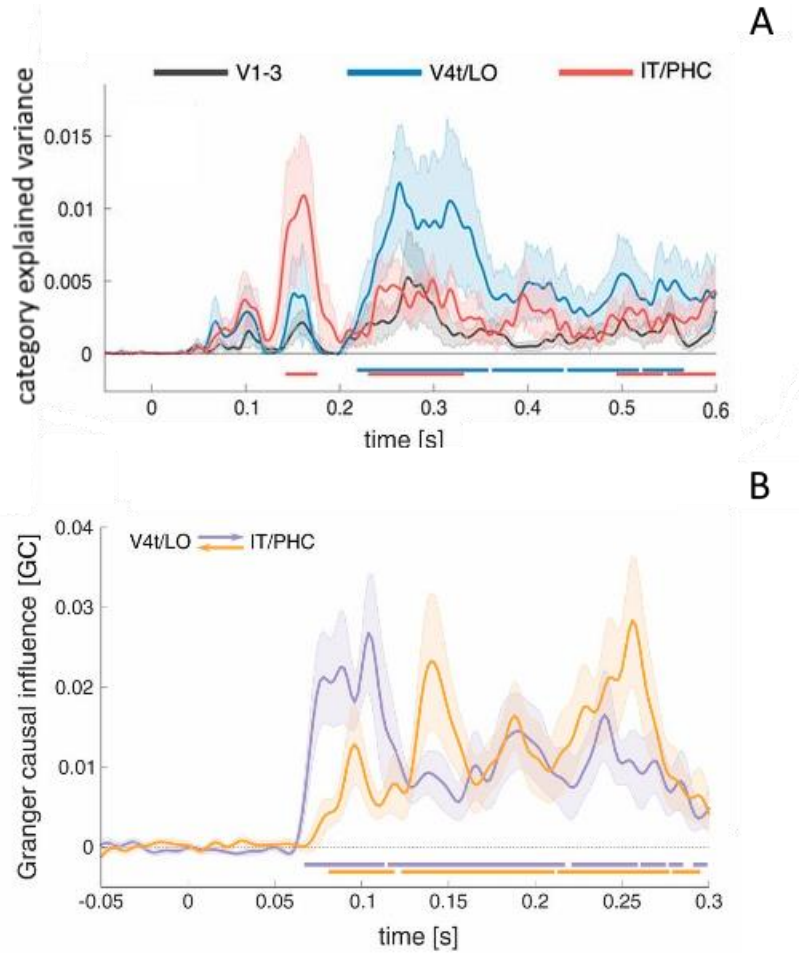


Figure 1: Recurrent processing in human visual cortex.
1.A Amount of variance in linear model explained over time by animal-vs.-object category for different regions of interest after stimulus onset. (Black curve) V1–V3. (Blue curve) V4t/lateral occipital cortex [LOC]. (Red curve) inferotemporal cortex [IT]/parahippocampal cortex [PHC]. The horizontal bars indicate time points with effects significantly exceeding pre-stimulus baseline. SE across participants shown as shaded area.
1.B Representational similarity Granger causality analysis was performed to estimate information flow between ventral-stream areas. (A) Feedforward (purple) and feedback (orange) direction of Granger causal influence between V4t/LOC and IT/PHC. The horizontal bars indicate time points with causal interactions exceeding effects during pre-stimulus baseline. Data is shown baseline corrected. Figure adapted from Kietzmann et al. (2019) with permission (CC BY-NC-ND).

Recurrent processing is functional in that it can increase robustness of visual information extraction (Kar, Kubilius, Schmidt, Issa, & DiCarlo, 2019; Rajaei, Mohsenzadeh, Ebrahimpour, & Khaligh-Razavi, 2019; Spoerer, McClure, & Kriegeskorte, 2017). However, the question remains, where exactly is information being extracted in the recurrent process described above, and by what neural

system? The most plausible explanation seems to be that neither V4t nor IT represented animals for each other. Rather, only through their interaction were they able to create the stable pattern of activity the experimenter was able to map onto the category of animal. This interaction process can be considered a computation interpreting lower-level representations. But can the stable pattern of activity emerging after some time be considered a representation?

In our experimental example of information transfer above, we focused on occipital and ventral temporal cortices. It is well known, however, that recurrence can be identified as a core principle of the architecture of our central nervous system. This becomes apparent when considering the cortico-basal ganglia-thalamo-cortical loops (Haber & Knutson, 2010; Parent & Hazrati, 1995; Sherman 2016), the default network (Buckner, Andrews-Hanna, & Schacter, 2008), and other "attentional" networks (Dixon et al., 2017, Dixon et al., 2018). Visual perception cannot be considered as completely independent of these larger networks as temporary blindness has been shown to be produced by emotion, surprisal and task related distractors (in the same order: Most, Chun, Widders, & Zald, 2005; Christopher Asplund, Todd, Snyder, Gilbert, & Marois, 2010; Michael, Hesselmann, Kiefer, & Niedeggen, 2011; Shapiro, Raymond, & Arnell, 1997). If the local pattern of activity in the temporal cortex feeds into a global pattern which itself feeds back onto the local pattern and only through their interaction keep each other stable, this is known as circular causality (Noble, R, Tasaki, Noble, P, & Noble, D, 2019; Witherington, 2011).

However, information cannot flow in circles. Information must flow unidirectionally from representation R to interpreting system C. Importantly, unidirectional information flow is compatible with R sending information to C, while C sends information to R. That is, information can be sent simultaneously in both directions. Only, the message that goes from R to C has to have different content than the message that goes from C to R. That is because, like Shannon (1948), I take information to mean reduction in uncertainty about the state of affairs in the world. In other words, something has to be learned. Hence, if R talks about E and C talks about E there is no information moving at all.

Consequently, the sandwich model needs a fundamental overhaul. We can keep the first low-level representation layer and the primary-motor-cortex action layer because they are the only direct connections of the brain with the environment. In between these two, however, the model I will employ from now on is a set of nodes reciprocally connected to each other.

To achieve are more general understanding of what is going on in such a recurrent network, let us consider our new model without representation or action layer. Let all nodes be initiated with a random activity pattern and then left alone. Only node R is never left alone, but rather is driven by a periodic pattern. At first, we can describe all other nodes as receiving information from R about this periodic driving force. Eventually, however, through continuous feedback and feedforward interaction, dynamic patterns will emerge in the system. Interestingly, depending on the exact setup of the system, using linear pattern analysis we could still decode the driving frequency, at least from some regions. Also, if we would lesion any of these identified regions, the pattern of the system can be expected to change, in general. Not fitting to our prior definition of representation is, however, that in this scenario even linear decodable regions take part in the complete, highly complex "computation." The issue that linearity does not separate computation and representation (*for* the system) anymore is separate from the issue discussed above, that information cannot flow in circles.

Therefore, the question arises if internal information transfer is the best way to understand the dynamics of an externally driven, recurrent system. Note that what we have described here shows great similarity to a classic resonance phenomenon. A guitar, or even a simple drum, can develop intricate resonance patterns along its body. The shape of these patterns depends in equal part on the shape of the instrument and the external driving frequency. A resonance pattern generally exhibits both temporal synchrony and spatial structure. Thus, understandably, talk of resonance in neural tissue is mostly limited to phenomena which exhibit synchrony-driven oscillatory patterns in the temporal domain (Buzsáki & Draguhn, 2004; Raj et al., 2020; Spiegler, Knösche, Schwab, Haueisen, & Atay, 2011; Thompson & Varela, 2001; van der Groen & Wenderoth, 2016).

One can, however, have resonance without oscillatory synchrony. Neurons, for example, are also called resonant when they act as frequency filters (Hutcheon & Yarom, 2000). Indeed this type of resonance is a cornerstone of Adaptive Resonance Theory introduced by Stephen Grossberg (1976a, 1976b). Very roughly, Grossberg speaks of resonance when, given a certain stimulus, from a feedforward perspective later layers tune the properties of earlier layers and vica versa, such that a stable, global pattern emerges. The underlying framework of that theory has been further refined to provide explanations for a wide range of cognitive phenomena (Carpenter, Grossberg, Markuzon, Reynolds, Rosen, 1992; Grossberg, 2012). Interestingly, there is another theory about general brain function which has a very similar approach. In predictive coding a stable percept for example can only develop if the feedfoward signal matches the "expectation" of the system (Friston & Kiebel, 2009; Hohwy, 2013). In mechanistic terms, the expectation translates to changes in the system's architecture which are happening through feedback signals. Naturally, the feedback signal is caused by earlier interaction of feedforward and feedback. Thus, although in predictive coding a stable percept is not called a resonance phenomenon, the terminology seems to be applicable here, as well.

The experiment by Kietzmann discussed above found a spatial pattern emerging from the interaction of recurrent neural networks and environmental input. This spatial pattern emerged once and then reemerged later, holding its shape for about 100ms. Importantly, however, Kietzmann did not report on a possible temporal structure within the spatially stable activity pattern. Temporal fluctuations were either random or overlooked. In either case, we have evidence for stable spatial patterns which emerge out of the interaction of internal feedforward and feedback processes, driven by environmental stimuli. Hence, I believe it fruitful to include such emergent stable patterns in the category of resonance phenomena in the brain.

This brings us back to the question of representation. In engineering, or physics, resonance is usually not described in terms of internal information exchange. It seems that it would not add anything to the explanation. As Chemero (2009, p.77) puts it, with particular reference to dynamical systems theory: "[T]he representational description of the system does not add much to

our understanding. […] despite the fact that one can cook up a representational story once one has the dynamical explanation, the representational gloss does not predict anything about the system's behavior that could not be predicted by dynamical explanation alone." Similarly, for Hutto and Myin (2013), there is simply nothing to represent. The brain interacts with the world, and together they make up a mind. The idea that parts of the brain should message other parts of the brain about states of the world only emerges with social cognition and public symbol systems (Hutto & Myin, 2017). Brette (2018) is defending a similar point when calling neural code a metaphor deriving from language encryption methods. Furthermore, as Ramsey (2017) points out, it might unnecessarily constraint our theorizing to presuppose representation as necessary for cognition. Indeed, why *should* we consider neuron R as having to inform neuron C in order to understand the brain's role in cognition?

I.5. Resonance in Fuzzy Nodes

Due to the large amount of feedback connections, as embrained agents, we are bound to work with resonance patterns. We have seen that if these resonance patterns span the whole system, like a standing wave pattern on a drum, the notion of information transfer becomes obsolete. As we will see below, however, the brain does show a certain modularity. Thus, if we want to stay with the analogy, it might be better described as an ensemble of musical instruments. Furthermore, there is a time delay between patterns which typically form in these modular nodes. The instruments are part of a concert. Finally, this time delay is functional. We are not hosting a symphony orchestra, but rather an improvising jazzband.

Circular causality can either be a property of a closed system, or a stable state in an open system. If we put the jazzband onto an open stage which is slowly driven through the country side, then whatever player first notices the change in scenery is likely to give an impulse for a new musical theme to her band members. As discussed above, two nodes can exchange information simultaneously in both directions if the messages have different content. Recall that our new model of the brain does not only consist of a recurrent network but that we can still identify a low-level representation and an action layer. Thus, given that the

embrained agent is in a changing environment, modularity and functional time delay of resonance might suffice to justify the talk of message passing in her brain.

Concerning the claim of modularity, while visual perception cannot be considered completely independent of the activity of larger networks (see above), it can also not be considered completely dependent on their activity. Typical interaction states in networks involving the occipital and ventral temporal cortex can develop independently of a significant amount of varying activity in other networks. The typical EEG signature during the perception of pictures of faces while sitting unmoving in front of a computer monitor, for example, is comparable to the signature recorded during face perception while naturally interacting with another human (once the recording has been cleaned of artifact-related variance) (Soto, 2019). Thus, there is local invariance of resonance which we can also describe as network modularity. From that perspective, these *somewhat* independent network modules can be thought of as single network nodes in an even larger network. In fact, the section on lesions above makes a strong point for modality of the brain. All and only the local networks that affect the function of a specific larger node, quantified by task performance, can be considered its subnodes.

One prominent example for the second claim, the functional time delay between nodes, is that action needs to lag behind perception. I define a time delay as functional if an experiment shows that performance is significantly worse without it. Note that “I need to open my eyes before I can grasp the glass in front of me” is not sufficent evidence for this claim. Remember that one reason we introduced linear decodability as condition for representation was that, otherwise, a future scientist might be able to decode the next upcoming bodily action from all over the brain. Resonance through recurrence does not make this distinction obsolete. On the contrary, besides verbal reporting, it is only our linear decoder allowing us to distinguish which stimulus we presented to the particpant before a linearly decodable population vector emerges in the motor cortex, which “proves” that perception comes before bodily action.

One experimental example which features time delay, functionality and linear decodability has been conducted by Wilming and

colleagues (2020). Using MEG, they decoded a participant's upcoming decision about the average contrast of a sequence of visually-presented gratings from frontal and parietal areas. Since participants had to wait for all gratings in the sequence to be presented before they could act, Wilming could map stable activity in the visual cortex to a particular grating sequence before decoding their final behavorial decision in the primary motor cortex. As a sidenote, to remind of the fact that network modularity is fuzzy, Wilming also found that primary motor cortex decisions changed visual cortex activity for subsequently presented gratings. Nonetheless, in tasks where participants have to wait for a sequence of events before they can make a decision, we can temporally dissociate the motor decision from some of the (necessary) perceptions. One node, which is always a network of subnodes, reaches a stable resonant state. The stable activity pattern influences the interaction among subnodes in another network. This network can, again, be considered a node which eventually reaches a steady resonant state and so until the apropriate motor action emerges. The perceptual states might, thus, be interpreted as informing the cognitive system about the state of the world. Our model again features a switching of its inner elements from computation to representation and back.

My argument for the sensibility of the notion of representation in the brain has relied heavily on perception, so far. Resonance and perception, however, need not be identical. Consider, for example, human spatial navigation. Ekstrom and colleagues (2003) recorded intracranially from the human medial temporal and frontal lobes while participants explored a virtual town. The experimenters were able to decode location in the virtual town, independently of viewpoint, from cells in hippocampus, amygdala, parahippocampal region, and frontal cortex. Ekstrom also recorded neural activity prior to exploration, when participants were presented buildings in the town without further context. Only two percent of variance in neural activity recorded during navigation could be explained by visual perception alone. Thus, the spatial navigation node can be dissociated from the visual perception node. Place cell activity is part of the resonance pattern of the spatial navigation node. In fact, all of the above regions are likely candidates for hosting subnodes of the navigation node. However, only lesion studies could show for sure. The greater navigation node plays an

important part in enabling the sequence of motor patterns which leads you on the quickest way to goal locations outside of your current perceptual space E. Hence, it can be considered as having the function to inform other nodes about a complementary state of the world E', which is not part of E. It fulfills this function in addition and in interaction with the perceptual node. Seperate nodes/networks can show different resonance patterns simultaneously. Cognition is not just one resonance pattern followed by the next. Therefore, the metaphor of the improvising jazzband still applies.

I.6. My Representation

Let me sum up my stance on representation in the brain, as developed so far. I model the brain as containing one low-level representational layer, one bodily-action layer, and, in between, a network of reciprocally connected nodes. These nodes can switch between representation R and computation or interpretation C.

In order for the activity of a node to count as a representation for another node, it needs to fulfill two properties. First, it needs to exhibit a temporally stable, spatial pattern that is linearly decodable. Second, the state of stable activity stretching over the node is necessary to perform a task involving the condition E which is supposed to be represented by that stable, linearly decodable pattern.

Due to recurrent processing, linear decodability does not imply linear processes in the decodable region. Linear decodability remains, however, a remarkable property without which talk of representation becomes inflationary. The stable pattern in its entirety and, hence, the possible representation is fuzzily stretching over a large network which can involve multiple cortical and subcortical regions. We call this network a node due to its invariant activity patterns given changes in other nodes. Fuzziness refers to overlap with other nodes which simultaneously represent or compute, and can be quantified using the noise-limit approach.

The above definition of representation seems compatible with most modern approaches on internal information transfer. Two prominent examples introduced already are Adaptive Resonance Theory and predictive coding, or processing, approaches (for

predictive processing see also Bastos, Usrey, Adams, Mangun, Fries, & Friston, 2012; Dora, Bohte, & Pennartz, 2020).

My account of representation still leaves some questions unanswered, however. One would be the notion of stability. What is the minimal amount of time the activity pattern has to keep its shape in order to be considered stable? This question also relates to the (as of yet) unanswered question concerning the scale of measurement precision discussed before. What spatial resolution suffices, or what spatial variance is acceptable, to define a shape of neuronal activity? Another question concerns the relation of subnodes to each other. Can we talk about one subnode informing another? Finally, I discussed the representational quality of activity patterns, but what about other decodable neuronal properties? Given time constraints, I will not further discuss the question of stability and subnode relation in this work. They remain potential issues of my definition of representation which need to be clarified elsewhere, eventually. Concerning potentially decodable properties, however, I will now specifically address the issue of decodable neuronal connectivity.

The synaptic configuration which influences the amount of ions flowing inside downstream neurons following an action potential is known as neuronal connectivity. Variations of Donald Hebb's postulate (1949), which is customary abbreviated as "Neurons that fire together wire together," are thought to be a cornerstone of both long-term (Dudai, 2004) and, according to a more recent proposal, short-term memory (D'Esposito & Postle, 2015; Jackman & Regehr, 2017; Mongillo, Barak, & Tsodyks, 2008). Consider the following example of short-term memory. Rose et al. (2016) have shown that an activity pattern which originally emerged as response to a certain stimulus, could be reactivated by a magnetic pulse. The method known as transcranial magnetic stimulation or TMS (see above) can lead to a general boost of neuronal activity in a localized area of the cortex. That an unspecific activity boost can lead to a specific activation pattern is largely due to the synaptic configuration of the involved network. Even after neural activity had returned to baseline, the pattern of connectivity and, thus, the configuration of the synapses remained stable long enough in order to shape the flux of ions prompted by the magnetic pulse, such that the neuronal activity pattern relating to the external stimulus could be "replayed."

Similar to neuronal activity, synaptic configurations can change in the order of ten milliseconds (Jackman & Regehr, 2017). Consequently, relating to our definition of representation for neuronal activity, i.e. firing rate, such a stable decodable connectivity pattern might be considered a representation. Also at least in principle, this stable pattern of connectivity can be considered decodable. Maybe not linearly, but given the right methods a researcher might find a mapping from the synapses and concentration of neurotransmitters to the previously presented stimulus. But is the representation for the experimenter or for the participant? The potential for an actionpotential is not an actionpotential. Consequently, in line with Klein's view on the overuse of memory in cognitive science (2015), I will not consider neuronal connectivity alone to represent for the participant.

The above analysis exemplifies again the role of functional action for representation. Only if neural activity plays a role in enabling action relating to what is supposed to be represented E, can it be considered a representation R for the participant. We have seen in the beginning of this chapter that an overt act of bodily behavior related to E can be measured through experiment. Given the above analysis of the difference between representation and computation in the brain, I now define a covert/mental act on E as any neural computation C influenced by a node R (which is representing E). Insofar as the computation takes part in a sequence of brain processes eventually leading to a functional change in motor activity A, it should be identifiable through experiment.

Only representation and mental action, however, would not lead to a very long life. For the embrained agent to keep its own form stable it needs to engage in bodily action involving E, at least sometimes, for example to eat. Action makes sense because of our body. But what is the relationship between brain and body beyond the circular causal bond of keeping each other stable?

Chapter II. Embodiment

In the second half of the 20th century a movement gained momentum initiated by specialists across disciplines from philosophy to robotics reevaluating the role of action for cognition (Bonner & Epstein, 2017; Brooks, 1991; Cisek & Kalaska, 2010; Engel, Maye, Kurthen, & König, 2013; Etzel, Gazzola, & Keysers, 2008; Glenberg, 1997; Glenberg & Kaschak, 2002; Gibson, 1950; Gibson, 1979; Hutto & Myin, 2013; Koffka, 1935; Lakoff & Johnson, 1999; Melnik, Hairston, Ferris, & König, 2017; Merleau-Ponty, 1945; O'Regan & Noë, 2001; Prinz, 1997; Wilson, 2002; Varela, Thompson, & Rosch, 1991). Motoric bodily action, so they argue, is *not* a somewhat uninteresting end product of cognitive processes. Rather, it needs to be considered a key player at the very core of cognition itself.

The central, nontrivial role of bodily action is one of the defining properties of what has become known as enacted embodiment. *Enactivism* emphasizes the creative aspect of interaction with the environment. Relating to the notion of circular causality introduced above, the environment shapes the action of the agent while the agent shapes its environment. Action makes sense. *Embodiment* emphasizes the structural aspect of that interaction. Action is performed by and affects specific bodies. Bodily structure is concrete. Consequently, enactivism and embodiment are somewhat complementary terms which only taken together can unfold their full potential for research in cognitive science. In the remainder of this work, whenever I mention embodied or enacted cognition, always read enacted *and* embodied cognition.

II.1. The End of a Hierarchy

A major pillar of the embodied framework is the belief that a hierarchical approach can only take us so far in our understanding of cognition (Varela et al., 1991). To see how cognition could work without higher processes ruling over lower processes, consider the robots developed by Rodney Brooks (1991). These robots had functional modules in the form of task specific layers. However, each layer had only to fulfill a simple task and was sparsely connected to the other layers. Brooks setup, thus, shows similarity to the idea of fuzzy modularity discussed above. Importantly, there

was no necessity for a highest layer taking in information from all other layers, computing the optimal action and relaying that to a hardware-(body-)specific action module. A parallel, but not independent, modular distribution of processing tasks is, indeed, one of the major goals of neuromorphic electronic circuits engineers on their way to building autonomous cognitive systems (Chicca, Stefanini, Bartolozzi, & Indiveri, 2014). In conclusion, modern robotics has already begun the uprising against the central processing unit.

Evidence for specialist modules which transform sensory influx into bodily action without the need for higher reasoning processes can also be found in animals. When the gannet, a large seabird, goes for a dive, the change in visual flow due to the approaching sea leads to a continuous folding of its wings which are closed exactly upon entering the water (Lee & Reddish, 1981). Human infant stepping is a more complex, multi-factor example. When infants who can still not walk on their own are prompted on a treadmill, typical synchronous dependencies between leg movements emerge, even before that movement becomes voluntary (Thelen, Ulrich, & Niles, 1987). Hence, here a specialist module emerges out of a sensorimotor feedback loop.

In conclusion, enactivists believe that complex cognition can emerge from the interaction of an ensemble of specialized modules which have learned simple cause-and-effect relations concerning their direct environment. Again, there is no need for a most-complex module which controls the lower ones.

II.2. Sensorimotor Relations

Similar reasoning leads enactivists to question the distinction between representation, computation and motor action. Their view is known as the sensorimotor, or ideomotor theory of cognition. Proponents of the theory emphasize that although representation depends on the changes in sensory activity, the structure of these changes depends not entirely on the agent's environment. The sensory changes are also fundamentally shaped by the agent's bodily shape and actions (Lenay, Canu, & Villon, 1997; Merleau-Ponty, 1945; Nagel, Christine, Kringe, Märtin & König, 2005; O'Regan & Noë, 2001; Prinz, 1997; Schumann &

O'Regan, 2017). Imagine, for example, someone standing on your left talking to you. When you turn to face the person, the neuronal firing patterns in your visual sensory areas will evolve quite differently from the firing patterns in your auditory cortex. We discussed before that the evolution of the neuronal activity depends on the environmental, but also on the local and the large-scale brain structure. The information flux affecting your central nervous system, however, is also shaped by the rotational movement and the geometry of the retina and the cochlear (as well as eyes and ears and in fact the general shape of your body). In other words, the relation of low-level sensory to motor patterns depends not only on neural network and environmental structure but also on bodily structure.

Lenay et al. (1997) used a very simple experimental setup to investigate the importance of bodily action for perception. They attached a photodiode to the index finger of the right hand of a blind (congenital or blindfolded) participant. Whenever the diode would register incoming light (above a certain threshold of intensity) a tactile stimulator held in the left hand vibrated. The experimenters further prepared light-emitting objects of different sizes, which they placed around the laboratory. After a habituation phase, participants would "feel" the position and volume of the objects in the room when freely moving their arm. However, when their movements were restricted to a rotational sweeping gesture with the hand without being able to change the length of their arm, participants did not perceive the depth but only the direction, that means, the angle of the object in space. Importantly, without right arm or hand movement, participants only felt a buzzing sensation in their left hand.

The experimental results by Lenay and colleagues corroborated findings obtained through a much more complex setup by White, Saunders, Scadden, Bach-y-Rita and Collins (1970). These researchers employed a 20 by 20 matrix of vibrators on the blind participant's back. The matrix transformed the visual input of a video camera into a tactile sensation. Participants who were allowed to move the camera perceived objects extended in the space beyond their body. Without camera movement, the participants would only feel the buzzing sensation on their skin and

had a much harder time identifying the object represented by the sensation on their back.

Note that in the above example where we turn towards a person while the person speaks to us, as well as in the two experiments, the motor activity is a voluntary, goal-directed action. You turned because you wanted to; thus, you have some expectations about the relation between changes in motor patterns and changes in sensory pattern. O'Regan and Noë (2001) conclude that perception is, indeed, enacted knowledge of sensorimotor relations. They further emphasize that, in their framework, it follows naturally that different senses are actually perceived as being different, a feat which otherwise is hard to explain: "From the point of view of the brain, there is nothing that in itself differentiates nervous influx coming from retinal, haptic, proprioceptive, olfactory, and other senses, and there is nothing to discriminate motor neurons that are connected to extraocular muscles, skeletal muscles, or any other structures. Even if the size, the shape, the firing patterns, or the places where the neurons are localized in the cortex differ, this does not in itself confer them with any particular visual, olfactory, motor or other perceptual quality" (ibid., p. 941). In other words, sensory input is "unlabeled" (Pennartz, 2015, p. 107) as illustrated in Figure 2.

A change in sensory activity can acquire a functional label, however, through its relation to motor activity and, hence, bodily action. An example illustrating how action can make sense (adapted from O'Regan and Noë) is waking up in a submarine in the deep sea. You are in a room filled with all kinds of strange dials and flickering buttons. You hear strange cracking and screeching noises of the steel around you, and through a small window you can make out absolute darkness on the outside of the room. The situation calls for learning by doing. Only through pushing buttons (and maybe turning some wheels and pulling some levers) and comparing the change in the dials, as well as observing the small window, will you eventually reach sea level again.

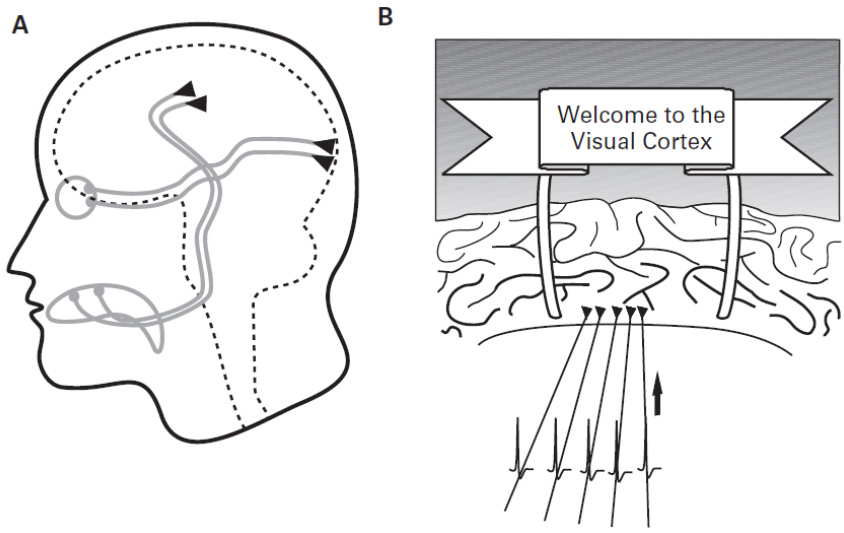


Figure 2: Unlabeled sensory input and unlabeled sensory cortices.
2.A Via several relay stations (not shown), taste receptors in the tongue project to the gustatory areas of the human neocortex, mainly located in the insula. Photoreceptors in the retina similarly project to the visual cortex. This external perspective on brain anatomy indicates how different brain areas are selectively provided with information from different peripheral sensors. At the level of cortical neurons, however, no information is available about the sensory origin of their inputs.
2.B When spike trains arrive along axons in the visual cortex, they come without "label" or "address of sender," and there is no "welcome" sign indicating that the spikes have entered the visual cortex, as opposed to cortical areas coding other modalities. This raises the question of how the brain itself identifies the modality to which incoming spike trains belong. Taken with permission from Pennartz (2015).

Indeed, if the meaning of changes in sensory cortex activity is not already hardwired from birth, identifying their content by their relation to motor activity seems ecological. It might even be considered necessary. Recall that a linear combination of the firing rate of single neurons in the primary motor cortex allows us to predict limb movement direction (Georgopoulos, et al. 1986). Consider that the experimenter as well as the participant can visually perceive this movement. Given that neuronal activity in V1 is not sufficient for this perception to arise, it seems rather curious that the movement trajectory in the perceptual space is a linear transformation of the motor neuron activity. The central nervous system is highly non-linear and there is no retinotopic "projector screen" after V1 anymore.

Admittedly however, O'Regan and Noë's claim about the structure of the nervous system not being sufficient for the different qualities of perception to arise, chases its own tail. "Knowledge" about sensory-motor relations can only be found in the structure of the brain. The structure is neuronal connectivity and activity. The interaction of connectivity and activity, given the right sensorimotor boundary conditions, can lead to the activity patterns necessary for a specific perception to arise. Therefore, while O'Regan and Noë make an important contribution in pointing to the importance of knowledge about sensorimotor relations for perception, it is not clear if that knowledge, i.e. the structure of the nervous system, really has to develop through learning by doing.

Note that our above examples were all of sensory substitution experiments trying to transform the skin into an eye. Natural selection, however, over millions of years, has led to the development of genetic code which carries specific (non-linear) information about how to build a human. Hence, intuitively at least, it makes sense that the structures important for survival are not only build through interaction with the environment after birth. On the other hand, its plasticity is one of the brain's most remarkable features (for a review see Fuchs & Flügge, 2014). Indeed, as we will see below, the middle way is the right way.

In an experiment by Held and Hein (1963), neonatal dark-reared cats were only exposed to visual structure in a vertically striped cylindrical arena. One cat was strapped in a gondola-like cart, which was pulled by another cat that moved otherwise freely through the arena. Although the cat in the gondola could move its head to look around, experimental results showed dramatic deficits in tasks involving visual perception. This well-known experiment is evidence that only the interplay of visual experience, motor experience and genetic structure is sufficient to develop visual perception.

The importance of motor experience becomes apparent also in the development of the human embryo. Already at nine weeks of prenatal development movement patterns involve arms and legs (Lüchinger, Hadders-Algra, van Kan, & de Vries, 2008). Mijna Hadders-Algra (2018) reviews evidence that prenatal early motor

activity significantly sculpts the developing nervous system. Thus, it comes as no surprise that early motor development correlates with general cognitive development (Bornstein, Hahn, Suwalsky, 2013; Ghassabian, Sundaram, Bell, Bello, Kus, & Yeung, 2016; Murray et al., 2006). In conclusion, volitional bodily action does seem to play an integral role concerning the ability of animals to attain knowledge of sensorimotor relations important for perception and cognition.

The importance of knowledge about sensorimotor relations for general cognition has also been argued influentially by James Gibson (1950). He thought these relations can help explain the perceived relative immobility of the space surrounding us. Gibson (1966) later extended these ideas and coined the term "affordances," by which he claims that our perception of an object in space is in fact the perception of what we can do with that object (cf. Koffka, 1935). In *The Ecological Approach to Visual Perception,* Gibson gives the following definition of the term: "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment" (Gibson 1979, p.127). The terminology coined by Gibson has since spread across disciplines, from social psychology to neuroscience to research on artificial intelligence and even product design (in the same order Costall, 1995; Jamone et al., 2016; Yamanobe et al., 2017; Norman, 1988).

An everyday example corroborating the claim that objects are perceived partly by identifying the actions they afford can be found when studying eye movements (Federico & Brandimonte, 2019). In a free-viewing task, when participants are presented with images depicting tools, they tend to look mostly at the grasp-able area of the tool. That is, if they are presented with a hammer, they rather look at the handle than the hammerhead. Furthermore, the amount of time participants spent looking at the grasp-able area depends on the context. If there is a nail next to the hammer, participants look longer at the handle than if the hammer is presented next to a piece of paper.

Evidence accumulates that perception and at least the planning of bodily action share a common neural code (Bonner & Epstein, 2017; Etzel et al., 2008; Melnik et al., 2017; Prinz, 1997). In a similar vein, Cisek and Kalaska (2010) describe cognition as a multitude of potential motor actions constantly interacting with each other and the current sensory input until a stable neural signature emerges which eventually leads to bodily action. Therefore, the concept of memory might best be understood as "the encoding of patterns of possible physical interaction with a three-dimensional world" (Glenberg, 1997, p. 1). Corroborating and further extending these ideas, Engel et al. (2013) present evidence from diverse areas of research culminating in their conjecture that even abstract reasoning is fundamentally grounded in bodily action (see also Lakoff & Johnson, 1999). In conclusion, the fundamental ideas behind enacted embodiment need to be taken seriously by modern cognitive science.

II.3. Acting Representations

Concerning the notion of information transfer in the brain, Hutto and Myin (2013) distinguish between conservative enacted cognition and radical enacted cognition. Conservative enactivists often believe that cognitive representations have both descriptive and imperative content simultaneously (Millikan, 1995; Wheeler, 2008). That is, representations both carry information about the structure of the world and initiate appropriate action at the same time. Radical enactivists do not believe that neurons exchange messages at all, at least not when considered as subparts of the complete embodied system (Chemero, 2009; Hutto & Myin, 2013). For radical enactivists, the structural differences between brain, body, and environment are not borders. The mind is not contained in the brain but rather literally extends into the world.

Considering the spectrum of representation in enacted theories of cognition, my account is very conservative. While I consider it a necessity for a representation to be involved in a process leading to an action on what is represented, I do not equate representation and action. Recall that I speak of action in two cases: either bodily action or mental action. My analysis led to the conclusion that only the computation, or interpretation, of the representation should be considered mental action. Bodily action follows after mental

action. As such, in my account, there is a strict border between representation and any kind of action.

Contrary to the conservative enacted approach, I do not believe it sensible to hold up a notion of representation if that which represents also acts simultaneously. As I argued before, a representation only exists for someone. Someone needs to interpret the representation, or there is no representation to begin with. Consequently, if the representation acts by itself, it cannot be considered a representation.

Interestingly, one can conceive of a situation in which the account I developed above makes me a radical enactivist. That is the special case when there is no sequence, but rather a necessary motor involvement in the resonant state, for example, of the perceptual node. Consider a certain perception which could only arise simultaneously with activity in the primary motor cortex. The perception in such a case cannot, without further argument, be considered a representation anymore. The reason is that the sensory areas would forward their latest changes to the motor areas, while the motor areas would feed their latest changes back to the sensory areas, and only that interaction would create the perception. Thus, the question arises where the interpreter of that perception is located.

Admittedly, since I allow for non-perceptual nodes, like a spatial navigation node, one could make the argument the perception informs that node. Furthermore, the navigation node will eventually inform the next goal-oriented action pattern. Therefore, only if all nodes were resonant simultaneously and that stable pattern would rely on motor activity would I become truly radical.

An interesting conclusion, it remains a hypothetical one. So far, I am not aware of evidence that the activity in the primary motor cortex is actually necessary for a specific perception to occur. Compare, for example, induced perception of self-movement (vection) in virtual or auditory environments (Andersen, Braunstein, 1985; Berthoz, Pavard, & Young, 1975; Riecke, Feuereissen, Rieser, & McNamara, 2015; Väljamäe, 2009).

In summary, embodied enactivists either believe that representation and action are identical or that there are no representations to be found in basic cognition. My account of

representation does not fit in either category because I split cognition into representation, computation and action. However, since representation can only be defined through action involving that which is represented, I am a very conservative enactivist.

II.4. Critique of Embodiment

There is general consensus that some basic cognitive processes are well described in the embodied framework. Nonetheless, critics as well as some proponents warn against an exaggerated embrace of that perspective (Adams & Aizawa, 2001; Andy Clark, 1999; Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016; Schlicht & Starzak, 2019). There are four major points of criticism.

First, the extra-cranial body is causally related to, but should not be considered part of, what constitutes cognition. Adams and Aizawa (2001) state that, while in principle cognition need not be brain bound, extra-cranial processes, given their structural differences, are best understood as tools of the cognitive processes within the cranium.

Secondly, some critics claim that the embodied framework does not help to illuminate the majority of findings cognitive science tries to explain. Goldinger et al. (2016), for example, discuss a plethora of classic findings in cognitive psychology for which they argue that a central role of bodily action does not do any explanatory work. One example is short-term memory scanning. Participants are asked to keep a number of items in memory and are subsequently prompted to recall if item X was in that list. Compare searching for your keys in your apartment. You naturally stop looking around when you found them. A proponent of embodiment might thus hypothesize that the search time decreases, on average, if the answer is yes. The search time, however, increases with the number of items independent of the answer being yes or no (Sternberg, 1966).

The third point of critique against claims of embodiment is that they either seem trivial or too vague. This point of critique is also picked up by Goldinger et al. who emphasize the importance of formal models and ask how to express embodiment in an equation.

The fourth and final point of critique states that more sophisticated cognitive machinery relies heavily on computations over stored representations. Andy Clark (1999) argues, for example, that the mark of the cognitive might be the ability to engage in something like off-line reasoning. That means a reasoning process in the absence of that which our thoughts concern. A hallmark of the enacted, embodied framework, however, is direct sensorimotor coupling as outlined above. Schlicht and Starzak (2019) point out that proponents of embodiment, so far, have fallen short of providing an adequate substitute for the notion of representation used in classical ("disembodied") cognitive science.

In the final chapter of this dissertation I will try to shed some light on the fourth point of critique. I will investigate how much embodiment influences representations which are used during off-line reasoning.

Chapter III. Embodiment Partly Structures Representations for Spatial Cognition

The interaction between representation and embodied action is still not well understood. One of the reasons is the complexity one adds to an experimental setup when allowing for real-world interactions. Consider, for example, spatial cognition. Finding one's way about in a large, cluttered environment seems to depend on knowledge of that large environment which one needs to build through exploration. Naturally, if one is not driven through that environment, bodily action is necessary to explore. But how can one dissociate mere necessity, that is the body as input and output channel, from bodily action playing an integral part in structuring the spatial representations which are part of that knowledge?

The following chapter will provide an answer to that question. However, since the two main studies I present rely on behavioral measures, the answer provided here can only be considered a rough sketch. That is because, as outlined in Chapter I and Chapter II above, talk of representation for the brain only makes sense if one can differentiate between representation, computation and action, in the brain. Naturally, bodily human behavior depends on all three aspects of neuronal activity. Without parallel measurements of the activity, I cannot be sure that the results I present affect representational structure at all. On the other hand, I also argued before that representation stretches across large-scale networks. Consequently, I assume that the main studies' results are at least an approximation of the effect of bodily behavior on spatial representations.

Spatial navigation in a large-scale environment can be considered a paradigmatic example of a "representation hungry" problem (Clark, 1999). In an unfamiliar (part of) town, we seldom start walking without first making a (mental) plan for how to quickly reach our goal's location. Recent developments in cognitive- and neuroscience further bolster the importance of this example. Numerous scientists have argued that the processes involving successful spatial navigation lie at the very heart of cognitive

reasoning itself (Bellmund, Gärdenfors, Moser & Doeller, 2018; Buzsáki & Moser, 2013; Constantinescu, O'Reilly & Behrens, 2016; Epstein, Patai, Julian & Spiers, 2017; Schiller et al., 2015). Indeed, the idea of a strong link between navigation and general problem solving in cognition is not new (Siegel & White, 1975; Tolman, 1948). In conclusion, spatial cognition seems an important field of research for understanding the role of representation for higher cognitive processes, like reasoning, in the absence of that which our thoughts concern.

III.1. Spatial Cognition

III.1.1. Egocentric versus Allocentric

Contemporary research on spatial cognition commonly distinguishes between two categories of spatial relations, namely egocentric and allocentric (Burgess, 2007; Feigenbaum & Rolls, 1991; Klatzky, 1998; Wolbers & Wiener, 2014). Spatial relations among parts of our extracranial body or among objects in the environment in relation to parts of our body are called egocentric (deictic, ideothetic). For egocentric representations our body, literally, is the frame of reference. Spatial relations exclusively among objects which are not part of our body are called allocentric (extrinsic, exocentric, geocentric). Consequently, only objects, or directions defined through objects in the environment can be reference frames for allocentric code.

One example for the distinctive coding of both sorts of spatial relations is provided by Feigenbaum and Rolls (1991). They recorded from the enthorinal cortex of the macaque monkey. The fixated monkey had to remember the position of objects on a screen in front of her. The researchers differentiated between three conditions, changing position of objects on the screen, changing position of the monkey relative to the screen and changing position of both monkey and screen together, relative to the room. For each of the three conditions, using a simple linear model Feigenbaum could identify specific neurons through change in firing rates independent of the respective remaining two conditions.

The frequency change of neurons specific to the position of the monkey relative to the screen can be considered egocentric code. This code was also shown to depend on head direction but not on the looking direction, i.e. retina. That means the head direction became the egocentric frame of reference in relation to which the neurons firing rate encoded the objects position. The change of firing rate relative to the position of the object on the screen, when balanced for head position, can be considered allocentric coding with the screen as local reference frame. Finally, some neurons anchored their allocentric code to the geometry of the room in which the experiment was set up.

Or at least so it seems. While the difference in the spatial relations can be easily defined, how it maps onto neuronal circuits is debated. There is overwhelming evidence that spatial relations among objects in a vista space, that is an area which can be overseen from one location, can be represented in both the egocentric and the allocentric frame of reference (Burgess, 2007). Klaus Gramann (2013) emphasizes differences in individual proclivities toward preferred reference frame during navigation. Furthermore, the neural bases of both representations strongly overlap as already became apparent in the above example and has been shown for multiple brain regions (Ekstrom, Arnold & Laria, 2014). These difficulties are further illustrated in the ongoing debate over how to interpret findings in numerous studies investigating egocentric or allocentric processing (Burgess, 2007; Banta Lavenex & Lavenex, 2009; Wolbers & Wiener, 2014). In fact, even the representational distinction itself, is disputed (Bennett, 1996; Filimon, 2015).

Even if the allocentric frame of reference only results from computations involving multiple egocentric frames of reference I still believe the distinction to be of value, however. If only to distinguish between the result of the computation, i.e. neurons which encode the position of an animal in a room marginalized over all possible paths it could have taken to get there, and the representations necessary for that result. That is, neurons which represent the position of an object in the animal's visual field, marginalized over all relative positions of animal and room. Naturally, both types of neurons only represent as part of larger networks.

Consider the task of finding a small object, like an apple, in a small cluttered environment, like an office, and picking it up. The fluidity and speed of humans in such simple picking challenges is still unmatched by modern-day robots (Lawrence, 2017; Leitner, 2018). Such tasks involve mostly egocentric spatial relations. Consequently, we can conclude that the involved representations have the property of allowing quick, goal-directed action. In contrast, usage of allocentric representations is often time consuming. Spatial decision latency, for example, has been reported to correlate with angular difference when participants had to point to objects located within a large environment which cannot be overseen from one point of view (Evans & Pezdek, 1980; Sholl, Kenny & DellaPorta, 2006).

The most remarkable property of allocentric representations is revealed in large environments which consist of multiple separated vista spaces, like a town. We are able to point in the direction of our goal location even if we neither saw nor walked this path before. Admittedly, the result is often only a very rough approximation of how the crow would fly (Warren, 2019). Nevertheless, this cognitive ability might still imply that representations of allocentric spatial relations are dissociable from embodied action. Consequently, spatial cognition might not (wholly) fit in an enacted framework.

III.1.2. Allocentric Spatial Knowledge

It is generally agreed upon that knowledge about spatial relations in a large environment is built up through, at least, three processes. Namely, memorizing landmarks, acquiring route knowledge, and connecting those to attain allocentric survey knowledge (McNaughton, Battaglia, Jensen, Moser & Moser, 2006; Siegel & White, 1975).

A landmark, here, does not need to be exceptionally large. Any immobile object or place which left an impression on the explorer such that she will remember it on subsequent appearance, or can represent it mentally even in its absence, can be considered a landmark. Representations underlying route knowledge combine landmarks and directional information. It is assumed that this

directional information comes from bodily action and/or visual flow experienced at the landmarks (Chrastil, 2013). Route knowledge is the ability to sequentially represent (already experienced) landmark-action pairs in the correct order to reach the goal landmark (Siegel & White, 1975). When the animal has acquired enough experience in the new environment to mentally deduce paths to the goal location which it never took before, we speak of a survey-like representation, also known as cognitive map (Tolman, 1948). It remains unclear, however, to what extent the processes on the way to survey knowledge are actually separate, and if these are the only processes involved (Chrastil, 2013; Ishikawa & Montello, 2006; Montello, 1998).

An additional factor that influences the allocentric spatial knowledge we acquire in large environments is the purpose of our experience. If we intent to learn spatial relations, for example to navigate to a goal more efficiently, we gather more comprehensive knowledge about our surroundings than when we do not care where we are. While this might seem trivial, the intention related difference in spatial knowledge assessed with behavioral measures like estimation of goal direction is remarkably low (for a review see Chrastil & Warren, 2011). Thus, contrary to intuition, allocentric spatial representations seem to form largely automatic, that is without the need for intention (Burte & Montello, 2017).

Research on large-scale spatial cognition has repeatedly found evidence for subgroups of spatial learners. Some people are very adept at acquiring survey knowledge of a new environment. Others are remarkably poor. These individual differences are not statistical noise but remain consistent over longitudinal studies. In one well-known study by Ishikawa and Montello, participants were driven along separate roads through the outskirts of a town (2006). Even after ten such drives a considerable number of participants still did not acquire comprehensive survey knowledge. Using large virtual environments Weisberg and colleagues (2014, 2016) could repeatedly identify three distinct groups of spatial aptitude. They further found that questionnaire data on self-assessed spatial ability, or “sense of direction,” correlates well with these groups. Burte and Montello (2017) as well as He also report self-assessed spatial aptitude to inform different groups of spatial learners (He, McNamara, & Brown, 2019). It is not clear how these large

individual differences fit into an embodied framework, given that motor ability in general would be comparable among participants.

III.1.3. Allocentric Spatial Knowledge and the Body

Proprioception in connection with motor activity can help to increase the quality of spatial knowledge acquired while exploring an environment (Chrastil & Warren, 2013; Ruddle, Volkova, Mohler & Buelthoff, 2011; Waller, Loomis & Haun 2004). The degree to which this is the case however is still debated. For instance, a recent study shows participants who change movement direction through whole body rotation, but only need to lean forward to increase speed, acquire spatial knowledge of a similar quality as free walking participants. Only bodily rotation as motor action, on the other hand, left participants performing worse (Nguyen-Vo, Riecke, Stuerzlinger, Pham & Kruijff, 2019). Evidently, even small proprioceptive and vestibular activity changes in combination with motor activity are sufficient to significantly increase the quality of allocentric spatial knowledge. An unexpected finding, it still corroborates the role of embodied processing. Generally, integrating information from different modalities into one coherent structure is a highly complex task (Cao, Summerfield, Park, Giordano, & Kayser, 2019; Chen, McNamara, Kelly, & Wolbers, 2017; Ernst & Bülthoff, 2004). If the brain is able to integrate many bodily cues simultaneously to create a more accurate and quicker accessible representation, it hints at neural dynamics shaped by the body in a nontrivial manner.

Ecology of representation also corroborates the claim of spatial processing being grounded in bodily action. Sensory substitution or augmentation devices, the usability of which relies on lawful sensorimotor relations, have been shown to have significant effects on both spatial perception and cognition (Bach-y-Rita, 1972; Lenay et al., 1997; Nagel et al., 2005; Schumann & O'Regan, 2017). Spatial judgements improve in accuracy for participants wearing a belt that provides constant sensory information about north. However, the participants' spatial reasoning processes seem to rely less on allocentric representational structures (Kaspar, König, Schwandt & König, 2014; König et al., 2016). Thus, representations might only be built if necessary to guide efficient action.

Recent results, both theoretical and experimental provide further evidence for a representational structure which is intimately linked to bodily action. Bicanski and Burgess, for example, developed a model for spatial memory and imagery which features pseudo motor signals as an important factor (2018). Bonner and Epstein (2017) reveal a bottom-up mechanism involving the visual cortex for perceiving potential paths for movement in one's immediate surroundings.

We want to experimentally investigate the effect of bodily action on allocentric representational structure using behavioral measures. Large-scale studies of spatial cognition have mostly been asking participants to estimate distances or judge directions. Participants that point from landmark A to landmark B, however, already believe to have found the shortest distance between both landmarks (at least implicitly). Therefore, these measures cannot be considered independent. We need to consider allocentric representations that clearly differ with respect to action relation. A second hurdle to overcome is that, in general, the distinction between egocentric and allocentric is not clear-cut, as described above.

III.1.4. The Pilot Study

Both hurdles have been largely overcome in a recent pilot study, the precursor to our current project. The "Osnabrück study" inquired into the character of allocentric spatial knowledge which participants had acquired while living in a real town (König, Goeke, Meilinger, & König, 2019). As outlined above a town is especially well suited to investigate the relationship of embodiment and allocentric representations because it is a large-scale environment made up of visually separated vista spaces.

In the Osnabrück study the researchers exclusively investigated judgement of direction. For this, three tasks were set up. Two tasks were of foremost importance because they involve spatial concepts related to largely independent forms of bodily movement. In the Absolute task participants judged the orientation of a housefront relative to cardinal north. In the Relative task participants judged the orientation of two houses relative to each other. The researchers argued that the spatial concept of cardinal

direction is categorically different from the spatial concept of house orientation concerning their relation to embodied action. While house orientation affords local action, like turning, north informs long-distance travel.

The researchers additionally set up the Pointing task in which participants had to judge the shortest direction from one house to another. Pointing from landmark to landmark is most familiar to the participant and a standard measure throughout literature on spatial cognition. Similar to knowledge about cardinal direction, it also enables the participant to remain on a straight line during travel. It is, however, the only one of the three tasks that requires knowledge of relative location to be solved. Since its familiarity might additionally confound embodiment effects, the focus of analysis in the Osnabrück study was the comparison of Absolute and Relative tasks.

As discussed earlier, we generally need more time to plan the shortest route in a large environment, than to identify and pick-up an object in front of us. The difference in reasoning time makes it interesting to compare task performance under time pressure to performance when the participant is free to reason. Consequently, in the Osnabrück study the spatial tasks were implemented for both time conditions.

The experimenters found that, under time pressure, participants were better at judging the angle between two housefronts, than judging the angle between house and cardinal direction of north. The analysis also revealed that without time pressure the relation reversed. When participants were allowed to reason indefinitely they excelled at the Absolute task. The interaction indicates two different forms of allocentric spatial representations. A possible explanation for the two forms could be the distinct actions afforded by the spatial concepts of house orientation and north, respectively.

III.2. The Present Study

The study I will describe in the following was a group project.
Therefore, from now on I will address the reader from the point of view of our research group and thus switch to “we”.

III.2.1. Introduction

One of the current study's goals was to reproduce the results of the Osnabrück study in a more controlled environment. Consequently, again, we needed to find an experimental environment of a size and structure sufficient to ensure minimal egocentric influence on task performance.

Older studies on spatial cognition have often relied on large storage houses or similar environments (Moser, 1988; Thorndyke & Hayes-Roth, 1980). However, such a setup severely limits reproducibility for other research groups. It gets even more problematic for large-scale navigation of real-city size. To control for prior knowledge, e.g. map knowledge, in such studies is complicated and costly. For this reason, modern studies mostly employ virtual environments. These range from simple PC gaming setups to large 180° encompassing screens (Richardson, Montello & Hegarty, 1999; Steck & Mallot, 1997). Generally, methods vary widely and, as of now, there is no established norm for a large-scale navigation setup.

However, in the last two decades virtual reality (VR) has become widely used among researchers across disciplines (Bohil, Alicea, & Biocca, 2011, Parsons, Gaggioli & Riva, 2017), also in spatial cognition (Hardless, Meilinger, & Mallot, 2015). Modern equipment combines stereoscopic 3D vision and largely unrestrained, full-body action possibilities with considerably low setup costs. Another advantage of a VR setup is that it allows access to coordinates of bodily movement in real time. Furthermore, VR allows for a feeling of "presence" of participants in the virtual environment, a reason why it is also used in research on emotion and therapy (Diemer, Alpers, Peperkorn, Shiban, & Mühlberger, 2015; Price & Anderson, 2007). Taking all of the above into consideration, we decided to repeat the pilot study by letting participants explore a complete city in VR. Our custom-made VR city contains roughly 200 houses spanning about 450 by 500 meters, which allows for extensive embodied exploration.

The main goal of the present study was to reproduce evidence for the two distinct forms of allocentric encodings found in the pilot study. As mentioned above, these distinct representations seem to

result from the respective relation of the represented spatial concept to embodied action. We, thus, sought to not only reproduce but also expand the findings by investigating spatial representations acquired through different means of embodied exploration. Consequently, we set up three experimental embodiment conditions. Participants in each condition navigate through the same town, only the bodily action possibilities differ. After city exploration participants complete the same three spatial tasks with and without time pressure as in the pilot study. We, additionally, employ a longitudinal in-between design in order to observe the development of the spatial encodings. Exploration in each embodiment condition is repeated three times. In summary, this allows us to compare the performances in three spatial tasks, times two decision-time conditions, times three exploration sessions, for each of the three modes of embodiment, respectively.

III.2.1.1. The VR Exploration Hypothesis

In the VR condition participants explore the city sitting on a swivel chair, controlling movement speed with a handheld controller. Participants change direction through head movements, which are unrestrained, as are torso rotations around their longitudinal axis.

The VR embodiment condition was mainly set up to reproduce, or call into question, the findings of the Osnabrück study. Our hypothesis (H1) for the results after VR exploration reads as follows: Performance of Absolute and Relative tasks interact with time pressure. With unlimited decision time judging house alignment to north is more accurate than judging relative orientation of two houses and vice versa under time pressure. Houses can become decision points that afford local action involving rotation. Access of cardinal direction allows the navigator to move long distances along one direction. Thus, finding the purported interaction after VR exploration suggests two forms of allocentric representations, which develop because they involve spatial concepts that differ in their associated actions.

Compared to the other two exploration conditions we will introduce below, VR most closely resembles embodied navigation without augmenting devices (e.g., a map or smartphone). Thus, the VR results might also shed light on the general development of allocentric knowledge in large environments. While we have no

concrete hypothesis, we are interested to see whether performance in the separate spatial tasks develops in parallel or shows interactions with increasing number of sessions.

III.2.1.2. The Map Exploration Hypothesis

For the map condition we created an interactive city map. Participants sit on an immobile chair in front of a setup of computer monitors. The monitor screens together display one large birds-eye view map of the VR city, with north being on top. Using a mouse cursor, participants can click on an individual house on the map, then a picture appears of the housefront view. In this setup, exploration of the VR city involves only a minimum of bodily movements.

Our hypothesis (H2) for the results after map exploration reads as follows: Performance in the Absolute task is better than in the Relative task. Since the map follows a north-up layout and the participants upright body is aligned with that axis for the entire exploration period, we expect the structure of the acquired representation to favor performance in the Absolute task in both time conditions.

III.2.1.3. The VR with Belt Exploration Hypothesis

The VR condition with belt resembles the VR condition except that the participant is additionally equipped with a sensory augmentation device. The device is a belt worn around the waist with equally spaced vibrators of which always only the one closest to north is active. That means, if participants rotate clockwise on the swivel chair, the vibration switches to an element in the counter-clockwise rotation direction such that the location of vibration on the body remains closest to north.

Participants in the present study were not trained with the belt prior to VR city exploration and, thus, accustoming effects might confound results. Nevertheless, our hypothesis (H3) for the results after VR with belt exploration reads as follows: Performance in the Relative task is better than in the Absolute task. We assume that the belt reduces the ecological value of building a representation of north which can be used in the absence of the belt's signal. Reasoning processes involving representations of the angle

between house fronts on the other hand, will not be negatively affected by the sensory augmentation device.

III.2.1.4. General expectations

We set out to design a study which incorporates three different embodiment conditions each allowing us to answer a different question, i.e. hypotheses one (H1) to three (H3). Consequently, we have no explicit hypotheses purporting effects in between those three conditions. Indeed, comparing hypotheses in-between different exploration conditions can come with a reduction in power due to possible in-between factors (e.g. getting accustomed to the belt might lower performance in all tasks) which are not the primary concern of this study. We do, however, have three expectations all of which should hold independent of the means of embodiment during exploration.

First (H4), we do expect to find better performance in Pointing than in the Relative and Absolute tasks. Pointing from one location to another represents direct knowledge about the (theoretically) shortest path to a goal. This kind of knowledge seems to be most important for navigation. In addition, of the three spatial tasks the Pointing task is probably by far the most familiar to our participants.

We already mentioned that a considerable part of people are poor spatial learners. Despite the apparent difficulty in generating large-scale survey knowledge however, we nevertheless expect the following hypothesis (H5) to hold for the results after exploration in any of the three embodiment conditions: Performance increases with session. The intervals between sessions range from one to seven days. Consequently, even for participants which are not adept at spatial navigation and acquired little knowledge during the first exploration, we believe long-term changes in neuronal structure and activity will facilitate learning during subsequent sessions.

The final hypothesis (H6) we believe to hold for the task performance results after exploration in all of the three embodiment conditions reads as follows: Performance under time pressure is worse than without. Egocentric representations have a structure that allows quick action, both mental and bodily.

Allocentric representations require more time to maximize accuracy of spatial judgements.

III.2.2. Methods

III.2.2.1. Participants

Overall, 259 young and healthy adults took part in our study. They were randomly assigned to three different exploration groups: virtual reality (VR), map and VR with sensory augmentation belt. We had to discard a total of 33 participants (8 VR, 7 map, 18 VR with belt) due to physiological issues like motion sickness in VR, as well as technical problems with data collection. The remaining 226 participants were distributed as follows: In the VR condition without belt we counted 82 participants (51 female; mean age 23 ± 3 years standard deviation). A total of 74 participants (50 female; age 24 ± 4) only explored a map of the city. Exploring the VR city while wearing a sensory augmentation belt were 70 participants (40 female; age 23 ± 4).

These numbers were informed by a power analysis based on the results of the pilot study mentioned in the introduction (König, et al., 2019). We calculated a sample size of about 60 participants to find the smallest effect in between two embodiment groups with 80% probability (G*Power (Faul, Erdfelder, Buchner & Lang, 2009): Two groups, one tailed, planned-power 80%, Cohen's d=0.46 ($\mu1=19.4$, $\mu2=18$, $\sigma=3$)). As a sidenote, while we did not have explicit hypotheses comparing embodiment conditions, we still believed the power reasonable in order not to miss effects during exploratory data analysis.

The number of participants necessary to only miss an effect in 20% of experiments was performed for only one session. Initially, we had planned one exploratory session per participant per embodiment condition. Unexpectedly, performance remained close to chance in most tasks across conditions. Considering constraints on time and resources we settled for a design with three sessions for each experimental group.

Of the VR participants 28 (16 female; age 23 ± 3) took part in the longitudinal study. Of these 22 showed complete data sets in each of the three sessions. The remaining six had to discard one or two sessions each due to technical error. For example, a participant's

complete data set of the first session in the VR embodiment condition had to be discarded but her data was recorded correctly in the second and third session. We included all 28 longitudinal VR participants as well as the 54 VR participants who only explored the city once, for statistical analysis.

For the map condition again 28 participants (15 female; age 24 ± 3) took part in the longitudinal study. Here 26 had complete data sets after three sessions. We included all 28 longitudinal as well as 46 single-session participants for analysis. For VR with belt 27 participants (13 female; age 24 ± 4) returned for multiple experimental sessions. We found 23 of these to have complete data sets. All 27 longitudinal as well as 43 single-session participants were included in the analysis. Table 1 gives an overview of the total number of participants in each experimental session. The detailed distribution of participants and data is shown in the Appendix section 1.

The participants could choose between either monetary compensation (9 Euros per hour) or ‘‘participant hours,’’ which are a requirement in most students’ study programs. All participants gave written informed consent. Also, only healthy participants (questionnaire data) were admitted for the experiment. The study was approved by the ethics committee of the Osnabrück University.

# Participants	First Session	Second	Third
VR	82	28	28
Map	74	28	28
VR with Belt	70	27	27

Table 1. Number of participants per experiment per session.

III.2.2.2. Virtual Reality City

We built a VR city using Unity (version 5.6.3 and 2017.2.0), a software development environment to build video games. We named the city Seahaven. It makes up an island spanning 450 times 500 m². Seahaven contains 213 houses ranging from wooden huts, garages and supermarkets, to villas, churches and office buildings.

To prevent participants from relying on simple navigation strategies Seahaven does not contain a building which is visible

from all parts of town. In addition, there are no specific city districts, and the street system does not follow an ordered grid (i.e., “Manhattan style”). Finally, the angle between housefronts and north is roughly equally distributed in steps of 30°.

The VR environment includes a moving sun to provide natural lighting conditions and an indication of the direction of north. The sky over Seahaven is always cloudless, so the sun is clearly visible as well as its movement from sunrise to sunset (East to West) during the course of the day. Each participant experienced one such VR day in her 30-minute exploration period in the city. The only available cues to infer cardinal direction in Seahaven, that is most importantly the cardinal direction of north, are the sun’s position as well as the shadows cast by objects in the city. Figure 3 depicts the city from different perspectives.

Figure 3. Virtual reality city Seahaven.
The figure depicts the VR city from different points of view. 1.1 Seahaven from a bird’s-eye perspective. 1.2 Panoramic view of the eastern part of the city. 2.3 2D example of how a VR participant experiences Seahaven.

III.2.2.3. Tasks

The task design is very similar to the setup used in the Osnabrück pilot study (König et al., 2019). Each participant is tested on three judgement-of-relative-direction tasks, once with and once without time pressure. The tasks follow a two-alternative-forced-choice (2AFC) design with 36 trials for each of the three kinds of spatial relations with each time condition. That makes for six task blocks each holding 36 trials (altogether 216 trials). The six task blocks were presented in random order. Trials per block were also randomized anew for each experimental session.

III.2.2.3.1. Absolute Task

In the Absolute task we assessed the participants' ability to estimate the orientation of single houses in relation to north. In a 2AFC setup the participant was presented two pictures of the same house. Both pictures were overlaid with an arrow inside an ellipse. Only one of these arrows pointed to the cardinal north (as defined inside the VR environment) direction. The other arrow pointed in a direction that diverged from north by some amount between 30° and 330° (in steps of 30°). The pictures were presented at the same time and one above the other on separate screens (details on task setup below). The participant had to choose the picture with the

Figure 4. Absolute task.
In an Absolute task trial the participant had to choose the picture on which the arrow correctly pointed in the cardinal direction of north. Not depicted are the grey screens shown in between Absolute task trials. We presented 36 trials which had to be answered in three seconds and 36 trials without decision time limit.

arrow pointing towards north. She could choose the upper image or the lower image respectively by pressing the upper or lower button on a response box. Figure 4 depicts the task schematically.

Altogether, 36 2AFC trials with five second breaks (gray screens) in between each trial were presented sequentially in a block. There were two Absolute task blocks. In one block the participant had unlimited decision time. In another block the decision had to be made within three seconds, amounting to a total of 72 Absolute task trials

III.2.2.3.2. Relative Task

In the Relative task, we assessed the participant's ability to estimate the orientation of houses in relation to each other. We employed a 2AFC setup again, similar to the setup for the Absolute task, only for the Relative task the participant saw three houses per trial. First the lower and upper screen both showed the same house (without an arrow). Then the upper and lower screen both showed two different houses. These are called target houses. One of these target houses faced in the same direction as the prime house. The other house of the target-house-pair faced in a direction that diverged from the prime house's orientation by some amount

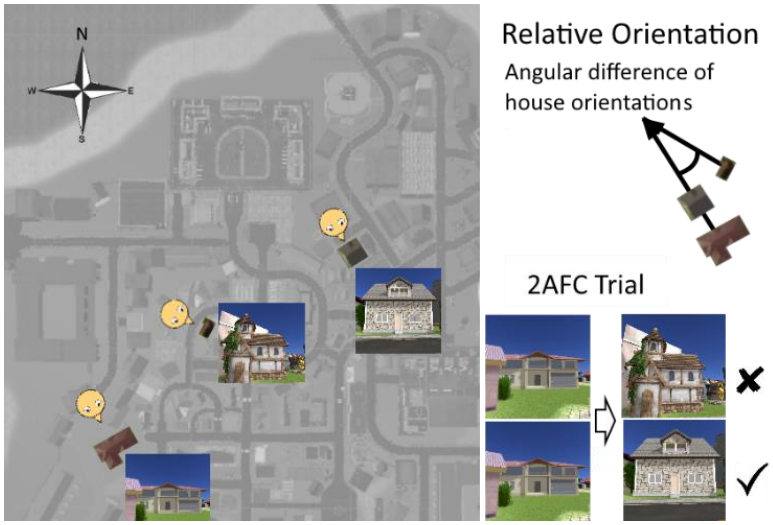


Figure 5. Relative Task.
In a Relative task trial the participant first saw one house. She then saw two different houses and had to decide which of these was oriented in the same direction as the first house. We presented 36 trials which had to be answered in three seconds and 36 trials without decision time limit.

between 30° and 330° (in steps of 30°). The participant's task was to select the target house with the same orientation as the prime house. As in the Absolute task the participant chose for the upper image by pressing the upper button and for the lower image by pressing the lower button on the response box. Figure 5 depicts the task schematically.

We presented 36 2AFC trials, each trial consisting of the prime house being shown for five seconds and then followed by the target houses, sequentially in a block. There were two Relative task blocks. In one block the participant had unlimited time to decide for the correct target. In another block the decision had to be made within three seconds, amounting to a total of 72 Relative task trials.

III.2.2.3.3. Pointing Task

In the Pointing task, we assessed the participants' ability to estimate the orientation of a house relative to the direction towards another house. We again employed a 2AFC setup, similar to the setup for the Absolute and the Relative task. Also similar to the Relative task, each trial started with the upper and lower screen both showing the same house (no arrow). This prime house which was shown for five seconds. Then, as in the absolute task, the participant was presented two pictures of the same house. These

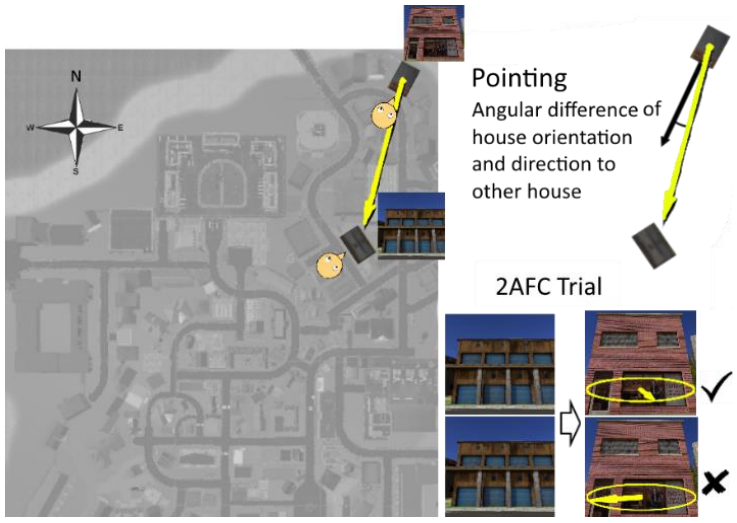


Figure 6. Pointing Task.
In a Pointing task trial the participant first saw one house. She then saw another house with two different arrows. The participant had to choose the arrow which pointed to the first house. We presented 36 trials which had to be answered in three seconds and 36 trials without decision time limit.

pictures of the one target house were both overlaid with an arrow inside an ellipsoid. One of these arrows pointed from the target towards the prime house. The other arrow pointed in a direction that diverged by some amount between 30° and 330° (in steps of 30°). The participants' task was to select the target house picture with the arrow that pointed towards the prime house. As in the Absolute and the Relative task the participant chose for the upper image by pressing the upper button and for the lower image by pressing the lower button on the response box. Figure 6 depicts the task schematically.

We presented each participant with 36 2AFC trials. Each trial consisted of the prime house being shown for five seconds, followed by the target house with a different arrow on the upper and the lower screen. There were two Pointing task blocks. In one 36 trial block the participant had unlimited time to decide for the correctly pointing arrow. In another block the decision had to be made within three seconds, amounting to a total of 72 Pointing task trials.

III.2.2.4. Stimuli preparation

The set of stimuli was entirely made up of screenshots of houses in the VR city. The screenshots were taken from the minimal distance each house could be wholly captured, with little else in the image that could serve as distractions. The pictures were taken during the course of a VR day. Since the sky in Seahaven is always cloudless each picture had good lighting conditions. Most importantly each screenshot was taken facing the building's front. Therefore, the orientation of each house in the picture was towards the participant sitting in front of the monitors.

Most of the screenshots were overlaid with 3D stylized arrows lying inside an ellipse, similar to a compass. The ellipse and arrows were chosen bright green to increase salience and together took up the lower third of the picture (Ellipse and arrow in Figures 4 to 6 were slightly modified for better readability). Each picture had a resolution of 2160x1920 pixels.

So as not to create any bias for a certain direction, we wanted the orientation of stimuli used in the tasks to be equally distributed. Additionally, if the correct and wrong options in a 2AFC trial differ

in their orientation by only 30°, it might be harder to judge correctly than if they are 60° apart. Consequently, we wanted the distribution of differences between choice options to be identical across all tasks. To not exhaust the participants too much after the already demanding exploration of the city we decided for only 36 trials per task and time condition. Using such a relatively small number of stimuli, we did not want to rely on simple randomization. We instead developed an algorithm which fulfils the above constraints, thereby creating a set of stimuli with homogeneous properties for each task. The algorithm is explained in detail in the Appendix section 2.

We used the algorithm to create one set of stimuli which was employed after each exploration independent of embodiment condition. Consequently, within each block all stimuli are identical across participants. The order of trials as well as the screen on which each stimulus is presented (upper or lower screen) during each task block, as well as the order of the six blocks themselves, are newly randomized for each participant.

III.2.2.5. Procedure

We partitioned each experimental session into three main parts. First the participant was introduced to the tasks. Then she explored the city. Finally, the participant completed the tasks. All three sessions over all three exploration conditions shared the same introduction and the same task blocks (task order was randomized). The conditions only differed in the embodied city exploration possibilities. A graphical overview is given in Figure 7.

III.2.2.5.1. Task introduction

The introduction consisted of three parts. It started with the experimenter explaining the spatial tasks, in which one example picture for each of the spatial tasks was shown to the participant. These pictures were taken from houses of the city of Osnabrück. After this explanation the participant sat at about 80 cm distance in front of a six-screen monitor setup in a 2x3 arrangement. Only the middle column of monitors was turned on. Each of the two screens showed an arrow inside an ellipse. The participant had to choose the arrow which pointed more upwards. She chose for the arrow on the upper screen by pressing the upper button labelled “U” and for the arrow on the lower screen by pressing the lower

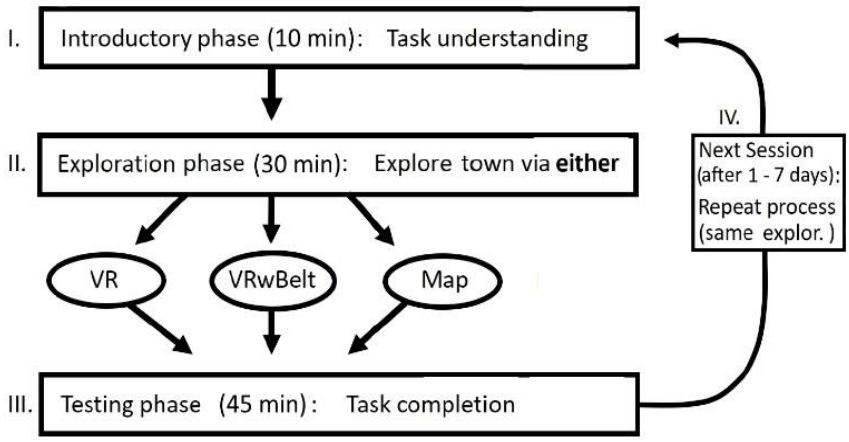


Figure 7. Overview of the complete study design.
Each experimental session started with an introductory phase. Here the participant was (re-)introduced to the tasks and completed a time pressure response training. The introduction was followed by the exploration phase. Here the participant explored the city for 30 minutes using one of three different bodily means for exploration. After the exploration phase the participant had to complete the six task blocks. Task completion was followed by questionnaires on navigational behavior and aptitude. Some participants returned for multiple sessions, always in one and the same bodily exploration condition.

button labelled “D” on a four-button response box (Black Box ToolKit USB response pad). The box was placed approximately 30 cm in front of her and left and right button were unmarked and had no function. If the participant did not decide after three seconds, the screens turned red. If she chose the wrong arrow the screens turned blue. Upon providing the correct answer within three seconds the screens turned green. Arrow pairs appeared sequentially until the participant correctly decided in 48 out of 50 trials. Following this response training the participant did one trial for each of the three spatial tasks, once with time limit and once without. For these six test trials, we again used houses from the city of Osnabrück. In these trials the participant, only received feedback if she did not answer in the time limit condition. Then the screens turned red. The complete introduction phase lasted approximately 10 minutes.

III.2.2.5.2. Exploration

Participants were randomly assigned into one of three groups, each defined through its specific means of embodiment of exploration.

If the participant returned for repeated measurements, she stayed in her initial embodiment group.

III.2.2.5.2.a VR Embodiment

The participant sat upright on a swivel chair wearing the VR head-mounted display and headphones. The headphones played a sound loop of ocean surf to prevent external environmental sounds from distracting or confusing the participant. The participant also held a VR controller in one hand.

Each participant got accustomed to the VR experience on a small island set up only for training purposes. Touching a disc on the controller with her thumb the participant was able to regulate both speed (max. 3.3 m/s) and direction with which she moved across the island. The participant was specifically instructed not to change direction using the handheld controller. To decrease the risk of motion sickness, change of movement direction should happen exclusively by rotating the chair.

We then proceeded with calibrating the eye tracker implemented in the VR goggles. During calibration the island disappeared. The participant had to focus her gaze onto ten sequentially appearing targets spread over the VR goggles' viewing field. This process was repeated for validation. Calibration and validation were repeated until gaze location could be detected with an error smaller 2°. The participant was then teleported to the VR town. A session always started at the same location, central in the city and facing east.

We instructed each participant that she could freely explore the city for half an hour and that this would amount to one virtual day, that is from virtual morning to evening. Every five minutes the validation process was repeated, but considerably shorter, with only one validation point. Only when the gaze location error became larger than 2° was the complete ten-point calibration repeated. At the end of the half-hour exploration time the participant was asked to turn towards north. Note that throughout exploration of the city, north could only be inferred from the sun's position and the direction of shadows. The complete VR exploration procedure took approximately 45 minutes. The setup is depicted in Figures 8.1 and 8.2.

III.2.2.5.2.b Map Embodiment

For the exploration of the city map we used the same six-screen monitor setup in a 2x3 arrangement as during the introductory phase. The left column of monitors was turned off. We used a 2D map of the VR city (see Figure 3.1, only houses and streets with white background), which stretched across the middle column of monitors. The participant sat about 80 cm away from the monitors and could use a mouse to move a cursor across the city map.

When the cursor hovered over a house, the house showed a red dot (only for 193 out of all 213 houses). After clicking on a house with a red dot, the right column of monitors showed, on both screens, the same frontal view of the house. The position of the dot on the bird's-eye-view house outline on the map indicated from which side the frontal-view picture of the house was taken. Thus, we ensured that participants knew the orientation of houses.

After the participant was made familiar with the interactive map design, we instructed her that she could freely explore the city for half an hour. After 15 minutes of exploration time had passed, we reminded the participant that 15 minutes were left. The complete map exploration procedure took 30 minutes. The setup is depicted in Figure 8.5.

III.2.2.5.2.c VR with Belt Embodiment

In this group the participant wore the sensory augmentation belt while exploring the city in VR. Before exploring the VR environment, the participant was introduced to the belt. She then walked around the lab with the activated belt vibrating north (real world direction) until she felt accustomed. The belt was then turned off and the participant introduced to the VR setup as described above.

Note that this is a very short introduction period, only sufficient to get familiar with the basic principle of the sensory augmentation device. Prior to this experiment, our participants did not have any experience with the belt. Different from past studies, they are not trained belt wearers (Nagel et al., 2005; Kaspar et al., 2014; König et al., 2016).

After the participant had been teleported into the VR town (to the same location as in the VR experiment), she was asked to turn towards a specific building. This way the participant directly faced

in the direction of VR north. We then activated the belt again and asked the participant to hold her arm out in the direction of the vibration. We changed the belt's vibration direction until it matched the VR city's north, that is until the participant pointed directly forward. Each belt wearing participant then freely explored the city for half an hour, from virtual morning to evening.

As in the VR condition, we performed a validation on eye tracking accuracy every five minutes. Likewise, at the end of the exploration time the participant was asked to turn north. The complete VR with belt procedure took approximately 55 minutes. The belt is depicted in Figure 8.3 and 8.4.

III.2.2.5.3. Task Completion

After the exploration period during which the participant became familiar with the spatial layout of the VR city, she did a series of spatial tasks with and without time pressure. The tasks were the same as introduced at the beginning of the experiment, but the spatial relations now asked for were among houses from the VR city.

The participant had to complete a block of 36 trials for the Absolute task with a time limit of three seconds for each trial. She also had to complete a block of 36 trials for the Absolute task with unlimited decision time for each trial. She did the same for the Relative task as well as the Pointing task. These six task blocks were presented in random order to each participant. Before each block a screen appeared explaining the task. As in the introductory phase, the participant sat approximately 80 cm away from the six-screen setup, and only the two middle screens were used. The task completion took around 45 minutes.

After the participant completed the tasks, she was asked to fill in the "Questionnaire of Spatial Strategies" (translated from German, "Fragebogen Räumlicher Strategien", or FRS; Münzer & Hölscher, 2011) asking about her own perceived everyday navigation strategies and skills. Finally, all participants that completed three repeated measurements were asked to sketch a map of the VR town from memory. The complete experiment took approximately two hours.

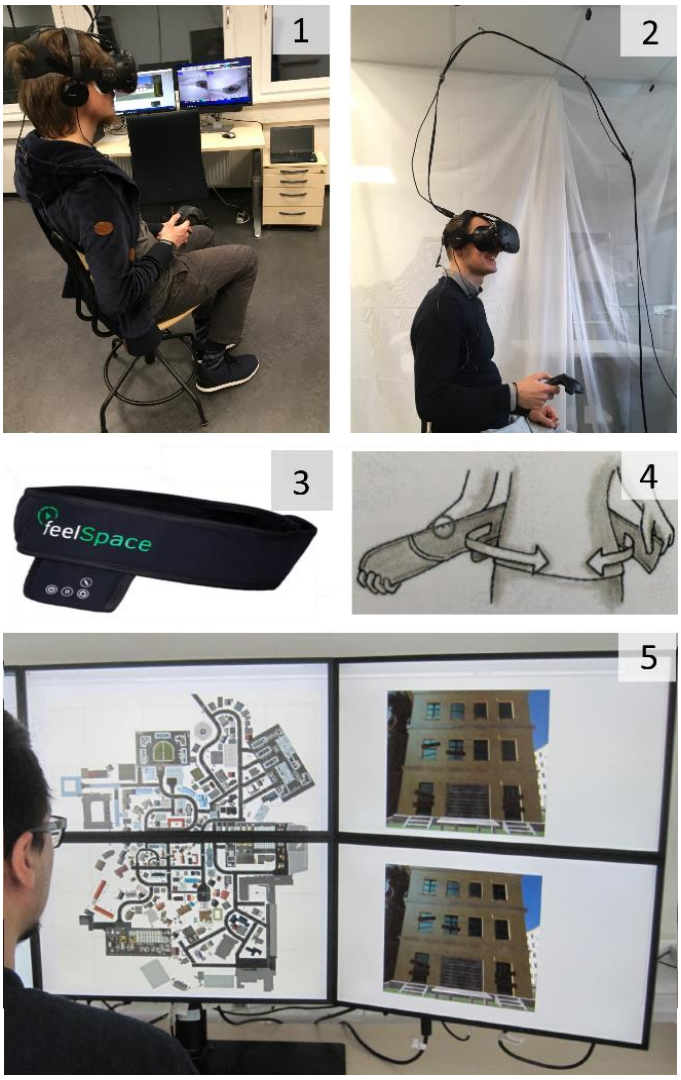


Figure 8. Experimental setup.
8.1 Participant with head-mounted VR display, headphones and VR controller during an experimental session. He uses the controller to regulate his movement speed and rotates on the chair to change movement direction in the VR environment. In the background one sees the experimenter station. **8.2** Experimenter with head-mounted display, headphones and controller. The cable management allows for free rotation. **8.3** The sensory augmentation belt. Visible is the feelSpace logo and the control unit (small rectangular bag with four symbols). The 16 evenly spaced vibration elements are covered by the outer layer of cloth and are, therefore, not visible. During the experiment, only the northernmost (VR north) vibration element is active. Image provided courtesy of feelSpace GmbH. **8.4** The belt is strapped around the waist. The participant can then walk around or (for the experiment) sit down on the VR swivel chair (depicted in Figure 8.1). **8.5** Experimenter sitting in front of the interactive map setup as used during a regular experimental session. With the mouse (not visible) he clicked on a house in the map to see a picture of the house.

III.2.2.6. Materials

III.2.2.6.1. Task Setup

Each of the six screens had width 53cm times height 31cm and was used at a resolution of 1920x1080 pixels. The 2x3 setup was synchronized with Samsung SyncMaster supported by three NVIDIA Quadro K620 graphics cards. The PC processor was an Intel Xeon E5620 running at 2.4 GHz with 12 GB RAM.

III.2.2.6.2. Virtual Reality Setup

For our experimental setup we used the HTC Vive Virtual Reality System. We installed two base stations on the ceiling of the lab. These tracked the position of both the head-mounted display, which we will refer to as VR goggles, as well as the VR controller at a rate of 60 Hz. The VR goggles had an image-refresh rate of 90 Hz and a binocular field of view of about 110°. The device used two OLED panels, one per eye, each having a display resolution of 1080×1200 (technical specifications as provided by HTC).

Inside the VR goggles we installed the Vive Pro Binocular Add-on eye tracker from Pupil Labs (Kassner, Patera, & Bulling, 2014). This eye tracker has a gaze accuracy of 1° and updates at a maximal frequency of 120 Hz (technical specifications as provided by Pupil Labs). For performance reasons we only tracked the gaze location at a frequency of 30 Hz. For more details on how we implemented eye tracking in VR see Clay, König and König (2019).

The software interface linking the PC and the Vive VR System was the game development software Unity (version 5.6.3 and 2017.2.0). The PC processor was an Intel Xeon E5 running at 3.7 GHz with 16 GB RAM. The graphics card was a NVIDIA GeForce GTX TITAN X. During a typical experimental session in the VR city, the rate of frames sent from the PC to the VR goggles varied between 90 and 30 Hz. We used Sony MDR-ZX310B headphones.

III.2.2.6.3. Sensory Augmentation Belt

For our experimental setup we used the naviGürtel Beta'17.Q1 in sizes C, M and L from feelSpace. The sensory augmentation device provides directional information of magnetic north via vibration around the waist.

The belt's control circuit board contains a compass module as well as 16 vibromotors uniformly distributed over its length. Each vibration unit covers an angle of 22.5°. The vibration frequency is 170 Hz. The belt is powered by a lithium-ion battery allowing for approximately eight hours of continuous usage. If the belt is used without an additional device (e.g. smartphone), it functions like a compass in that only the vibration unit that is closest to north vibrates.

The feelSpace belt can also be used in combination with a Bluetooth capable device, allowing users to recalibrate north. We used the belt in combination with the smartphone Samsung Galaxy J3 (2017), which has Bluetooth capabilities. The Android application Luftlinie beta then enabled us to gauge the belt's control circuit onto the VR city's internal north.

III.2.2.6.4. Spatial Strategy Questionnaire

The FRS questionnaire is a self-report measure to assess proclivity and ability of participants to use three main spatial navigation strategies (Münzer & Hölscher, 2011; Münzer, Fehringer & Kühl, 2016). The three strategies are differentiated in global-egocentric, survey-allocentric and cardinal-allocentric. The global-egocentric questions are about general navigational ability as well as knowledge of directions and knowledge of routes and usage thereof, it comprises ten items. The survey-allocentric factor includes items specifically asking about mental map formation and usage, it comprises seven items. The cardinal-allocentric scale finally only includes questions about the knowledge of cardinal directions, it comprises two items. Each question is a Likert item with a score ranging from 1 (“I disagree strongly.”) to 7 (“I agree strongly.”).

III.2.2.6.5. Data

The resulting data is available at https://osf.io/kyp83/.

III.2.2.7. Methods of Analysis

The experiment follows a longitudinal, within-subject design for each exploration group. Consequently, we decided for linear

mixed-effect models with the participant as random effect grouping variable as our method of analysis.[1]

Note that our analysis includes all participants even though about 2/3 did not complete the repeated sessions. While we excluded some data due to technical error, the total imbalance is almost entirely due to a change in the original study design, which only called for one session instead of three sessions per participant. Session is therefore nested within subgroups of subjects: One subgroup only received one session and the other received three sessions. The experimental procedures were identical for each subgroup. Also, t-tests did not reveal significant differences between the one-session and three-session participants in their respective first sessions. We thus decided to ignore subgroup as possible factor and treat the missing data as completely missing at random. Then, including all participants should result in the estimate closest to the population value (Ibrahim & Molenberghs, 2009; Rubin, 1976).

Former studies investigating spatial cognition abilities found that males reacted faster and performed better than females in a variety of spatial tasks (Masters & Sanders, 1993; Moffat, Hampson, & Hatzipantelis, 1998; Newhouse, Newhouse, & Astur, 2007). Preliminary analysis of the data collected in our study did not reveal evidence that gender differences affect our general claims. We therefore do not include gender as a factor in our models.

We do not include item (house stimulus) as a random effect since we specifically designed an algorithm to ensure homogeneous distribution of stimuli properties across tasks (see above and Appendix section 2). In addition, we find that accumulated over participants, for each task a large fraction of the items is learned already in the first session (see Results). For these reasons, we take the house stimuli in each task to be an exhaustive sample of the underlying population, that is the town Seahaven, which makes item a fixed variable (Tukey & Green, 1960).

[1] It has been argued that log-transformed odds are better suited for analysis of percentage values due to floor- and ceiling effects (Jaeger, 2008). Since the results presented here all lie in a range close to 50 % and can be well approximated as normal, we employ linear regression.

We differentiate in models for best data fit and models for plotting purposes. The best-fit models are selected following a stepdown procedure. First the maximum model is run. All possible fixed terms, that means also all possible interactions are included. Following Barr et al. (2013) but adjusting for sample size, we also include all these terms except interactions as random effects. Then a likelihood ratio test is run comparing the maximum model to a nested model with one less fixed term (random structure unchanged). The comparison is done for all fixed terms. Afterwards, the least likely coefficient is dropped from both the fixed and (given it is part of it) the random effect structure. This cleaning step leaves us with the maximum-1 model. The procedure is repeated until a maximum-k model is reached with only significant fixed effects. This is the best-fit model.

In order to use likelihood ratio tests, we employ a maximum likelihood fit algorithm (not restricted). Throughout the analysis, we fit a heterogeneous, but due to convergence issues, diagonal covariance matrix, i.e. possible covariance between random effects is not represented.[2]

The effects of a posteriori model selection on Type I and Type II errors are complex and no universal policy is agreed upon (Altman & Andersen, 1989; Heinze, Wallisch, & Dunkler, 2018; Murtaugh, 2009; Tibshirani, 1996). For mixed models, the situation becomes even more complicated (Müller, Scealy, & Welsh, 2013). To minimize potential model selection effects on alpha and beta errors, here we only interpret significant coefficients at odds with, or in favor of, our a priori hypotheses. We provide all model design matrices and results in Appendix section 4.

Model selection also makes the level at which to extract nonsignificant coefficients for plotting purposes somewhat arbitrary. If, for example, in the best-fit model the difference between some task conditions over sessions turns out not to be significant, do we depict the average as the best linear-fit line? To

[2] These occurred for both, the balanced dataset of only repeatedly measured and the complete set of all participants. That is, the convergence issues were not due to imbalanced data. We also compared the mixed-model algorithms of MATLAB (R2018b) and R (lme4 v1.1-21). Both showed similar convergence issues. Convergence is in fact a common problem in mixed modelling (Eager & Roy, 2017).

allow for qualitative trends to show in our plots, we decided to fit each plot line separately on the data of exclusively one of the six task conditions over sessions. For example, the Absolute task, three second decision time condition was fitted independently. This separate plot-fit approach is simple and systematic. The only exceptions were the plots depicting linear-fit lines where we marginalize over the three spatial tasks. Details are provided in Appendix section 3.

III.2.3. Results

We define performance or accuracy in each task condition as the fraction of correct answers in percent. A performance of fifty percent corresponds to chance in the no-time-limit task condition. With time limit, if participants do not answer in the three seconds the trial is counted as wrong. Thus, here chance performance can be below fifty percent if the participant hesitates too long.

III.2.3.1. VR Exploration

Following hypothesis one (H1) we expect participants under time pressure to perform better in the Relative than in the Absolute task. Linear mixed-model (Appendix 4, Model s-VR) likelihood comparison however does not provide evidence for such an effect. The best fit model does not differentiate between task performances. It only hints at an increase in mean performance over three sessions by two percent, which however remains insignificant (coefficient session=2, standard error (SE)=1, confidence interval (CI)= -0.1, 4.2 (all in percent correct trials, that is performance); likelihood ratio statistics $\chi 2(1)=3.4$, $p=0.065$). For a qualitative comparison see Figure 9.5.

A second result to be expected if H1 were to hold would be that, with unlimited decision time judging house alignment to north is more accurate than judging relative orientation of two houses. That is the opposite effect as expected with time pressure. The model fit (u-VR) does not corroborate this claim, at least not one-to-one. It rather seems that the difference in performance between Absolute and Relative task reverses with session. Absolute starts out worse than Relative but outperforms it with more sessions. The model reveals a significant change of performance difference from session one to session three of about six percent (interaction Abs. to Rel. with session= -5.6, SE=2.1, CI=

-9.7, -1.5; $\chi 2(1)=7.18$, $p=0.007$). Extrapolating this interaction, we can surmise that H1 holds for large environments only for explorers with sufficient experience.

We do also find a main effect of task in the no-time-limit condition. The model fit (u-VR) shows that Pointing outperforms the Absolute task by two percent (Abs. to Poi. = 2.2, SE=0.9, CI=0.4, 4; $\chi 2(1)=5.5$, $p=0.018$). This result corroborates hypothesis four (H4) which simply claims that participants should perform best in the Pointing task. For a qualitative comparison see Figure 9.6.

To investigate participants' overall development in the three tasks across sessions, comparing also both time conditions, we fit a model (a-VR) that incorporated the complete VR data and time as an additional potential factor. Likelihood comparisons did not show any task to be significantly above overall task average, implying that Pointing does not outperform both tasks in both time conditions. That is, VR embodiment does not show conclusive support for H4. As purported by hypothesis five (H5) however, the model did reveal a performance increase by five percent when participants answered without time pressure (time=5, SE=0.7, CI=1.9, 5.2; $\chi 2(1)=41.6$, $p<0.001$).

It also becomes evident that knowledge of spatial relations in our VR city accumulates over several visits as stated in hypothesis six (H6). We find an increase over the three sessions by about four percent. (session=3.6, SE=0.8, CI=1.9, 5.2; $\chi 2(1)=18$, $p<0.001$). However, the difference in performance from three-seconds to unlimited decision time increases over sessions by three percent as well (time with session=3.2, SE=1.5, CI=0.2, 6.1; $\chi 2(1)=4.5$, $p=0.035$). This interaction reveals that the accumulated knowledge after three exploration sessions is mostly of a form that requires processing time to be accessed. For a qualitative comparison see Figure 9.4.

Visual comparison of task performances in each session which are detailed in Figures 9.1 to 9.3 does not give a coherent picture without model guidance. Performance is higher in the condition with no time pressure compared to the condition with a three-second time limit. Much the same way, performance is higher in later experimental sessions compared to earlier experimental sessions. However, we find that only the Pointing performance increases linearly with the number of sessions in both time

conditions. When free to reason, participants show a strong increase in performance in the Absolute task from the first to the second session. No increase in performance is found for the Relative task. Under time pressure, the performance in both tasks stays at chance level.

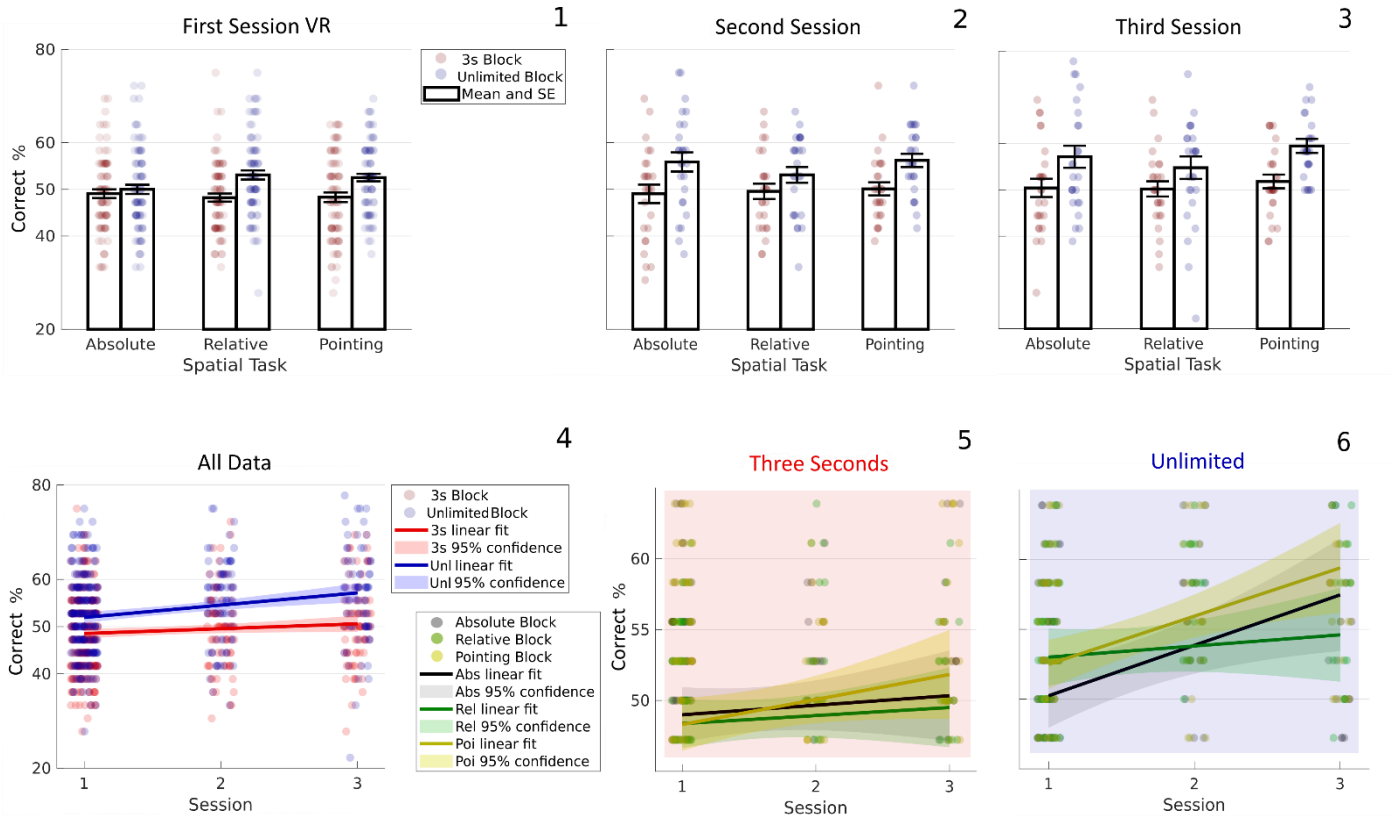


Figure 9. VR Exploration**.**

9.1-3 (Black bars) Barplots of mean performance and standard error (SE) in all six trial blocks of VR participants in their first, second and third session, respectively. (Red dots) Single participant's performance in block with three seconds decision time per trial. (Blue dots) Single participant's performance in block with unlimited decision time.

9.4 (Red dots) Single participant's performance in the three-second condition for each spatial task, respectively. (Blue dots) Single participant's performance for unlimited decision time. (Red line) Linear mixed-effect penalized least squares fit with participant as grouping variable across all three-second blocks. (Blue line) Linear fit across all unlimited decision time blocks. (Red area) The 95% confidence interval for the three-second condition linear fit. That is, 95 of 100 intervals each from a different sample include the actual population fit. (Blue area) The 95% confidence interval for the unlimited condition. **9.5** (Black, Green, Yellow dots) Single participant's performance in the Absolute, Relative and Pointing spatial task trial block, respectively, only for trials with time limit. (Black, Green, Yellow line) Linear fit across sessions in the condition with time limit, for Absolute, Relative and Pointing task, respectively. (Black, Green, Yellow area) The 95% confidence interval for the three-second condition linear fit for Absolute, Relative and Pointing, respectively. **9.6** (Black, Green, Yellow dots) Single participant's performance in the Absolute, Relative and Pointing spatial task trial block, respectively, only for trials without time pressure. (Black, Green, Yellow line) Linear fit across sessions in the free-reasoning condition, for Absolute, Relative and Pointing task, respectively. (Black, Green, Yellow area) The 95% confidence interval for the unlimited time condition linear fit for Absolute, Relative and Pointing, respectively.

All model design matrixes and detailed results can be found in the Appendix sections 3 and 4.

III.2.3.2. Map Exploration

Hypothesis two (H2) predicts that performance in the Absolute task will be better than in the Relative task when exploring the 2D city map. Linear mixed-model (Model s-Map) likelihood comparison however does not provide evidence for such an effect under time pressure. We instead find that participants perform about five percent better in the Relative task compared to Pointing (Poi. to Rel.=5.4, SE=1.3, CI=2.9, 7.8; $\chi2(1)=16.3$, $p<0.001$). Participants also outperform Pointing in the Absolute task by four percent (Poi. to Abs.=3.9, SE=1.2, CI=1.6, 6.1; $\chi2(1)=10.9$, $p<0.001$).

Note that these results are largely due to a learning process over multiple sessions. This is evident as difference in performance from Relative to Pointing increases by eight percent over the three sessions (Poi. to Rel. with session=8, SE=2.1, CI=4, 11.9; $\chi2(1)=14.9$, $p<0.001$). Similarly, Absolute to Pointing increases by six percent over sessions (Poi. to Abs. with session=6.1, SE=2, CI=2.3, 10; $\chi2(1)=9.6$, $p=0.002$). In statistical terms, these are two interactions of task and session. Importantly, these interactions were both unexpected. For a qualitative comparison see Figure 10.5.

For unlimited reasoning time the situation is more complicated. Model (u-Map) results hint at Relative to be two percent above Pointing, this effect however is not significant (Rel. to Poi.= -1.9, SE=1, CI=-3.8, 0; $\chi2(1)=3.7$, $p=0.054$). Without this term in the model, a significant change of about five percent (Rel. to Abs.=4.6, SE=2, CI=0.3, 8.8; $\chi2(1)=4.38$, $p=0.036$) of the difference of Absolute and Relative performance over session reduces to four percent and is, therefore, no longer significant (Rel. to Abs.=3.7, SE=2.2, CI=-.5, 7.9; $\chi2(1)=2.94$, $p=0.086$; not in Appendix). Given these results we can only speak about a qualitative interaction resembling the significant interaction of Absolute and Relative in the VR condition. That is, qualitatively, the difference in performance between Absolute and Relative task reverses with session leading to Absolute eventually outperforming Relative with more experience.

We would not interpret this as qualitative evidence for map knowledge favoring Absolute task performance, i.e. not for H2. We rather see the qualitative interaction as corroborating the idea of two separate forms of spatial representations, as expressed in H1. For a qualitative comparison see Figure 10.6.

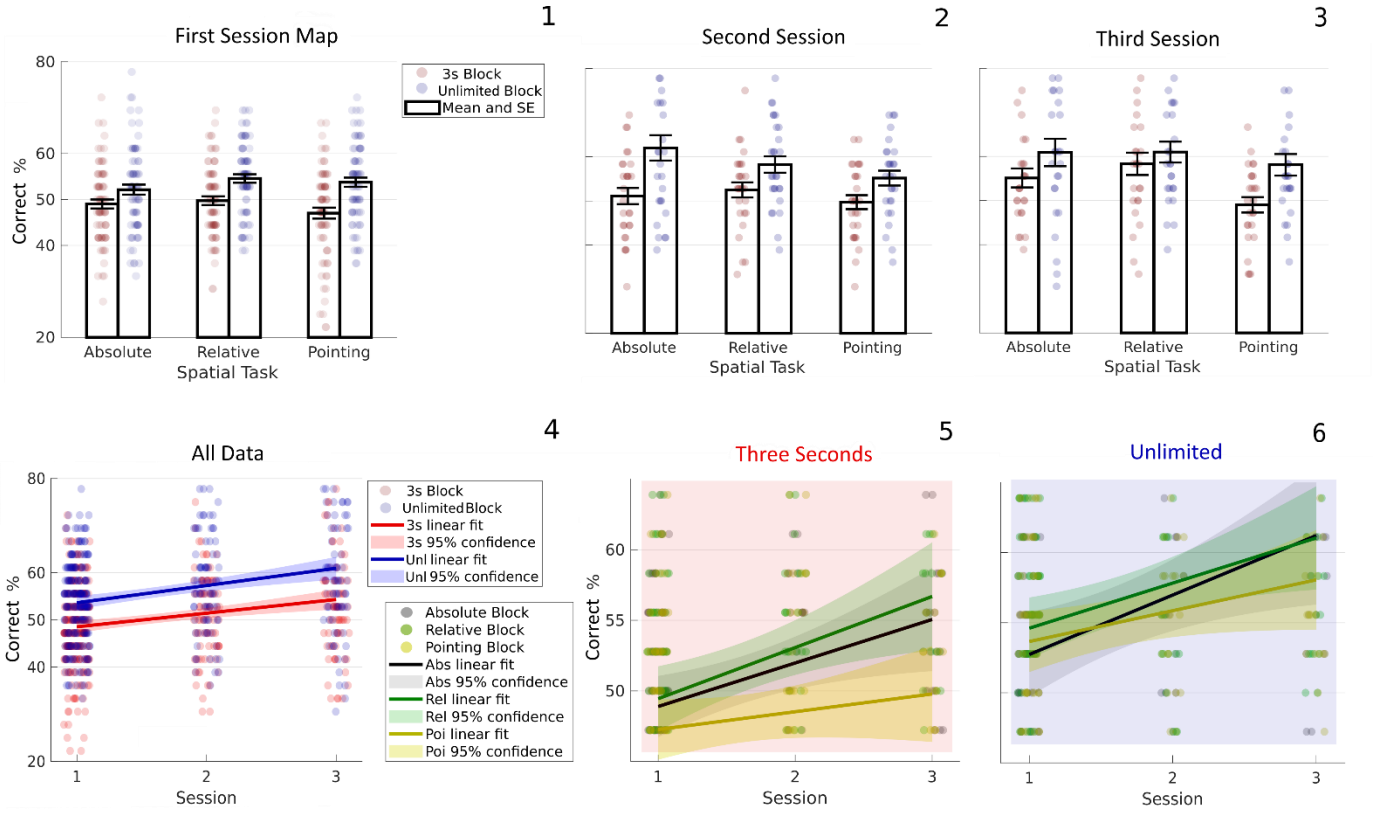


Figure 10. Map Exploration**.**

10.1-3 (Black bars) Barplots of mean performance and standard error (SE) in all six trial blocks of map participants in their first, second and third session, respectively. (Red dots) Single participant's performance in block with three seconds decision time per trial. (Blue dots) Single participant's performance in block with unlimited decision time.

10.4 (Red dots) Single participant's performance in the three-second condition for each spatial task, respectively. (Blue dots) Single participant's performance for unlimited decision time. (Red line) Linear mixed-effect penalized least squares fit with participant as grouping variable across all three-second blocks. (Blue line) Linear fit across all unlimited decision time blocks. (Red area) The 95% confidence interval for the three-second condition linear fit. That is, 95 of 100 intervals each from a different sample include the actual population fit. (Blue area) The 95% confidence interval for the unlimited condition.

10.5 (Black, Green, Yellow dots) Single participant's performance in the Absolute, Relative and Pointing spatial task trial block, respectively, only for trials with time limit. (Black, Green, Yellow line) Linear fit across sessions in the condition with time limit, for Absolute, Relative and Pointing task, respectively. (Black, Green, Yellow area) The 95% confidence interval for the three-second condition linear fit for Absolute, Relative and Pointing, respectively. **10.6** (Black, Green, Yellow dots) Single participant's performance in the Absolute, Relative and Pointing spatial task trial block, respectively, only for trials without time pressure. (Black, Green, Yellow line) Linear fit across sessions in the free-reasoning condition, for Absolute, Relative and Pointing task, respectively. (Black, Green, Yellow area) The 95% confidence interval for the unlimited time condition linear fit for Absolute, Relative and Pointing, respectively.

To investigate participants' overall development in the three tasks across sessions, comparing also both time conditions, we fit a model (a-Map) that incorporated the complete map data and time as an additional potential factor. To be expected, the model does not provide evidence for H2, rather performance in the Pointing task is significantly below task average (mean to Poi.= -2.2, SE=0.7, CI=-3.6, -0.7; $\chi2(1)=7.1$, $p=0.008$;). This result speaks against H4 which purports Pointing to be the best task across embodiment conditions. Further likelihood comparisons do provide evidence for H5, that is performance increases without time pressure (time= 5.4, SE=0.8, CI=3.7, 7; $\chi2(1)=31$, $p<0.001$). Also, H6 seems to hold since performance increases by about six percent over the three sessions (session=6.6, SE=1.2, CI=4.3, 8.8; $\chi2(1)=26.9$, $p<0.001$). For a qualitative comparison see Figure 10.4.

Comparable to the VR embodiment results, visually assessing systematic differences in performance development from Figures 10.1 to 10.3 is barely possible. We see participants performing worse with limited decision time and performance increasing with session. When free to reason, participants show a strong increase in Absolute task performance from the first to the second session. Also discernible is that performance in the Pointing task increases with session but remains below both Absolute and Relative performance in both time conditions. All in all, performance in the Relative task increases the most.

All model design matrixes and detailed results can be found in Appendix sections 3 and 4.

III.2.3.3. VR with Belt Exploration

As stated in hypothesis three (H3) we expect participants to perform better in the Relative spatial task asking for house-to-house orientations than in the Absolute task asking for house-to-north orientations if equipped with the north belt during the exploration period.

Performance under time pressure does not corroborate our claim. Linear mixed-model (Model s-VRwB) likelihood comparison show instead that both Absolute and Relative start out well below

Pointing but draw even in the third session. The model makes evident that the difference in performance from Pointing to Absolute decreases by five percent over the three sessions (Poi. to Abs. with session=5.2, SE=1.9, CI=1.5, 9; $\chi2(1)=7.6$, p=0.006). Similarly, Pointing to Relative decreases four percent over sessions (Poi. to Rel. with session=3.9, SE=1.9, CI=0.2, 7.6; $\chi2(1)=4.2$, p=0.04). For a qualitative comparison see Figure 11.5. Extrapolating those results hints at both Absolute and Relative outperforming Pointing with more exploration time. The result implies that unexperienced belt wearers are not able to integrate the information it provides into a quickly accessible format.

Without limit on reasoning time the participants' decisions provide evidence for H3. The best model fit (u-VRwB) reveals that Relative outperforms Absolute by two percent across sessions (Abs. to Rel.=2, SE= 0.9, CI=0.2, 3.5; $\chi2(1)=4.6$, p=0.03). For a qualitative comparison see Figure 11.6.

To investigate participants' overall development in the three tasks across sessions, comparing also both time conditions, we fit a model (a-VRwB) that incorporated the complete VR with belt data and time as an additional potential factor. The model provides evidence for H4 because it shows Pointing to be significantly better than the task average (mean to Poi.=1.1, SE=0.4, CI=0.3 1.9; $\chi2(1)=7$, p=0.008).

Furthermore, as in the other two conditions, we find evidence for H5 and H6. That is performance increases without time pressure (time=3.3, SE=0.9, CI=1.6, 5.1; $\chi2(1)=12.4$, $p<0.001$) and over repeated sessions (session=3.4, SE=0.8, CI=1.8 5; $\chi2(1)=17.23$, $p<0.001$). For a qualitative comparison see Figure 11.4.

Also, similar to the other two exploration conditions, visually assessing systematic differences in performance development from Figures 11.1 to 11.3 is barely possible. In the three-second spontaneous-decision condition participants perform worse than with reasoning time and performance increases with session. Unprecedented across exploration condition is that for decisions without time pressure, Absolute performance does not show a large increase from the first to the second session.

All model design matrixes and detailed results can be found in Appendix sections 3 and 4.

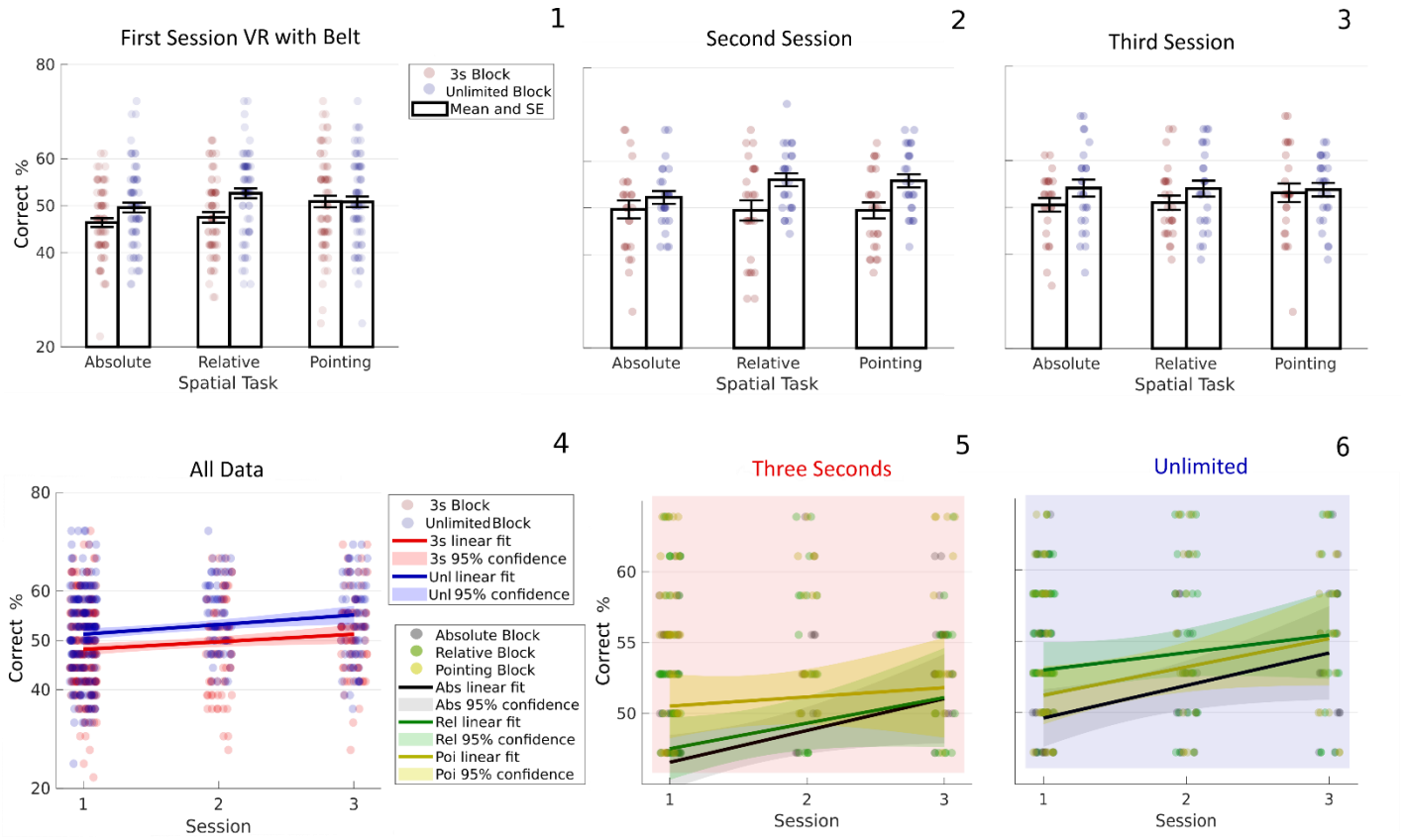


Figure 11. VR with Belt Exploration**.**

11.1-3 (Black bars) Barplots of mean performance and standard error (SE) in all six trial blocks of VR with belt participants in their first, second and third session, respectively. (Red dots) Single participant's performance in block with three seconds decision time per trial. (Blue dots) Single participant's performance in block with unlimited decision time.

11.4 (Red dots) Single participant's performance in the three-second condition for each spatial task, respectively. (Blue dots) Single participant's performance for unlimited decision time. (Red line) Linear mixed-effect penalized least squares fit with participant as grouping variable across all three-second blocks. (Blue line) Linear fit across all unlimited decision time blocks. (Red area) The 95% confidence interval for the three-second condition linear fit. That is, 95 of 100 intervals each from a different sample include the actual population fit. (Blue area) The 95% confidence interval for the unlimited condition.

11.5 (Black, Green, Yellow dots) Single participant's performance in the Absolute, Relative and Pointing spatial task trial block, respectively, only for trials with time limit. (Black, Green, Yellow line) Linear fit across sessions in the condition with time limit, for Absolute, Relative and Pointing task, respectively. (Black, Green, Yellow area) The 95% confidence interval for the three-second condition linear fit for Absolute, Relative and Pointing, respectively. **11.6** (Black, Green, Yellow dots) Single participant's performance in the Absolute, Relative and Pointing spatial task trial block, respectively, only for trials without time pressure. (Black, Green, Yellow line) Linear fit across sessions in the free-reasoning condition, for Absolute, Relative and Pointing task, respectively. (Black, Green, Yellow area) The 95% confidence interval for the unlimited time condition linear fit for Absolute, Relative and Pointing, respectively.

III.2.3.4. VR as Reference Condition

III.2.3.4.1. VR versus Map

As outlined in the introduction, the embodiment condition while exploring the city in VR only, is closest to navigating an unknown large-scale environment without additional help from external devices. It can, thus, be considered a baseline or reference for acquisition of allocentric representations. Through using a map as navigation device, a participant can gather all information necessary to solve the spatial tasks, while performing only a minimal amount of bodily movement.

In the following we directly compare VR and map results using linear mixed models through simply adding exploration condition as new possible fixed effect. Note that this comparison is mostly of an exploratory nature. Only H4, H5 and H6 are supposed to hold across all three embodiment conditions during exploration. Since H5 and H6 have already been shown to fulfill this expectation we are not reporting the combined data results for these hypotheses here.

Modelling (s-VRvMap) the task performance with time limit of both exploration conditions creates no new insights. We refer the interested reader to the Appendix section 4.

In the decision condition without time pressure the comparison (u-VRvMap) reveals that the difference between Absolute and average task performance reverses over the three sessions (Abs. to Rel. with session= -3.2, SE=1.5, CI=-6.1, -0.2; $\chi2(1)=4.49$, p=0.034). This result further strengthens the only qualitative evidence for the same interaction appearing in both the VR and map condition over sessions. It thus provides additional evidence for two robust forms of allocentric representations which, given enough exploration experience, will satisfy H1. That is Relative outperforms Absolute with time limit and vice versa without.

The largest model (a-VRvMap) likelihood comparisons incorporating both time conditions reveal that performance after map exploration is better than performance after exploring the city in VR (exploration=2.3, SE=0.8, CI=0.7, 3.9; $\chi2(1)=8.2$, p=0.004). This is despite the fact that Pointing performance is below overall task average after map exploration but above average after VR exploration (mean to Poi. with explor.= -3.1, SE=0.7, CI=-4.4, -1.8;

$\chi2(1)=20.4$, $p<0.001$). We have seen before that in the map embodiment condition Pointing is significantly worse than the other two tasks. Taken together the findings are evidence for an effect of exploration condition on the validity of H4, which claims that spatial reasoning processes improve in accuracy when they involve a highly familiar action.

In order to find out if Pointing performance itself is worse after map than after VR exploration, and not just in relation to other tasks, we fit a model (p-VRvMap) with only the Pointing task data of both groups. We do not find a main effect of experimental exploration condition on Pointing performance (exploration= -0.9, SE=0.9, CI=-2.6, 0.9; $\chi2(1)=0.87$, $p=0.35$). We do also not find evidence that such an effect develops with longer exploration time, i.e. number of repeated sessions (explor. with session= -1.8, SE=1.9, CI= -5.6, 2; $\chi2(1)=0.88$, $p=0.35$). For qualitative results compare Figure 9.5 and 9.6 to Figure 10.5 and 10.6.

All model design matrixes and detailed results can be found in Appendix section 4.

III.2.3.4.2. VR versus VR with Belt

The sensory augmentation belt provides the participant with egocentric information about cardinal direction, thereby making allocentric representations somewhat obsolete. In the following we directly compare VR- and VR with belt results through linear mixed models, adding exploration condition as new possible fixed effect. Only H4, H5 and H6 are supposed to hold across all three embodiment conditions. Again, we will not report combined data results for H5 and H6 here.

Neither model (s-VRvB) likelihood comparisons of performance resulting under time pressure, nor modelling (u-VRvB) the combined free-reasoning results reveal any differences between both groups. On the contrary, combining both VR and VR with belt, free-reasoning results reveal that the difference between Absolute and Relative reverses over the three sessions (Abs. to Rel. with session= -4.7, SE=1.4, CI= -7.5, -1.8; $\chi2(1)=10.3$, $p=0.001$). The interaction is the same as in the VR condition and when we combine VR and map results above. That is again evidence for the robustness of the distinction between two forms of allocentric representations as purported in H1.

The polar histogram in Figure 12 depicts the orientation of participants when asked to turn towards north while in the VR city (at the end of exploration). It is apparent that participants with belt can distinguish north significantly better than participants without belt. The fact that belt participants, after exploration, perform qualitatively worse (compare Figure 11.6 and Figure 9.6) when judging the direction of north than VR-only participants speaks for H3. We do not build stable representations usable in the absence of that which they represent when we do not need to. Representations are ecological.

Finally, incorporating both time conditions in one large model (a-VRvBelt) provides evidence that Pointing performance is above overall task average (mean to Poi.=0.8, SE=0.3, CI=0.1, 1.4; $\chi 2(1)=5.6$, $p=0.018$). The stability of high Pointing performance relative to the other two spatial tasks, as claimed in H4, emphasizes the impeding effect of map learning, discussed before.

All model design matrices and detailed results can be found in Appendix 4.

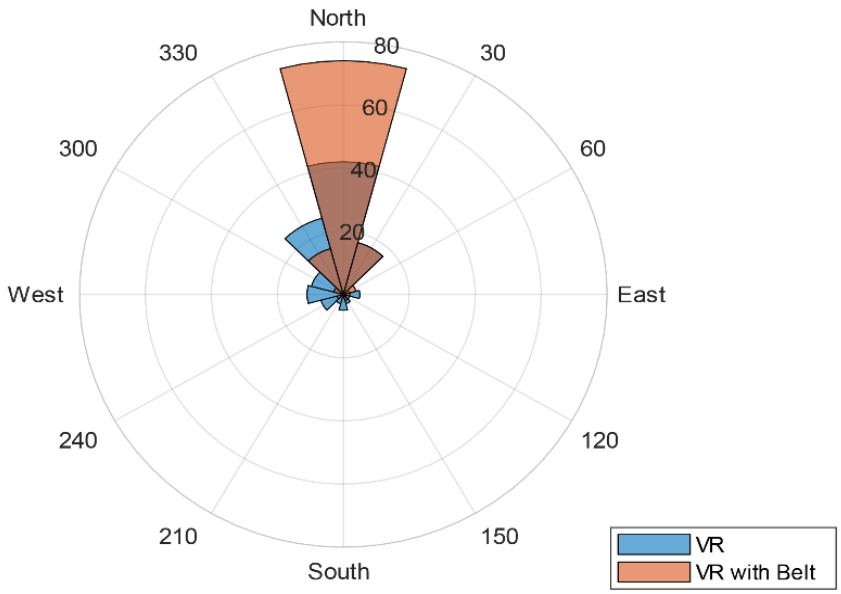


Figure 12. Polar coordinate histogram depicting the direction participants faced after having been asked to turn towards North. The command was given at the end of each exploration session. Each bin covers an angle of 30°. The numbers increasing radially depict the number of participants in that bin. All participants that only explored VR without belt are in blue. All participants that only explored VR wearing the sensory augmentation belt are in red.

III.2.3.5. Behavior of Participants and Items

Further interpretation of the results presented above faces two major obstacles. First, spatial performance is generally only ten percent above chance level and often below. That undermines the generalizability of the results as floor effects, that means artefacts due to overall very limited acquired knowledge, might potentially confound our results. That is not to say that our results are not reproducible, the statistical analysis is sound, but their interpretation becomes significantly more involved.

Past studies have repeatedly found evidence for differences in spatial aptitude in humans. Some people just seem to build more comprehensive representations or can use these more effectively than others (Ishikawa & Montello, 2006; Weisberg et al., 2014; Weisberg & Newcombe, 2016). Optimally, we need to provide evidence for a similar distinction among our participants and show that the results discussed above are stable across those differences in spatial learning ability.

Second, although participants were instructed to "give their best" in all tasks, we never explicitly instructed them to search for certain buildings or follow certain routes while exploring the VR city. Especially given the size of our VR city, reproducibility or generalizability of our results seems not to be guaranteed. Ideally, our results should generalize to the entire population of houses in our city and to different cities. Since we only use 36 sets of house stimuli or items, in each task, we need to make sure that the results do not simply reflect the same set of houses learned by each participant.

III.2.3.5.1. Participant Behavior Indicates Spatial Aptitude

We will first investigate the possibility of confounding floor effects. As exemplified in the lower row of Figure 13 performances can be approximated as normal in all task and time conditions across exploration conditions and sessions. That is, task performance does not reveal participant clusters of low and high navigation aptitude.

An additional factor we need to consider is that each experimental session provides us with six dependent variables which change their values with each sessions. Subgroups might split into different combinations of these variables, i.e. being good at Pointing in the

first session could imply being bad at the Relative and Absolute task in later sessions. Also our participants might not split into subgroups all together, but rather fall into a more continuous spectrum of spatial ability.

A great advantage of our study design is the level of control. It allows analysis of a multitude of independent behavioral variables. These variables in turn might allow us to better identify individual spatial cognitive strategy and aptitude. We investigate the stability of our main results presented above in light of four behavioral factors. Percentage of houses seen, median looking distance, movement speed and self-report. For the following analysis we only use the values of the first session, respectively.

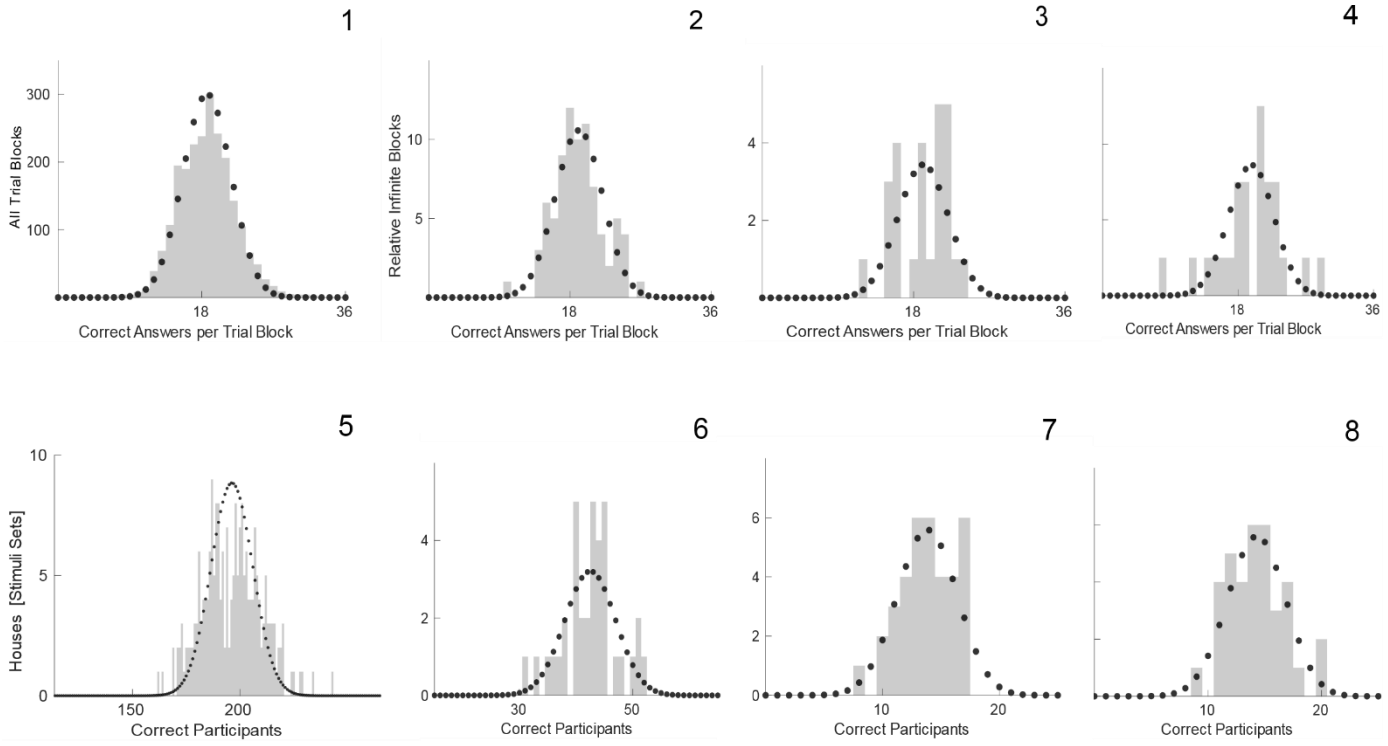


Figure 13. Performance Distribution.
13.1-4 (Gray bars) Histogram of correct choices summed over all trial blocks in 13.1 and a single exemplary condition trial block per session in 13.2-4. The single condition is the Relative spatial task with no time limit per trial and only for participants who explored VR without belt. Shown are the first, the second and the third sessions, respectively. (Black dots) Binomial distribution *B(p,n)* with *p* equal to the ratio of correct trials to all *n*=36 trials per block. **13.5-8** (Gray bars) Histogram of correct participants as "chosen" by individual house stimuli sets summed over all trial blocks in 13.5 and the VR, Relative task, no-time-limit trial blocks per session in 13.6-8. Shown is the first, the second and the third session, respectively. (Black dots) Binomial distribution *B(p,n)* with *p* equal to the ratio of correct participants to all *n* participants in one session.

III.2.3.5.1.a Percentage of Houses Seen

In Figure 14.1 to 14.3 we illustrate the percentage of houses seen or clicked for each participant in each embodiment condition. Each bin gives the number of participants that saw or clicked on that percentage of houses during one exploration session. For the VR and VR with belt conditions, we made use of the eye tracking data. We combined both VR goggle and eye tracking coordinates to compute a gaze ray and infer which VR house this ray has hit at any one moment. If the ray hit a house for at least 260ms we counted that house as seen. Details of this procedure can be found in the publication by Clay, König and König (2019).

All three exploration conditions show a left-skewed unimodal distribution with similar amplitude and mode at around 85%. Note that each distribution is accumulated over all three sessions. For the following analysis we only use the values of the first session. We transform each distribution to a normal form using the exponential function before adding it as factor in the design matrix of its respective model. To evaluate the effect of the percentage of houses seen we simply add the factor to the full models of the limited- and unlimited decision time respectively and perform the step-down procedure described in the Methods section.

We do not find any effect in the VR and the VR with belt condition. In the map condition however, the number of houses clicked-on during exploration proves to be a highly significant indicator of performance both with (Model s-Map.p) and without time limit (u-Map.p). With an increase from about 50% to 100% of clicked houses average performance rises with consecutive sessions by 16%, given three seconds decision time (session with percent=16.3, SE=5, CI= 6.5, 26.2; $\chi2(1)=11$, $p=0.001$).

The behavioral factor also shows that the performance difference among Relative, Absolute and Pointing task over sessions is itself a differential effect. The difference in performance increases with the percentage of houses clicked-on (Poi. to Rel. with session with percent=16.4, SE=4, CI= 8.4, 24.3; $\chi2(1)=13.7$, $p<0.001$) (Poi. to Abs. with session with percent=11.8, SE=4, CI= 3.9, 19.7; $\chi2(1)=8.2$, $p=0.004$).

Similar to the time pressure condition, with unlimited decision time performance rises faster with sessions if the participant saw more houses. The effect is the largest which involves a behavorial variable in our study, with a change in performance of 35% (session with percent=35.4, SE=8.9, CI= 17.9, 53; $\chi 2(1)=15.1$, $p<0.001$).

Likewise, the performance difference between Pointing and Relative (and Absolute, not reported) decreases when the number of houses decreases (Poi. to Rel with percent= -22.6, SE=11.2, CI= -44.6, -0.5; $\chi 2(1)=4$, $p=0.04$). Figure 14.4 illustrates the difference

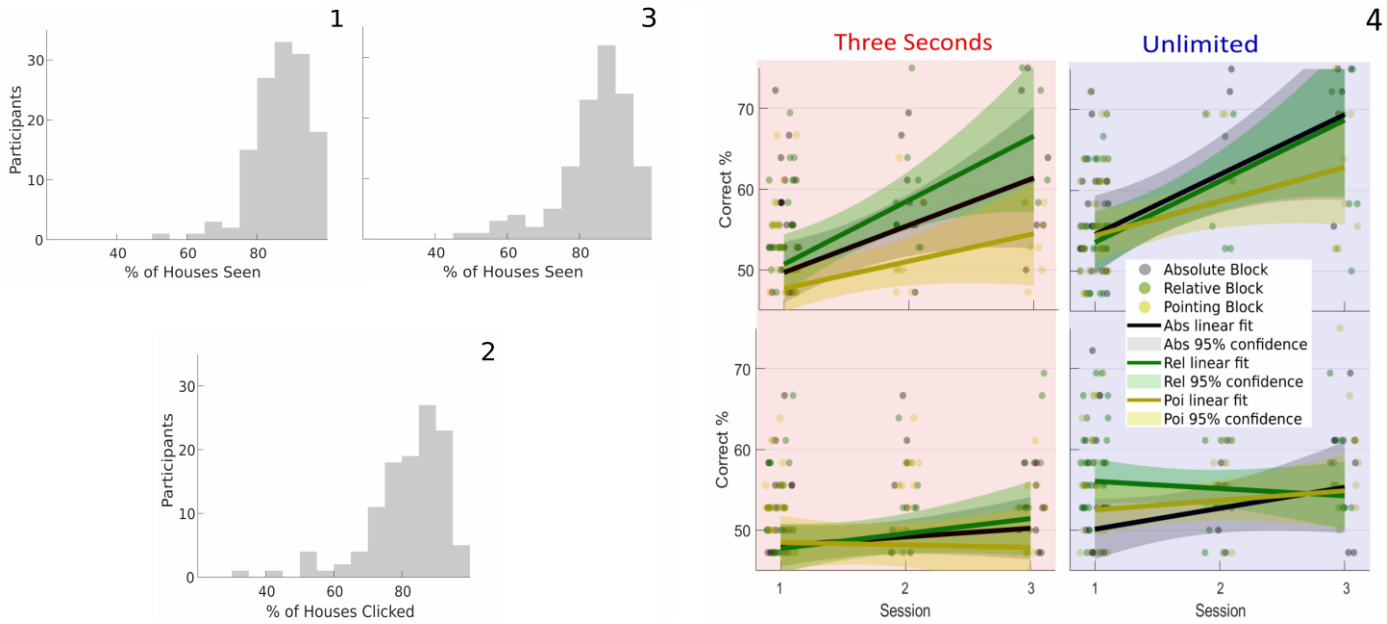


Figure 14. Percentage of Houses Seen.
14.1 Histogram of the percentage of houses seen during VR exploration. Each bin covers 5%. Its height denotes the number of participants that have seen that percentage of houses during one exploration session. A house is counted as "seen", when it has been looked at for 260ms. The total number of houses in the city is 213. **14.2** Histogram of the percentage of houses seen during map exploration. Each bin covers 5%. Its height denotes the number of participants that have clicked on that percentage of houses during one exploration session. After clicking on a house, a frontview of that house appears on the screen. The total number of clickable houses is 192. **14.3** Histogram of the percentage of houses seen during VR exploration with sensory augmentation belt. Each bin covers 5%. Its height denotes the number of participants that have seen that percentage of houses during one exploration session. A house is counted as "seen", when it has been looked at for 260ms. The total number of houses in the city is 213.
14.4 The left column depicts performances with three second decision time, the right column with unlimited time. The upper row shows the development of spatial task performance for map participants which clicked on more than 85% of houses during the first session. The lower row illustrates the same for participants with less than 85% of seen houses. (Black, Green, Yellow dots) Single participant's performance in the Absolute, Relative and Pointing spatial task trial block, respectively. (Black, Green, Yellow line) Linear fit across sessions for Absolute, Relative and Pointing task, respectively. (Black, Green, Yellow area) The 95% confidence interval for the linear fit for Absolute, Relative and Pointing, respectively.

between participants that saw less than 85% and participants that clicked on more than 85% of all houses during map exploration.

Finally, the percentage-seen factor also allows a more differentiated view on the qualitative evidence for H1 after map exploration we reported in the main results above. That is, Absolute does outperform Relative performance with more sessions, but only for those participants that have clicked on fewer houses (Rel. to Abs. with percent= -24.6, SE=11.2, CI= -46.6, -2.5; $\chi2(1)=4.8$, p=0.03). Only a part of the map participants, the ones that perform significantly worse, show evidence for two forms of allocentric representations.

III.2.3.5.1.b Looking Distance

The eye and head coordinates also allow us to infer the distance from which participants look at houses while exploring the city in the VR and the VR with belt condition. Median looking distance (weighted by the number of looking events per house, i.e. looking time) follows a log-normal distribution. We consequently use its logarithm as an additional factor in the limited and unlimited decision time models. For the following analysis we only use the looking distance values of the first session.

We find a strong effect of looking distance for belt wearers under time pressure (Model s-VRwB.d). Those participants that stay further away from the houses they see are better at pointing from one house to another. Closely observing participants, on the other hand are about 17% percent better at judging relative orientation of houses (Poi. to Rel. with distance=17.2, SE=5.5, CI= 6.4, 28; $\chi2(1)=9.7$, p=0.002). They also show an increase of 14% when judging the orientation of houses towards north (Poi. to Abs. with distance=14.3, SE=5.5, CI= 3.5, 25.1; $\chi2(1)=6.7$, p=0.01). That is, the factor of looking distance shows a three percent stronger increase in the Relative task compared to the Absolute task. This finding further corroborates H3.

Interestingly, we find no significant distance effects with unlimited decision time or for participants without belt.

III.2.3.5.1.c Movement Speed

We further analyzed movement speed across participants in the VR and the VR with belt exploration condition. Recall that participants

were able to regulate speed by moving their thumb up or down on a disc on a handheld controller. In contrast to other behavioral factors we find the distribution of median movement speed in both embodiment conditions to be bimodal (about one third of participants are slow). The bimodal form is consistent across sessions.

For the following analysis we only use the values of the first session. We perform a median split and map the participants' velocities below and above the median to a binary code. We then use this code as an additional factor in the limited and unlimited decision time models.

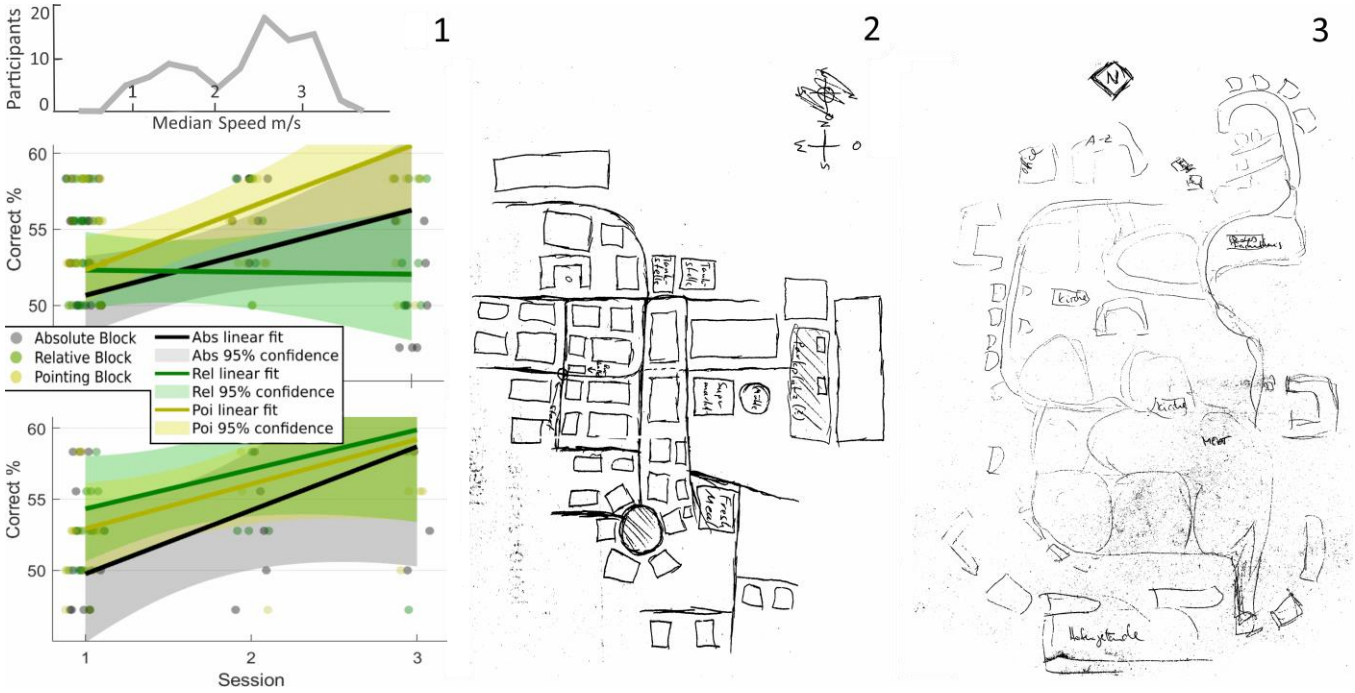


Figure 15. Speed Groups and Sketch Maps.
15.1 Upper panel, distribution of median movement speed per participant in the first session VR exploration without belt. Each edge in the graph marks the amplitude of a histogram bar binning all participants with median speed in an interval of 0.3 m/s. Middle panel, development of spatial task performance in the VR condition with unlimited decision time and median movement speed larger 2 m/s. Lower panel, same development for participants with median movement speed below 2 m/s. (Black, Green, Yellow dots) Single participant's performance in the Absolute, Relative and Pointing spatial task trial block, respectively. (Black, Green, Yellow line) Linear fit across sessions for Absolute, Relative and Pointing task, respectively. (Black, Green, Yellow area) The 95% confidence interval for the linear fit for Absolute, Relative and Pointing, respectively.
15.2 Exemplary sketch map drawn by a VR participant after the third exploration session. Participants were instructed to draw a quick sketch map in around five minutes and indicate the direction of North **15.3** Exemplary sketch map drawn by a map participant after the third exploration session. Participants were instructed to draw a quick sketch map in around five minutes and indicate the direction of North.

Modelling the unlimited decision time condition (Model u-VR.s) reveals that the difference in performance between the Absolute and Relative task depends on the movement speed (Abs. to Rel. with speed (speed difference 1.5 m/s) = -3.3, SE=1.7, CI= -6.6, 0; $\chi 2(1)=3.86$, p=0.049).

Importantly, the factor of walking speed does not interact with the effect of session on the differential learning of Relative and Absolute. Instead, adding the factor to the model increases the probability for H1, that is Absolute outperforming Relative with more experience (Abs. to Rel. with session= -6.2, SE=2.1, CI= -10.4, -2.1; $\chi 2(1)=8.5$, p=0.004). For a visual comparison of both groups see Figure 15.1.

The median movement speed of the belt wearing participants does not inform statistically significant changes in spatial task performance.

III.2.3.5.1.d Self-Report

As final behavioral factor we analyze self-report data on individual spatial aptitude and strategy. The data stems from the FRS questionnaire participants filled out in their first experimental session. The questionnaire assesses proclivity and ability of participants to use three forms of spatial knowledge. General spatial knowledge with a focus on egocentric strategies, survey knowledge and cardinal knowledge.

The distribution of general-egocentric answers shows a left-skew and was transformed with an exponential function. Self-assessed survey knowledge is normally distributed. Cardinal knowledge is perceived as being poor, that means it shows a strong right-skew and was consequently log-transformed.

Performance in the VR embodiment condition increases with self-assessed survey (Model s-VR.rs, u-VR.rs) and cardinal knowledge (s-VR.rc, u-VR.rc) in both time conditions. Survey knowledge shows the largest effect on average task performance without time pressure. Participants with the lowest self-assessed Likert score, one, perform eleven percent below participants with the highest Likert score, seven (survey=11.1, SE=2.7, CI= 5.7, 16.4; $\chi 2(1)=14.7$, $p<0.001$). Compared to the Relative task, participants with high cardinal self-esteem are better at the Absolute task than those with

low cardinal scores (Abs. to Rel. with cardinal=8.1, SE=2.3, CI= 3.5, 12.7; $\chi2(1)=11.8$, $p<0.001$).

Neither factor interacts with the effect of session on the differential learning of Relative and Absolute. Instead the probability for H1 rises when including survey knowledge as additional factor in the unlimited decision time VR model (Abs. to Rel. with session= -6.4, SE=2.2, CI= -10.6, -2.1; $\chi2(1)=8.7$, $p=0.003$). The same holds for cardinal knowledge as additional factor (Abs. to Rel. with session= -6.9, SE=2.2, CI= -11.2, -2.6; $\chi2(1)=9.9$, $p=0.002$).

In the map condition, in addition to survey (u-Map.rs) and cardinal knowledge (s-Map.rc, u-Map.rc), general navigation ability mixed with egocentric strategies (u-Map.rg) also results in higher performance (general=8.4, SE=3.2, CI= 2.1, 14.6; $\chi2(1)=6.5$, $p=0.01$). This could be an indicator for an influence of egocentric representations build up during map exploration.

In contrast to the above results, exploring in VR with belt only leads to an effect of self-assessed cardinal knowledge with limited decision time (s-VRwB.rc). Given that participants have to answer in three seconds, however, the effect is remarkably strong with an increase of ten percent in average performance over the three sessions (session with cardinal=10.1, SE=3.4, CI= 3.5, 16.7; $\chi2(1)=8.9$, $p=0.003$). That is, for participants with high affinity for cardinal directions the belt boosts building spatial representations which can be accessed quickly.

III.2.3.5.1.e Sketch Maps

A potential fifth independent behavioral variable would be the quality of sketch maps participants were instructed to draw after they completed three exploration sessions. Unfortunately, the idea came up late in the experiment, when most measurements were already complete. Hence, given little data and no prior hypotheses we refrain from a quantitative in depth analysis. However, in free drawing paradigms even sophisticated methods like regression coefficients between a participant's metric-transformed and the actual map coordinates rely on the experimenter to manually label landmarks (Gardony, Taylor, & Brunye, 2015; Huffman & Ekstrom, 2019). Furthermore, such comparisons of labeled landmarks become somewhat problematic when participants are allowed to

freely explore a large and complex environment such as ours. Some participants will focus on different landmarks than others.

Consequently, we believe qualitative analysis to still be of value in our case. Indeed, simple visual comparison makes apparent that maps drawn from memory after the third session of map exploration are of remarkably higher quality than the maps sketched after exploring the city in VR with and without belt. We present two example sketches in Figure 15. At first sight the VR sketch seems more detailed. Comparison with the bird's-eye view in Figure 3.1 however, shows that the sketch drawn by the map participant, albeit faint and with fewer rectangles, is a much more accurate depiction of the real VR city layout. In conclusion, the map embodiment condition allows for the best sketch maps.

III.2.3.5.1.f Concluding Remarks

All of the above results follow from an exploratory post-hoc analysis. For this reason, their significance needs to be interpreted with caution. Consequently, we refrain from adding more than one behavioral factor to a model at the same time although this would allow us to, for example, investigate interactions among behaviors. However, even if some of our findings will turn out to be spurious in future experiments, it is remarkable that many behavioral effects increase the significance of the marginalized results presented before. While behavioral differences clearly indicate a spectrum of spatial aptitude and strategy among our participants, all significant results for, or against, our initial hypotheses remain stable. The analysis thus provides evidence against the existence of confounding floor effects.

For all significant behavioral effects presented we provide details on their respective models and results in the Appendix section 4.

III.2.3.5.2. Item Behavior Indicates Invariant Structure

Participant behavior indicates consistent differences in spatial navigation aptitude and strategy among individuals. The same might hold for item behavior. Here, item behavior simply refers to the different properties of items. Some of these properties can be ordered along a dimension, for example angular difference or distance between items, and thus allow for statistical analysis. In

contrast to the participants' behavioral variables discussed above however, items vary during the course of a spatial task and among different tasks. Therefore, they also allow to identify differences among structures of spatial representations which might hold across all participants.

III.2.3.5.2.a Angular Alignment

Concerning angular difference, reported elsewhere we do find performance in the Absolute task after map exploration to excel if the orientation of stimulus house and north are aligned (König et al., 2019). Alignment, here, means that the stimulus house front faces south. Thus, the participants' egocentric perspective with respect to the item aligns with the cardinal direction of north. The map embodiment condition did not reveal alignment effects in Relative or Pointing task trials when the respective prime house faced south.

We tested the VR and VR with belt results for both north and east alignment effects. East was the direction all participants faced when starting VR city exploration. It has been shown that the direction from which one enters a new environment can become a reference direction which facilitates spatial tasks, e.g. pointing, even if that environment is large (Shelton & McNamara, 2004; He, McNamara, Bodenheimer, & Klippel, 2018).

For the analysis of each exploration condition we used repeated measures ANOVA with session, task, time condition and angular difference (0°, 30°, …, 180°) as independent variables. Neither of the two VR embodiment conditions, however, shows strong alignment effects. This lack of evidence for a dominant reference direction could be due to an interference of the task related importance of north and the starting direction of east.

III.2.3.5.2.b Distance

The Relative and the Pointing task allow an investigation of the effect of distance between prime and target house on trial performance. We use simple linear regression to capture the relation of the percentage of correct participants in a trial and the distance of the trial house pair.

With unlimited decision time we find no correlation of trial performance and house pair distance in any session. When

decisions have to be made in three seconds however, participants reveal a differential learning effect. Judging alignment of house fronts becomes significantly easier when both houses are close together compared to them being far apart.

In the VR condition this effect becomes strongest after three exploration sessions. In the third session participants are about ten percent worse when the correct house pair distance increases by 100m (100m distance= -9.16, SE=3.23, CI= -15.77, -2.55; t(34)=-2.82, p=0.008, R2=0.19). VR participants with belt show a similar effect after two sessions (100m distance= -10.17, SE=3.5, CI= -17.32, -3.02; t(34)=-2.89, p=0.007, R2=0.2). Even map participants show a qualitative decrease of six percent for every 10cm (on screen) increase of pair distance (10 cm distance= -5.93, SE=3.08, CI= -12.24, 0.37; t(34)=-1.91, p=0.064, R2=0.1).

Remarkably, performance in the Pointing task shows no dependence on distance between house pairs after exploration in either embodiment condition. All regression model results for both time conditions are presented in Table 2. The findings are illustrated in Figure 16.

Three Seconds		Distance	(Intercept)	SE	(SE Intercept)	CI		t(34)	p	R^2
VR	Relative	-9.16	(61.46)	3.23	(1.65)	-15.77	-2.55	-2.82	0.008	0.19
	Pointing	-1.67	(54.84)	2.65	(1.49)	-7.06	-3.73	-0.63	0.53	0.01
Map	Relative	-5.93	(65.62)	3.08	(1.58)	-12.24	0.37	-1.91	0.064	0.1
	Pointing	2.42	(44.64)	3.03	(1.66)	-3.58	8.42	0.82	0.42	0.02
Belt	Relative	-10.17	(61.92)	3.5	(1.8)	-17.32	-3.02	-2.89	0.007	0.2
	Pointing	2.95	(44.1)	2.01	(1.13)	-1.16	7.05	1.46	0.15	0.06
Unlimited										
VR	Relative	1.33	(52.66)	2.07	(1.17)	-2.87	5.53	0.64	0.52	0.01
	Pointing	3.08	(55.04)	2.54	(1.21)	-2.07	8.22	1.21	0.23	0.04
Map	Relative	-3.83	(67.21)	2.66	(1.48)	-9.22	1.56	-1.45	0.16	0.06
	Pointing	-0.5	(58.85)	2.45	(1.15)	-5.45	4.45	-0.2	0.84	0
Belt	Relative	-2.42	(59.92)	2.1	(1.17)	-6.71	1.87	-1.15	0.26	0.04
	Pointing	3.47	(50.74)	2.25	(1.07)	-1.1	8.03	1.55	0.13	0.07

Table 2. House Distance Regression Results.
Results of all simple linear regression models for the third session VR exploration condition, third session map exploration and second session VR with Belt exploration. Each model correlates trial distance of house pairs (prime and (correct) target) in the Relative or Pointing task and trial percentage of correct participants. The coefficient Distance is equal to the change in percentage of correct participants per 100 m increase in distance (except for Map, there it is 10cm). Intercept denotes the putative correct percentage for zero distance between house pairs. SE is the standard error of the estimate of the coefficient and CI the confidence interval. The degree of freedom is always 34 because there are 36 trials per task block.

As already mentioned during the analysis of participant behavior above, these results follow from an exploratory post-hoc analysis and thus are to be interpreted with caution. Also, an extrapolation of the linear fit leads to performance values eventually reaching zero percent which seems not plausible. An exponential decay model might describe the data more accurately. Because it is sufficient to show a distance effect and for simplicity of comparison, here we stick with the linear model.

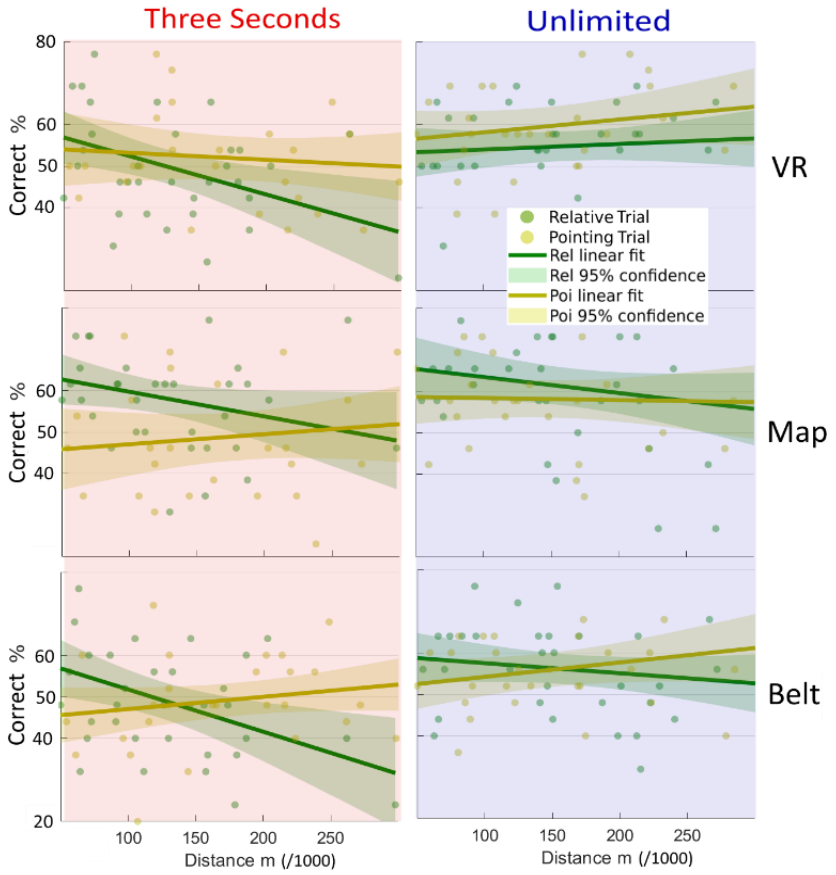


Figure 16. House Distance.
(Green dots) Percentage of correct participants for the Relative task trial with that distance between prime and correct target house. Left column depicts performance under time pressure, right column without. First row shows VR condition data of the third session. Second row shows third session, map condition and third row second session of VR with Belt. Distance is in meters for both VR experiments and in millimeters for Map. For easier comparison only distances from 50 m to 300 m are shown.
(Yellow dots) Similar for Pointing task trials with distance between prime and target house.
(Green line and area) Linear fit over performance and 95% confidence interval for the Relative task trials. (Yellow line and area) Linear fit over performance and 95% confidence interval for the Pointing task trials.

The similarity of the distance-time-pressure findings in all three experiments points towards a difference in the spatial representation accessed during the Relative task compared to the Pointing task.

III.2.3.5.3. Item Behavior Indicates Reproducibility

The second major potential confound we want to exclude are results resembling only a handful of learned houses, that is task items. Our results should generalize to the entire population of houses in our city, so we can expect them to be reproducible in other cities.

The behavioral variable, percentage of houses seen, discussed above and depicted in Figure 14 already provides evidence that participants explored large parts of our VR city. Such behavior speaks against participants only learning the same few houses. To double check this conclusion we simulated the decisions of participants in spatial task trials. Such a simulation is possible because if all participants had only learned the same few house stimuli, trials involving the unknown stimuli should follow a random Bernoulli process.

We focused on the first experimental session since it has the largest number of participants in each of the three exploration conditions, and thus the most weight in the respective linear regression analysis. For the simulation we chose the VR embodiment, Relative task and unlimited decision time results as an example. The simulation was, therefore, set up with 80 participants, each of which identified the same 36 houses. Here one house is one item or trial, i.e., it represents a set of houses. Each house was either classified correctly or incorrectly.

We simulated a Bernoulli process, once with probability $p = 0.5$ for 30 “unknown” houses and once with probability $p = 0.7$ for six “known” houses. The second time, we did the same with $p = 0.5$ for 18 houses and $p = 0.6$ for 18 houses. We iterated the simulation until the mean performance for both samples was identical to the VR example, that is 53%. We then created four histograms, two with as many bins as houses, i.e. 36, and two with as many bins as participants, i.e. 80. In the first house histogram, each bin holds the number of correct participants from the simulation with $p = 0.7$ for 6 houses and rest chance. The second house histogram holds the

number of correct participants from the simulation with p = 0.6 for 18 houses. Similarly for both 80-bin participant histograms, only here the bins contain the number of correctly identified houses. Furthermore, we added to each histogram a binomial distribution with p equal to the performance mean of the simulated sample, respectively.

The first row of histograms in Figure 17 shows that the performance of participants cannot be distinguished from a binomial distribution even if only a small fraction of houses is learned. The second row shows that resulting "house performances" can be distinguished. The lower left histogram can be identified as the sum of two binomial distributions, made up mostly of houses with chance performance and only a few that have been learned well. Consequently, if a large fraction of houses were learned in our experiment, the number of correctly identifying participants should follow a single binomial distribution, similar to the lower right panel of Figure 17.

To compare this prediction with our experimental data, we again create histograms with as many bins as number of participants. Each bin holds the number of trials, for which that many participants decided correctly. Note that each task trial consists of one stimulus set (e.g. three houses in the relative task), and that no two stimulus sets include the same items, that is houses. Like in the simulation above, we define house performance as the fraction of correct participants in that condition. Binomial distributions are overlaid on the histograms with p equal to the house performance. As illustrated in Figure 13 all resulting single binomial model distributions fit well to the sample distributions of both participant and item performance in the VR example for the Relative-task, unlimited-decision-time condition.

The (somewhat qualitative) comparison of simulation and real data further corroborates that participants do not only learn the same few houses. That is, despite the freedom given to participants during spatial navigation in the large VR city, our results should be invariant with respect to different stimuli and other towns of similar size.

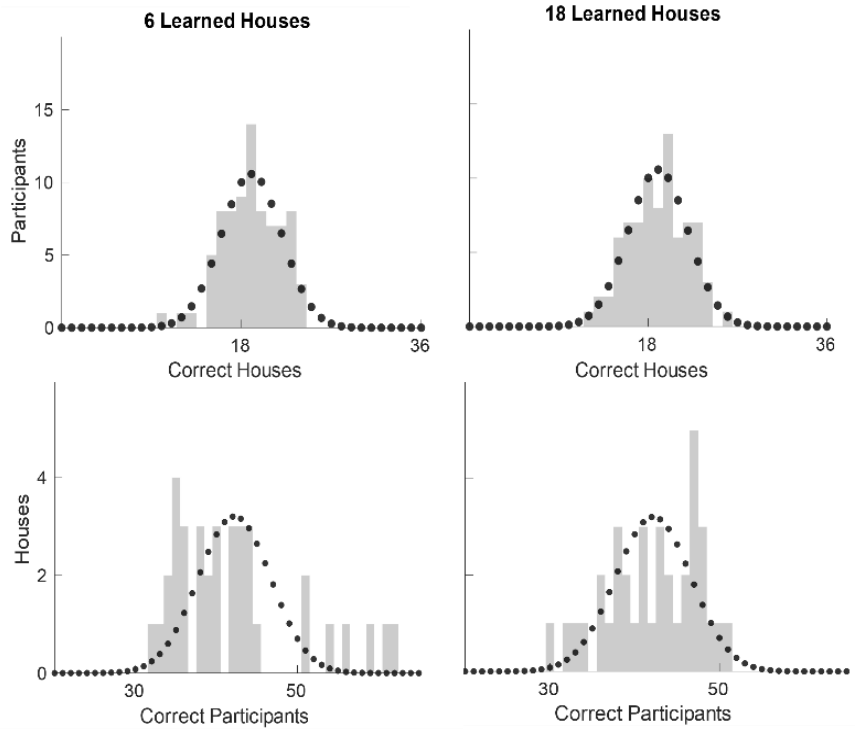


Figure 17. Performance Simulation.
(Gray bars) Histograms of simulated data with mean performance of 53%. The first column presents results if, of 36 houses, six have probability p=0.7 to be correctly identified by 80 participants and 30 houses have p=0.5, i.e. chance. The second column results if 18 houses have probability p=0.55 and 18 chance.
(Blacks dots) Binomial distribution models B(p,n) with p=0.53 and n=36 houses or n=80 participants.

III.2.4. Discussion

III.2.4.1. Spatial Cognition is Difficult. But Not for All.

As hypothesized, with accumulating exploration time average task accuracy rises above chance (H5). It nevertheless remains, in our view, surprisingly low. Comprehensive allocentric spatial representations of our VR city seem hard to acquire. In part, this might be due to a deprecated set of sensorimotor relations offered by our setup. Here the main factor is probably the lack of free walking (Chrastil & Warren, 2013; Ruddle et al., 2011) or at least the possibility to move forward through leaning forward with the upper body (Nguyen-Vo et al., 2019).

Another reason for the overall low task performance could be the size and complexity of the environment. After all, the city spans about 450 times 500 meters and consists of over 200 individual houses arranged at all possible angles along a road network which is not grid-like. The factor of size and complexity might also interact with the specific instruction to explore freely.

The lack of task-specific instructions might have led to additional variance in explorative behavior and performance. Some participants might have performed worse than they would have with more specific instructions. However, this potential variance is, in part, welcome. Since at least the Relative task was probably unfamiliar to most participants prior to this experiment, intentional instructions might have led to unnatural exploration strategies. We decided for a compromise to address this possibility. Participants were instructed to explore freely. Through the introductory part, however, they were also aware of the spatial tasks to be completed afterwards.

As noted by Chrastil and Warren (2011), spatial intention can come in varying degrees. In a study by van Asselen et al. (2005), for example, participants underperformed who were led from one room to another with the explanation that the first room was chosen by mistake. Burte and Montello (2017) did not claim a mistake, but rather asked their "incidental" group to pay attention to the architectural appearance and design of the neighborhood through which they were led. Their study does not reveal significant differences in spatial task performance between incidental and intentional learners. Given that our participants knew that they would be tested about spatial relations, we believe our compromise between natural behavior and spatial intention to be a sensible choice. More task-specific instructions might have led to higher performance, but at the cost of generalizability.

III.2.4.1.1. Individual Differences

Tests of behavioral variability during exploration reveal mixed results. Considering the percentage of houses seen, both VR embodiment conditions (one with belt) show very similar and strongly left-skewed distributions, which means that most houses were seen by all participants. Statistical analysis did not reveal an effect of the number of houses seen on performance.

For the map condition the percentage of house fronts seen/clicked makes a huge difference. A possible factor might be that only a deliberate click, that is a motor action in addition to just hovering with the mouse cursor over the house was necessary to see the house front.

Interestingly, we do not find a simple main effect of the number of clicks on average performance. Rather the difference in number of houses clicked informs differential development of the three spatial tasks over the three sessions in the map condition. The difference in clicking behavior might, thus, imply a difference in spatial strategy. A different spatial strategy could, in turn, imply a difference in spatial intention. Participants who clicked fewer times, might simply have been less motivated. Since a map is not a large-scale environment, we will discuss the findings in more detail below.

In the VR conditions the behavioral variable of movement speed might have been an indicator for spatial strategy and/or intention. Participants who moved more slowly through the city learned the Relative and Absolute tasks significantly better. Note, however, that the slow participants' average performance still only peaked around 60% after three exploration sessions. The performance is low even for slowly exploring candidates.

III.2.4.1.2. Embodied Roots of Spatial Cognition

Building a comprehensive overview over large environments is generally a difficult task. That becomes even more apparent in spatial judgements under time pressure. In our three embodiment conditions, only after map exploration did participants' judgements of spatial relations reach accuracy significantly above chance when asked to decide within three seconds. As predicted, participants generally needed to "take their time" to perform correct spatial judgements (H6).

The correlation can be interpreted as an argument against a close connection of bodily action and navigation skills. Consider, for example, competitions in robotics development, like the Amazon Picking Challenge (2017) or the ICRA Tidy Up My Room Challenge (2018). The embodied tasks tested in those challenges are of immense computational complexity, but easily completed by humans in a short span of time. Consequently, if knowledge about

allocentric spatial relations is grounded in bodily action, then why does it take so long to make use of it?

In recent years, it has been shown that spatial tasks result in neural activity which resembles the allocentric and egocentric activity during original perception of the spatial scene (Chadwick, Jolly, Amos, Hassabis, & Spiers, 2015; Marchette, Vass, Ryan, & Epstein, 2014; Vass & Epstein, 2017). Already in 2007, Byrne, Becker and Burgess developed a comprehensive model close to these findings. The model has been refined and extended by Bicanski and Burgess (2018). Remarkably, a hallmark of both versions is the importance of "mock" motor efferents for spatial reasoning (see also Thorndyke & Hayes-Roth, 1980). Possibly solving a task involving allocentric spatial representations takes so much longer because it involves a form of sensorimotor simulation or re-enactment. The importance of simulation in cognition has been argued before (Barsalou, 2009; Gallese & Goldman, 1998; Hesslow, 2002; Jeannerod, 2001). That allocentric information should be processed in such a way seems inefficient from a computational perspective and indicates sensorimotor relations to play a major role in the development of spatial cognition.

Also, even if spatial strategy and intention can explain variance in participants' performance, as long as these cognitive variables directly correlate with behavioral variables which constitute the act of spatial exploration itself, that still leaves room for an embodied explanation.

Admittedly, however, in the discussion so far we have left out an important behavioral result, namely large individual differences in spatial aptitude. Self-reports on spatial strategies assessed by the FRS questionnaire correlates with individual differences in performance. These differences occur in both time conditions and can range up to ten percent. It is not clear how an embodied approach can encompass these strong effects of self-perceived spatial ability, which have been found in many other studies as well (Burte & Montello 2017; He et al., 2019; Ishikawa & Montello, 2006; Weisberg et al., 2014; Weisberg & Newcombe, 2016). There seem to be cognitive processes involved in spatial reasoning which develop somewhat orthogonal to the embodied root.

Indeed, the findings concerning allocentric and egocentric simulation mentioned above focus mostly on how to transform the

allocentric position of the goal location into an egocentric code relating directly to limb position and vice versa. It is believed that a sequence of such processes plays an integral part in creating a movement trajectory towards the goal. Not much is stated, however, about how the allocentric representation of the goal location is recreated in the first place. This process seems too fast and localized to involve sensorimotor simulation (Pfeiffer & Foster, 2013). Indeed, modeling approaches expect attractor dynamics, shaped by the hippocampus' internal recurrent connectivity and activated by a given context, to suffice (Corneil & Gerstner, 2015; Gönner, Vitay, & Hamker, 2017). The models describe cognitive processes which rely on an internal connectivity structure for which the term embodiment seems misleading.

On the other hand, due to lack of statistical power, in our study we did not check for similarities between spatial questionnaire responses and exploration behavior. Some variance in spatial ability might directly relate to bodily movement patterns, after all. Eye- and general-body-motion tracking devices can allow future studies to elucidate this relationship. Modern machine learning approaches could also be used to compare the explanatory power of bodily behavior and mobile EEG in predicting spatial ability, strategy and task performance.

III.2.4.2. Egocentric Map Knowledge?

In both, the VR and the VR with belt conditions, participants can better judge how to point (or get) from one house to another than the orientation of a house relative to north or another house. These results are in line with our hypothesis that representations that allow us to find the shortest distance between two locations are important and that performance is favored by the familiarity of an action (H4).

Surprisingly, in the map condition, of all three tasks, pointing from house to house shows the worst performance in both time conditions. Note that this is not a main effect of experimental exploration condition on pointing performance. Although we find evidence that Pointing performance develops more slowly with map exploration than with VR, it is not significant. Rather, the interaction of exploration and tasks described above appears because map exploration boosts both Absolute and Relative performance while leaving the Pointing task unaffected. The

selective boost of performance becomes especially apparent under time pressure.

Pointing is one of the most prominent measures of spatial expertise in research on human cognition, but the literature remains inconclusive as to the difference between spatial knowledge acquired through studying a map versus full-body exploration. One study showed no main effect of embodiment condition on pointing accuracy (Richardson et al., 1999). Others find full-body experience to favor pointing (Shelton & McNamara, 2004; Thorndyke & Hayes-Roth, 1980). Again, others show the opposite effect (Fields & Shelton, 2006; Moser, 1988).

In a recent study Zhang, Zherdeva and Ekstrom (2014) investigate pointing accuracy after exploring a (desktop) VR city or it's map over multiple sessions. Importantly, they employ two different task setups when probing the ability to point to a house. Both setups have in common that participants have to rotate an arrow shown on a screen in front of them until it points to a certain building. However, in one setup, the screen only shows the arrow, the ground plane, and written instructions as to where participants are right now and where they should point towards. In the other setup, participants are first shown the VR city again, but without any task relevant houses. They walk through this version of the city until they feel "oriented." Then the pointing trials begin, this time only instructing them in which direction to point. Zhang et al. find that VR exploration facilitates accuracy in the latter "oriented" setup. Map exploration on the other hand facilitates accuracy in the ground-plane setup. They also find these differences to be transient effects. After sufficient exploration time, the differences vanish.

Our tasks are situated in between Zhang et al.'s oriented and ground-plane tasks. In addition, our interactive map allows participants to take on egocentric perspectives parallel to the VR ground plane. Consequently, the interaction Zhang et al. discovered might cancel a main effect of exploration condition on pointing accuracy in our study.

Their findings, however, cannot account for the boost of performance in the Absolute and Relative task after map exploration. To understand this difference between spatial knowledge acquired from a map versus spatial knowledge acquired

in VR, we will first try to explain why map exploration should boost task performance at all. For this, we present two (not mutually exclusive) explanations. The first explanation emphasizes the difference in computational complexity between both embodiment conditions. The load of information an agent must process during full-body immersive navigation in order to be able to later solve the three spatial tasks is massively downscaled in the map condition. Indeed, that humans show comparable performance after VR and map exploration exemplifies the degree to which our cognitive structures are specialized to process sensorimotor information efficiently.

The second explanation for the general improvement in performance after map exploration relative to VR exploration emphasizes that map participants can perceive a large part of the spatial relations necessary to solve the tasks from one point of view, i.e. the map view. Allocentric and egocentric representations could, therefore, inform each other, a potential factor which might especially favor performance under time pressure. Such an interaction of both types of reference frames confounds analysis of allocentric spatial knowledge.

The well-documented alignment effect speaks for the existence of this confounding factor. Learning spatial relations via map leads participants to form representations which allow them to better recall spatial relations aligned to the axes defined by the map (Evans & Pezdek, 1980; Richardson et al., 1999; Shepard & Hurwitz, 1984). Intuitively, this reads like a prime example of an allocentric frame of reference. Note, however, that the vertical north-south axis of the map aligns perfectly with the vertical bodily axis going from head to toe through the participant. Therefore, the effect can also be, at least in part, due to the alignment of real- and imagined egocentric heading direction while sitting in front of the map and facing a stimulus house that is oriented downwards on the north-up map (Shepard & Hurwitz, 1984; Richardson, et al., 1999).

Our hypothesis that map exploration should favor the Absolute task (H2) was partly informed by the alignment effect as well. Although we do not find such a general improvement, as reported elsewhere, we do find participants to excel in the Absolute task if the orientation of house and north are aligned (König et al., 2019). Note that alignment of a spatial reference system to, for example

initial heading, has also been reported after full-body exploration of a large new environment (He et al., 2018; Shelton & McNamara, 2004). Thus, the map alignment effect and allocentric representation do not necessarily exclude one another.

However, while both the lower computational complexity as well as the additional egocentric reference frame can explain a boost in average performance, only the egocentric approach can account for its selectivity, as we shall see below.

III.2.4.2.1. Visual Memory from One Perspective

So far, it remains unresolved that performance in the Absolute and Relative task excels after map exploration, while Pointing accuracy is not affected relative to task performance after VR exploration. To understand this effect, we first need to acknowledge that Pointing is fundamentally different from both the Absolute and the Relative task in our study. Only pointing necessitates knowledge of each house's location. This is of special importance for the map condition.

The orientation of a house relative to north or relative to another house can be learned in a direct association. During map exploration, participants can associate a VR house picture (later task stimulus) with its rectangular symbol on the map. They, then, only need to recall this rectangle's orientation in order to solve the Absolute task. They do not need to compare the orientation with the map because, as mentioned before, their bodily orientation suffices. North is up. Similarly, for the Relative task, participants can mentally compare two rectangles' orientations by accessing the single-point-of-view representation of the map two times. The Absolute and Relative tasks are largely solvable by accessing one and the same egocentric frame of reference multiple times. Shelton and McNamara (2004, p. 168) corroborate this idea: "the visual overlap for same-perspective and same-orientation images may allow participants to appeal to only the visual representation and base the judgment on image matching." In contrast to the VR and the VR with Belt experiments, no further spatial reasoning involving mental rotations is necessary.

When pointing from house to house, however, recall of rectangles' orientations becomes insufficient. Participants then need to represent the orientation of the prime house rectangle and both

the prime- and target-house symbols' locations in order to infer the correct pointing direction. Thus, the image-matching mechanism proposed above mostly favors judgements of angles among house rectangles relative to the reference direction defined by the image itself, especially under time pressure.

As a sidenote, qualitative comparison of the sketch maps drawn after VR or map exploration shows the latter to be much more accurate, a finding which fits well with the idea of an egocentric-visual-experience representation.

The behavioral variable "percentage of houses seen" reveals two encoding subgroups. People that clicked on more (different) houses were up to 30% better at answering the spatial tasks than those that clicked on fewer, although both clicked on enough houses to perform comparably. Furthermore, the more-houses group shows the differential boost of performance for Relative and Absolute task in both time conditions. The less-houses group shows no such boost under time pressure or without. Interestingly, this difference rather hints at Absolute outperforming the Relative task with consecutive exploration sessions like in the VR experiment, given unlimited decision time.

If we interpret the interaction of Absolute and Relative task performance with exploration experience as evidence for an allocentric representation, such a representation is consequently used by low-performing participants. The high-performance, more-houses group seems to have relied more strongly on an egocentric representation, which could also explain their ability to transform their spatial knowledge into action so quickly.

The differentiation into dominant forms of encoding might be due to individual differences in spatial intention and/or spatial strategy. If we associate the number of clicks with spatial intention, participants who clicked on fewer houses might simply have paid less attention to the spatial relations on the map. Therefore, the findings indicate the development of an allocentric frame of reference to be a more automatic process. We will go into more detail on the cognitive processes underlying the build-up of an allocentric frame of reference during map learning below.

For now, map learning shows evidence for a strong effect of egocentric representation, or at least a different kind of allocentric

representation than acquired in large environments. This effect should be considered in the interpretation of experiments analyzing knowledge acquisition through maps. Finally, we want to emphasize that a group distinction is an oversimplification. The variable of percentage of houses seen shows a continuous unimodal distribution. In reality, there is an interaction spectrum between both kinds of representations.

III.2.4.3. Different Representations for Different Actions

After VR exploration, three seconds of decision time is barely enough for our average participant to reach chance performance when judging house-to-north or house-to-house alignment (note that we counted no answer as a wrong answer). In other words, after three consecutive exploration sessions, which add up to 90 minutes in our VR city, the acquired allocentric spatial representations do not yet reliably inform intuitive decisions for most participants. Some, in particular those with high self-assessed survey or cardinal knowledge, however, have already formed representational structures which allow quick computation leading to action.

Averaging all VR (without belt) participants in the unlimited decision time condition, we find a significant interaction between task performance and overall exploration time. With little spatial experience, Relative performance is better than Absolute, and this difference reverses with more experience. Including behavioral variables strengthens this interaction. For example, people that move more slowly through the VR city show more improvement, both in the Relative task and the Absolute task, such that the interaction remains stable.

As a first explanation for a difference in task performance one should always consider the differences in task setup. Indeed, a participant needs to recall at least two houses to solve the Relative task, but only one to solve the Absolute task. Hence, the Relative task might be more difficult. On the other hand, it is not clear why the probability to be able to recall the direction towards north should be higher than the probability to have seen and memorized the orientation of the second house. In any case, the explanation can be dismissed because, as stated above, with little spatial experience participants perform better in the Relative task than in the Absolute task.

An explanation for low Absolute task performance after the first VR session could be that the participants simply did not find north. Only starting in the second session were they able to infer the cardinal direction of north during one virtual day using the movement of the sun and shadows alone. This would explain the exceptionally strong increase in performance from the first to the second session in the Absolute task.

However, this “overload” claim seems insufficient to explain why we find the same apparent jump from the first to the second session after map exploration. The rectangular map is north-up and at all times fully visible to the participant. We, therefore, do not believe the possibly confounding factor to be decisive for the differential task development in the VR condition. However, because performance development in most tasks and across all exploration conditions seemed to best fit a linear model, we did not focus our limited statistical power on the apparent non-linearity in the Absolute task. This effect remains to be elucidated by future research.

Also, if the interaction of house-to-north and house-to-house accuracy with exploration session were largely due to difficulties in finding north, the differential development should disappear for participants with high cardinal self-esteem. This is not the case. As mentioned above, including the behavioral variable of self-reported cardinal knowledge in the model heightens the probability for the interaction of Absolute and Relative performance over sessions.

Another argument against “overload” comes from the fact that extrapolating our results leads to similar findings as reported in the pilot study with residents of the city of Osnabrück (König et al., 2019). They found a reversal of task accuracy with decision time. Under time pressure, participants were better at judging house-to-house orientations, but with more decision time they excelled when aligning houses and north. We expected to find the interaction with decision time of the pilot study in the VR condition as well (H1). We cannot reproduce the Osnabrück study’s results exactly. However, given that finding north during exploration was not too difficult for the participants and the development of the Absolute task can be approximated as linear, we can extrapolate our interaction of task and session. The Absolute task would then

eventually outperform the Relative task, when deciding without time pressure, as expected.

One possible explanation for the need of extrapolation is that, while our study observed the developing structure of allocentric knowledge in a completely unfamiliar city, the pilot study probed the structure of spatial knowledge in a familiar environment. Pilot study participants had resided in the city of Osnabrück for at least a year, meaning that their spatial knowledge of the city was significantly further developed. As it stands, our study complements the pilot study's findings since our focus lies on the development of allocentric representations.

Relating our results to the Osnabrück study has to be done with caution. That is because we cannot exclude that the high performance in the Osnabrück-north-pointing task without time pressure was due to participants having seen maps of the town beforehand. Exploration with the help of a map can lead to representations aligned with the map layout (Meilinger, Frankenstein, Watanabe, Bülthoff, & Hölscher, 2015).

For the possible confound speaks the fact that participants in the Osnabrück study performed best when house orientation and south were aligned, that means, in trials where participants were oriented towards north. Similarly, Frankenstein et al. (2012) found that participants performed best when prompted to point towards a goal location which lay in the direction of north in their hometown.

On the other hand, it is not clear why such map knowledge should not also lead to better performance when comparing house and north under time pressure. The possibly confounding factor cannot explain the interaction of Absolute and Relative task with decision time. Consequently, while map knowledge might play a role in the Osnabrück study's findings, there remains a significant amount of variance which asks for a different explanation.

We argue that both the interaction of Absolute and Relative performance with exploration time, as well as their interaction with reaction time in the pilot study, are caused by the difference in the two tasks' relation to embodied action. Both tasks are equivalent in that they ask for angular differences between directions of orientation only. Knowledge about relative location is irrelevant,

and the action-relation of angle is rotation. The difference between the two tasks is that of house orientation and cardinal direction. What makes the difference significant is that these spatial concepts afford different actions.

III.2.4.3.1. Binary versus Survey Code

Cardinal direction is important when traversing long distances in a straight line. It allows the navigator to avoid obstacles while keeping the general direction of movement constant. As long as a landmark, like a large building, remains visible during travel, it can be used in the same manner.

When the navigator is situated within an environment where landmarks are themselves part of visually separated vista spaces, their role changes (Lynch, 1960). These landmarks connect routes (Siegel & White, 1975; Montello, 1998; Chrastil, 2013). They are the start and goal locations of a journey, but they can also become obstacles on the way. Furthermore, recognizable buildings along the way can help the navigator turn at the right moment. Consequently, such local landmarks become decision points on the navigator's journey, often associated with a change of direction, which implies whole-body rotation.

It might be ecological to code the relative orientation of these "nodes of the decision system" (Siegel & White, 1975, p. 29) in direct relation to all other already memorized landmarks (König et al., 2019). If the representational structure would be constituted by direct linkages between pairs of landmarks, that is pairs of houses, it could be learned and accessed quickly. We call this kind of representational structure a binary code. We will further refer to the landmarks the code relates to each other as nodes.

However, while the ecological value of survey knowledge (finding the shortest route) increases with the scale of the environment, information about relative orientation of landmark nodes becomes less important. In line with this reasoning, we find that when participants have to judge the relative orientation in three seconds, their performance decreases with increasing node-pair distance. The effect does not appear with unlimited decision time. The binary representational structure seems to be preferably accessed when quick judgement of rotation among local nodes is needed. The finding is in line with the Osnabrück study's results which revealed

Relative to outperform Absolute under time pressure (König et al., 2019).

Indeed, that connection strength should decrease as a function of distance seems a necessary property of the all-to-all linkage form of the representation. Otherwise, given any number of nodes N, it would allow access to the relative orientation of all N-squared node pairs (divided by two because the relations are symmetric and subtract N because each node is identical to itself). In other words, without a mechanism to weaken links between node pairs a linear increase in node number leads to an exponential increase in the information carried by the binary representation.

The defining characteristic of the cognitive map, deduction of short paths, implies knowledge of the relative location of the starting node and the goal node. That is knowledge of allocentric node location. We can refer to the underlying representation as survey code.

Remarkably, pointing from node to node does not reveal a qualitative tendency for a decrease of accuracy with increasing distance. The invariance speaks for survey code to be structured differently than representation of allocentric node orientation, that is binary code.

Weisberg et al. (2014) have shown that pointing accuracy does depend on distance, even with reasoning time. In their experiment, if participants point from a building in one neighborhood to a building in another neighborhood, pointing accuracy is lower than when participants point in one neighborhood from house to house. They identify participants who do not show the distance dependence as good spatial integrators. We posit that, without time pressure, no distance dependence only holds for well-integrated environments. The explanation also fits the lack of clear distance effects in hometown pointing studies (Frankenstein et al., 2012; König et al., 2019).

If we point to a node which lies in a certain direction, e.g. north, the further away the node is located, the closer our pointing direction becomes to north. Knowledge of cardinal direction can, thus, be considered the limit case of being able to point to a node as far away as possible. That means it enables us to reach the node on the shortest (in principle) possible path. Consequently,

knowledge of cardinal direction also depends on a survey-like representation.

Finally, as mentioned above, without time pressure we do not find an effect of distance on judgment of relative node orientation. Hence, the different encodings need not be strictly separate. The survey code might hold some information about the relative orientation of houses as well. Alternatively, given enough reasoning time survey- and binary code may be used together in order to solve a spatial task.

A possible test for the distinction of binary- and survey code on the neurophysiological level could be done as follows. There is ample evidence from experiments with rats that changes in local environmental cues lead to different changes in encoding than deforming the geometric shape of the environment (for a recent review, see Latuske, Kornienko, Kohler, & Allen, 2018). If we identify place cells as part of a survey code, then our above analysis predicts only a small effect on these cells when manipulating the relative orientation of landmarks inside the arena. We do not expect a complete remapping. Maybe the change in landmark orientation would have a greater effect on firing-rate coding (Leutgeb, C., Leutgeb, J.K., Barnes, Moser, E.L., McNaughton, & Moser, M. B., 2005). If a change in the location of landmarks would have a similar effect as the change in orientation our proposed distinction might be mistaken.

Were it not for the Absolute task, participants might not have learned the direction of north in our VR city. Nevertheless, there is evidence that survey knowledge of large environments is aggregated along (at least) one global reference direction (Shelton & McNamara, 2004; Meilinger, Riecke, & Bülthoff, 2014; He et al., 2018). Given sufficient exploration experience, alignment of node orientation and such a reference direction should become part of survey code. Admittedly, however, environments are nested in larger environments and we cannot assume one all-encompassing survey representation (Meilinger,2008; Wolbers & Wiener, 2014. Thus, the term “global” for a reference direction of any one environment might be misleading.

In our VR condition (also with belt), we do not find alignment of house orientation and the cardinal direction of north to significantly improve task performance. The reference direction,

however, often coincides with the first perspective from which one enters a new environment (Shelton & McNamara, 2004; He et al., 2018). Since our participants always started from the same (central) position in the city facing east, we also tested that cardinal direction for alignment effects, again without significant results. It is possible that knowledge about the cardinal direction of north and initial starting orientation interfered destructively. In order to firmly establish the effect of a reference direction in future experiments, both task direction and starting orientation should coincide. Alternatively, one of the two should be randomized.

Even if one were to accept the distinction between binary- and survey code outlined above, one might ask if it is indeed the difference in bodily action relating to house orientation and cardinal direction of north respectively which cause the different forms of representation. The pilot study revealed an additional argument for the distinction in coding being due to the actions associated with the spatial concept coded. Recall that the Osnabrück researchers found the Relative task to outperform the Absolute task under time pressure. However, when judging the direction of streets relative to one another or aligned to north, participants were quicker in the north task (König et al., 2019). This finding speaks for the orientation of streets being *primarily* encoded in a structure that develops along a reference direction, that is survey code. Indeed, while streets can also serve as landmarks, they afford different actions than the kind of landmarks we considered above as nodes. Streets do not afford turning. One does not change direction on a street. On the contrary, streets afford long, straight-line movement. Therefore, the example is further evidence for the representational structure of a spatial concept depending on the embodied action it affords.

Again, however, corroborating our hypothesis by referring to the pilot study's results has to be done with caution because these are possibly confounded by map-knowledge. That is because streets make up prominent features in maps. Since maps are oriented toward north, three seconds might have been sufficient for participants to tap into that knowledge when relating streets of their hometown.

In conclusion, notwithstanding the possible confounding factors, our embodied interpretation can explain a considerable amount of

the combined variance of both the Osnabrück and the VR exploration results. We will see in the following sections how it can also relate to the map and especially to the VR with belt results.

We are not the first to argue that spatial representations are to a large extent grounded in action possibilities. Greene and Olivia (2009) have shown that cognitive models classifying natural scenes according to global and action-related properties like depth or navigability are better than classifiers attending to local geometric scene properties. Computational models for human navigation consistently reveal the importance of pseudo-motor signals during spatial reasoning (Byrne, Becker, & Burgess, 2007; Bicanski & Burgess, 2018). Physiological evidence for the coding of navigational affordances in the human visual system has recently been put forward by Bonner and Epstein (2017). They conclude that there exists a bottom-up mechanism for perceiving potential paths for movement in one's immediate surroundings.

III.2.4.3.2. Similarities between VR and Map Exploration

We argued above that map participants who click on more houses rely on a different representational system than those who click on fewer houses. While this is a continuous distinction, not an either-or, fewer-click participants show the same interaction of Absolute and Relative task with session as we find after VR exploration. Similar development of spatial-task performance might result from similar allocentric-representation build during VR and map exploration. Consequently, we have reason to believe that participants who clicked on fewer houses relied more heavily on allocentric processing.

Intuitively, this similar development in both embodiment conditions does not fit with the affordance theory. Exploration of the VR city via map does not allow the participant to (visually) move her body forward, following a straight line, or to stop and turn. The participant is sitting, looking at the computer screens in front of her depicting the map. However, exploring the map allows for other forms of movement.

We are not so interested in the hand movements that control the mouse cursor for house-picture selection, but rather in the movements of the eyes. The participants are sitting at a distance of

about 80 cm from the two monitors on which the map is presented. Visual acuity declines exponentially outside the 2° solid angle dictated through the geometry of the retina and eyeball. That makes for an area with high visual acuity of about 3cm diameter. Hence, at any moment in time, the participant can only focus on a couple of houses (rectangles) simultaneously. Likewise, if the participant were to explore the city in 3D, house walls would obstruct vision in the ground plane, allowing explorers only to perceive parts of their environment at a time. In that way, visual, 2D-city-map exploration is somewhat similar to full-body movement through a 3D city.

Evidence for similarity between navigation in the real world and navigation on maps also comes from an eye-tracking study by Ooms, De Maeyer and Fack (2013). Both expert and novice map users fixated mostly on approximately linear reference structures like roads and rivers (their maps were too large to make out single houses). Furthermore, Timm and Papenmeier (2019) found eye movements to affect working memory for spatial configurations. While we did not track eye movements in the map condition, we find that participants clicked on most of the 200 houses multiple times during each session. The resulting distribution of clicked houses is barely distinguishable from the distribution of houses viewed during VR exploration.

We argue that houses on the 2D map plane can become landmarks of the visual map exploration process. Similar to the VR experience, or navigation in the real world, these landmarks become nodes of decision. The participant's gaze wanders from one node to another node, stops there, changes direction and eventually comes to a halt at the next "interesting" node. Also, similar to navigating in the VR city, the concept of cardinal direction can be associated with straight-line movement. On its way north, the participant's gaze wanders upwards on the map, passing multiple landmarks on its way. Also, if the participant recalls a certain node she wants to revisit, she will rotate her eyeballs such that her gaze will follow a straight line ending at the goal node. As if she were to point from one node to another stretching her arm, her gaze will move along the shortest path. In conclusion, exploring a city map via eye movements could lead to both binary- and survey code.

Neurophysiological evidence for survey code building during visual exploration of images has been found by Killian, Jutras and Buffalo in 2012. They found cells which increased their firing rate whenever the gaze of a monkey moved over locations on the two-dimensional image plane, which would, taken together, make up a regular hexagonal grid-like pattern. Similar grid cells are found in humans and other animals during spatial navigation in natural (3D) environments (Doeller, Barry, & Burgess, 2010; Hafting, Fyhn, Molden, Moser, M. B., & Moser, E. I., 2005).

In addition to the egocentric representation posited before, which allows simple image matching to solve certain spatial tasks, the limited angle of visual acuity at any one moment during map exploration might make it ecological to represent spatial relationships among map symbols in an allocentric frame of reference as well. Parallel development of egocentric and allocentric representations is well documented (Burgess, 2007; Gramann, 2013). However, research systematically investigating such a differentiation during map learning is sparse (Zhang et al., 2014; Zhang, Copara, & Ekstrom, 2012). For future experiments, the relative strength of the development of grid cells during visual exploration might be an indicator for a preferred frame of reference.

III.2.4.4. Augmentation versus Representation

Without time pressure, as expected, participants in the VR with belt exploration condition perform significantly better in the Relative task than in the Absolute task (H3). The sensory augmentation device indicates north via vibrotactile stimulation. Therefore, naturally, when asked to turn towards north at the end of the VR and the VR with belt exploration session, respectively, participants were much more accurate with the belt. Why then, should we expect performance to increase when participants judge house-to-house alignment, rather than when they judge the relative orientation of house front and cardinal direction, after exploration?

Prior research investigating changes in spatial perception and navigation strategies in long-term belt wearers reported “automated navigation without mental reflection” (Kaspar et al., 2014). The additional sensorimotor loop provides constant information about cardinal direction in an egocentric reference

frame. It, thereby, makes it unnecessary to represent north in a cognitive map. Indeed, building survey code itself becomes less ecological, at least in its former structure. That is because the belt enables the wearer to move in a straight line in any direction with minimal cognitive effort.

Neurophysiological findings corroborate the claim that the belt impedes development of long-term survey knowledge. The hippocampus is a brain region important for allocentric spatial memory recall as evidenced, for example, by Maguire et al. (1998). In their study the hippocampus shows larger activity when participants have to mentally deduce a new route due to obstacles appearing in an otherwise familiar environment. That is compared to neuronal activity in the hippocampus when participants can simply follow a known path. Sabine König and colleagues found reduced activity in the hippocampus in participants who had trained with the feelSpace belt over the course of seven weeks (König, et al., 2016).

Note that, if cardinal direction is part of survey code, so is relative location. Therefore, lower performance in the Absolute task indicates lower performance in the Pointing task as well. We did not perform a statistical analysis because our original hypothesis focused on the effect of the belt comparing Relative and Absolute task. However, at least qualitatively, Pointing performance for belt wearers is worse than for those participants who explored the VR city without belt, as expected.

In addition to the decline in quality of survey code we expected the information provided by the belt to enhance the binary representation linking house-nodes. The binary representation, as posited above, is the primary form of representation for relative house orientations. We hypothesized that the additional vibrotactile experience of the direction of north while a belt-participant encountered a new landmark would add to the strength of the binary linkages. However, while we did find the expected effect of Relative outperforming Absolute, Relative task performance was not better than after VR exploration without belt.

Unconscious integration of the information provided by the new sensorimotor loop into a spatial representation that can be recalled in the absence of that which it represents seems not effective after 90 minutes of wearing time. Participants might have had to

consciously map the vibration around their waist to the concept of north in order to make use of it. Some direct form of understanding of the lawful change in neural activity caused by the interaction of VR north, the belt, and the participant's body was missing. The augmentation belt had not developed into a truly a new sense, yet.

That sensorimotor relations have to be learned before a sense can be fully used as such is well known (Bach-y-Rita, 1972; O'Regan & Noë, 2001; von Senden, 1960). Indeed, only long-term belt wearers reported changes in spatial perception (Kaspar et al., 2014). Also neural signatures indicating changes in sensorimotor processing in belt wearers have been recorded over several weeks (König et al., 2016).

The lack of belt training somewhat impedes our results. On the other hand, it allowed us to collect a sample large enough to investigate subgroups of spatial learning. Distance from which houses were looked at and self-reporting on knowledge of cardinal direction both correlate with the speed at which belt participants were able to acquire spatial knowledge. For participants with high cardinal self-esteem, the belt boosts their average performance by ten percent compared to three percent for cardinal participants without belt. It indicates a sort of integration has already taken place in some participants even after only three training sessions. Remarkably, the effect only appears when prompted under time pressure. The result might, therefore, hint at a representational structure different from both survey and binary code. Alternatively, the newly exploitable sensorimotor relation might lead to a structural change in one or both forms of representation.

So far, the VR with belt embodiment condition indicates that new sensorimotor relations can lead to profound changes in the development of allocentric spatial representations by changing their ecological value. Also, that judgments of the relative orientation of buildings are not affected by the belt, is further evidence for the difference between the encoding of house orientation and cardinal direction as posited above. That is evidence of the existence of two forms of allocentric representations, namely binary and survey code.

Schumann and O'Regan (2017) have presented a sensory augmentation device indicating the direction of north via auditory signal. After manipulating the device to indicate north too early or

too late during self-rotation, they observed long-term recalibration in vestibular rotation judgements. That is evidence for fast integration of the form of spatial information provided by the new sensorimotor loop. Comparing the effects of the unmanipulated device on the formation of allocentric knowledge would be a possible test for the ideas presented here.

III.2.4.5. Development of Allocentric Spatial Knowledge

Knowledge about spatial relations among objects in a large environment is believed to build through memorizing landmarks, learning routes, and connecting those to attain survey knowledge (McNaughton et al., 2006; Siegel & White, 1975). However, the extent to which these constitute three separate processes, and whether these are the only processes involved, remains unclear (Chrastil, 2013; Ishikawa & Montello, 2006; Montello, 1998).

Montello for example, emphasizes the continuous and parallel acquisition of all three kinds of knowledge (Ishikawa & Montello, 2006; Montello, 1998). Chrastil (2013) adds the formation of topological graph knowledge as a process on the way from simple route to metric survey knowledge (see also Poucet, 1993). She further argues that there does not exist a clear one-to-one mapping between neural activity and the purported milestones of spatial knowledge acquisition. Rather, there exists a multitude of sub processes the interaction of which depends on higher cognitive and/or environmental factors.

The idea that graph knowledge is part of a developmental process towards survey knowledge is debated. Some authors believe the idea of a cognitive map, which generally stands for survey knowledge, should be completely substituted by cognitive graphs (Meilinger, 2008; Warren, Rothman, Schnapp, & Ericson, 2017; Warren, 2019). A node in such a graph is made up of a salient landmark. To get from one node to the other, we need to recall the visual and/or motor transformations experienced while travelling between both. This sort of route knowledge accessible at a certain node is called a label and makes up the edge connecting two nodes. Survey knowledge then simply consists of accumulating new nodes and edges while, at the same time, strengthening the ability to access well-known and important ones (Warren, 2019).

Even in such a labeled cognitive graph, however, the nodes need to hold some kind of local reference frame relative to which the edges can be defined. The landmarks have to possess an orientation relative to which the label of rotation is defined, for example. That is because an angle only exists in relation to a reference direction.

If the node itself makes up a local reference frame, we find ourselves facing a cognitive map again, only a smaller version than before. On the other hand, if the local reference frame is itself a labelled graph then each node in the smaller graph, again, needs its own local reference frame relative to which something can rotate, and so on.

Tobias Meilinger (2008) proposes a synthesis of cognitive graph and cognitive map to account for survey knowledge. He hypothesizes that a node does not simply represent a landmark. Rather, each node represents the landmark embedded in part of its surrounding area, i.e. the local vista space. He furthermore claims that grid cells, the example par excellence of cognitive map proponents, only extend over such vista spaces, or nodes. Indeed, opaque borders have been shown to lead to compartmentalization of grids in humans and other animals (Derdikman, Whitlock, Tsao, Fyhn, Hafting, Moser, M. B., & Moser, E.I., 2009; He & Brown, 2019).

The issue of cognitive map versus graph remains debated. We believe, however, that most authors would agree with Chrastil's emphasis on the importance of sub-processes, or cognitive processes in between the acquisition of landmark, route, and survey representations. Our findings corroborate her emphasis and indicate a major role for the different actions associated with, or afforded by, the spatial concepts represented. We, thereby, complement past research on the effect of bodily cues on spatial integration (Chrastil & Warren, 2013; Nguyen-Vo et al., 2019; Ruddle et al., 2011; Waller et al., 2004).

A factor which demands clarification is the individual difference in spatial aptitude. How does it come about that some people are consistently so much better at integrating spatial knowledge than others (Ishikawa & Montello, 2006; Weisberg et al., 2014; Burte & Montello, 2017; He et al., 2019)? These findings point towards differences in neural structure which cannot be readily explained by sensorimotor effects alone.

Finally, the question arises how much of the processes involved in developing survey knowledge of large-scale environments can be mapped to developing knowledge in other domains. Finding the shortest route towards a goal in an environment which cannot be "overlooked" in its entirety shows interesting parallels to general problem solving (Siegel & White, 1975; Tolman, 1948). Furthermore, recent experimental findings have found parallels between navigating environmental- and mental space relating to the activity of grid cells (Bellmund et al., 2018; Buzsáki & Moser, 2013; Constantinescu et al., 2016; Epstein et al., 2017; Schiller et al., 2015); Thus, spatial cognition is a promising research direction to disentangle sensorimotor effects on representational structure from other, complementary factors, e.g. genetically predetermined, neurophysiological boundary conditions.

III.2.5. Conclusion

Our study investigates the relationship of embodied enactivism and cognitive representation. We focus on the development of allocentric spatial knowledge. The representations underlying this knowledge are particularly interesting because they represent relations among objects which are not part of the human body.

Low performance in judging allocentric spatial relations even after repeated exploration of a VR city indicates that spatial cognition involving large environments is difficult. Further evidence comes from the fact that these judgments require a considerable amount of time. Other research indicates cognitive simulation to play an important part in spatial reasoning. We conclude that the difficulties posed by spatial cognition in large environments are due to its rootedness in bodily action.

On the other hand, independent behavioral variables allow us to distinguish individual participants who excel at spatial reasoning. Their outstanding spatial skills imply orthogonal cognitive structures supplementing the embodied processes under investigation.

Map learning boosts performance both with and without time pressure, but only in some spatial tasks. The differential boost suggests an effect of embodiment during exploration on the structure of the acquired knowledge. We illustrate that the effect is largely due to an egocentric frame of reference, or at least an

allocentric representational structure different from that attained in large environments. Individual differences, again, allow for a more differentiated picture, providing evidence that some participants do acquire allocentric knowledge similar to that acquired in large environments.

Reference direction and relative location of landmarks in a large environment, which cannot be overlooked from any one location, can still be learned to greater precision than relative orientation of landmarks. This is due to the former two allocentric spatial relations being more important to move on the shortest path to a goal than the latter.

The contrast in representational quality also resembles the different affordances of the spatial concepts represented. Landmark orientation affords rotation, that is change of direction. Allocentric reference direction and landmark location allow for straight-line travel.

Landmark orientation is possibly integrated in the same representational structure as reference direction and location. A survey code also known as cognitive map. The primary form of encoding of landmark orientation, however, is a binary structure of landmark pairs. The encoding strength decreases with distance between landmarks.

Participants equipped with a sensory augmentation device which provides constant information about reference direction while they explore the VR city, develop impaired survey knowledge. In the spatial tasks after the exploration, relative orientation of landmarks is judged with greater precision than alignment of landmark and reference direction.

Sensory augmentation can transform an allocentric spatial concept, like reference direction, into an egocentric experience. The transformation does not affect the ecological value the spatial concept affords for navigation, namely, to quickly move to the goal location. It does, however, make it less ecological to build allocentric knowledge which allows participants to reason about the spatial concept in the absence of its environmental counterpart. Consequently, cognitive resources are shifted to other processes, leaving the participant with impaired survey knowledge if she disconnects from the sensory augmentation device.

The shift in cognitive resources does not affect judgment of relative landmark orientation and thus corroborates the claim of the two forms of allocentric spatial representations.

Our findings illustrate that, even representations of spatial relations among objects which are not part of the human body, have a large part of their roots in bodily action. However, exemplified by the large individual differences in spatial aptitude which do not seem to stem from bodily differences, we cannot fully close the gap between allocentric spatial cognition and embodiment.

As a closing statement, we would like to stress the importance of testing cognition in the wild. Our brains are part of a body which constantly interacts with the environment. Closing it off from most of these interactions reduces variance in our results, but it might also hide more general principles of cognition.

Epilogue

What does it mean to be lost?

It means not knowing how to act.

In our everyday lives we seldom really get lost. Nevertheless, most humans have experienced a moment in which they were not sure what to do next. Such a situation can become especially pressing in a large, unfamiliar environment. Remarkably, at the moment directly after we have decided upon the next action, or so I claim, we do not feel lost anymore. Action makes sense.

I have argued that action comes after representation; therefore, in order for sensible bodily action to arise, the goal first has to be represented in my brain. We also have seen that in between representation and action there lies computation. Sometimes the goal is located in a large environment containing buildings and walls we cannot see- or pass through. In such cases that demand complex navigation, the computation leading to the next action can take a long time.

If, however, the goal lies within our perceptual space, meaning that it is not shielded by opaque borders, the reasoning process seems to magically reduce to “simply” moving towards the represented. Our brain seems to forget that it still needs to navigate through the myriad of possible motor processes in order to decide for the one which will bring it closer to the goal.

I discussed above that it is remarkably hard to describe the effect a lesion of a sensory modality has without making a tautological statement. The effect of the belt is similar. What does the north belt do? It points towards north. However, it also enormously widens our perceptual space, thereby allowing us “simply” to move on the shortest path in any direction. On the other hand, if the belt vibrates north and we hear a car coming from that direction, we can “simply” evade the car. The representation of north and the representation of the car together create a new boundary surface. We find, given our bodily shape and action possibilities, the optimal trajectory on the new surface without further thought. Human perception seems to have evolved for quick and flexible bodily action. Again, only when the goal location is outside of our perceptual space are we in need of mental representation crunching.

I provided evidence that even large-scale spatial reasoning relies partly on cognitive processes which developed to instigate local bodily action. It also became clear however, that there remains variance, the explanation for which cannot be found in bodily structures alone. More precisely, there remains variance which depends on the structure of the intracranial body.

Indeed, while the brain is well known for its architectural flexibility, its long-term survival also depends on some "hard" features. The extensive feedback connections, but also the layered cortex being separated from subcortical structures, provide necessary boundary conditions for functional sequences of representation, computation and action.

It was the structure of the intracranial body which led me to compare the brain to an ensemble of musical instruments. Again, only through the hard, boundary conditions can the emerging resonance patterns become functional and, thus, be considered representations. Furthermore, if we want to explain, i.e. predict, what the brain will do next, we cannot make do without assuming some modularity. Otherwise, our description of the system's dynamics becomes too complex for us to understand.

I denoted the stable activity distinguishing representation from computation a resonance pattern because it results from an interaction of feedforward and feedback connections. Each component in a circular causal system causes- and is affected by the structure of the system at the same time. Therefore, the reciprocal relationship of networks in the brain, but also between bodily action and representation, make it hard to distinguish cause and effect. The dilemma dissolves, however, if either representation or bodily action can exist without the other. Indeed, we are able to hear birds chirping while sitting at the desk and while going for a run. Even though our auditory experience in both cases might not be identical, bodily action does not seem to play a causal role in our ability to listen to the birds' song.

Motor activity and related bodily movement might be necessary for the structures to develop that allow general human cognition, including perception. However, at a certain point the brain learns how to deal with uninteresting variance in environmental stimuli. It learns to *somewhat* dissociate the sensorimotor relations of the

two outermost layers of motor action and early sensory activity from literally more central, reciprocal relations of neural action.

There is still a lot to learn about the relationship of body, representation, and (spatial) cognition. In doing so, it might be fruitful to try and understand the brain as a body as well. Predictive coding approaches are already treating the brain as a system that wants to keep its structure stable by acting to predict what it wants to predict. How exactly one distinguishes an action in the brain from anything else going on there, however, is still to be determined. In this work I developed a rough taxonomy. Hopefully it can inspire more to come.

Let us explore the environmental interaction with the brain and body, as well as the brain in interaction with itself, on equal footing. Following Varela, Thompson and Rosch, who were, in turn, inspired by Eastern philosophy, let us walk the middle way.

Words of Gratitude

Most of my gratitude goes to my supervisor Peter König. The passion he feels and conveys for his work is truly inspiring. Over the last three years I came to very much appreciate his honest feedback, which (although scary at times) is always to the point. He treats his students as equals and regularly meets with all of them, a quality which I hope to be able to adapt in the future of my academic career.

The second-largest thank-you goes to the Research Training Group Situated Cognition. Achim Stephan and Albert Newen played key roles in the organization, and Beate Krickel was a mentor for all the PhD students. The open dialogue among the students and between philosophy and empirical sciences proved rewarding both personally and professionally.

I also want to extend my thanks to Sabine König. Her positive energy was very motivating. In addition, I learned from her about experiment management and the writing process. We shared many helpful discussions concerning results and interpretation.

The work presented here would not have been possible without the help of a great team of junior researchers. In particular, I would like to thank Viviane Clay, who played a fundamental role in the design and construction of Seahaven. I am also indebted to Justine Winkler, Debora Nolte, Raul Sulaimanov, Laura Duesberg, Kirsten Rittershofer, Markéta Bečevová, Shadi Derakshan, Lara Syrek, Jasmin Walter, Lucas Essmann, Paula Eisenhauer, and Valerie Meyer. Also, there would be no experiments without the willingness and effort of participants, whom I want to thank as well.

Generally, I want to thank the team of the Neurobiopsychology lab for making me feel at home, especially thanks to the heartfelt, literal embraces of Artur Czeszumski.

To my girl in Leipzig, the wonderfully honest, beautiful, and quick-witted Kaitlyn Kennedy, thank you for your integral feedback on my writing endeavors, as well as clearly stating when you do not understand something! I extend my kisses to you.

I also want to thank the two additional reviewers, Tobias Meilinger and Fred Hamker, for taking the time to try to make sense of what I have created here. Thank you also, Tobias, for reaching out to me spontaneously, providing comfort in the dark times of endless writing and rewriting. Thank you, Fred, for the upfront trust!

Last but not least, I want to thank my family and friends who are always by my side and give me so much strength through love.

Appendix: Supplementary Material

A.1. Distribution of Participants and Data

A.1.1. VR Exploration

Overall number of participants	82		
Only participated in first session	54	Participated in all three sessions	28
Data sets	54	Participated with complete data sets	22
		Participated missing only first session data	2
		Missing only second session data	2
		Missing only third session data	2
		Data Sets	78
Total number of data sets (optimal number of data sets)	122 (246)		

Table A1. Distribution of participants and data sets in the VR embodiment condition. One data set includes all results in the spatial tasks but also head position and eye tracking data of the participant recorded during city exploration in one session. Optimally one participant produces one data set for each of the three experimental sessions.

A.1.2. Map Exploration

Overall number of participants	74		
Only participated in first session	46	Participated in all three sessions	28
Data sets	46	Participated with complete data sets	26
		Participated missing only third session data	2
		Data Sets	80
Total number of data sets (optimal number of data sets)	126 (222)		

Table A2. Distribution of participants and data sets in the Map embodiment condition. One data set includes all results in the spatial tasks but also mouse (house) clicking data of the participant recorded during city exploration in one session. Optimally one participant produces one data set for each of the three experimental sessions.

A.1.3. VR with Belt Exploration

Overall number of participants	70		
Only participated in first session	43	Participated in all three sessions	27
Data sets	43	Participated with complete data sets	23
		Participated missing only second session data	1
		Missing only third session data	2
		Missing first and second session data	1
		Data Sets	76
Total number of data sets (optimal number of data sets)	119 (210)		

Table A3. Distribution of participants and data sets in the VR with Belt embodiment condition. One data set includes all results in the spatial tasks but also head position and eye tracking data of the participant recorded during city exploration in one session. Optimally one participant produces one data set for each of the three experimental sessions.

A.2. Task Stimuli Algorithm

A.2.1. Definition of House Orientation

When a participant in the VR city is facing VR north, she has an orientation of 0°. After a 90° clockwise rotation her orientation is 90°. The orientation of a house (front) is defined as the participant's orientation when she faces that house.

A.2.2. Absolute Stimuli Algorithm

In the Absolute task we show the same house twice, once with an overlaid arrow pointing to north, once with an arrow not pointing to north. We have to distribute 36 trials with equal distribution of house orientations and equal distribution of compass needle arrows. The left/right difference is irrelevant.

We bin the orientations of all houses in 12 bins of 30° width centered on n*30°, n=0,1, 2, …11. We call this set BIN12 and denote the orientation of these houses o_n (o_0=0, o_1=30, o_2=60, ...). Note that o_n need not be the exact orientation of that house. The exact orientation of house x is oE(x) = n_x*30° + Δ_x where n_x is the number of the bin in BIN12 the house is in and Δ_x is the difference to the bin center. Δ_x can be negative or positive. Then we create a second index by pooling opposing bins, i.e. differing by 180°. This gives us 6 bins. We call this set BIN6. Bin 0 in BIN6, for example, includes houses centered on o_0 and o_6. Bin 1 includes houses centered on o_1 and o_7. For the absolute task we randomly draw 6 houses from each of the 6 bins of BIN6. We number these houses i=0, 1, …5 for the bins they were sampled from and j=0,1, …5 for the count of the samples. This creates a matrix h_{ij} which is depicted in Table A4.

h_{00} first house drawn randomly from bin 0 in BIN6	h_{01} second house drawn randomly from bin 0 in BIN6	h_{02} third house...			
h_{10} first house drawn randomly from bin 1 in BIN6a	h_{11} second house...				
...	...				

					...
				...	h_{55} sixth house drawn randomly from bin 5 in BIN6a

Table A4. House stimuli distribution matrix for the Absolute task.

We check that for each set of 6 houses out of bin i in BIN6 at most 4 came from the same bin in BIN12. That is at most 4 with o_n=i or o_n=i+6 (see definition of BIN6). If not, we sample again until the criterion is reached. This procedure ensures that in the Absolute task as many houses are oriented either north/south or east/west as the oblique orientations. Exact orientations of the correct arrows are then $arrow_{ij}$ = 360° - $oE(h_{ij})$. Then we draw for each house h_{ij} of orientation o_n=i or o_n=i+6 a wrong orientation oW_{ij} (o_n ± (j+1)) modulo (12). Whether it is added or subtracted is determined by chance. Again we check that in the set of j=0, 1, …5 the ± is not more unbalanced than 4:2. In words, we take the orientation of the house, add or subtract between 1 and 6 multiples of 30° and calculate the modulo 12, as we have only 12 bins. This gives the orientation for the second (wrong) option in the 2AFC absolute orientation task as shown in Table A5. Note that the wrong orientation is not equal the wrong arrow orientation. The orientation of the wrong arrow is 360°- (oW_{ij}+Δ_{ij}).

We now have 36 stimuli pairs which we show in a random sequence for either the 3s or unlimited time condition. We decide that randomly as well. For the other condition we use the same algorithm to create again 36 pairs which we show in a random sequence. We exclude the houses from the first condition in the second condition if possible (the algorithm still finds houses of the needed orientation). If not, possible we choose a random house of the needed orientation. Finally, for each pair it is determined by chance if the wrong oriented arrow is shown on the upper or lower screen.

$oW_{00} = h_{00}$ +/- 30°	$oW_{01}=h_{01}$ +/- 60°	$oW_{01}=h_{02}$ +/- 90°	...		
$oW_{10} = h_{10}$ +/- 30°	$oW_{11}=h_{11}$ +/- 60°	...			
...	...				
					...
				...	$oW_{55}=h_{55}$ +/- 60°

Table A5. Wrong north-direction distribution for each house stimuli in the Absolute task.

A.2.3. Relative Stimuli Algorithm

In the Relative task we first show a house. We then show two different houses, one with the same orientation as the first house, the other with a different orientation (no arrows are shown). For the Relative task we use 18 houses as prime houses. That are the houses which are shown first. These houses have been chosen because they were most visited by participants in a pilot experiment. The primes are chosen such that each bin in BIN6 contains three. The relation of orientation for o_n=i and o_n=i+6 is either 2:1 or 1:2.

For each bin in BIN6 we draw randomly a 2*3 prime sequence. That is, for bin i we draw the first 3 without putting back, then put all back and draw the second three. We call this the prime sequence i. For each prime sequence i, we parse through the sequence and randomly draw a target j (a house with the same orientation) from all not prime houses, without putting back.

We cannot use exact orientations here, as we do not have enough target houses with the same exact orientation as each prime (due to 18 primes and 2 conditions, if no target were to repeat, we would need 4 exact targets per prime). To keep the error within a 15° margin if Δ prime is positive we randomly draw a target x for which Δ_x is positive and analogous for negative Δ prime. If there is no such target, we take a target from the bin center. These draws

are all without putting back. This gives us a matrix r_{ij} shown in Table A6.

r_{00} pair of first house from prime sequence 0 and random house of same (max error 15°) orientation (target)	r_{01} pair of second house from prime sequence 0 and random house of same orientation (target)	r_{02} pair of third house...			
r_{10} pair of first house from prime sequence 1 and random house of same orientation (target)	r_{11} pair of second house...				
...					
					...
				...	r_{55} pair of sixth house from prime sequence 6 and random house of same orientation (target)

Table A6. Stimuli distribution matrix for the Relative task defining each pair of prime and correct target house.

Table A7. Wrong target house distribution matrix for the Relative task.

$oW_{00} = r_{00}$ +/- 30°	$oW_{01}=r_{01}$ +/- 60°	$oW_{02}=r_{02}$ +/- 90°	...		
$oW_{10} = r_{10}$ +/- 30°	$oW_{11}=r_{11}$ +/- 60°	...			
...	...				
					...
				...	$oW_{55}=r_{55}$ +/- 60°

We compute the second, wrong target house, just as we computed the wrong orientation for the Absolute task above. Then we get the matrix depicted in the Table A7. We draw from all not-prime and not-target houses, without putting back. To keep the n*30° distance between correct and wrong oriented houses, we apply the same target rule as defined above. That is, if the pair denotes houses with positive deviation from their center bin, we take a wrong oriented house which positively deviates from its center bin, etc.

We now have 36 triples which we show in a random sequence for either the unlimited or the 3s condition. We decide that at random as well. For the other condition we use the same algorithm to create again 36 pairs which we show in a random sequence. We exclude the targets from the first condition in the second condition if possible (the algorithm still finds targets of the needed orientation). If not possible we choose a random target with the needed orientation. Finally, for each triple it is determined by chance if the wrong oriented house is shown on the upper or lower screen.

A.2.4. Pointing Stimuli Algorithm

In the Pointing task we first show a house. Then we show a second house with two different arrows, one of which points to the first, the prime house. For each prime house p, we create a BIN12 set, with orientations now being all pointing directions of other houses (targets) to prime p. We call this set of BIN12 sets BIN12p. From BIN12p we create analogously the set of BIN6 sets for each prime p, BIN6p. If we now proceed to just randomly draw primes and then order the respective pointing target such that we get a matrix with pointing directions like h_{ij}, we will get a lot of targets pointing into the same direction as the orientation of a prime, or e.g. parallel to its long side. Such symmetrical prime pointing target configurations might be easier memorized. We therefore try to spread pointing directions evenly across all prime orientations. For i from 0:5, for each prime p in the sequence i (see Relative Task Algorithm), we draw a target pointing to it, without putting back, where the pointing direction is drawn from bin j in BIN6p. Here j is the number of that prime in the sequence i - 1. That is essentially the transpose of matrices h_{ij} and r_{ij}. It gives us matrix p_{ij} depicted in Table A8.

We compute the second, wrong orientation, just as we computed the wrong orientation for the Absolute and Relative task. Except

that we need to transpose rows and columns of the angle addition and subtraction. This time the pointing direction of the arrow for each correct and wrong choice can be directly used on the pictures. Table A9 gives an overview of the resulting distribution. We now have 36 triples which we show in a random sequence for the unlimited or the 3s condition. We can randomly decide for which. For the other condition we use the same algorithm to create again 36 pairs which we show in a random sequence. We exclude the targets from the first condition in the second condition if possible (the algorithm still finds targets of the needed pointing direction). If not possible we choose a random target with the needed pointing direction. Finally, for each triple it is determined by chance if the wrong pointing arrow is shown on the upper or lower screen.

p_{00} pair of first house p from prime sequence 0 and house (with exact arrow) drawn randomly from bin 0 in BIN6p	p_{01} pair of second house p from prime sequence 0 and house drawn randomly from bin 1 in BIN6p	p_{02} pair of third house p...			
p_{10} pair of first house p from prime sequence 1 and house drawn randomly from bin 0 in BIN6p	p_{11} pair of second house p from prime sequence 1 and house drawn randomly from bin 1 in BIN6p...	...			
...	...				
					...
				...	p_{55} pair of 6th house p from prime sequence 5 and house drawn randomly from bin 5 in BIN6p

Table A8. Stimuli distribution matrix for the Pointing task defining each pair of prime and target house.

$oW_{00} = p_{00}$ +/- 30°	$oW_{01}=p_{01}$ +/- 30°	$oW_{02}=p_{02}$ +/- 30°	...		
$oW_{10} = p_{10}$ +/- 60°	$oW_{11}=p_{11}$ +/- 60°	...			
$oW_{20} = p_{20}$ +/- 90°	...				
					...
				...	$oW_{55}=p_{55}$ +/- 180°

Table A9. Wrong pointing direction distribution for the Pointing task target house stimuli.

A.3. Plot Results

A.3.1. Bar Plot Results

In this section we present the values of mean performance and standard errors in all experimental conditions depicted in the bar plots in Results (Figures 9.1-3, 10.1-3 and 11.1-3). Table A10 details the results in the VR condition (Figures 9.1-3). Table A11 contains the results in the map condition (Figures 10.1-3). Table A12 details the results in the VR with belt condition (Figures 11.1-3).

Session 1	Absolute	Relative	Pointing
3s	49.1 ± 0.9	48.2 ± 0.9	48.3 ± 1
Inf	50 ± 1	53.1 ± 1	52.5 ± 0.8
Session 2	Abs	Rel	Poi
3s	49 ± 2	49.6 ± 1.6	50.1 ± 1.4
Inf	55.9 ± 2.1	53.1 ± 1.7	56.2 ± 1.4
Session 3	Abs	Rel	Poi
3s	50.4 ±2	50.2±1.6	51.8±1.5
Inf	57.2 ±2.4	54.8±2.4	59.5±1.5

Table A10. Mean performance plus/minus standard error in percent in all conditions after VR exploration

Session 1	Absolute	Relative	Pointing
3s	49 ±1	49.7±1	47±1.2
Inf	52.1 ±1.1	54.6±0.9	53.8±1
Session 2	Abs	Rel	Poi
3s	51.1 ±1.8	52.5±1.7	49.7±1.6
Inf	62 ±2.9	58.2±1.9	55.2±1.7
Session 3	Abs	Rel	Poi
3s	55.1 ±2.1	58.3±2.5	49±1.7
Inf	60.9 ±3.1	61±2.4	58.1±2.4

Table A11. Mean performance plus/minus standard error in percent in all conditions after map exploration

Session 1	Absolute	Relative	Pointing
3s	46.4 ±0.9	47.5±1.1	50.9±1.2
Inf	49.6 ±1.1	52.6±1.1	50.8±1.1
Session 2	Abs	Rel	Poi
3s	49.6 ±1.9	49.4±2.2	49.4±1.7
Inf	52.2 ±1.4	56±1.4	55.8±1.4
Session 3	Abs	Rel	Poi
3s	50.6 ±1.4	51±1.5	53.1±2
Inf	54.1 ±1.8	54±1.7	53.8±1.4

Table A12. Mean performance plus/minus standard error in percent in all conditions after VR exploration with belt.

A.3.2. Model Plot Results

A.3.2.1. Example Model

For readers not familiar with the MATLAB R2018b syntax we give one complete example of the linear mixed-effects model (LMEM) and script structure we used for fitting a single condition across sessions. The example fits the performances in the VR condition, Relative task, unlimited time trial blocks over sessions (green line in Results, Figure 9.6).

First we create a table 'RelInf_Table_VR', holding in the first column all performances in the Relative, unlimited condition 'RelInf_VR' ordered underneath each other. We start with all participants in the first session, then all participants in the second session and finally the third session. The second column holds the coding used for the predictor session 'C_Session_VR'. In the third column we store the respective participant's code number 'C_Subject_VR'. An exemplary depiction of the resulting data structure is given in Table A13.

RelInf_Table_VR	RelInf_VR	C_Session_VR	C_Subject_VR
Session 1	49	0	1
	50	0	2
	50	0	3
Session 2	52	.5	1
	53	.5	3

Session 3	54	1	1
	56	1	3

Table A13. the data structure input to the fitlme() function in MATLAB to compute the LMEM. The first column holds the performances in the Relative unlimited condition. The second column holds the coding used for the predictor session. In the third column we store the respective participant's code number.

Before feeding the table into the LMEM we define the general LMEM structure as:
LMEM_Results_RelInf_VR=fitlme(RelInf_Table_VR,
'RelInf_VR~1+C_Session_VR+(1|C_Subject_VR)')

The model structure is defined through the formula bracketed by the apostrophes (notation is similar to Wilkinson and Rogers, 1973). The term '(1|C_Subject_VR)' specifies the random effects model- and covariance matrix. The sparse number of single condition trials/observations in session two and three allows only a fit of one parameter for each subject. Consequently, the covariance matrix becomes a scalar. We decided for intercept as random effect because it leads to an overall better fit than slope. The model results are given in Table A14.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	132	2	82	2
Fixed effects	β0(Intercept) ± SE | CI	53.01 ± 1.02	50.98 55.04		
	β1(Session) ± SE | CI | χ2(1) | p	1.6 ± 2.01	-2.38 5.57	0.63	0.43
Random effects	Intercept σ | CI	3.32	1.44 7.64		
	Residual variance | CI	8.81	7.52 10.33		

Table A14. Results of the LMEM fit over sessions of the Relative task, unlimited time condition after VR exploration. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.

A.3.2.2. VR Exploration Model Plots

We provide details on the models we used for creating the VR condition linear fit plots in Results (Figures 9.4-6). The fixed effect design matrix is identical across the single-condition-over-session plots for all embodiment conditions. This also holds for the time plots that marginalize over the tree spatial tasks (Figure 9.4). The design matrix is shown in Table A15.

Table A16 holds model results for the three second condition, marginalizing spatial tasks (Figure 9.4). Table A17, A18 and A19 show results for the Absolute -, Relative - and Pointing task, respectively. All computed for the three second condition (Figure 9.5).

Table A20 shows the model for the unlimited condition, marginalizing spatial tasks (Figure 9.4). Table A21, A22 and A23 depict results for the Absolute -, Relative - and Pointing task, respectively. Here computed for the unlimited condition (Figure 9.6).

Intercept	Session
1	0
1	.5
1	1

Table A15. The fixed effect design matrix used in all plot LMEMs across embodiment conditions.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	396	2	164	3
Fixed effects	β0(Intercept) ± SE | CI	48.53 ± 0.54	47.46 49.59		
	β1(Session) ± SE | CI | χ2(1) | p	2.06 ± 1.3	-0.49 4.61	2.42	0.12
Random effects	Intercept σ | CI	1.83	0.69 4.84		
	Session σ | CI	3.89	1.8 8.4		
	Residual variance | CI	8.01	7.42 8.64		

Table A16. Results of the LMEM fit over sessions marginalizing over all three spatial tasks but only the 3s time conditions after VR exploration. #obs means

total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1+Session|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	132	2	82	2
Fixed effects	β0(Intercept) ± SE | CI	48.99 ± 0.97	47.06 50.92		
	β1(Session) ± SE | CI | χ2(1) | p	1.35 ± 1.8	-2.31 5	0.53	0.47
Random effects	Intercept σ | CI	4.26	2.58 7.04		
	Residual variance | CI	7.86	6.66 9.27		

Table A17. Results of the LMEM fit over sessions of the 3s Absolute task condition after VR exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	132	2	82	2
Fixed effects	β0(Intercept) ± SE | CI	48.37 ± 0.86	46.67 50.07		
	β1(Session) ± SE | CI | χ2(1) | p	1.14 ± 1.56	-1.95 4.22	0.51	0.47
Random effects	Intercept σ | CI	4.48	3.05 6.57		
	Residual variance | CI	6.45	5.42 7.66		

Table A18. Results of the LMEM fit over sessions of the 3s Relative task condition after VR exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with

1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	132	2	0	1
Fixed effects	β0(Intercept) ± SE | CI	48.31 ± 0.91	46.5 50.12		
	β1(Session) ± SE | CI | χ2(1) | p	3.53 ± 1.84	-0.12 7.18	3.62	0.057
Random effects	Residual variance | CI	8.44	7.48 9.53		

Table A19 Results of the LMEM fit over sessions of the 3s Pointing task condition after VR exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect intercept σ was close to zero ($2*10^{-6}$) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The LMEM formula is: 'CondResult~1+Session'.
(We could have used a simple linear model, but results are similar and staying in LMEM analysis style is simpler).

Degrees of freedom (df)	#obs | #fix | #rand | #cov	396	2	82	2
Fixed effects	β0(Intercept) ± SE | CI	51.94 ± 0.54	50.89 52.99		
	β1(Session) ± SE | CI | χ2(1) | p	5.22 ± 1.56	2.14 8.29	9.51	0.002
Random effects	Session σ | CI	5.86	3.81 9		
	Residual variance | CI	8.57	8 9.21		

Table A20. Results of the LMEM fit over sessions marginalizing over all three spatial tasks but only the unlimited time conditions after VR exploration. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect intercept σ was close to zero (0.008)

which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The LMEM formula is: 'CondResult~1+Session+(-1+Session|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	132	2	82	2
Fixed effects	β0(Intercept) ± SE | CI	50.24 ± 1.07	48.13 52.36		
	β1(Session) ± SE | CI | χ2(1) | p	7.23 ± 2.03	3.22 11.24	12.14	0.0005
Random effects	Intercept σ | CI	4.69	2.84 7.77		
	Residual variance | CI	8.62	7.3 10.17		

Table A21. Results of the LMEM fit over sessions of the unlimited Absolute task condition after VR exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	132	2	82	2
Fixed effects	β0(Intercept) ± SE | CI	53.01 ± 1.02	50.98 55.04		
	β1(Session) ± SE | CI | χ2(1) | p	1.6 ± 2.01	-2.38 5.57	0.63	0.43
Random effects	Intercept σ | CI	3.32	1.44 7.64		
	Residual variance | CI	8.81	7.52 10.33		

Table A22. Results of the LMEM fit over sessions of the unlimited Relative task condition after VR exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	132	2	82	2
Fixed effects	β0(Intercept) ± SE | CI	52.49 ± 0.8	50.92 54.07		
	β1(Session) ± SE | CI | χ2(1) | p	6.91 ± 1.5	3.93 9.89	19.38	0.00001
Random effects	Intercept σ | CI	3.55	1.98 6.36		
	Residual variance | CI	6.38	5.31 7.66		

Table A23. Results of the LMEM fit over sessions of the unlimited Pointing task condition after VR exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

A.3.2.3. Map Experiment Model Plots

We provide details on the models we used for creating the Map experiment linear fit plots in Results (Figures 10.4-6). Table A24 holds model results for the three second condition, marginalizing spatial tasks (Figure 10.4). Table A25, A26 and A27 show results for the Absolute -, Relative - and Pointing task, respectively. All computed for the three second condition (Figure 10.5).

Table A28 shows the model for the unlimited condition, marginalizing spatial tasks (Figure 10.4). Table A29, A30 and A31 depict results for the Absolute -, Relative - and Pointing task, respectively. Here computed for the unlimited condition (Figure 10.6).

Degrees of freedom (df)	#obs | #fix | #rand | #cov	384	2	148	3
Fixed effects	β0(Intercept) ± SE | CI	48.55 ± 0.63	47.32 49.78		
	β1(Session) ± SE | CI | χ2(1) | p	5.79 ± 1.62	2.61 8.97	11.06	0.00088

Random effects	Intercept σ | CI	2.21	0.91 5.39		
	Session σ | CI	5.75	3.52 9.4		
	Residual variance | CI	8.82	8.14 9.56		

Table A24. Results of the LMEM fit over sessions marginalizing over all three spatial tasks but only the 3s time conditions after VR exploration. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1+Session |Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	128	2	74	2
Fixed effects	β0(Intercept) ± SE | CI	48.9 ± 1.04	46.85 50.95		
	β1(Session) ± SE | CI | χ2(1) | p	6.17 ± 1.97	2.27 10.06	9.04	0.0026
Random effects	Intercept σ | CI	3.26	1.27 8.38		
	Residual variance | CI	8.59	7.26 10.17		

Table A25. Results of the LMEM fit over sessions of the 3s Absolute task condition after map exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	128	2	74	2
Fixed effects	β0(Intercept) ± SE | CI	49.45 ± 1.05	47.37 51.54		
	β1(Session) ± SE | CI | χ2(1) | p	7.24 ± 1.82	3.64 10.84	14.96	0.00011

Random effects	Intercept σ | CI	5.45	3.8 7.83		
	Residual variance | CI	7.5	6.33 8.87		

Table A26. Results of the LMEM fit over sessions of the 3s Relative task condition after map exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	128	2	74	2
Fixed effects	β0(Intercept) ± SE | CI	47.27 ± 1.06	45.16 49.37		
	β1(Session) ± SE | CI | χ2(1) | p	2.52 ± 2.08	-1.61 6.64	3.62	0.057
Random effects	Intercept σ | CI	1.32	0 525.94		
	Residual variance | CI	9.38	7.93 11.1		

Table A27. Results of the LMEM fit over sessions of the 3s Pointing task condition after map exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect intercept σ was very close to zero which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include. We therefore did not include it in the model.
The LMEM formula is: 'CondResult~1+Session'.
(We could have used a simple linear model, but results are similar and staying in LMEM analysis style is simpler).

Degrees of freedom (df)	#obs | #fix | #rand | #cov	384	2	148	3
Fixed effects	β0(Intercept) ± SE | CI	53.65 ± 0.66	52.36 54.94		
	β1(Session) ± SE | CI | χ2(1) | p	7.31 ± 2.21	2.97 11.64	9.61	0.0019
Random effects	Intercept σ | CI	2.94	1.66 5.21		
	Session σ | CI	9.71	6.64 14.21		
	Residual variance | CI	8.63	7.98 9.34		

Table A28. Results of the LMEM fit over sessions marginalizing over all three spatial tasks but only the unlimited time conditions after map exploration. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1+Session|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	128	2	74	2
Fixed effects	β0(Intercept) ± SE | CI	52.73 ± 1.36	50.03 55.44		
	β1(Session) ± SE | CI | χ2(1) | p	8.49 ± 2.07	4.38 12.59	15.7	0.00007
Random effects	Intercept σ | CI	8.63	6.76 11.03		
	Residual variance | CI	8.23	6.95 9.74		

Table A29. Results of the LMEM fit over sessions of the unlimited Absolute task condition after map exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	128	2	74	2
Fixed effects	β0(Intercept) ± SE | CI	54.61 ± 1.04	52.56 56.66		
	β1(Session) ± SE | CI | χ2(1) | p	6.44 ± 1.88	2.72 10.16	11.18	0.00083
Random effects	Intercept σ | CI	4.54	2.83 7.31		
	Residual variance | CI	7.94	6.73 9.37		

Table A30. Results of the LMEM fit over sessions of the unlimited Relative task condition after map exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	128	2	74	2
Fixed effects	β0(Intercept) ± SE | CI	53.66 ± 1.07	51.54 55.78		
	β1(Session) ± SE | CI | χ2(1) | p	4.4 ± 2	0.43 8.37	4.64	0.031
Random effects	Intercept σ | CI	3.93	1.98 7.81		
	Residual variance | CI	8.64	7.32 10.21		

Table A31. Results of the LMEM fit over sessions of the unlimited Pointing task condition after map exploration.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

A.3.2.4. VR with Belt Experiment Model Plots

We provide details on the models we used for creating the VR with Belt experiment linear fit plots in Results (Figures 11.4-6). Table A32 holds model results for the three second condition, marginalizing spatial tasks (Figure 11.4). Table A33, A34 and A35 show results for the Absolute -, Relative - and Pointing task, respectively. All computed for the three second condition (Figure 11.5). Table A36 shows the model for the unlimited condition, marginalizing spatial tasks (Figure 11.4). Table A37, A38 and A39 depict results for the Absolute -, Relative - and Pointing task, respectively. Here computed for the unlimited condition (Figure 11.6).

Degrees of freedom (df)	#obs | #fix | #rand | #cov	357	2	70	2
Fixed effects	β0(Intercept) ± SE | CI	48.19 ± 0.67	46.88 49.51		
	β1(Session) ± SE | CI | χ2(1) | p	3.07 ± 1.22	0.67 5.46	6.25	0.012
Random effects	Intercept σ | CI	2.65	1.5 4.7		
	Residual variance | CI	8.75	8.07 9.49		

Table A32. Results of the LMEM fit over sessions marginalizing over all three spatial tasks but only the 3s time conditions after VR exploration with belt. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect session σ was close to zero (0.01) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	119	2	70	2
Fixed effects	β0(Intercept) ± SE | CI	46.56 ± 0.93	44.7 48.4		
	β1(Session) ± SE | CI | χ2(1) | p	4.49 ± 1.7	1.13 7.86	6.73	0.0095
Random effects	Intercept σ | CI	3.71	2.05 6.71		

	Residual variance | CI	7.03	5.89 8.41		

Table A33. Results of the LMEM fit over sessions of the 3s Absolute task condition after VR exploration with belt.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. $\chi2(1)$ is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	119	2	70	2
Fixed effects	β0(Intercept) ± SE | CI	47.52 ± 1.05	45.35 49.68		
	β1(Session) ± SE | CI | $\chi2(1)$ | p	3.61 ± 1.95	-0.25 7.47	3.35	0.067
Random effects	Intercept σ | CI	4.79	2.69 8.54		
	Residual variance | CI	7.99	6.57 9.71		

Table A34. Results of the LMEM fit over sessions of the 3s Relative task condition after VR exploration with belt.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. $\chi2(1)$ is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	119	2	70	2
Fixed effects	β0(Intercept) ± SE | CI	50.52 ± 1.14	48.27 52.78		
	β1(Session) ± SE | CI | $\chi2(1)$ | p	1.31 ± 2.14	-2.94 5.55	0.37	0.55
Random effects	Intercept σ | CI	3.46	1.33 9.02		
	Residual variance | CI	9.1	7.65 10.83		

Table A35. Results of the LMEM fit over sessions of the 3s Pointing task condition after VR exploration with belt.

#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect intercept σ was very close to zero which leads to numerical instabilities and means one should (and can) ignore this effect. We therefore did not include it in the model.
The LMEM formula is: 'CondResult~1+Session'.
(We could have used a simple linear model, but results are basically identical and staying in LMEM analysis style was considered less confusing).

Degrees of freedom (df)	#obs | #fix | #rand | #cov	357	2	70	2
Fixed effects	β0(Intercept) ± SE | CI	51.26 ± 0.69	49.91 52.62		
	β1(Session) ± SE | CI | χ2(1) | p	3.89 ± 1.12	1.68 6.09	8.87	0.0029
Random effects	Intercept σ | CI	3.8	2.7 5.34		
	Residual variance | CI	7.64	7.03 8.3		

Table A36. Results of the LMEM fit over sessions marginalizing over all three spatial tasks but only the unlimited time conditions after VR exploration with belt. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect session σ was close to zero (0.01) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	119	2	70	2
Fixed effects	β0(Intercept) ± SE | CI	49.62 ± 0.98	47.67 51.57		
	β1(Session) ± SE | CI | χ2(1) | p	4.56 ± 1.74	1.11 8.01	6.62	0.01
Random effects	Intercept σ | CI	4.46	2.58 7.71		

	Residual variance | CI	7.09	5.82 8.64		

Table A37. Results of the LMEM fit over sessions of the unlimited Absolute task condition after VR exploration with belt.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	119	2	70	2
Fixed effects	β0(Intercept) ± SE | CI	53 ± 0.98	51.06 54.92		
	β1(Session) ± SE | CI | χ2(1) | p	2.44 ± 1.8	-1.13 6.01	1.73	0.19
Random effects	Intercept σ | CI	3.55	1.5 8.42		
	Residual variance | CI	7.54	6.18 9.2		

Table A38. Results of the LMEM fit over sessions of the unlimited Relative task condition after VR exploration with belt.
#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	119	2	70	2
Fixed effects	β0(Intercept) ± SE | CI	51.22 ± 1.01	49.22 53.22		
	β1(Session) ± SE | CI | χ2(1) | p	3.97 ± 1.83	0.34 7.6	4.47	0.034
Random effects	Intercept σ | CI	4.05	1.76 9.31		
	Residual variance | CI	7.59	6.08 9.46		

Table A39. Results of the LMEM fit over sessions of the unlimited Pointing task condition after VR exploration with belt.

#obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. $\chi 2(1)$ is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

A.4. Best-Fit Model Results

In this section of the Appendix we provide detailed results of all best-fit models. That are the models selected according to the step-down procedure described in Methods. We also provide information on the maximum models, before step-down. For an introductory example of how to set up a linear mixed effect model (LMEM) in MATLAB, please see above.

A.4.1. Best-Fit VR Models

We provide details on all linear models used for the quantitative analysis of the VR embodiment condition as presented in Results. Table A40 and A41 hold model results and design matrix, respectively, of the fit over session, task and time. Table A42 and A43 depict results and design matrix of the fit over session, task and three seconds decision time. Table A44 and A45 contain results and design matrix of the fit over session, task and unlimited decision time.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	792	4	164	3
Fixed effects	β0(Intercept) ± SE | CI	52.01 ± 0.48	51.08 52.94		
	β1(Session) ± SE | CI | χ2(1) | p	3.57 ± 0.83	1.94 5.2	18.02	0.000022
	β2(Time) ± SE | CI | χ2(1) | p	5 ± 0.67	1.94 5.2	41.13	1.4*10⁻10
	β1*β2 ± SE | CI | χ2(1) | p	3.16 ± 1.49	0.23 6.08	4.45	0.035
Random effects	Intercept σ | CI	2.52	1.85 3.43		
	Time σ | CI	0.79	0 321.05		
	Residual variance | CI	8.35	7.91 8.8		

Table A40. Model a-VR.
Results of the best LMEM fit of the VR condition over session, task and time. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect session σ was close to zero (0.02)

which leads to numerical instabilities and means one can and should ignore this effect. It is therefore not included in the best fit model.
The best fit LMEM formula is:
'CondResult~1+Time+Session+Time*Session+(1+Time|Subject)'.
In short: 'CondResult~Time*Session+(1+Time|Subject)'.

Condition	β0(Intercept)	β1(Session)	β2(Time)	β3(Mean to Rel)	β4(Mean to Poi)	β1(Session)* β2(Time)
S1 Abs 3s	1	-.5	-.5	-1	-1	.25
S1 Rel 3s	1	-.5	-.5	1	0	.25
S1 Poi 3s	1	-.5	-.5	0	1	.25
S1 Abs Inf	1	-.5	.5	-1	-1	-.25
S1 Rel Inf	1	-.5	.5	1	0	-.25
S1 Poi Inf	1	-.5	.5	0	1	-.25
S2 Abs 3s	1	0	-.5	-1	-1	0
S2 Rel 3s	1	0	-.5	1	0	0
S2 Poi 3s	1	0	-.5	0	1	0
S2 Abs Inf	1	0	.5	-1	-1	0
S2 Rel Inf	1	0	.5	1	0	0
S2 Poi Inf	1	0	.5	0	1	0
S3 Abs 3s	1	.5	-.5	-1	-1	-.25
S3 Rel 3s	1	.5	-.5	1	0	-.25
S3 Poi 3s	1	.5	-.5	0	1	-.25
S3 Abs Inf	1	.5	.5	-1	-1	.25
S3 Rel Inf	1	.5	.5	1	0	.25
S3 Poi Inf	1	.5	.5	0	1	.25

Table A41. Model a-VR Matrix.
The fixed effect design matrix used to fit the VR condition over session, task and time. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult~Time*Session*MeanRel+Time*Session*MeanPoi+(1+Time+Session+MeanRel+MeanPoi | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	396	4	164	3
Fixed effects	β0(Intercept) ± SE | CI	49.55 ± 0.58	48.41 50.69		
	β1(Session) ± SE | CI | χ2(1) | p	2.03 ± 1.09	-0.12 4.18	3.4	0.065 (ns)
Random effects	Intercept σ | CI	2.62	1.72 3.99		
	Residual variance | CI	8.03	7.44 8.66		

Table A42. Model s-VR.
Results of the LMEM fit including the last, not significant term of the VR condition over session and task only for the 3 second time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect session σ was close to zero (0.008) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The LMEM formula is: 'CondResult~1+Session+(1|Subject)'.

Condition	β0(Intercept)	β1(Session)	β4(Abs to Rel)	β3(Abs to Poi)
S1 Abs 3s	1	-.5	0	0
S1 Rel 3s	1	-.5	1	0
S1 Poi 3s	1	-.5	0	1
S2 Abs 3s	1	0	0	0
S2 Rel 3s	1	0	1	0
S2 Poi 3s	1	0	0	1
S3 Abs 3s	1	.5	0	0
S3 Rel 3s	1	.5	1	0
S3 Poi 3s	1	.5	0	1

Table A43. Model s-VR Matrix.
The fixed effect design matrix of the LMEM fit including the last, not significant term of the VR condition over session and task only for the 3 second time condition. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors

depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects. The maximum LMEM formula is:
'CondResult~Session*AbsRel+Session*AbsPoi+(1+Session+AbsRel+AbsPoi| Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	396	4	164	3
Fixed effects	β0(Intercept) ± SE | CI	53.82 ± 0.68	52.48 55.16		
	β1(Session) ± SE | CI | χ2(1) | p	7.14 ± 1.35	4.49 9.79	27.07	$2*10^{-7}$
	β3(Abs to Poi) ± SE | CI | χ2(1) | p	2.2 ± 0.93	0.37 4	5.54	0.018
	β1*β4(Abs to Rel) ± SE | CI | χ2(1) | p	-5.6 ± 2.1	-9.69 -1.51	7.18	0.0074
Random effects	Intercept σ | CI	2.65	1.69 4.16		
	Residual variance | CI	8.51	7.89 9.18		

Table A44. Model u-VR.
Results of the best LMEM fit of the VR condition over sessions and tasks for the unlimited time condition results. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here both random effects session σ and absolute to pointing task difference σ were close to zero (0.008, 0.01) which leads to numerical instabilities and means one can and should ignore these effects. We therefore did not include them in the model. The best fit LMEM formula is:
'CondResult~1+Session+AbsPoi+AbsRel:Session+(1|Subject)'.
Note 'AbsRel:Session' means only the interaction of 'AbsRel' and 'Session', while 'AbsRel*Session' would also include both 'AbsRel' and 'Session' in addition to their interaction.

Condition	β0(Intercept)	β1(Session)	β3(Abs to Poi)	β4(Abs to Rel)	β1(Session)* β4(Abs to Rel)
S1 Abs Inf	1	-.5	0	0	0
S1 Rel Inf	1	-.5	0	1	-.5
S1 Poi Inf	1	-.5	1	0	0
S2 Abs Inf	1	0	0	0	0
S2 Rel Inf	1	0	0	1	0
S2 Poi Inf	1	0	1	0	0
S3 Abs Inf	1	.5	0	0	0
S3 Rel Inf	1	.5	0	1	.5
S3 Poi Inf	1	.5	1	0	0

Table A45. Model u-VR Matrix.
The fixed effect design matrix of the best fit LMEM fit of the VR condition over session and task only for the unlimited time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is: 'CondResult ~
Session*AbsRel+Session*AbsPoi+(1+Session+AbsRel+AbsPoi | Subject)'.

A.4.2. Best-Fit Map Models

We provide details on all linear models used for the quantitative analysis of the map embodiment condition as presented in Results. Table A46 and A47 hold model results and design matrix, respectively, of the fit over session, task and time. Table A48 and A49 depict results and design matrix of the fit over session, task and three seconds decision time. Table A50 and A51 contain results and design matrix of the fit over session, task and unlimited decision time.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	768	4	296	5
Fixed effects	β0(Intercept) ± SE | CI	55.11 ± 0.7	53.73 56.48		
	β1(Session) ± SE | CI | χ2(1) | p	6.55 ± 1.16	4.27 8.83	26.87	$2.2*10^{-7}$
	β2(Time) ± SE | CI | χ2(1) | p	5.35 ± 0.84	3.71 7	30.99	$2.6*10^{-8}$

	β3(Mean to Poi) ± SE | CI | χ2(1) | p	-2.15 ± 0.74	-3.59 -0.7	7.08	0.0079
Random effects	Intercept σ | CI	3.74	2.94 4.77		
	Session σ | CI	4.71	3.03 7.31		
	Time σ | CI	4.53	3.09 6.66		
	Mean to Poi σ | CI	2.65	1.15 6.13		
	Residual variance | CI	8.52	8 9.06		

Table A46. Model a-Map.
Results of the best LMEM fit of the map condition over session, task and time. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is: 'CondResult ~ 1+Time+Session+MeanPoi+(1+Time+Session+MeanPoi | Subject)'.

Condition	β0(Intercept)	β1(Session)	β2(Time)	β3(Mean to Poi)	β4(Mean to Abs)
S1 Abs 3s	1	-.5	-.5	0	1
S1 Rel 3s	1	-.5	-.5	-1	-1
S1 Poi 3s	1	-.5	-.5	1	0
S1 Abs Inf	1	-.5	.5	0	1
S1 Rel Inf	1	-.5	.5	-1	-1
S1 Poi Inf	1	-.5	.5	1	0
S2 Abs 3s	1	0	-.5	0	1
S2 Rel 3s	1	0	-.5	-1	-1
S2 Poi 3s	1	0	-.5	1	0
S2 Abs Inf	1	0	.5	0	1
S2 Rel Inf	1	0	.5	-1	-1
S2 Poi Inf	1	0	.5	1	0
S3 Abs 3s	1	.5	-.5	0	1
S3 Rel 3s	1	.5	-.5	-1	-1

S3 Poi 3s	1	.5	-.5	1	0
S3 Abs Inf	1	.5	.5	0	1
S3 Rel Inf	1	.5	.5	-1	-1
S3 Poi Inf	1	.5	.5	1	0

Table A47. Model a-Map Matrix.
The fixed effect design matrix used to fit the map condition over session, task and time. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult ~
Time*Session*MeanPoi+Time*Session*MeanAbs+(1+Time+Session+MeanPoi+ MeanAbs | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	384	5	222	4
Fixed effects	β0(Intercept) ± SE | CI	48.09 ± 0.85	46.43 49.75		
	β1(Poi to Abs) ± SE | CI | χ2(1) | p	3.86 ± 1.15	1.59 6.13	10.94	0.00094
	β2(Poi to Rel) ± SE | CI | χ2(1) | p	5.38 ± 1.25	2.93 7.83	16.34	0.000053
	β1*β3(Session) ± SE | CI | χ2(1) | p	6.14 ± 1.96	2.29 9.98	9.59	0.002
	β2*β3(Session) ± SE | CI | χ2(1) | p	7.97 ± 2.01	4.01 11.9	14.88	0.00011
Random effects	Intercept σ | CI	2.91	1.75 4.85		
	Poi to Abs	1.42	0 91.29		
	Poi to Rel	3.56	1.68 7.55		
	Residual variance | CI	8.59	7.86 9.38		

Table A48. Model s-Map.
Results of the best fit LMEM of the map condition over session and task only for the 3 second time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is

the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'CondResult~1+PoiAbs+PoiRel+PoiAbs:Session+PoiRel:Session+(1+PoiAbs+PoiRel | Subject)'.

Condition	β0(Intercept)	β1(Poi to Abs)	β2(Poi to Rel)	β3(Session)	β1(Poi to Abs)* β3(Session)	β2(Poi to Rel)* β3(Session)
S1 Abs 3s	1	1	0	-.5	.-5	0
S1 Rel 3s	1	0	1	-.5	0	-.5
S1 Poi 3s	1	0	0	-.5	0	0
S2 Abs 3s	1	1	0	0	0	0
S2 Rel 3s	1	0	1	0	0	0
S2 Poi 3s	1	0	0	0	0	0
S3 Abs 3s	1	1	0	.5	.5	0
S3 Rel 3s	1	0	1	.5	0	.5
S3 Poi 3s	1	0	0	.5	0	0

Table A49. Model s-Map Matrix.
The fixed effect design matrix of the best fit LMEM fit of the map condition over session and task only for the 3 second time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult~Session*PoiAbs+Session*PoiRel+(1+Session+PoiAbs+PoiRel | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	384	4	148	3
Fixed effects	β0(Intercept) ± SE | CI	57.65 ± 0.93	55.83 59.47		
	β1(Session) ± SE | CI | χ2(1) | p	5.21 ± 1.6	2.04 8.38	10.14	0.0015
	β2(Rel to Poi) ± SE | CI | χ2(1) | p	-1.86 ± 0.96	-3.76 0	3.7	0.054 (ns)
	β1*β3(Rel to Abs) ± SE | CI | χ2(1) | p	4.59 ± 2.18	0.3 8.8	4.38	0.036

Random effects	Intercept σ | CI	5.12	4 6.52		
	Session σ | CI	4.3	1.89 9.8		
	Residual variance | CI	8.71	8.03 9.45		

Table A50. Model u-Map.
Results of the LMEM fit including the two last, not significant terms (β1*β3 interaction is only significant because of β2) of the map condition over session and task only for the unlimited time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Random effect Relative to Pointing task difference σ was close to zero (0.007) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The best fit LMEM formula is:
'CondResult~1+Session+RelPoi+RelAbs:Session+(1+Session | Subject)'.

Condition	β0(Intercept)	β1(Session)	β2(Rel to Poi)	β3(Rel to Abs)	β1(Session)* β3(Rel to Abs)
S1 Abs Inf	1	-.5	0	1	.-5
S1 Rel Inf	1	-.5	0	0	0
S1 Poi Inf	1	-.5	1	0	0
S2 Abs Inf	1	0	0	1	0
S2 Rel Inf	1	0	0	0	0
S2 Poi Inf	1	0	1	0	0
S3 Abs Inf	1	.5	0	1	.5
S3 Rel Inf	1	.5	0	0	0
S3 Poi Inf	1	.5	1	0	0

Table A51. Model u-Map Matrix.
The fixed effect design matrix of the LMEM fit including the two last, not significant terms of the map condition over session and task, only for the unlimited time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.

The maximum LMEM formula is:
'CondResult~Session*RelPoi+Session*RelAbs+(1+Session+RelPoi+RelAbs | Subject)'.

A.4.3. Best-Fit VR with Belt Models

We provide details on all linear models used for the quantitative analysis of the VR embodiment condition with belt as presented in Results. Table A52 and A53 hold model results and design matrix, respectively, of the fit over session, task and time. Table A54 and A55 depict results and design matrix of the fit over session, task and three seconds decision time. Table A56 and A57 contain results and design matrix of the fit over session, task and unlimited decision time.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	714	4	210	4
Fixed effects	β0(Intercept) ± SE | CI	51.43 ± 0.45	50.54 52.32		
	β1(Session) ± SE | CI | χ2(1) | p	3.4 ± 0.81	1.8 4.99	17.23	0.000055
	β2(Time) ± SE | CI | χ2(1) | p	3.3 ± 0.89	1.55 5.05	12.43	0.00042
	β3(Mean to Poi) ± SE | CI | χ2(1) | p	1.11 ± 0.41	0.31 1.91	7	0.0082
Random effects	Intercept σ | CI	2.17	1.43 3.27		
	Time σ | CI	5.11	3.48 7.5		
	Mean to Poi σ | CI	1.31	0.47 3.64		
	Residual variance | CI	8.07	7.59 8.57		

Table A52. Model a-VRwB.
Results of the best LMEM fit of the VR with belt condition over session, task and time. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Random effect session σ was close to zero

(0.009) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The best fit LMEM formula is:
'CondResult~1+Time+Session+MeanPoi+(1+Time+MeanPoi | Subject)'.

Condition	β0(Intercept)	β1(Session)	β2(Time)	β3(Mean to Poi)	β4(Mean to Rel)
S1 Abs 3s	1	-.5	-.5	-1	-1
S1 Rel 3s	1	-.5	-.5	0	1
S1 Poi 3s	1	-.5	-.5	1	0
S1 Abs Inf	1	-.5	.5	-1	-1
S1 Rel Inf	1	-.5	.5	0	1
S1 Poi Inf	1	-.5	.5	1	0
S2 Abs 3s	1	0	-.5	-1	-1
S2 Rel 3s	1	0	-.5	0	1
S2 Poi 3s	1	0	-.5	1	0
S2 Abs Inf	1	0	.5	-1	-1
S2 Rel Inf	1	0	.5	0	1
S2 Poi Inf	1	0	.5	1	0
S3 Abs 3s	1	.5	-.5	-1	-1
S3 Rel 3s	1	.5	-.5	0	1
S3 Poi 3s	1	.5	-.5	1	0
S3 Abs Inf	1	.5	.5	-1	-1
S3 Rel Inf	1	.5	.5	0	1
S3 Poi Inf	1	.5	.5	1	0

Table A53. Model a-VRwB Matrix.
The fixed effect design matrix used to fit the map condition over session, task and time. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult ~
Time*Session*MeanRel+Time*Session*MeanRel+(1+Time+Session+MeanRel+ MeanPoi | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	357	3	70	2
Fixed effects	β0(Intercept) ± SE | CI	49.72 ± 0.62	48.51 50.93		
	β1(Poi to Abs) ± SE | CI | χ2(1) | p *β3(Session)	5.24 ± 1.89	1.53 8.96	7.63	0.0057
	β2(Poi to Rel) ± SE | CI | χ2(1) | p *β3(Session)	3.88 ± 1.89	0.17 7.6	4.21	0.04
Random effects	Intercept σ | CI	2.7	1.55 4.7		
	Residual variance | CI	8.68	8 9.42		

Table A54. Model s-VRwB.
Results of the best LMEM fit of the VR with belt condition over sessions and tasks for the 3 second time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect session σ was close to zero (0.02) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The best fit LMEM formula is:
'Condition~1+PoiAbs:Session+PoiRel:Session+(1|Subject)'.
Different from 'PoiAbs*Session', 'PoiAbs:Session' only includes the interaction in the model.

Condition	β0(Intercept)	β1(Poi to Abs)	β2(Poi to Rel)	β3(Session)	β1*β3	β2*β3
S1 Abs 3s	1	1	0	-.5	.-5	0
S1 Rel 3s	1	0	1	-.5	0	-.5
S1 Poi 3s	1	0	0	-.5	0	0
S2 Abs 3s	1	1	0	0	0	0
S2 Rel 3s	1	0	1	0	0	0
S2 Poi 3s	1	0	0	0	0	0
S3 Abs 3s	1	1	0	.5	.5	0
S3 Rel 3s	1	0	1	.5	0	.5

S3 Poi 3s	1	0	0	.5	0	0

Table A55. Model s-VRwB Matrix.
The fixed effect design matrix of the best fit LMEM fit of the VRwB condition over session and task only for the 3 second time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult~Session*PoiAbs+Session*PoiRel+(1+Session+PoiAbs+PoiRel | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	357	3	70	2
Fixed effects	β0(Intercept) ± SE | CI	52.6 ± 0.75	51.13 54.07		
	β1(Session) ± SE | CI | χ2(1) | p	3.89 ± 1.11	1.7 6.08	11.96	0.00054
	β2(Abs to Rel) ± SE | CI | χ2(1) | p	1.83 ± 0.85	0.16 3.51	4.6	0.032
Random effects	Intercept σ | CI	3.84	2.74 5.37		
	Residual variance | CI	7.58	6.98 8.23		

Table A56. Model u-VRwB.
Results of the best LMEM fit of the VR with belt condition over sessions and tasks only for the unlimited time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here both the random effects session σ and difference from Absolute to Relative task σ were close to zero (0.01, 0.02) which leads to numerical instabilities and means one can and should ignore these effects. We therefore did not include them in the model.
The best fit LMEM formula is: 'Condition~1+Session+AbsRel+(1|Subject)'.

Condition	β0(Intercept)	β1(Session)	β2(Abs to Rel)	β3(Abs to Poi)
S1 Abs Inf	1	-.5	1	0
S1 Rel Inf	1	-.5	0	0

S1 Poi Inf	1	-.5	0	1
S2 Abs Inf	1	0	1	0
S2 Rel Inf	1	0	0	0
S2 Poi Inf	1	0	0	1
S3 Abs Inf	1	.5	1	0
S3 Rel Inf	1	.5	0	0
S3 Poi Inf	1	.5	0	1

Table A57. Model u-VRwB Matrix.
The fixed effect design matrix of the best fit LMEM fit of the VR with belt condition over session and task only for the unlimited time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is: 'CondResult ~ Session*AbsRel+Session*AbsPoi+(1+Session+AbsRel+AbsPoi | Subject)'.

A.4.4. Best-Fit VR-as-Reference Models

A.4.4.1. Best-Fit VR versus Map Models

We provide details on all linear models used for the quantitative analysis of the comparison of VR and Map condition as presented in Results. Table A58 and A59 hold model results and design matrix, respectively, of the fit over experimental exploration, session, task and time. TableA60 and A61 show the results and design matrix, fitted on exploration, session and time for only the Pointing task.

Table A62 and A63 depict results and design matrix of the fit over exploration, session, task and three seconds decision time. Table A64 and A65 contain results and design matrix of the fit over exploration, session, task and unlimited decision time.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	1560	10	468	4
Fixed effects	β0(Intercept) ± SE | CI	53.15 ± 0.4	52.4 53.9		
	β1(Exploration) ± SE | CI | χ2(1) | p	2.29 ± 0.79	0.73 3.85	8.16	0.0043

	β2(Session) ± SE | CI | χ2(1) | p	4.96 ± 0.7	3.58 6.33	43.48	$4.3*10^{-11}$
	β3(Time) ± SE | CI | χ2(1) | p	4.82 ± 0.52	3.81 5.84	65.9	$4.4*10^{-16}$
	β1*β4(Mean to Rel) ± SE | CI | χ2(1) | p	2.13 ± 0.68	0.8 3.45	9.82	0.002
	β1*β5(Mean to Poi) ± SE | CI | χ2(1) | p	-3.07 ± 0.68	-4.4 -1.75	20.44	$6.2*10^{-6}$
	β1*β2 ± SE | CI | χ2(1) | p	2.79 ± 1.4	0.04 5.54	3.9	0.048
	β1*β2*β4 ± SE | CI | χ2(1) | p	3.26 ± 1.52	0.28 6.25	4.6	0.032
	β1*β2*β5 ± SE | CI | χ2(1) | p	-4.39 ± 1.52	-7.37 -1.41	8.31	0.004
	β2*β3*β4 ± SE | CI | χ2(1) | p	-2.8 ± 1.18	-5.12 -0.49	5.61	0.018
Random effects	Intercept σ | CI	3.15	2.62 3.8		
	Session σ | CI	3.16	1.89 5.3		
	Time σ | CI	3.25	2.2 4.8		
	Residual variance | CI	8.48	8.16 8.83		

Table A58. Model a-VRvMap.
Results of the best LMEM fit of both, VR and Map condition, over session, task and time . #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'Condition~1+ Exploration +Session+Time+ Exploration:MeanRel+ Exploration:MeanPoi+Time:Session+Exploration:Session:MeanRel+Exploration: Session:MeanPoi+Session:Time:MeanRel+(1+Session+Time|Subject)'.
Different from 'Exploration *MeanRel', 'Exploration:MeanRel' only includes the interaction in the model.

Condition	β0 (Int.)	β1 (Expl.)	β2 (Sess.)	β3 (Time)	β4(Mean to Rel)	β5(Mean to Poi)	β1* β4	β1* β5	β1* β2	β1* β2* β4	β1* β2* β5	β2* β3* β4
S1 Abs 3s VR	1	-.5	-.5	-.5	-1	-1	.5	.5	.25	-.25	-.25	-.25
S1 Rel 3s VR	1	-.5	-.5	-.5	1	0	-.5	0	.25	.25	0	.25
S1 Poi 3s VR	1	-.5	-.5	-.5	0	1	0	-.5	.25	0	.25	0
S1 Abs Inf VR	1	-.5	-.5	.5	-1	-1	.5	.5	.25	-.25	-.25	.25
S1 Rel Inf VR	1	-.5	-.5	.5	1	0	-.5	0	.25	.25	0	-.25
S1 Poi Inf VR	1	-.5	-.5	.5	0	1	0	-.5	.25	0	.25	0
S2 Abs 3s VR	1	-.5	0	-.5	-1	-1	.5	.5	0	0	0	0
S2 Rel 3s VR	1	-.5	0	-.5	1	0	-.5	0	0	0	0	0
S2 Poi 3s VR	1	-.5	0	-.5	0	1	0	-.5	0	0	0	0
S2 Abs Inf VR	1	-.5	0	.5	-1	-1	.5	.5	0	0	0	0
S2 Rel Inf VR	1	-.5	0	.5	1	0	-.5	0	0	0	0	0
S2 Poi Inf VR	1	-.5	0	.5	0	1	0	-.5	0	0	0	0
S3 Abs 3s VR	1	-.5	.5	-.5	-1	-1	.5	.5	-.25	.25	.25	.25
S3 Rel 3s VR	1	-.5	.5	-.5	1	0	-.5	0	-.25	-.25	0	-.25
S3 Poi 3s VR	1	-.5	.5	-.5	0	1	0	-.5	-.25	0	-.25	0
S3 Abs Inf VR	1	-.5	.5	.5	-1	-1	.5	.5	-.25	.25	.25	-.25
S3 Rel Inf VR	1	-.5	.5	.5	1	0	-.5	0	-.25	-.25	0	.25
S3 Poi Inf	1	-.5	.5	.5	0	1	0	-.5	-.25	0	-.25	0

VR												
S1 Abs 3s Map	1	.5	-.5	-.5	-1	-1	-.5	-.5	-.25	.25	.25	-.25
...												

Table A59. Model a-VRvMap Matrix.
The fixed effect design matrix used to fit both, VR and Map condition, over session, task and time. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult ~
Expl*Time*Session*MeanRel+Expl*Time*Session*MeanPoi+(1+Time+Session+MeanRel+MeanPoi | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	520	5	468	4
Fixed effects	β0(Intercept) ± SE | CI	52.6 ± 0.44	51.73 53.47		
	β1(Session) ± SE | CI | χ2(1) | p	4.31 ± 0.96	2.42 6.2	18.17	0.00005
	β2(Time) ± SE | CI | χ2(1) | p	5.9 ± 0.85	4.24 7.57	39.42	$3.4*10^{-10}$
	β3(Explore) ± SE | CI | χ2(1) | p	-0.85 ± 0.88	-2.58 0.88	0.87	0.35 (ns)
	β1*β3 ± SE | CI | χ2(1) | p	-1.83 ± 1.93	-5.61 1.96	0.88	0.35 (ns)
Random effects	Intercept σ | CI	1.85	0.85 4.03		
	Session σ | CI	3.53	1.16 10.76		
	time σ | CI	5.42	3.24 9.08		
	Residual variance | CI	7.96	7.14 8.87		

Table A60. Model p-VRvMap.
Results of the LMEM fit including the two last, not significant terms (visual inspection allows both β3 and β1*β3 as effects), of the VR and Map condition

over session and time only for the pointing task. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. $\chi2(1)$ is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. The fitted LMEM formula is: 'CondResult~1+Session+Time+ Exploration + Exploration:Session+(1+Session+Time | Subject)'.

Condition	β0(Intercept)	β1(Session)	β2(Time)	β3(Exploration)	β1(Session)* β3(Exploration)
S1 Poi 3s VR	1	-.5	-.5	-.5	.25
S1 Poi Inf VR	1	-.5	.5	-.5	.25
S2 Poi 3s VR	1	0	-.5	-.5	0
S2 Poi Inf VR	1	0	.5	-.5	0
S3 Poi 3s VR	1	.5	-.5	-.5	-.25
S3 Poi Inf VR	1	.5	.5	-.5	-.25
S1 Poi 3s Map	1	-.5	.-5	.5	-.25
...					

Table A61. Model p-VRvMap Matrix.
The fixed effect design matrix of the LMEM fit including the two last, not significant terms, of the VR and Map condition over session and time only for the pointing task results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is: 'CondResult~Session*Time* Exploration +(1+Session+Time | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	780	7	468	4

Fixed effects	β0(Intercept) ± SE | CI	51.02 ± 0.49	50.05 51.98		
	β1(Session) ± SE | CI | χ2(1) | p	3.92 ± 0.85	2.25 5.6	19.47	0.00001
	β2(Explore) ± SE | CI | χ2(1) | p	3.51 ± 1.01	1.54 5.49	12	0.0005
	β3(AbsPoi) ± SE | CI | χ2(1) | p	-1.52 ± 0.64	-2.79 -0.26	5.36	0.021
	β1*β2 ± SE | CI | χ2(1) | p	5.87 ± 2	1.93 9.8	8.26	0.004
	β2*β3 ± SE | CI | χ2(1) | p	-4.81 ± 1.44	-7.63 -1.99	10.11	0.001
	β1*β2* β3 ± SE | CI | χ2(1) | p	-6.35 ± 3.18	-12.59 -0.11	3.91	0.048
Random effects	Intercept σ | CI	2.89	2.14 3.9		
	Session σ | CI	2.64	0.73 9.63		
	Abs to Poi σ | CI	1.39	0.06 34.36		
	Residual variance | CI	8.31	7.78 8.87		

Table A62. Model s-VRvMap.
Results of the best LMEM fit of the VR and the Map over sessions and tasks for the 3 second time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'Condition~1+Session+ Exploration +AbsPoi+ Session: Exploration + Exploration:AbsPoi+Session:Exploration:AbsPoi+(1+Session|Subject)'.

Condition	β0 (Inter.)	β1 (Session)	β2 (Expl.)	β3(Abs to Poi)	β4(Abs to Rel)	β2*β3	β1*β2	β1*β2*β3
S1 Abs Inf VR	1	-.5	-.5	0	0	0	...	...
S1 Rel Inf VR	1	-.5	-.5	0	1	...		
S1 Poi Inf VR	1	-.5	-.5	1	0			
S2 Abs Inf	1	0	-.5	0	0			

VR								
S2 Rel Inf VR	1	0	-.5	0	1			
S2 Poi Inf VR	1	0	-.5	1	0			
S3 Abs Inf VR	1	.5	-.5	0	0			
S3 Rel Inf VR	1	.5	-.5	0	1			
S3 Poi Inf VR	1	.5	-.5	1	0			
S1 Abs Inf Map	1	-.5	.5	0	0			
...								

Table A63. Model s-VRvMap Matrix.
The fixed effect design matrix used to fit both, VR and Map condition, over session, task and the 3 second time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult ~
Expl*Session*AbsPoi+Expl*Session*AbsRel+(1+Session+AbsRel+AbsPoi | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	780	5	312	3
Fixed effects	β0(Intercept) ± SE | CI	55.78 ± 0.53	54.74 56.81		
	β1(Session) ± SE | CI | χ2(1) | p	7.01 ± 1.04	4.97 9.04	42.28	$7.89*10^{-11}$
	β2(Exploration) ± SE | CI | χ2(1) | p	3.16 ± 1.04	1.12 5.21	9.02	0.003
	β2*β3(Abs to Poi) ± SE | CI | χ2(1) | p	-3.1 ± 1.32	-5.69 -0.51	5.5	0.019
	β1*β4(Abs to Rel) ± SE | CI | χ2(1) | p	-3.15 ± 1.48	-6.06 -0.24	4.49	0.034

Random effects	Intercept σ | CI	4.08	3.34 5		
	Session σ | CI	2.34	0.46 11.9		
	Residual variance | CI	8.69	8.2 9.2		

Table A64. Model u-VRvMap.
Results of the best LMEM fit of the VR and the Map over sessions and tasks for the unlimited time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. $\chi2(1)$ is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'Condition~1+Session+ Exploration +Session:AbsRel+
Exploration:AbsPoi+(1+Session|Subject)'.

Condition	β0 (Interc.)	β1 (Session)	β2 (Expl.)	β3(Abs to Poi)	β4(Abs to Rel)	β1*β4	β2*β3
S1 Abs Inf VR	1	-.5	-.5	0	0	0	…
S1 Rel Inf VR	1	-.5	-.5	0	1	…	
S1 Poi Inf VR	1	-.5	-.5	1	0		
S2 Abs Inf VR	1	0	-.5	0	0		
S2 Rel Inf VR	1	0	-.5	0	1		
S2 Poi Inf VR	1	0	-.5	1	0		
S3 Abs Inf VR	1	.5	-.5	0	0		
S3 Rel Inf VR	1	.5	-.5	0	1		
S3 Poi Inf VR	1	.5	-.5	1	0		
S1 Abs Inf	1	-.5	.5	0	0		

Map							
...							

Table A65. Model u-VRvMap Matrix.
The fixed effect design matrix used to fit both, VR and Map condition, over session, task and the unlimited time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult ~
Expl*Session*AbsRel+Expl*Session*AbsPoi+(1+Session+AbsRel+AbsPoi|
Subject)'.

A.4.4.2. Best-Fit VR versus VR with Belt Models

We provide details on all linear models used for the quantitative analysis of the comparison of VR and VR with belt condition as presented in Results. Table A66 and A67 hold model results and design matrix, respectively, of the fit over experimental exploration, session, task and time. Table A68 and A69 depict results and design matrix of the fit over exploration, session, task and three seconds decision time. Table A70 and A71 contain results and design matrix of the fit over exploration, session, task and unlimited decision time.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	1506	7	760	5
Fixed effects	β0(Intercept) ± SE | CI	51.74 ± 0.33	51.1 52.4		
	β1(Session) ± SE | CI | χ2(1) | p	3.48 ± 0.57	2.36 4.6	36.39	$1.6*10^{-9}$
	β2(Time) ± SE | CI | χ2(1) | p	3.76 ± 0.52	2.75 4.78	44.15	$3*10^{-11}$
	β3(Mean to Abs) ± SE | CI | χ2(1) | p	-0.89 ± 0.34	-1.55 -0.22	6.84	0.0089
	β4(Mean to Poi) ± SE | CI | χ2(1) | p	0.75 ± 0.31	0.14 1.35	5.59	0.018
	β1*β2*β3 ± SE | CI | χ2(1) | p	2.64 ± 1.13	0.41 4.86	5.4	0.02

	β2*β4*β5(Exploration) ± SE | CI | χ2(1) | p	-2.12 ± 1.01	-4.11 -0.14	4.39	0.036
Random effects	Intercept σ | CI	2.45	1.94 3.09		
	Time σ | CI	3.49	2.41 5.07		
	Mean to Abs σ | CI	1.89	1.28 2.8		
	Mean to Poi σ | CI	1.1	0.37 3.29		
	Residual variance | CI	8.01	7.67 8.37		

Table A66. Model a-VRvB.
Results of the best LMEM fit of both, VR and VR with Belt condition, over session, task and time. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Random effect session σ was close to zero (0.02) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The best fit LMEM formula is:
'Condition~1+Session+Time+MeanAbs+MeanPoi+
Session:Time:MeanAbs+Time:MeanPoi:Exploration +
(1+Time+MeanAbs+MeanPoi|Subject)'.
Different from 'Session*Time*MeanAbs', 'Session:Exploration:MeanAbs' only includes the interaction.

Condition	β0 (Inter.)	β1 (Session)	β2 (Time)	β3(Mean to Abs)	β4(Mean to Poi)	β5(Expl.)	β1*β2*β3	β2*β4*β5	β2*β4*β5
S1 Abs 3s	1	-.5	-.5	1	0	-.5	.25	...	...
S1 Rel 3s	1	-.5	-.5	-1	-1	-.5	...		
S1 Poi 3s	1	-.5	-.5	0	1	-.5			
S1 Abs Inf	1	-.5	.5	1	0	-.5			
S1 Rel Inf	1	-.5	.5	-1	-1	-.5			
S1 Poi Inf	1	-.5	.5	0	1	-.5			
S2 Abs 3s	1	0	-.5	1	0	-.5			
S2 Rel 3s	1	0	-.5	-1	-1	-.5			

S2 Poi 3s	1	0	-.5	0	1	-.5			
S2 Abs Inf	1	0	.5	1	0	-.5			
S2 Rel Inf	1	0	.5	-1	-1	-.5			
S2 Poi Inf	1	0	.5	0	1	-.5			
S3 Abs 3s	1	.5	-.5	1	0	-.5			
S3 Rel 3s	1	.5	-.5	-1	-1	-.5			
S3 Poi 3s	1	.5	-.5	0	1	-.5			
S3 Abs Inf	1	.5	.5	1	0	-.5			
S3 Rel Inf	1	.5	.5	-1	-1	-.5			
S3 Poi Inf	1	.5	.5	0	1	-.5			
S1 Abs 3s	1	-.5	-.5	1	-0	.5			
...									

Table A67. Model a-VRvB Matrix.
The fixed effect design matrix used to fit both, VR and VR with Belt condition, over session, task and time. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult ~
Expl*Time*Session*MeanAbs+Expl*Time*Session*MeanPoi+(1+Time+Session+MeanAbs+MeanPoi | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	753	3	304	3
Fixed effects	β0(Intercept) ± SE | CI	49.15 ± 0.48	48.2 50.1		
	β1(Session) ± SE | CI | χ2(1) | p	2.51 ± 0.81	0.92 4.1	9.54	0.002
	β2(Abs to Poi) ± SE | CI | χ2(1) | p	1.39 ± 0.67	0.07 2.7	4.14	0.042
Random effects	Intercept σ | CI	2.62	1.8 3.7		
	Abs to Poi	2.18	0.6 8.15		
	Residual variance | CI	8.28	7.78 8.78		

Table A68. Model s-VRvB.

Results of the best LMEM fit of the VR and the VR with Belt condition over sessions and tasks for the 3 second time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. $\chi2(1)$ is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effect session σ was close to zero (0.01) which leads to numerical instabilities and means one can and should ignore this effect. We therefore did not include it in the model.
The best fit LMEM formula is:
'Condition~1+Session+AbsPoi+(1+AbsPoi|Subject)'.

Condition	β0 (Intercept)	β1 (Session)	β2(Abs to Poi)	β3(Abs to Rel)	β4 (Explor.)
S1 Abs Inf VR	1	-.5	0	0	-.5
S1 Rel Inf VR	1	-.5	0	1	-.5
S1 Poi Inf VR	1	-.5	1	0	-.5
S2 Abs Inf VR	1	0	0	0	-.5
S2 Rel Inf VR	1	0	0	1	-.5
S2 Poi Inf VR	1	0	1	0	-.5
S3 Abs Inf VR	1	.5	0	0	-.5
S3 Rel Inf VR	1	.5	0	1	-.5
S3 Poi Inf VR	1	.5	1	0	-.5
S1 Abs Inf VR with Belt	1	-.5	0	0	.5
...					

Table A69. Model s-VRvB Matrix.
The fixed effect design matrix used to fit both, VR and VR with Belt condition, over session, task and the 3 second time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which

dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult ~
Expl*Session*AbsPoi+Expl*Session*AbsRel+(1+Session+AbsPoi+AbsRel | Subject)'.

Degrees of freedom (df)	#obs | #fix | #rand | #cov	753	4	304	3
Fixed effects	β0(Intercept) ± SE | CI	53.41 ± 0.51	52.41 54.4		
	β1(Session) ± SE | CI | χ2(1) | p	6.06 ± 0.94	4.22 7.9	40.59	$1.9*10^{-10}$
	β2(Abs to Poi) ± SE | CI | χ2(1) | p	1.37 ± 0.64	0.11 2.64	4.53	0.033
	β1*β3(Abs to Rel) ± SE | CI | χ2(1) | p	-4.66 ± 1.44	-7.5 -1.83	10.34	0.001
Random effects	Intercept σ | CI	3.2	2.43 4.2		
	Abs to Poi σ | CI	1.67	0.24 11.43		
	Residual variance | CI	8.08	7.61 8.57		

Table A70. Model u-VRvB.
Results of the best LMEM fit of the VR and the VR with Belt condition over sessions and tasks for the unlimited time condition. #obs means total number of experimental observations. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. Here the random effects session σ (0.02) which leads to numerical instabilities and means one can and should ignore these effects. We therefore did not include them in the model.
The best fit LMEM formula is:
'Condition~1+Session+AbsPoi+Session:AbsRel+(1|Subject)'.

Condition	β0 (Interc.)	β1 (Session)	β2(Abs to Poi)	β3(Abs to Rel)	β4 (Explor.)	β1*β3
S1 Abs Inf VR	1	-.5	0	0	-.5	0
S1 Rel Inf VR	1	-.5	0	1	-.5	-.5
S1 Poi Inf VR	1	-.5	1	0	-.5	0
S2 Abs Inf VR	1	0	0	0	-.5	0
S2 Rel Inf VR	1	0	0	1	-.5	0
S2 Poi Inf VR	1	0	1	0	-.5	0
S3 Abs Inf VR	1	.5	0	0	-.5	0
S3 Rel Inf VR	1	.5	0	1	-.5	.5
S3 Poi Inf VR	1	.5	1	0	-.5	0
S1 Abs Inf VR with Belt	1	-.5	0	0	.5	0
...						

Table A71. Model u-VRvB Matrix.
The fixed effect design matrix used to fit both, VR and VR with Belt condition, over session, task and the unlimited time condition results. The grey columns denote all main or simple effect predictors and their matrix coding, which dropped out in the process of model selection. Note that the maximum model before step down selection included all predictors depicted here as well as all their interactions as fixed effects. The maximum model also included all fixed effect terms except interactions as random effects.
The maximum LMEM formula is:
'CondResult ~
Expl*Session*AbsRel+Expl*Session*AbsPoi+(1+Session+AbsRel+AbsPoi | Subject)'.

A.4.5. Best-Fit Behavioral Models

We provide details on the linear models used for evaluating the effect of independent behavioral variables on performance in all spatial tasks and time conditions and for each embodiment condition. More specifically we only provide details for those models that show a significant effect for the specific behavioral variable. That is, we report those best-fit models that still include the behavioral factor after step down.

The design matrix for each model before step-down was identical to the model for that specific condition without the behavioral factor. Each behavioral variable was either directly normalized and centered or it was first transformed to a normal distribution and then that distribution was normalized and centered before being added to the design matrix. For a simplified example see Table A72.

Condition	β0(Intercept)	β1(Task)	β2(Behavior)	β1(Task)* β2(Behavior)
Task 1 Participant 1	1	0	≈ -.5	0
Task 2 Participant 1	1	1	≈ -.5	≈ -.5
Task 1 Participant 2	1	0	≈ -.43	0
Task 2 Participant 2	1	1	≈ -.43	≈ -.43
…				
Task 1 Participant N	1	0	≈ .5	0
Task 2 Participant N	1	1	≈ .5	≈ .5

Table A72. Simple example of a normalized and centered behavioral factor being added to a design matrix.

A.4.5.1. Best-Fit Map Model Including Percentage of Houses Seen

We provide details on the linear model used for evaluating the effect of the independent behavioral variable “percentage of houses seen” as presented in the Results section. The design matrix was identical to that used for the s-Map model (u-Map model

below), only the behavioral factor was added. Table A73 and A74 hold the model results and transformation of the behavioral variable.

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	348 (36)	8	210	4
Fixed effects	β0(Intercept) ± SE | CI	48.47 ± 0.84	46.8 50.13		
	β1(Poi to Abs) ± SE | CI | χ2(1) | p	4.01 ± 1.24	1.56 6.45	10.19	0.0014
	β2(Poi to Rel) ± SE | CI | χ2(1) | p	6.29 ± 1.24	3.84 8.74	20.88	$5*10^{-6}$
	β1*β3(Session) ± SE | CI | χ2(1) | p	7.0 ± 2.07	2.91 11.09	11.01	0.0009
	β2*β3(Session) ± SE | CI | χ2(1) | p	10.62 ± 2.08	6.53 14.72	23.63	$1*10^{-6}$
	β1*β3* β4(Percent) ± SE | CI | χ2(1) | p	11.79 ± 4.01	3.89 19.68	8.21	0.0041
	β2* β3* β4 ± SE | CI | χ2(1) | p	16.36 ± 4.03	8.43 24.29	13.73	0.00021
	β3* β4 ± SE | CI | χ2(1) | p	16.32 ± 4.99	6.49 26.15	11.01	0.0012
Random effects	Intercept σ | CI	2.53	1.33 4.81		
	Poi to Abs | CI	2.24	0.38 13.08		
	Poi to Rel | CI	2.33	0.49 11.03		
	Residual variance | CI	8.36	7.59 9.2		
Transform function	exponential				

Table A73. Model s-Map.p.
Results of the best fit LMEM of the map condition over session and task and percent of houses seen only for the 3 second time condition. #obs means total number of experimental observations. # miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:

'CondResult~1+PoiAbs+PoiRel+PoiAbs:Session+PoiRel:Session+PoiAbs:Percent+ PoiRel:Percent+Session:Percent+(1+PoiAbs+PoiRel | Subject)'.

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	348 (18)	6	210	4
Fixed effects	β0(Intercept) ± SE | CI	57.51 ± 0.8	55.97 59.04		
	β1(Session) ± SE | CI | χ2(1) | p	8.05 ± 1.46	5.16 10.94	25.93	$4*10^{-7}$
	β2(Percent) ± SE | CI | χ2(1) | p	13.68 ± 3.53	6.73 20.62	13.46	0.00024
	β1*β2 ± SE | CI | χ2(1) | p	35.43 ± 8.93	17.86 53	15.12	0.0001
	β2*β3(Rel to Abs) ± SE | CI | χ2(1) | p	-24.57 ± 11.21	-46.62 -2.51	4.75	0.029
	β2*β4(Rel to Poi) ± SE | CI | χ2(1) | p	-22.59 ± 11.21	-44.64 -0.54	4.03	0.044
Random effects	Intercept σ | CI	4.13	3.05 5.58		
	Session σ | CI	2.87	0.52 15.78		
	Residual variance | CI	8.77	8.04 9.56		
Transform function	exponential				

Table A74. Model u-Map.p.
Results of the best fit LMEM of the Map condition over session and task and percent of houses seen only for the unlimited time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'CondResult~1+Session+Percent+Session:Percent+RelAbs:Percent:Session+RelP oi:Percent:Session+ (1+Session | Subject)'.

A.4.5.2. Best-fit VR Model Including Movement Speed

We provide details on the linear model used for evaluating the effect of the independent behavioral variable "movement speed" as presented in the Results section. The design matrix was identical to that used for the u-VR model, only the behavioral factor was added. Table A75 holds the model results and transformation of the behavioral variable.

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	384 (12)	5	80	2
Fixed effects	β0(Intercept) ± SE | CI	54.13 ± 0.71	52.73 55.53		
	β1(Session) ± SE | CI | χ2(1) | p	7.44 ± 1.37	4.74 10.15	28.17	$1*10^{-7}$
	β2(Abs to Poi) ± SE | CI | χ2(1) | p	2.14 ± 0.95	0.26 4.01	5	0.025
	β1*β3(Abs to Rel) ± SE | CI | χ2(1) | p	-6.24 ± 2.12	-10.41 -2.06	8.5	0.0035
	β3*β4(Speed) ± SE | CI | χ2(1) | p	-3.29 ± 1.66	-6.57 -0.01	3.86	0.049
Random effects	Intercept σ | CI	2.63	1.66 4.18		
	Residual variance | CI	8.48	7.85 9.16		
Transform function	Speed > 2 = .5 Speed < 2 = -.5				

Table A75. Model u-VR.s.
Results of the best fit LMEM of the VR condition over session and task and movement speed only for the unlimited time condition. #obs means total number of experimental observations. # miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'CondResult~1+Session+AbsPoi+AbsRel:Session+AbsRel:Speed+(1 | Subject)'.

A.4.5.3. Best-Fit VR with Belt Model Including Looking Distance

We provide details on the linear model used for evaluating the effect of the independent behavioral variable "looking distance" as presented in the Results section. The design matrix was identical to that used for the s-VRwB model, only the behavioral factor was added. Table A76 holds the model results and transformation of the behavioral variable.

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	354 (3)	5	69	2
Fixed effects	β0(Intercept) ± SE | CI	49.92 ± 0.6	48.73 51.25		
	β1(Poi to Abs) ± SE | CI | χ2(1) | p *β3(Session)	5.38 ± 1.87	1.7 9.1	8.2	0.0042
	β2(Poi to Rel) ± SE | CI | χ2(1) | p *β3(Session)	3.71 ± 1.87	0.04 7.39	3.93	0.047
	β1*β4(Distance) ± SE | CI | χ2(1) | p	14.28 ± 5.49	3.47 25.08	6.68	0.0097
	β2*β4(Distance) ± SE | CI | χ2(1) | p	17.21 ± 5.49	6.4 28.01	9.67	0.0018
Random effects	Intercept σ | CI	2.58	1.51 4.43		
	Residual variance | CI	8.54	7.87 9.25		
Transform function	none				

Table A76. Model s-VRwB.d.
Results of the best LMEM fit of the VR with belt condition over sessions and tasks and looking distance for the 3 second time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'Condition~1+PoiAbs:Session+PoiRel:Session+PoiAbs:Distance+PoiRel:Distance +(1|Subject)'.

A.4.5.4. Best-Fit Models Including Self-Report

We provide details on the linear models used for evaluating the effect of the independent behavioral variable "self-report" as presented in Results section. Self-report splits into three factors, namely general-egocentric, survey and cardinal. We provide those models in which one of the factors showed a significant effect. The order is general, survey and cardinal and three seconds to unlimited for each.

A.4.5.4.1. Best-Fit VR Model Including Self-Report Survey

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	348 (48)	3	74	2
Fixed effects	β0(Intercept) ± SE | CI	48.92 ± 0.48	47.96 49.87		
	β1(Survey) ± SE | CI | χ2(1) | p	5.45 ± 2.52	0.48 10.41	4.40	0.035
	β1* β2(Session)* β3(Abs to Poi) ± SE | CI | χ2(1) | p	21.78	7.18 36.38	8.44	0.0036
Random effects	Survey σ | CI	8.4	4.44 15.9		
	Residual variance | CI	8.12	7.51 8.77		
Transform function	none				

Table A77. Model s-VR.rs.
Results of the best fit LMEM of the VR condition over session and task and self-report survey only for the three seconds time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is:
'CondResult~1+Survey+Survey:Session:AbsPoi+(Survey|Subject)'.

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	348 (48)	6	74	2
Fixed effects	β0(Intercept) ± SE | CI	53.14 ± 0.65	51.85 54.43		
	β1(Session) ± SE | CI | χ2(1) | p	6.37 ± 1.38	3.65 9.09	20.32	$7*10^{-6}$
	β2(Abs to Poi) ± SE | CI | χ2(1) | p	2.81 ± 0.97	0.89 4.73	8.18	0.0042
	β3(Survey) ± SE | CI | χ2(1) | p	11.07 ± 2.70	5.74 16.39	14.69	0.00012
	β1*β4(Abs to Rel) ± SE | CI | χ2(1) | p	-6.39 ± 2.15	-10.64 -2.14	8.65	0.0032
	β1*β3 ± SE | CI | χ2(1) | p	12.36 ± 5.0	2.51 22.21	6.01	0.014
Random effects	Survey σ | CI	7.56	3.86 14.83		
	Residual variance | CI	8.31	7.69 8.97		
Transform function	none				

Table A78. Model u-VR.rs.
Results of the best fit LMEM of the VR condition over session and task and self-report survey only for the unlimited time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'CondResult~1+Survey+Session+AbsPoi+AbsRel:Session+AbsRel:Survey+(Survey | Subject)'.

A.4.5.4.2. Best-Fit VR Model Including Self-Report Cardinal

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	348 (48)	2	74	2
Fixed effects	β0(Intercept) ± SE | CI	48.91 ± 0.50	47.91 49.91		
	β1(Cardinal) ± SE | CI | χ2(1) | p	3.39 ± 1.57	0.29 6.49	4.43	0.035
Random effects	Cardinal σ | CI	6.18	3.28 11.67		
	Residual variance | CI	8.17	7.54 8.85		
Transform function	natural logarithm				

Table A79. Model s-VR.rc.
Results of the best fit LMEM of the VR condition over session and task and self-report cardinal only for the three seconds time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The LMEM formula is: 'CondResult~1+Cardinal+(Cardinal|Subject)'.

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	348 (48)	5	74	2
Fixed effects	β0(Intercept) ± SE | CI	53.40 ± 0.69	52.02 54.77		
	β1(Session) ± SE | CI | χ2(1) | p	6.57 ± 1.41	3.79 9.35	20.85	$5*10^{-6}$
	β2(Abs to Poi) ± SE | CI | χ2(1) | p	2.82 ± 0.98	0.89 4.76	8.16	0.0042
	β1*β3(Abs to Rel) ± SE | CI | χ2(1) | p	-6.89 ± 2.17	-11.17 -2.61	9.87	0.0016
	β3*β4(Cardinal) ± SE | CI | χ2(1) | p	8.10 ± 2.32	3.54 12.67	11.82	0.00058
Random effects	Intercept σ | CI	2.16	1.14 4.09		

	Residual variance | CI	8.36	7.71 9.06		
Transform function	natural logarithm				

Table A80. Model u-VR.rc.
Results of the best fit LMEM of the VR condition over session and task and self-report cardinal only for the unlimited time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. $\chi 2(1)$ is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'CondResult~1+Session+AbsPoi+AbsRel:Session+AbsRel:Cardinal+(1 | Subject)'.

A.4.5.4.3. Best-Fit Map Model Including Self-Report General

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	366 (18)	3	144	3
Fixed effects	β0(Intercept) ± SE | CI	57.06 ± 0.87	55.35 58.78		
	β1(Session) ± SE | CI | $\chi 2(1)$ | p	6.88 ± 1.5	3.92 9.83	19.03	0.000013
	β2(General) ± SE | CI | $\chi 2(1)$ | p	8.36 ± 3.17	2.12 14.59	6.49	0.01
Random effects	Intercept σ | CI	4.79	3.68 6.24		
	Session σ | CI	4.46	1.98 10.04		
	Residual variance | CI	8.83	8.12 9.6		
Transform function	exponential				

Table A81. Model u-Map.rg.
Results of the best fit LMEM of the map condition over session and task and self-report general only for the unlimited time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95%

confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is: 'CondResult~1+Session+General+(1+Session | Subject)'.

A.4.5.4.4. Best-Fit Map Model Including Self-Report Survey

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	366 (18)	4	144	3
Fixed effects	β0(Intercept) ± SE | CI	57.59 ± 0.87	55.88 59.31		
	β1(Session) ± SE | CI | χ2(1) | p	7.57 ± 1.5	4.60 10.54	22.51	$2*10^{-6}$
	β2(Survey) ± SE | CI | χ2(1) | p	11.31 ± 3.59	4.24 18.38	9.08	0.0025
	β1*β2 ± SE | CI | χ2(1) | p	12.26 ± 6.13	0.19 24.32	3.89	0.048
Random effects	Intercept σ | CI	4.66	3.56 6.11		
	Session σ | CI	4.10	1.62 10.35		
	Residual variance | CI	8.84	8.13 9.61		
Transform function	none				

Table A82. Model u-Map.rs.
Results of the best fit LMEM of the map condition over session and task and self-report survey only for the unlimited time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'CondResult~1+Session+Survey+Session:Survey+(1+Session | Subject)'.

A.4.5.4.5. Best-Fit Map Models Including Self-Report Cardinal

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	366 (18)	6	288	5
Fixed effects	β0(Intercept) ± SE | CI	48.04 ± 0.8	46.45 49.63		
	β1(Poi to Abs) ± SE | CI | χ2(1) | p	3.61 ± 1.18	1.29 5.93	9.2	0.0024
	β2(Poi to Rel) ± SE | CI | χ2(1) | p	5.28 ± 1.21	2.89 7.68	16.07	$6*10^5$
	β1*β3(Session) ± SE | CI | χ2(1) | p	5.56 ± 2.07	1.48 9.65	6.96	0.0083
	β2*β3(Session) ± SE | CI | χ2(1) | p	7.61 ± 2.09	3.49 11.74	12.09	0.0005
	β2*β4(Cardinal) ± SE | CI | χ2(1) | p	6.76 ± 2.8	1.25 12.27	5.32	0.021
Random effects	Intercept σ | CI	2.46	1.28 4.71		
	Session | CI	6.89	4.46 10.64		
	Poi to Abs | CI	3.18	1.3 7.75		
	Poi to Rel | CI	3.71	1.95 7.09		
	Residual variance | CI	8.59	7.86 9.38		
Transform function	natural logarithm				

Table A83. Model s-Map.rc.
Results of the best fit LMEM of the map condition over session and task and self-report cardinal only for the three seconds time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'CondResult~1+PoiAbs+PoiRel+PoiAbs:Session+PoiRel:Session+PoiRel:Cardinal+(1+PoiAbs+PoiRel+Session | Subject)'.

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	366 (18)	4	144	3
Fixed effects	β0(Intercept) ± SE | CI	57.56 ± 0.8	55.97 59.14		
	β1(Session) ± SE | CI | χ2(1) | p	7.52 ± 1.37	4.82 10.23	25.93	$4*10^{-7}$
	β2(Cardinal) ± SE | CI | χ2(1) | p	11.97 ± 2.41	7.23 16.71	21.38	$4*10^{-6}$
	β1*β2 ± SE | CI | χ2(1) | p	15.36 ± 4.05	7.39 23.33	13.22	0.00027
Random effects	Intercept σ | CI	4.26	3.21 5.67		
	Session σ | CI	2.54	0.31 20.67		
	Residual variance | CI	8.81	8.1 9.58		
Transform function	natural logarithm				

Table A84. Model u-Map.rc.
Results of the best fit LMEM of the map condition over session and task and self-report cardinal only for the unlimited time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation.
The best fit LMEM formula is:
'CondResult~1+Session+Cardinal+Session:Cardinal+(1+Session | Subject)'.

A.4.5.4.6. Best-Fit VR with Belt Model Including Self-Report Cardinal

Degrees of freedom (df)	#obs (#miss) | #fix | #rand | #cov	336 (21)	4	67	2
Fixed effects	β0(Intercept) ± SE | CI	49.6 ± 0.64	48.33 50.87		

	β1(Poi to Abs) ± SE | CI | χ2(1) | p *β3(Session)	5.54 ± 1.93	1.73 9.35	8.1	0.0044
	β2(Poi to Rel) ± SE | CI | χ2(1) | p *β3(Session)	3.73 ± 1.93	-0.07 7.54	3.69	0.054 (ns)
	β3*β4(Cardinal) ± SE | CI | χ2(1) | p	10.11 ± 3.36	3.49 16.74	8.89	0.0028
Random effects	Intercept σ | CI	2.77	1.64 4.7		
	Residual variance | CI	8.59	7.9 9.34		
Transform function	natural logarithm				

Table A85. Model s-VRwB.rc.
Results of the best LMEM fit of the VR with belt condition over sessions and tasks for the 3 second time condition. #obs means total number of experimental observations. #miss is the number of missing observations due to participants excluded missing behavioral data. #fix means number of fixed effect coefficients. #rand means total number of random effect coefficients. #cov means number of covariance parameters. SE is standard error and CI 95% confidence interval borders. χ2(1) is the likelihood ratio statistic with 1 degree of freedom difference to the model without this fixed effect. σ is the random effect standard deviation. The best fit LMEM formula is:
'Condition~1+PoiAbs:Session+PoiRel:Session+Session:Cardinal+(1|Subject)'.

References

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 43–64.

Altman, D., & Andersen, P. (1989). Bootstrap investigation of the stability of a Cox regression model. *Statistics in medicine*, 771-783.

Andersen, G. J., & Braunstein, M. L. (1985). Induced self-motion in central vision. *Journal of Experimental Psychology: Human Perception and Performance*, 122–132.

Anderson, M. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behavorial Brain Sciences*, 245-313.

Armstrong, K., Fitzgerald, J., & Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron*, 791-798.

Asplund, C., Todd, J., Snyder, A., Gilbert, C., & Marois, R. (2010). Surprise-induced blindness: a stimulus-driven attentional limit to conscious perception. *The Journal of Experimental Psychology: Human Perception and Performance*, 1372-1381.

Bach-y-Rita, P. (1972). *Brain mechanisms in sensory substitution.* San Diego: Academic Press Inc.

Banta Lavenex, P., & Lavenex, P. (2009). Spatial memory and the monkey hippocampus: not all space is created equal. *Hippocampus*, 8-19.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 771-783.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B*, 1281-1289.

Bastos, A., Usrey, W., Adams, R., Mangun, G., Fries, P., & Friston, K. (2012). Canonical microcircuits for predictive coding. *Neuron*, 695-711.

Bellmund, J. L., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*.

Bennett, A. T. (1996). Do animals have cognitive maps? *Journal of Experimental Psychology*, 219-224.

Berthoz, A., Pavard, B., & LR, Y. (1975). Perception of linear horizontal self-motion induced by peripheral vision (linearvection) basic

characteristics and visual-vestibular interactions. *Experimental Brain Research*, 471-489.

Bicanski, A., & Burgess, N. (2018). A neural-level model of spatial memory and imagery. *eLife*.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, 752–762.

Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences*.

Bornstein, M. H., Hahn, C. S., & Suwalsky, J. T. (2013). Physically developed and exploratory young infants contribute to their own long-term academic achievement. . *Psychological science*, 1906–1917.

Brette, R. (2018). Is coding a relevant metaphor for the brain? *Behavioral and Brain Sciences*, 1-44.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 139-159.

Buckner, R., Andrews-Hanna, J., & Schacter, D. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1-38.

Burgess, N. (2007). Spatial Memory: How Egocentric and Allocentric Combine. *Trends in Cognitive Sciences*, 551-557.

Burte, H., & Montello, D. R. (2017). How sense-of-direction and learning intentionality relate to spatial knowledge acquisition in the environment. *Cognitive Research: Principles and Implications*, 2-18.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 1926-1929.

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Review nature neuroscience*, 130-138.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychological review*, 340-75.

Candidi, M., Urgesi, C., Ionta, S., & Aglioti, S. (2008). Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible actions. *Social Neuroscience*, 388-400.

Cao, Y., Summerfield, C., Park, H., Giordano, B. L., & Kayser, C. (2019). Causal Inference in the Multisensory Brain. *Neuron*, 1076-1087.

Carpenter, G., Grossberg, S., Markuzon, N., Reynolds, J., & Rosen, D. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 698-713.

Chadwick, M. J., Jolly, A. E., Amos, D. P., Hassabis, D., & Spiers, H. J. (2015). A Goal Direction Signal in the Human Entorhinal/Subicular Region. *Current Biology*, 87-92.

Chemero, A. (2009). *Radical embodied cognitive science.* Cambridge: MIT Press.

Chen, X., McNamara, T. P., Kelly, J. W., & Wolbers, T. (2017). Cue combination in human spatial navigation. *Cognitive Psychology*, 105-144.

Chicca, E., Stefanini, F., Bartolozzi, C., & Indiveri, G. (2014). Neuromorphic Electronic Circuits for Building Autonomous Cognitive Systems. *Proceedings of the IEEE*, 1367-1388.

Chrastil, E. R. (2013). Neural evidence supports a novel framework for spatial navigation. *Psychonomic Bulletin & Review*, 208–227.

Chrastil, E. R., & Warren, W. H. (2011). Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review*, 1-23.

Chrastil, E. R., & Warren, W. H. (2013). Active and passive spatial learning in human navigation: Acquisition of survey knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1520-1537.

Churchland, P. S., Ramachandran, V. S., & Sejnowsk, T. J. (1994). A critique of pure vision. In C. K. Davis, *Computational neuroscience. Large-scale neuronal theories of the brain* (pp. 23-60). Cambridge: The MIT Press.

Cisek, P., & Kalaska, J. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience*, 269-298.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Science*, 345-351.

Clay, V., König, P., & König, S. (2019). Eye Tracking in Virtual Reality. *Journal of Eye Movement Research*.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* New Jersey: Lawrence Erlbaum Associates Publishers.

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing Conceptual Knowledge in Humans with a Grid-like Code. *Science*, 1464–1468.

Corneil, D., & Gerstner, W. (2015). Attractor network dynamics enable preplay and rapid path planning in maze-like environments. *Proceedings of the 28th International Conference on Neural Information Processing Systems* (pp. 1684–1692). Montreal: MIT Press.

Costall, A. (1995). Socializing affordances. *Theory & Psychology*, 467–481.

Daniel, D. H., & Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content.* Cambridge: MIT Press.

deCharms, R. C., & Zador, A. (2000). Neural representation and the cortical code. *Annual Review of Neuroscience*, 613–647.

Dechent, P., Merboldt, K.-D., & Frahm, J. (2004). Is the human primary motor cortex involved in motor imagery? *Cognitive Brain Research*, 138–144.

Dennett, D. (1991). *Consciousness Explained.* New York: Little, Brown and Co.

Derdikman, D., Whitlock, J., Tsao, A., Fyhn, M., Hafting, T., Moser, M.-B., & Moser, E. I. (2009). Fragmentation of grid cell maps in a multicompartment environment. *Nature Neuroscience*, 1325–1332.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. . *Annual review of psychology*, 115–142.

Devereux, B., Clarke, A., Marouchos, A., & Tyler, L. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, 18906-18916.

Diemer, J., Alpers, G., Peperkorn, H., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology*.

Dixon, M. L., De La Vega, A., Mills, C., Andrews-Hanna, J., Spreng, R. N., Cole, M. W., & Christoff, K. (2018). Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. *Proceedings of the National Academy of Sciences*, 1598-1607.

Dixon, M., Andrews-Hanna, J., Spreng, R., Irving, Z. C., Mills, C., Girn, M., & Kalina, C. K. (2017). Interactions between the default network and dorsal attention network vary across default

subsystems, time, and cognitive states. . *Neuroimage*, 632-649.

Doeller, C., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, 657-661.

Dora, S., Bohte, S. M., & Pennartz, C. (2020). Deep predictive coding accounts for emergence of complex neural response properties along the visual cortical hierarchy. *bioRxiv*.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, 51-86.

Eager, C., & Roy, J. (2017). Mixed Effects Models are Sometimes Terrible. *arXiv:1701.04858*.

Eger, E., Sterzer, P., Russ, M., Giraud, A., & Kleinschmidt, A. (2003). A supramodal number representation in human intraparietal cortex. *Neuron*, 719-725.

Ekstrom, A. D., Arnold, A. E., & Laria, G. (2014). A critical review of the allocentric spatial representation and its neural underpinnings: toward a network-based perspective. *frontiers in Human Neuroscience*.

Ekstrom, A., Kahana, M., Caplan, J., Fields, T. A., Isham, E. A., Newman, E., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, 184-188.

Engel, A., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Science*, 202-209.

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nature Neuroscience*, 1504-1513.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 162-169.

Etzel, J. A., Gazzola, V., & Keysers, C. (2008). Testing Simulation Theory with Cross-Modal Multivariate Classification of fMRI Data. *PLoS One*.

Evans, G. W., & Pezdek, K. (1980). Cognitive mapping: Knowledge of real-world distance and location information. *Human Learning and Memory*, 13-24.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 1149–1160.

Federico, G., & Brandimonte, M. (2019). Tool and object affordances: An ecological eye-tracking study. *Brain and Cognition*.

Feigenbaum, J. D., & Rolls, E. T. (1991). Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. *Psychobiology*, 21-40.

Felleman, D., & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1-47.

Fields, A. W., & Shelton, A. L. (2006). Individual Skill Differences and Large-Scale Environmental Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, pp. 506-515.

Filimon, F. (2015). Are All Spatial Reference Frames Egocentric? Reinterpreting Evidence for Allocentric, Object-Centered, or World-Centered Reference Frames. *frontiers in Human Neuroscience*.

Frankenstein, J., Mohler, B. J., Bülthoff, H. H., & Meilinger, T. (2012). Is the Map in Our Head Oriented North? *Psychological science*, 120-125.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 1211-1221.

Fuchs, E., & Flügge, G. (2014). Adult neuroplasticity: More than 40 years of research. *Neural Plasticity*.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Science*, 493-501.

Gallivan, J., McLean, D., Valyear, K., Pettypiece, C., & Culham, J. (2011). Decoding action intentions from preparatory brain activity in human parieto-frontal networks. *Journal of Neuroscience*, 9599-9610.

Gardony, A. L., Brunyé, T. T., & Taylor, H. A. (2015). Navigational aids and spatial memory impairment: The role of divided attention. *Spatial Cognition and Computation*, 246–284.

Georgopoulos, A., Schwartz, A., & Kettner, R. (1986). Neuronal population coding of movement direction. *Science*, 1416-1419.

Ghassabian, A., Sundaram, R., Bell, E., Bello, S., Kus, C., & Yeung, E. (2016). Gross Motor Milestones and Subsequent Development. *Pediatrics*.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. T. (1950). *The Perception Of The Visual World.* Boston: Houghton Mifflin.

Gibson, J. T. (1979). The Theory of Affordances. In J. T. Gibson, *The Ecological Approach to Visual Perception* (pp. 127-137). Boston: Houghton Mifflin.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences* , 1-19.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 558-565.

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, 959–978.

Gönner, L., Vitay, J., & Hamker, F. (2017). Predictive Place-Cell Sequences for Goal-Finding Emerge from Goal Memory and the Cognitive Map: A Computational Model. *Frontiers in Computational Neuroscience*.

Gramann, K. (2013). Embodiment of Spatial Reference Frames and Individual Differences in Reference Frame Proclivity. *Spatial Cognition & Computation*, 1-25.

Grattan, L. M., Bloomer, R. H., Archambault, F. X., & Eslinger, P. J. (1994). Cognitive flexibility and empathy after frontal lobe lesion. *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, 251-259.

Greene, M., & Oliva, A. (2009). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive psychology*, 137-176.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 121-134.

Grossberg, S. T. (2012). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control.* Berlin: Springer Science & Business Media.

Haber, S., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, 4-26.

Hadders-Algra, M. (2018). Early human motor development: From variation to the ability to vary and adapt. *Neuroscience & Biobehavioral Reviews*, 411-427.

Hadjikhani, N., Liu, A., Dale, A., Cavanagh, P., & Tootell, R. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience*, 235-241.

Hafting, T., Fyhn, M., Molden, S., Moser, M., & Moser, E. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 801-806.

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature*, 147-150.

Hardless, G., Meilinger, T., & Mallot, H. (2015). Virtual Reality and Spatial Cognition. In J. Wright, *International Encyclopedia of the Social Behavioral Sciences* (pp. 133-137). Amsterdam: Elsevier Science.

He, Q., & Brown, T. I. (2019). Environmental Barriers Disrupt Grid-like Representations in Humans during Navigation. *Current Biology*, 2718-2722.

He, Q., McNamara, T. P., & Brown, T. I. (2019). Manipulating the visibility of barriers to improve spatial navigation efficiency and cognitive mapping. *Scientific Reports*.

He, Q., McNamara, T. P., Bodenheimer, B., & Klippel, A. (2018). Acquisition and transfer of spatial knowledge during wayfnding. *Journal of experimental phsychology: Learning, memory and cognition*.

Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory.* Hoboken: Wiley.

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal*, 431-449.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 872–876.

Hesslow, G. (2002). Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Science*, 242-247.

Hohwy, J. (2013). *The predictive mind.* Oxford: Oxford University Press.

Huffman, D., & Ekstrom, A. (2019). Which way is the bookstore? A closer look at the judgments of relative directions task. *Spatial Cognition and Computation*, 93-129.

Hurley, S. (1998). Vehicles, Contents, Conceptual Structure, and Externalism. *Analysis*, 1-6.

Hutcheon, B., & Yarom, Y. (2000). Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends in Neuroscience*, 216-222.

Hutto, D. D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content.* Cambridge: MIT Press.

Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content.* Cambridge: MIT Press.

Ibrahim, J. G., & Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. *TEST*, 1-43.

Ishikawa, T., & Montello, D. (2006). Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive psychology*, 93-129.

Jackman, S., & Regehr, W. (2017). The Mechanisms and Functions of Synaptic Facilitation. *Neuron*, 447-464.

Jackman, S., & Regehr, W. (2017). The Mechanisms and Functions of Synaptic Facilitation. *Neuron*, 447-464.

Jacquet, P. O., & Avenanti, A. (2015). Perturbing the action observation network during perception and categorization of actions' goals and grips: State-dependency and virtual lesion TMS effects. *Cerebral Cortex*, 598–608.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 434-446.

Jamone, L., Ugur, E., Cangelosi, A., Fadiga, L., Bernardino, A., Piater, J., & Santos-Victor, J. (2016). Affordances in Psychology, Neuroscience, and Robotics: A Survey. *IEEE Transactions on Cognitive and Developmental Systems*.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, 103-109.

K, O. J., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* , 939 –1031.

Kadosh, R. C., Kadosh, K. C., Kaas, A., Henik, A., & Rainer, G. (2007). Notation-Dependent and -Independent Representations of Numbers in the Parietal Lobes. *Neuron*, 307-314.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, 4302-4311.

Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T. I., & Logothetis, N. K. (2020). Decoding the contents of consciousness from prefrontal ensembles. *bioRxiv*.

Kar, K. K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, 974–983.

Kaspar, K., König, S., Schwandt, J., & König, P. (2014). The experience of new sensorimotor contingencies by sensory augmentation. *Consciousness and Cognition*, 47-63.

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile. *arXiv*.

Kayser, C., Petkov, C., & Logothetis, N. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, 1560-1574.

Keller, H. (1929). *Midstream: My Later Life.* New York: Doubleday, Doran & Company.

Kietzmann, T. C., Spoerer, C., Sörensen, L. K., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences of the United States of America*, 21854-21863.

Killian, N., Jutras, M., & Buffalo, E. (2012). A map of visual space in the primate entorhinal cortex. *Nature*, 761-764.

Klatzky, R. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In *Spatial Cognition* (pp. 1-17). Berlin: Springer.

Klein, S. (2015). What memory is. *WIREs Cognitive Science*, 1-38.

Koffka, K. (1935). *Principles of Gestalt Psychology.* New York: Harcourt, Brace.

König, P. (2017, Jun. 13). *Embodied Cognition* [Keynote Talk]. RTG Situated Cognition Opening Workshop, Bochum, Germany.

König, S. U., Clay, V., Nolte, D., Duesberg, L., Kuske, N., & König, P. (2019). Learning of Spatial Propiertes of a Large-Scale Virtual City With an Interactive Map. *frontiers in Human Neuroscience*.

König, S. U., Goeke, C., Meilinger, T., & König, P. (2019). Are allocentric spatial reference frames compatible with theories of Enactivism? *Psychological Research*, 498-513.

König, S., U. Schumann, F., Keyser, J., Goeke, C., Krause, C., Wache, S., & König, P. (2016). Learning new sensorimotor contingencies: Effects of long-term use of sensory augmentation on the brain and conscious perception. *PLoS ONE*.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience. *Frontiers in Systems Neuroscience*.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., . . . Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 1126-41.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought.* New York: Basic Books.

Latuske, P., Kornienko, O., Kohler, L., & Allen, K. (2018). Hippocampal Remapping and Its Entorhinal Origin. *Frontiers in Behavorial Neuroscience*.

Lawrence, C. (2017). *Amazon Picking Challenge*. Retrieved from https://amazonpickingchallenge.org/.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. *Nature*, 293–294.

Leitner, J. (2018). *ICRA 2018 Tidy Up My Room Challenge*. Retrieved from http://juxi.net/challenge/tidy-up-my-room/.

Lenay, C., Canu, S., & Villon, P. (1997). Technology and perception: the contribution of sensory substitution systems. *Proceedings Second International Conference on Cognitive Technology Humanizing the Information Age* (pp. 44-53). Aizu-Wakamatsu: IEEE Computer Society.

Leutgeb, S., Leutgeb, J., Barnes, C., Moser, E., McNaughton, B., & Moser, M. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 619-623.

Lüchinger, A., Hadders-Algra, M., van Kan, C., & de Vries, J. (2008). Fetal onset of general movements. *Pediatric Research*, 191-195.

Lynch, K. (1960). *The Image Of The City.* Cambridge: The M.I.T. Press.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S., Frith, C. D., & O'Keefe, J. (1998). Knowing Where and Getting There: A Human Navigation Network. *SCIENCE*, 921-923.

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2014). Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nature neuroscience*, 1598–1606.

Markov, N. T., Ercsey-Ravasz, M., Lamy, C., Ribeiro Gomes, A. R., Magrou, L., Misery, P., . . . Van Essen, D. (2013). The role of

long-range connections on the specificity of the macaque interareal cortical network. *Proceedings of the National Academy of Sciences*, 5187-5192.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, 337–341.

McNaughton, B., Battaglia, F., Jensen, O., Moser, E., & Moser, M. (2006). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 663-678.

Meilinger, T. (2008). The Network of Reference Frames Theory: A Synthesis of Graphs and Cognitive Maps. In C. Freksa, N. Newcombe, P. Gärdenförs, & S. Wölfl, *Spatial Cognition VI. Learning, Reasoning, and Talking about Space: International Conference Spatial Cognition 2008, Freiburg, Germany, September 15-19, 2008. Proceedings.* Freiburg: Springer.

Meilinger, T. (2008). The Network of Reference Frames Theory: A Synthesis of Graphs and Cognitive Maps. In P. Gärdenfors, N. Newcombe, C. Freksa, & S. Wolfl, *Spatial Cognition VI. Learning, Reasoning, and Talking about Space* (pp. 344-360). Freiburg: Springer.

Meilinger, T., & Vosgerau, G. (2010). Putting Egocentric and Allocentric into Perspective. In C. Hölscher, T. F. Shipley, M. O. Belardinelli, J. A. Bateman, & N. S. Newcombe, *Spatial Cognition VII. Spatial Cognition 2010. Lecture Notes in Computer Science* (pp. 207-221). Berlin, Heidelberg: Springer.

Meilinger, T., Frankenstein, J., Watanabe, K., Bülthoff, H. H., & Hölscher, C. (2015). Reference frames in learning from maps and navigation. *Psychological Research*, 1000–1008.

Meilinger, T., Riecke, B. E., & Bülthoff, H. H. (2014). Local and global reference frames for environmental spaces. *The Quarterly Journal of Experimental Psychology*, 542-569.

Meister, I., Wilson, S., Deblieck, C., Wu, A., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 1692-1696.

Meletti, S., Benuzzi, F., Cantalupo, G., Rubboli, G., Tassinari, C., & Nichelli, P. (2009). Facial emotion recognition impairment in chronic temporal lobe epilepsy. *Epilepsia*, 1547-1559.

Melnik, A., Hairston, D. W., Ferris, D., & König, P. (2017). EEG correlates of sensorimotor processing: Independent Components involved in sensory and motor processing. *Scientific Reports*.

Merleau-Ponty, M. (1945). *Phenomenology of Perception.* Paris: Éditions Gallimard, Routledge & Kegan Paul.

Michael, L., Hesselmann, G., Kiefer, M., & Niedeggen, M. (2011). Distractor-induced blindness for orientation changes and coherent motion. *Vision Research*, 1781–1787.

Millikan, R. (1991). Perceptual content and Fregean myth. *Mind*, 439-459.

Millikan, R. (1995). Pushmi-pullyu representations. *Philosophical Perspectives*, 185-200.

Moffat, S. D., Hampson, E., & Hatzipantelis, M. (1998). Navigation in a ‘‘Virtual’’ maze: sex differences and correlation with psychometric measures of spatial ability in humans. *Evolution and Human Behavior*, 73–87.

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 1543-1546.

Montello, D. R. (1998). A New Framework for Understanding the Acquisition ofSpatial Knowledge in Large-Scale Environments. In M. &. R.G.Golledge, *Spatial and temporal reasoning in geographic information systems* (pp. 143-154). New York: Oxford University Press.

Moser, S. (1988). Cognitive Mapping in a Complex Building. *Environment and Behaviour*, 21–49.

Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review*, 654–661.

Müller, S., Scealy, J. L., & Welsh, A. H. (2013). Model Selection in Linear Mixed Models. *Statistical Science*, pp. 135-167.

Münzer, S., & Hölscher, C. (2011). Entwicklung und validierung eines fragebogens zu räumlichen strategien. *Diagnostica*, pp. 111-125.

Münzer, S., Fehringer, B., & Kühl, T. (2016). Standardized norm data for three self-report scales on egocentric and allocentric environmental spatial strategies. *Data in Brief*, pp. 803-811.

Müri, R., & Nyffeler, T. (2008). Using transcranial magnetic stimulation to probe decision-making and memory. *Progress in Brain Research*, 413-418.

Murray, G., Veijola, J., Moilanen, K., Miettunen, J., Glahn, D. C., Cannon, T. D., . . . Isohanni, M. (2006). Infant motor development is associated with adult cognitive categorisation in a longitudinal birth cohort study. *Journal of Child Psychology and Psychiatry*, 25-29.

Murtaugh, P. A. (2009). Performance of several variable-selection methods applied to real ecological data. *Ecology Letters*, pp. 1061-1068.

Nagel, S. K., Carl, C., Kringe, T., Märtin, R., & König, P. (2005). Beyond sensory substitution – Learning the sixth sense. *Journal of Neural Engineering*, 13-26.

Newhouse, P., Newhouse, C., & Astur, R. S. (2007). Sex differences in visual-spatial learning using a virtual water maze in prepubertal children. *Behavioural Brain Research*, 1–7.

Nguyen-Vo, T., Riecke, B. E., Stuerzlinger, W., Pham, D., & Kruijff, E. (2019). NaviBoard and NaviChair: Limited Translation Combined with Full Rotation for Efficient Virtual Locomotion. *IEEE Transactions on Visualization and Computer Graphics*, 1-12.

Niedeggen, M., Michael, L., & Hesselmann, G. (2012). Closing the gates to consciousness: distractors activate a central inhibition process. *Journal of Cognitive Neuroscience*, 1294-1304.

Noble, R., Tasaki, K., Noble, P. J., & Noble, D. (2019). Biological Relativity Requires Circular Causality but Not Symmetry of Causation: So, Where, What and When Are the Boundaries? *Frontiers in physiology*.

Norman, D. (1988). *The Design of Everyday Things.* New York: Basic Books.

Nunn, J., Gregory, L., Brammer, M., Williams, S. C., Parslow, D. M., Morgan, M. J., . . . Gray, J. A. (2002). Functional magnetic resonance imaging of synesthesia: activation of V4/V8 by spoken words. *Nature Neuroscience*, 371-375.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavior and Brain Sciences*, pp. 939 – 973.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 171-175.

Ooms, K., De Maeyer, P., & Fack, V. (2013). Study of the attentive behavior of novice and expert map users using eye tracking. *Cartography and Geographic Information Science*, pp. 37-54.

Parent, A., & Hazrati, L. (1995). Functional anatomy of the basal ganglia. II. The place of subthalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Research Reviews*, 128-154.

Parkin, A. J. (1996). *Explorations in Cognitive Neuropsychology.* Hove: Psychology Press.

Parsons, T., Gaggioli, A., & Riva, G. (2017). Virtual Reality for Research in Social Neuroscience. *Brain Science*.

Pearce, J., Roberts, A., & Good, M. (1998). Hippocampal lesions disrupt navigation based on cognitive maps but not heading vectors. *Nature*, 75-77.

Pennartz, C. (2015). *The Brain's Representational Power.* Cambridge: MIT Press.

Perkel, D. H., & Bullock, T. H. (1968). Neural coding. *Neurosciences Research Program Bulletin*, 221–348.

Pfeiffer, B., & Foster, D. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 74-79.

Pierce, C. S. (1906). *The Basis of Pragmaticism.* Harvard: Robin Catalogue.

Popal, H., Wang, Y, & Olson, I. (2019). A Guide to Representational Similarity Analysis for Social Neuroscience. *Social Cognitive and Affective Neuroscience*, 1243–1253.

Poucet, B. (1993). Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. *Psychological Review*, 163–182.

Price, M., & Anderson, P. (2007). The role of presence in virtual reality exposure therapy. *Journal of anxiety disorders*, 742–751.

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, 129-154.

Raj, A., Cai, C., Xie, X., Palacios, E., Owen, J., Mukherjee, P., & Nagarajan, S. (2020). Spectral graph theory of brain oscillations. *Human Brain Mapping*, 2980-2998.

Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R., & Khaligh-Razavi, S.-M. (2019). Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *Plos Computational Biology*.

Ramsey, W. (2017). Must cognition be representational? *Synthese*, 4197-4214.

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition*, 741-750.

Riecke, B. E., Feuereissen, D., Rieser, J. J., & McNamara, T. P. (2015). More than a cool illusion? Functional significance of self-

motion illusion (circular vection) for perspective switches. *Frontiers in Psychology*.

Rodriguez, E., George, N., Lachaux, J., Martinerie, J., Renault, B., & Varela, F. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 430-433.

Rose, N., LaRocque, J., Riggall, A., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, 1136-1139.

Roth, R. (2006). *Introduction to Coding Theory.* Cambridge: Cambridge University Press.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, pp. 581-592.

Ruddle, R., Volkova, E., Mohler, B., & Buelthoff, H. H. (2011). The effect of landmark and body-based sensory information on route knowledge. *Memory and Cognition*, 686 - 699.

Rump, B. R., & McNamara, T. P. (2013). Representations of interobject spatial relations in long-term memory. *Memory & Cognition*, 201–213.

S., G. (1976). Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. *Biological Cybernetics*, 187-202.

Save, E., Cressant, A., Thinus-Blanc, C., & Poucet, B. (1998). Spatial firing of hippocampal place cells in blind rats. *Journal of Neuroscience*, 1818-1826.

Schiller, D., Eichenbaum, H., Buffalo, E., Davachi, L., Foster, D., Leutgeb, S., & Ranganath, C. (2015). Memory and Space: Towards an Understanding of the Cognitive Map. *THE JOURNAL OF NEUROSCIENCE*, 13904-13911.

Schlicht, T., & Starzak, T. (2019). Prospects of enactivist approaches to intentionality and cognition. *Synthese*.

Schumann, F., & O'Regan, J. K. (2017). Sensory augmentation: integration of an auditory compass signal into human perception of space. *Scientific Reports*.

Schwartz, A., Moran, D., & Reina, G. (2004). Differential representation of perception and action in the frontal cortex. *Science*, 380-383.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 379–423.

Shapiro, K., Raymond, J., & Arnell, K. (1997). The attentional blink. *Trends in Cognitive Science*, 291-296.

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

Shelton, A. L., & McNamara, T. P. (2004). Orientation and Perspective Dependence in Route and Survey Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 158-170.

Shelton, F., & Reding, M. (2001). Effect of lesion location on upper limb motor recovery after stroke. *Stroke*, 107-112.

Shepard, R. N., & Hurwitz, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition*, 161-193.

Sherman, S. (2016). Thalamus plays a central role in ongoing cortical functioning. *Nature neuroscience*, 533-541.

Sholl, M. J., Kenny, R. J., & DellaPorta, A. K. (2006). Allocentric-heading recall and its relation to self-reported sense-of-direction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 516-533.

Siegel, A. W., & White, S. H. (1975). The Development of Spatial Representations of Large-Scale Environments. *Advances in Child Development and Behavior*, 9-55.

Sievers, B., Parkinson, C., Kohler, P. J., Hughes, J., Fogelson, S. V., & Wheatley, T. (2018). Visual and auditory brain areas share a neural code for perceived emotion. *bioRxiv*.

Soto, V. (2019). *Examining Face-sensitive Brain Potentials in Natural Environments Using Mobile EEG.* Liverpool: University of Liverpool.

Spiegler, A., Knösche, T., Schwab, K., Haueisen, J., & Atay, F. (2011). Modeling brain resonance phenomena using a neural mass model. *PLoS Computational Biology*.

Spoerer, C., McClure, P., & Kriegeskorte, N. (2017). Recurrent Convolutional Neural Networks: A Better Model of Biological Object Recognition. *Frontiers in Psychology*.

Steck, S. D., & Mallot, H. A. (1997). *The Role of Global and Local Landmarks in Virtual Environment Navigation.* Tuebingen.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 652–654.

Stuss, D., Floden, D., Alexander, M., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: dissociation of

processes and frontal lobe lesion location. *Neuropsychologia*, 771-786.

Teuber, H.-L. (1955). Physiological psychology. *Annual Review of Psychology*, 267-296.

Thelen, E., Ulrich, B. D., & Niles, D. (1987). Bilateral coordination in human infants: Stepping on a split-belt treadmill. *Journal of Experimental Psychology: Human Perception and Performance*, 405–410.

Thompson, E., & Varela, F. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Science*, 418-425.

Thorndyke, P. W., & Hayes-Roth, B. (1980). *Differences in spatial knowledge acquired from maps and navigation.* Office of Naval Research.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 267-288.

Timm, D., & Papenmeier, F. (2019). Reorganization of spatial configurations in visual working memory. *Memory & Cognition*, 1469–1480.

Tolman, E. C. (1948). Cognitive Maps in Rats and Men. *The Psychological Review*, 189-208.

Tukey, J. W., & Green Jr., B. W. (1960). Complex Analyses of Variance: General Problems. *Psychometrika*, 127-152.

Väljamäe, A. (2009). Auditorily-induced illusory self-motion: a review. *Brain Research Reviews*, 240-255.

van Asselen, M., Fritschy, E., & Postma, A. (2006). The influence of intentional and incidental learning. *Psychological Research*, 151-156.

van der Groen, O., & Wenderoth, N. (2016). Transcranial Random Noise Stimulation of Visual Cortex: Stochastic Resonance Enhances Central Mechanisms of Perception. *Journal of Neuroscience*, 5289-5298.

van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M., Poort, J., van der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proceedings of the National Academy of Sciences*, 14332-14341.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge: MIT Press.

Vass, L., & Epstein, R. (2017). Common Neural Representations for Visually Guided Reorientation and Spatial Imagery. *Cerebral Cortex*, 1457-1471.

von Senden, M. (1960). *Space and sight: the perception of space and shape in the congenitally blind before and after operation.* Oxford: Free Press of Glencoe.

Waller, D., Loomis, J. M., & Haun, D. B. (2004). Body-based senses enhance knowledge of directions in large-scale environments. *Psychonomic Bulletin & Review*, 157-163.

Warren, W. H. (2019). Non-Euclidean navigation. *Journal of Experimental Biology*.

Warren, W. H., Rothman, D. B., Schnapp, B. H., & Ericson, J. D. (2017). Wormholes in virtual space: From cognitive maps to cognitive graphs. *Cognition*, 152-163.

Warren, W., Rothman, D., Schnapp, B., & Ericson, J. (2017). Wormholes in virtual space: From cognitive maps to cognitive graphs. *Cognition*, 152-163.

Weisberg, S. M., & Newcombe, N. S. (2016). How do (some) people make a cognitive map? Routes, places, and working memory. *Journal of experimental psychology. Learning, memory and cognition.*, 768-785.

Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Epstein, R. (2014). Variations in Cognitive Maps: Understanding Individual Differences in Navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 669-682.

White, B. W., Saunders, F. A., Scadden, L., Bach-Y-Rita, P., & Collins, C. C. (1970). Seeing with the skin. *Perception & Psychophysics*, 23–27.

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Journal of Royal Statistics Society*, 392–399.

Wilming, N., Murphy, P. R., Meyniel, F., & Donner, T. H. (2020). Disentangling Decision-related Feedforward and Feedback Signals in Human Cortex. *bioRxiv*.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 625-636.

Witherington, D. C. (2011). Taking emergence seriously: The centrality of circular causality for dynamic systems approaches to development. *Human Development*, 66–92.

Wolbers, T., & Wiener, J. M. (2014). Challenges for identifying the neural mechanisms that support spatial navigation: the impact of spatial scale. *frontiers in Human Neuroscience*, 1-12.

Yamanobe, N., Wan, W., Ramirez-Alpizar, I. G., Petit, D., Tsuji, T., Akizuki, S., . . . Harada, K. (2018). A Brief Review of Affordance in Robotic Manipulation Research. *Journal of the Robotics Society of Japan*.

Zhang, H., Zherdeva, K., & Ekstrom, A. D. (2014). Different “routes” to a cognitive map: dissociable forms of spatial knowledge derived from route and cartographic map learning. *Memory & Cognition*, 1106-1117.