Eberhard Karls University of Tübingen
Faculty of Science

Bachelor Thesis Cognitive Science

Is there a canonical orientation in novel scenes?

Gwen Hirsch

March 2018

Reviewer

Prof. Dr. Hanspeter A. Mallot
Cognitive Neuroscience
Department of Biology
University of Tübingen

Supervisor

Banafsheh Grochulla
Cognitive Neuroscience
Department of Biology
University of Tübingen

Hirsch, Gwen:
Is there a canonical orientation in novel scenes?
Bachelor Thesis Cognitive Science
Eberhard Karls University of Tübingen
Period: December 1, 2017 – March 31, 2018

Erklärung

Hiermit erkläre ich, dass ich diese schriftliche Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe.

Datum, Ort, Unterschrift

Acknowledgements

I would like to thank my reviewer Prof. Dr. Hanspeter A. Mallot who helped me develop the question of my thesis and design the experiment and who resolved any unclarities I had. Furthermore, I thank my supervisor Banafsheh Grochulla for showing me how to work with an Oculus Rift headset and for readily answering all my questions about programming and conducting the experiment while still granting me autonomy in my decisions.

I am grateful that I got the chance to complete my bachelor thesis in the Cognitive Neuroscience Lab and to gain insights about the organization and dynamics of this work group. The advice of other group members and the stimulating inputs and discussions in the weekly Students- and Lab Meeting helped and taught me a lot as well.

Last, but not least, I would like to thank my fellow bachelor students for the helpful conversations we had in the bachelor meeting and my friends and family for all the ways in which they supported me.

Abstract

In order for humans to be able to orient themselves as quickly and accurately as possible, a categorization of the visible surroundings is necessary. One characteristic of spatial representations that has been proposed to facilitate scene recognition is the canonicality of specific orientations. Canonical orientations might result from intrinsic cues in a scene, such as the layout geometry, salient features, landmarks or main axes. In this case, they might be identifiable in novel scenes. Our prediction was that, in a task where participants were asked to memorize goal orientations in close proximity to a previously identified canonical orientation, there would be a bias towards this canonical orientation in later recall. Participants wore an Oculus Rift headset and were able to freely turn around inside a virtual reality, where we presented full 360° panorama pictures folded into cylinders around them to make orientation feel as natural as possible. The results revealed no bias towards the preidentified canonical orientation, although participants' own accounts on what they used for orientation seemed to be in line with the hypothesis. Two explanations for this result pattern which we considered to be equally probable were (1) that canonical orientations in novel scenes do exist, but our prediction of a bias towards them was wrong, or (2) that canonical orientations in novel scenes do not exist and that the phenomenon of canonicality only emerges in familiar scenes as a characteristic of spatial representations in long–term memory, formed on the basis of prior experiences with the locations.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Theoretical and empirical background

In order to be able to successfully orient themselves or navigate through various kinds of environments, people have to perceive, process and evaluate their surroundings. These evaluations partly depend on external cues of the environment itself, but also on the conscious knowledge of ones current position and how to achieve the goal of reaching ones destination. To integrate this knowledge with ones perception, cognitive categorization processes are required to form some sort of representation of the surrounding space and frame of reference based on the perceptual input. We are interested in investigating the nature of these spatial representations. Some important questions to this end are: How do we select visual input in order to form a spatial representation? What is captured by early scene recognition processes to allow for scene categorizations? How is memory involved, i.e., are there different stages of spatial representations? In what ways do the formed representations influence practical applications like orientation and navigation? And what are essential characteristics of spatial representations that affect our interaction with the outside world? In the following, we will summarize some of the research that has been done in the field of spatial cognition so far with the objective of finding answers to these questions, draw connections to the topic of our own study and specify our hypothesis.

In the feature integration theory, put forth by Treisman and Gelade (1980), the perception of a whole scene is thought to be succesively built up by first filtering separate features from our visual input, then identifying associated components and shapes to recognize individual objects and finally integrating these objects through focused attention to arrive at the complexity level of the scene. This traditional conceptualization has since been challenged by findings where the categorization of scenes is sometimes faster than would be expected by this theory: Oliva and Schyns (1997) proposed that the categorization of complex visual stimuli like scenes does not necessarily have to take place after recognition at the object level is completed. They distinguish between precise and more coarse visual categorization processes, which operate at fine spatial scales that have a high spatial frequency profile or coarse spatial scales with a low spatial frequency profile, respectively. They argue that the categorization task at hand might be able to influence perception top–down to flexibly work at the most informative spatial scale, which they refer to as the *diagnostic scale*. So for example, the task of precisely discriminating between similar objects relies on detecting fine scale boundary edges, while categorization of the type of scene one is looking at requires the processing of coarser resolutions, i.e., blobs. The opposing view to this is that coarse structure is recognized before fine structure in general, analogous to the time course of perceptual availability of coarse before fine structure (for psychophysical evidence see Schyns and Oliva, 1994). In their experiments, Oliva and Schyns were able to show that indeed, categorization interacted with the perceptual processing level in the way that a previously sensitized spatial scale will bias the processing of a stimulus towards favoring information at that scale for recognition, even when the stimulus ambiguously contains multiple spatial scales with meaningful information. In other words, expectations in a high–level categorization task can influence the selection of a diagnostic scale for further processing from the multiple spatial scales which are simultaneously registered in low–level perception.

The fact that object recognition is not essential for a reliable scene categorization is further elaborated by Oliva (2005), where she discusses what she calls the *gist* of a scene. Even short presentations of novel scenes, permitting only a glimpse, allow the observer to extract enough perceptual and semantic information to comprehend its meaning and consistently classify it into a category. Oliva and Torralba (2006) introduce the concept of global image feature detection as a mechanism that works in parallel to the local image analysis and actually facilitates it by narrowing down the number of different kinds of objects to be expected for processing in a particular environment. With this, they provide a model for how a scene's structural layout can be represented independently of object recognition in early processing stages in a similar fashion to the hypothesized precedence of global before local feature analysis in the two–dimensional plane of an image (Navon, 1977). The basic idea is that a global image feature encodes the spatial relationships of a scene in the form of activity configurations of local feature detectors, thereby summarizing the feature distributions of an image. From the representations built out of global image features, observers are able to derive the so–called *spatial envelope* of the scene. It includes information about the spatial layout in terms of properties such as openness, expansion or naturalness and surface characteristics (e.g., smooth, rough) together with a semantic description that enables categorization both on higher–order levels (general categories, e.g., natural, urban) and on more basic levels (specific categories, e.g., forest, town square). The model of the spatial envelope gives an intuitive understanding of how gist of a scene perception in humans might function.

Basten, Meilinger and Mallot (2012) focus their attention on the distinction between *ego-* and *allocentric* spatial representations in working or long–term memory and on whether or not these are oriented. Representations of space in working memory constructed from the perceptual input are *egocentric*, meaning they are defined with respect to ones position and orientation relative to the surroundings. *Allocentric* representations in long–term memory on the other hand are defined by environmental spatial relations, i.e., do not depend on a persons current location or orientation. The authors let participants of their study draw sketches of a familiar place either from memory or with the task to imagine travelling to this place, thereby priming the orientation of the mentally travelled path. Without mental travel, there was a consistently preferred orientation which the authors hypothesized might result from intrinsic cues of the layout geometry, a salient landmark (a church in this case) or a geographical slant. With mental travel, the primed orientation was sketched most often. They concluded that orientation priming led to the retrieval of egocentric views from the long–term into the working memory. Röhrich, Hardiess and Mallot (2014) used a similar sketching task in their study: Passersby at various locations were asked to draw a nearby target location from memory. Results revealed that drawn orientations of the target place strongly depended on participants' current location, which was interpreted as recall being based on oriented working memory representations. The authors also proposed a view-based model of spatial long-term memory extending the concept of the *view graph* proposed by Schölkopf and Mallot (1995). In this model, allocentric representations of spaces consist of a collection of snapshot–like views of a space from previous egocentric experiences with it, which can be loaded into working memory if required. Furthermore, views can be represented multiple times resulting in a population code of the different views of a scene. A view will be represented more often due to salience; and the most salient view or a number of most salient views might be refered to as the *canonical* view(s) – we will elaborate more on this later on. Possibilities for selection criteria determining canonicality of a view given by Röhrich et al. were the presence and distribution of landmarks, geometric layout, salience of visible objects and distinctiveness to other places.

Mou and McNamara (2002) connect the concept of ego– and allocentric representations to that of frames of reference as a means of interpreting and representing our surrounding environment. Their experimental results were in line with their hypothesis that environments have identifiable intrinsic axes which are more salient than others and are used to organize our spatial representations, as opposed to egocentric experience imposing a frame of reference onto the environment. These intrinsic reference systems are selected based on the spatial structure of the surroundings (e.g., the walls of a building or a room), certain spatial and nonspatial properties of objects contained in ones view (e.g., axes of bilateral symmetry or grouping effects such as proximity and similarity) and also based on perception of salience due to egocentric experience. Participants of their study were instructed to learn arrangements of objects inside either a rectangular or round room from a nonegocentrical perspective and make judgements of relative directions afterwards. The judgments of relative directions that were aligned or orthogonal to the intrinsic reference frame were faster and more accurate than for other headings and participants were able to represent the layout nonegocentrically. Two subsequent studies replicated these general findings: (1) for a large-scale space (Mcnamara, Rump, & Werner, 2003) and (2) in the absence of the explicit identification of the intrinsic axes through instructions, proving that intrinsic cues of the layout geometry suffice for the selection of an intrinsic frame of reference and, remarkably, the first egocentric perspective was not the dominant cue to this end (Mou, Zhao, & McNamara, 2007). These findings, among others (e.g., Mallot, Röhrich, & Hardiess, 2014), support the notion of orientation dependence of spatial representations.

In an influential study from the related field of object recognition, the phenomenon of an interindividually consistent preference of certain views of objects over others coined the term of the canonical view (Palmer, Rosch, & Chase, 1981). In Blanz, Tarr and Bülthoff (1999), some of the specific attributes characterizing canonical views were further investigated in a design that allowed participants to adjust their viewpoint of three–dimensional computer models of common and nonsense objects nearly without restrictions. In the first experiment, participants were given the task to adjust the view as if they wanted to take a picture of the object for use in a brochure and in the second experiment they should adjust the view as they had imagined the object. The results showed clusterings of viewpoints mostly at three–quarter views that avoided occlusions in the first experiment and more frontal and side view clusterings in the second. Factors that contributed the most to the canonicality of a view were geometrical factors, including the salience, relevance and visibility of features and the stability of a given view with respect to small shifts, as well as the contextual factors familiarity (frequency of encounters) and functionality (how well a view illustrates an objects functional part); although functionality was less predictive of canonicality than the other factors. The contribution of the contextual factors was confirmed by the absence of a canonical view in nonsense objects, since they were unfamiliar and had no clear function. The authors concluded that canonical views, formed on the basis of observers' experiences and geometrical properties of the objects, play an important role in understanding the nature of object representations.

The idea of canonical views was later picked up by Ehinger and Oliva (2011), who attempted to apply the same concept to scenes. In reference to analyses of photo databases that largely depicted the same views for the same locations, they thought it likely that certain views in a scene are particularly characteristic and facilitate recognition. Their approach was to present full 360° panorama pictures in an on–screen window, which could be dragged with a computer mouse to change the view. Participants were told to select the view they thought was the best snapshot of the scene in each image and designate a name to the image. The authors' hypothesis was that participants would try to maximize the visible space of a scene and prioritize views which are best suited for navigational purposes (i.e., streets, paths); either directly because of this functional value, or because of a high familiarity due to it, or both. The results showed a high agreement on the best view between participants, which was interpreted as generally pointing towards the existence of canonical views. The agreement was higher for scenes containing variations in viewing distances than for large open spaces and if subjects agreed on the name, they tended to agree on the best view as well. As Ehinger and Oliva had expected, subjects tried to maximize the visible space, but the presumed bias towards navigationally relevant views was not confirmed. They speculated that navigation could be too general as a function and views with functional aspects that are more context–sensitive to the scene might overlap more with canonical views. Other possible reasons for view selection they considered were the amount or variety of visible objects, the diagnosticity of certain objects inside a given scene or aesthetic reasons (e.g., the degree of symmetry). In summary, according to the authors' findings, scenes and spaces are likely to be represented in terms of prefered – *canonical* – views.

After this introductory review of existing conceptualizations and research regarding spatial representations, we have learned several things: We considered how scene recognition processes might operate (flexibly, at a task–dependent diagnostic scale) and the terms in which scene representations might be organized (in terms of a spatial envelope). Furthermore, we considered the role played by egocentric working memory and allocentric long–term memory representations and how humans utilize these representations to perform orientation and navigation tasks. We saw evidence that identifiable intrinsic frames of reference in scenes organize our spatial representations, which supports the idea of orientation dependence in our representations. At last, we had a look at findings regarding the canonicality of views both in objects and scenes. On the basis of this theoretical and empirical background, we now want to pose a new question that has not yet been thoroughly explored. In the next section, we will proceed by formulating our hypothesis, predicting possible consequences we think might result from it and define the goal we try to achieve in this study.

1.2 Hypothesis and goal of this study

Our interest in this study lies in further investigating the nature of spatial representations; more precisely, whether specific characteristics that influence how we interact with surrounding spaces can be singled out. Some of the aforementioned studies assume there to be some sort of privilege of certain orientations over others in the representation of a location (Basten et al., 2012; Ehinger and Oliva, 2011; Mallot et al., 2014; McNamara et al., 2003; Mou and McNamara, 2002; Mou et al., 2007; Röhrich et al., 2014). A singular most privileged orientation in a scene has been called canonical – analogous to canonical views in mental representations of objects (Palmer et al., 1981; Blanz et al., 1999). It would make sense if different kinds of mental representations based on visual input would share similar characteristics such as canonicality, but they likely would not depend equally upon the same factors, because of differences in what makes certain contents more informative than others for a successful object or scene recognition and because different processes seem to be at work for global and local image analysis (Oliva and Torralba, 2006). Another difference between object and scene recognition is that views of objects are centered in the object in the sense that a change in view is typically accompanied by a positional shift of the observer or turns of the object, whereas scene recognition is centered in the observer in the sense that changes in orientation are accomplished by turns of the observer inside of the scene. To account for these potential differences in canonicality in objects and scenes, we decided to use the term *canonical orientation* for scenes.

We pose the question whether canonical orientations can be identified in novel scenes. One possibility of how a privilege for an orientation might manifest itself, suggested by Röhrich et al. (2014), is that more copies of this orientation are stored in long–term memory. But it is not clear whether the canonicality of an orientation emerges due to the number of stored copies of one orientation or the other way around – that intrinsic qualities of certain orientations lead to the storing of more copies. The question then comes down to whether canonicality is already established in early perception based on the salience or diagnosticity of certain orientations or if canonicality is a characteristic of allocentric representations, thus no privilege for certain orientations over others should exist for novel scenes yet. Because unfamiliar places do not yet have allocentric representations in long–term memory, they are suitable to distinguish between these possibilities. We think that the formation of allocentric spatial representations is driven by intrinsic qualities of the scenes themselves, in other words, that perception will be biased towards sampling orientations which maximize the information content or are particularly salient more often. This would explain the high levels of agreement on the best snapshot of a scene in the experiment conducted by Ehinger and Oliva (2011). One possible problem of their expriment is that people's understanding of what constitutes a good snapshot of a scene might not necessarily be equal to the canonical orientation in case of its existence, as Blanz et al. (1999), who used the same direct measure, also pointed out. Answers might have been biased due to aesthetic reasons or photographic contents encountered in everyday life (e.g., inside newspapers, brochures or internet websites).

Our hypothesis is that canonical orientations exist and can be identified for novel scenes, given that some orientations are more salient (e.g., contain prominent landmarks). If this is true, we predict that there will be a bias towards this canonical orientation in a memory task. In our experiment, we used an indirect measure to test this hypothesis. Participants had the task to memorize goal orientations other than a previously identified canonical orientation in novel scenes and recall them afterwards. The measure was the bias in degrees towards or away from this canonical orientation. In order to create an environment which would feel as realistic and natural as possible for participants to interact with, we used 360° panorama pictures put onto a cylinder in a virtual reality. Participants wore an Oculus Rift headset that allowed them to freely look around inside the virtual cylinders simply by turning their heads. This created the impression of actually being at the locations. The criteria we used to identify the canonical orientations in each scene were based on the suggested factors of the above reviewed studies, especially Ehinger and Oliva (2011), Mou et al. (2007) and Röhrich et al. (2014), namely available prominent landmarks, salient features, layout geometry and intrinsic axes. The used panorama pictures differed in the amount of visible space, meaning that the most diagnostic scale which would be optimal for scene recognition was not the same fore each scene. We added a noise pattern to the images in the encoding phase, in which participants were supposed to memorize a goal orientation. We did this to make fine spatial scales harder to detect and prime visual processing towards favouring courser spatial scales to make participants rely more on forming a spatial envelope of the scene than on local feature analysis. The goal orientations were chosen such that the identified canonical orienation was visible in the encoding phase and should ideally have been represented in participants' perception of the gist of the scene, leading to a bias in this direction in the recall phase. Goal orientations were either a short or wide angle away from the canonical orientation and we expected the bias to point consistently towards the canonical orientation in both conditions, albeit maybe a stronger bias for the goal orientations closer to the canonical orientation.

Prior to the actual experiment, we conducted a pilot study that had the same basic setup as the final experiment with the intent of checking if everything worked according to our desire and to see whether first results seemed to support our idea, while still being able to modify and fine–tune some details if necessary. The expected bias did not seem to emerge, but since the number n of participants was very small ($n = 3$), the data could not be analyzed in detail, although the experience of the pilot study helped us find some factors that we thought might have confounded the results (they are mentioned in 2.2). A few alterations of some specifics of the experiment were made to eliminate them. The goal of the present study is to see if there will be a bias towards a previously identified canonical orientation in novel scenes. This would (1) support the idea of the existence of canonical orientations in itself and (2) provide evidence that canonicality is based on intrinsic factors of a scene affecting the formation of allocentric representations by guiding gaze behaviour to sample prefered orientations more often (which might be how canonicality manifests itself in long–term memory), as opposed to canonicality emerging as a result of the formation of allocentric representations.

2 Methods

2.1 Participants

All except one of the subjects participating in the study were students of the University of Tübingen. Of ten participants in total, four were female and six male. Their age ranged from 21 to 28 years old ($M = 23.9$ years). All but one person lived or had previously lived inside of Tübingen or in close proximity for an average duration of 2.65 years. All had normal or corrected to normal vision. There was no monetary compensation for the participation in the experiment, but if they so desired, subjects could get course credit for the time the experiment took for certain psychology or cognitive science modules, but none of them made use of this. All subjects were naive as to the purpose of the study.

2.2 Materials and Stimuli

In preparation of the experiment, several panorama pictures were taken in the surrounding area of Tübingen. Three of those, together with an additional artificial computer image, were selected to be one of four stimuli presented in the experiment. Each image had a resolution of 24576×3919 pixels. This rather small number of chosen images was due to the need for an appropriate number of repetitions in each condition per location and the external limitation that the experiment should not take too long or else signs of fatigue might have confounded the results. Two of the four panorama pictures served as controls; one of them showed a homogenous forest scene with a bridge on one side, from now on refered to as the Forest location (FO). The four tested orientations (see 2.3.2 for further explanation) were directed away from this bridge, where no one outstanding feature was visible, only trees. For this image we expected no bias, because the tested orientation was not identified as being canonical for the scene. The other control was the computer image which was comprised of a black triangle on a blue background, from now on refered to as the Pyramid location (PY). The tip of the pyramid was identified as the canonical orientation. In case of an existing bias, we expected it to turn out the strongest for this image, because the pyramid, or more precisely, the tip and bottom of the pyramid, were the only features participants could use for orientation. We added this picture after we found no bias in the pilot study with the idea that all possible confounding factors of the expected effect caused by intrinsic image features that might arise in natural images would be eliminated in this artificial image. The remaining two locations were selected based on how unambiguously and consistently we were able to identify a direction we considered to be the canonical orientation of the scene. In both cases we settled for scenes which contained church towers as distinctive landmarks, which we identified as the canonical orientations – one of them from inside of the village Bebenhausen (BE), the other outside of Hagelloch (HA) from a viewpoint overlooking the village with the church tower sticking out as the largest structure. We expected the bias to be the second largest for the HA location after the PY location, because the village Hagelloch looked like a rather homogenuous patch apart from the church tower that stood out, making it the most probable feature to use for orientation. The bias was expected to be the smallest for the BE location, because there were many salient edges, for example from church windows, and more distinguishable objects like trees, streets and fences that could be used for orientation by participants.

In the encoding phase of a trial (see 2.3.2 for detailed description), a noise pattern, specifically Gaussian white noise with a mean value of 0.4 and a variance of 0.8, was added to the images to make it more difficult for participants to only focus on prominent basic features like vertical edges in the center instead of being forced to take in and memorize the whole presented view. This degradation was also added after the pilot study, because the memorization seemed to be too easy a task for subjects without some further impediment, so an existing bias towards the canonical orientation might have been missed. The four images together with their noisy counterparts, which served as the stimuli of the experiment, are depicted in figure 2.1.

The experiment was conducted using a virtual environment created with the gaming engine Unity, version 5.5.1f1, in conjunction with an Oculus Rift headset. The panorama pictures were put onto the inner surface of a cylinder that was $200 \times 200 \times 133$ Unity length units in size. Some of the code for the experiment used parts from previous experiments from the Cognitive Neuroscience work group. Participants were seated on a revolving chair, which they could rotate without obstructions. The laptop that ran the experiment was mounted on a metal platform attached to the backrest of the chair. The right arm of the participants rested on the computer mouse on another metal platform attached to the right side of the chair; by clicking the left mouse button, they proceeded from instruction texts to experimental trials. The subjects wore the Oculus Rift headset and inside the virtual environment were placed into the center of one of the cylinders where they were able to see the stimulus on its inner surface. They could look around by turning their heads or the chair. The environment was made more realistic by adding a blue cloudy sky above the cylinder. The bottom was left plain to prevent participants from using ground features for orientation.

Figure 2.1: Control and test stimuli presented in the encoding and recall phase of a trial (see 2.3.2 for an explanation). Acronyms are added to the top left corners. The preidentified canonical orientations were the respective centers of these images. The noise pattern was added only in the encoding phase. In the image versions of the recall phase, one can see the number scales at the top and bottom which were used to indicate remembered orientations.

At the beginning of the experiment, participants were instructed about the experimental procedure after which they were able to ask questions and afterwards gave their written consent to participate. In the end they answered a questionnaire (see appendix) that gathered personal information about participants' current residence, about the duration they had (previously) lived inside Tübingen and about familiarity with the three natural locations to establish that they were sufficiently novel to the participants. Finally, they were asked if they had consciously devised strategies to solve the tasks or actively used certain features for orientation. The total duration of the experiment was approximately one hour.

2.3 Procedure

2.3.1 Familiarization phase

The experiment consisted of two phases. In the familiarization phase, participants had to find small colored geometric shapes (a circle, a square and a triangle) that were placed somewhere inside the scenes and report the corresponding numbers on a numberscale that was added to the top and bottom of the images. Each of the four locations was presented two times resulting in a total of eight trials and three shapes had to be found per image. This was done simply to make sure the participants saw the whole of each novel scene by making them fully turn and look around in each scene in order to find the shapes, because in the pilot study participants did not have an incentive to take in the whole scenes, leading to incomplete representations. There was no time limit for finding the three shapes in an image.

2.3.2 Experimental phase

The second phase was the experimental phase, where the task was to memorize orientations and reproduce them afterwards. One trial likewise consisted of two phases – the encoding and the recall phase. In the encoding phase, at the beginning of each trial, an instruction to memorize the following scene was shown. By a leftclick of the mouse, the instruction disappeared and one of the four noisy images was presented in one of four orientations for a duration of 500 milliseconds (ms). This short exposure was chosen to make memorization more challenging for participants. The four orientations were defined relative to the previously identified canonical orientation as either a short angle of $\pm 15°$ distance away from the canonical orientation or a wide angle of $\pm 30°$ distance away from it. The angle distances were chosen to be in close proximity to the canonical orientation in hope of exposing the expected bias even if the effect size would be small. The other reason was that the canonical orientation should still be visible from the goal orientation, and the field of vision when wearing an Oculus Rift headset is limited to 80° in width. The stimulus was presented statically, meaning that when participants would turn their heads, they would not have been able to look around the environment so that the goal direction was defined unambiguously. Because of the short presentation duration, this made no big difference from the participants' perspective, but they were still asked to not turn their heads in the encoding phase to prevent symptoms of motion sickness which may occur while wearing an Oculus Rift headset. The task in this trial phase was to memorize the displayed orientation for later recall.

In a second stage of the encoding phase, participants had to complete a secondary working memory task. The image from the start was followed by a blackscreen and after 500 ms one of the capital letters "E" or "F" appeared centered in participants' field of vision for the duration of 100 ms, again followed by a blackscreen for 500 ms. The task was to memorize the letter identity in addition to the previous orientation in the scene. There were three main reasons a secondary task was added: First of all, there was the need to have a separation between the encoding and recall phase, otherwise finding the goal orientation would have been too simple and there would have been a larger probability of overlooking the first presented image. The second reason was to make it impossible for participants to complete the task by holding an afterimage in their minds and then simply finding the orientation that matched this afterimage, thereby eliminating the need to orient themselves in the scene. This could have also been achieved by displaying a mask after the first view presentation, but participants might have devised the strategy to actively close their eyes after the first view was shown after realizing that their memory performance would benefit from not paying attention to the mask. The last reason was to put additional load on the subjects working memories, for one to make the task more difficult, but also to prompt them to actively orient themselves in the scenes and not rely on keeping local features of the center of the presented view in their working memories.

In the recall phase of a trial, participants were first given the instruction to find the former orientation, then, after a leftclick of the mouse, were placed inside the same location as before. This time they were able to look around freely and no noise pattern was added to the pictures. The shown orientation was defined relative to the goal direction as ±25° away from it ± a 5° jitter which ensured that subjects would not turn their heads a fixed distance every time, which they might have come to realize. Here, their task was to recall the goal orientation and verbally report the number that corresponded to the center of the previously displayed orientation on a numberscale added to the top and bottom of the images. There were strokes for each of 360° on the numberscale and every fifth number was indicated. In addition, the numberscale had one of 16 possible shifts, to prevent participants from remembering the

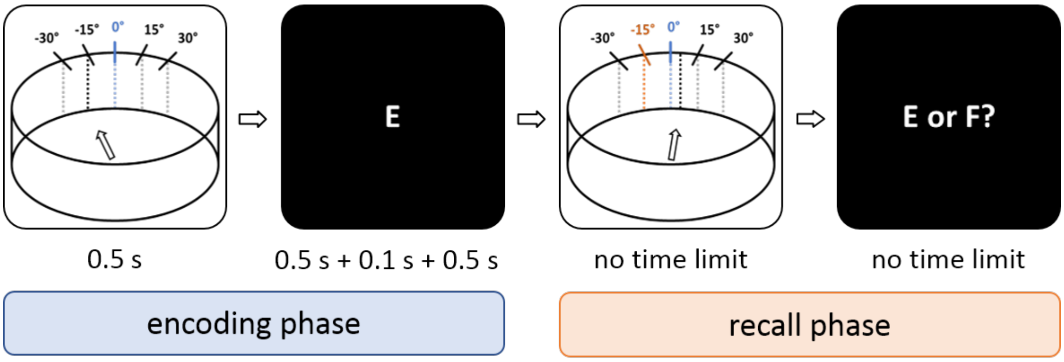

Figure 2.2: Schematic course of one trial. In the encoding phase, participants were supposed to remember one of four possible orientations ($-15°$ in this example) relative to the preidentified canonical orientation (defined as $0°$). One of the letters "E" or "F", presented for 0.1 s before and after the presentation of a blackscreen for 0.5 s, had to be remembered as well. In the recall phase, participants had to reproduce the memorized orientation and letter identity.

numbers corresponding to the four goal orientations. Finally, after giving their number responses, the participants should verbally indicate the identity of the letter which was shown. A typical course of one trial is depicted schematically in Figure 2.2. In order to make sure that participants payed attention to the letters and did not make random guesses instead, they were asked to try to keep their accuracy level above 90% correct answers.

2.4 Design

The experiment had a 4×2–within–subjects design that had the independent variables location with the four conditions FO, BE, HA and PY and angle of the encoding phase view with the two conditions short and wide angle. For each of the eight combinations of location and angle times two sides of the angle condition relative to the canonical orientation, there were eight repetitions – four on each side of the participants viewing direction in the recall phase relative to the goal orientation. Four of those eight repetitions had a small recall angle distance from the goal orientation of $\pm 20°$ and the other four a large angle distance of $\pm 30°$. This resulted in $4 \times 2 \times 2 \times 2 \times 4 = 128$ total trials per person plus four additional practice trials at the beginning to learn the task. The 128 test trials were divided into four blocks with 32 trials each. After completing a block, participants could take a short break and were provided feedback of the percentage of accurate letter responses so far. Every block contained all of the 32 combinations of location, angle, side in relation to the canonical orientation and side relative to the goal orientation in randomized order.

3 Results

3.1 Questionnaire results

In a questionnaire after the experiment, participants were asked to rate their familiarity of the three natural scenes FO, BE and HA on a scale from 1, meaning "I have not seen this place before." to 7, meaning "I have seen this place before and know where it is.". If they thought they knew the places, they were asked to tell which places were depicted in a follow–up question. In case the answers were incorrect, we adjusted their own rating to 4. This stood for maximal uncertainty in either direction as opposed to knowing that one knows or does not know the place, but still accounted for participants' impression of familiarity. We wanted to make sure that they were sufficiently unfamiliar with the locations to exclude possible allocentric long–term memory representations to interfere with our endeavor of investigating the canonicality in novel scenes. A Mann–Whitney test revealed familiarity ratings to be significantly lower than 4 for the FO location ($Mdn = 1$), $U = 4$, $p = .008$, the BE location ($Mdn = 3$), $U = 5$, $p = .020$, and the HA location ($Mdn = 2.5$), $U = 3$, $p = .035$. The results are illustrated in Figure 3.1.

Participants were also asked whether they had consciously devised strategies to solve the tasks during the experiment or used specific details of the scenes for orientation. Two said they had just in general looked for prominent landmarks or vertically oriented objects such as trees, windows or houses. In the FO scene, seven out of ten said they had tried to recognize prominent trees or certain branches. For both the BE and the HA scenes, seven participants stated they had consciously used the church tower for orientation. In the BE scene, participants additionally said to have used features like parts of the church building, fences or trees and in the HA scene they used certain distinctive houses of the village or bushes in the distance. In the PY scene, two said they had oriented themselves after the tip of the pyramid. For the letter memorization, one person used the strategy to internally repeat it, another had laid their hand on the right or left leg depending on the letter identity.

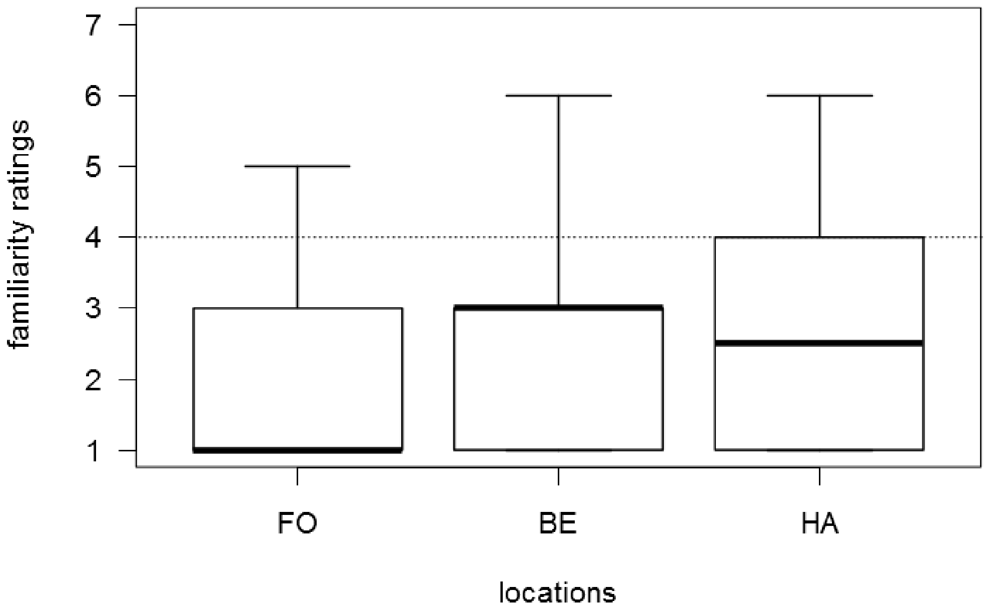

Figure 3.1: Boxplots of familiarity ratings of the three natural scenes. A rating of 1 meant that participants did not know the place and 7 that they knew it and could tell where it is. We tested whether the ratings were significantly lower than 4, which corresponded to maximal uncertainty.

3.2 Data Analysis of the experiment

The number responses from the familiarization phase were not analyzed, because the purpose of this phase was simply to make participants actively look around and familiarize themselves with the scenes. Nonetheless we exmined the responses to ensure that all participants were able to properly report numbers on the number scale that corresponded to the directions of three geometric shapes. This turned out to be the case for all participants (disregarding small deviations of one or two degrees).

The accuracy of letter responses was recorded to check whether performances were above the chance level of 50 % correct responses. Under the null hypothesis of participants randomly guessing the letters we calculated a one–tailed binomial test for the dichotomous variable of letter correctness for each participant. All accuracy rates were significantly higher than chance level with an average of 88.4 % correct letter responses. The accuracy rates and detailed test results for each participant are enlisted in the appendix, Table A1.

Angle responses that were outside a range of ±40° in regards to the goal orientation (outside the field of vision when wearing an Oculus Rift headset) were excluded from further analysis, with the reasoning that orientations outside of this range could not have been remembered, because they were not visible in the encoding phase. This led to the exclusion of 3.75 % of trials from further analysis. The remaining data was then tested for a normal distribution in each experimental condition using Shapiro-Wilk tests. Some of these tested significant, indicating a violation of the assumption of a normal distribution; the detailed results are appended in Table A2. Because of this we decided to use non-parametric tests for the following analysis.

We performed a Friedman test of differences among the mean biases relative to the canonical orientation for the four locations and two angle conditions. The test result turned out not significant ($\chi^2(3) = 5.4$, $p = .144$). After the comparison between the experimental conditions, we also tested whether the mean biases of each condition differed significantly from 0°, using Mann–Whitney tests. Even though the Mann–Whitney test is nonparametric, the analysis was with respect to the mean biases as opposed to the median of the biases, because biases measured in degrees are on an interval–scale. The test results are presented in Table 3.1 and the mean biases are visualized in Figure 3.2. Surprisingly, out of five significant biases, only two point in the direction of the canonical orientation, namely in the cases of the wide angle conditions of the FO location and the BE location. The case of the FO location is noteworthy because we expected there to be no bias at all for this location. Another result that was especially surprising, was that there was not only no significant bias towards the canonical orienation in the PY location, but a significant bias

Table 3.1: Results of the Mann–Whitney tests to see whether mean biases towards or away from the canonical orientation were significant. Positive mean values correspond to a bias towards the canonical orientation, negative mean values correspond to a bias away from it. Significant mean biases are marked with one star for $p \leq .05$, two stars for $p \leq .01$ and three stars for $p \leq .001$.

location	angle condition	U	p	M (degrees)	SD (degrees)
FO	short	19	.215	-1.2	3.6
	wide	52	.004**	2.4	2.7
BE	short	7	.018*	-1.2	2.2
	wide	50	.009**	1.1	1.5
HA	short	10	.041*	-2.2	4.6
	wide	15	.116	-1.1	2.5
PY	short	0	<.001***	-5.8	4.7
	wide	43	.063	0.6	1.8

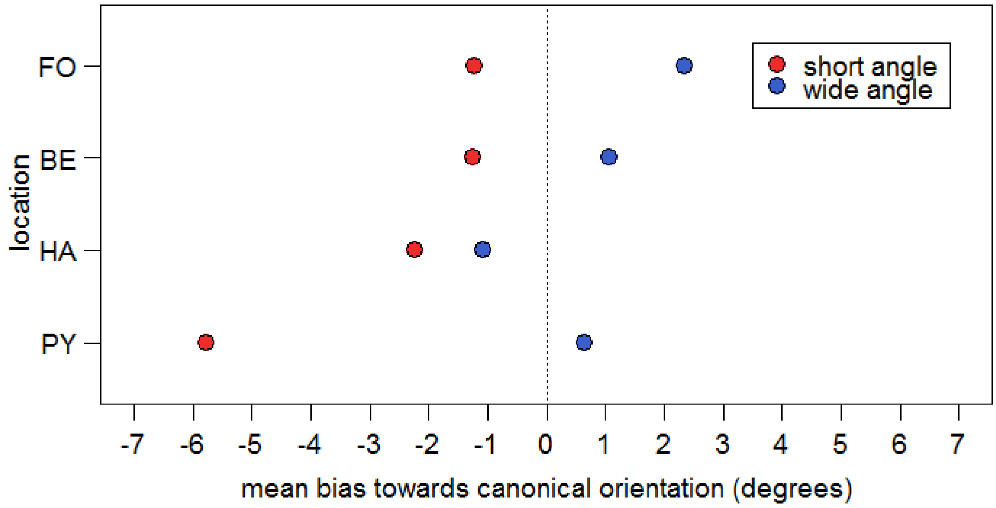

Figure 3.2: For each location the mean bias towards the canonical orientation is shown. Which of these are significant is detailed in Table 3.1. The differences between locations are not significant.

away from it for the short angle condition with a mean bias of -5.6°, which also happened to be the largest mean deviation of all. At last we wanted to see whether and how the two angle conditions short and wide differed for each side of the canonical orientation. Again, we used Mann–Whitney tests to check whether the mean biases for each of the four individual goal orientations (left wide, left short, right short and right wide) differed significantly from 0°. The angle response distributions and which biases were significant are illustrated using boxplots in Figures 3.3–3.6 and the detailed test results can be found in the appendix (Table A3).

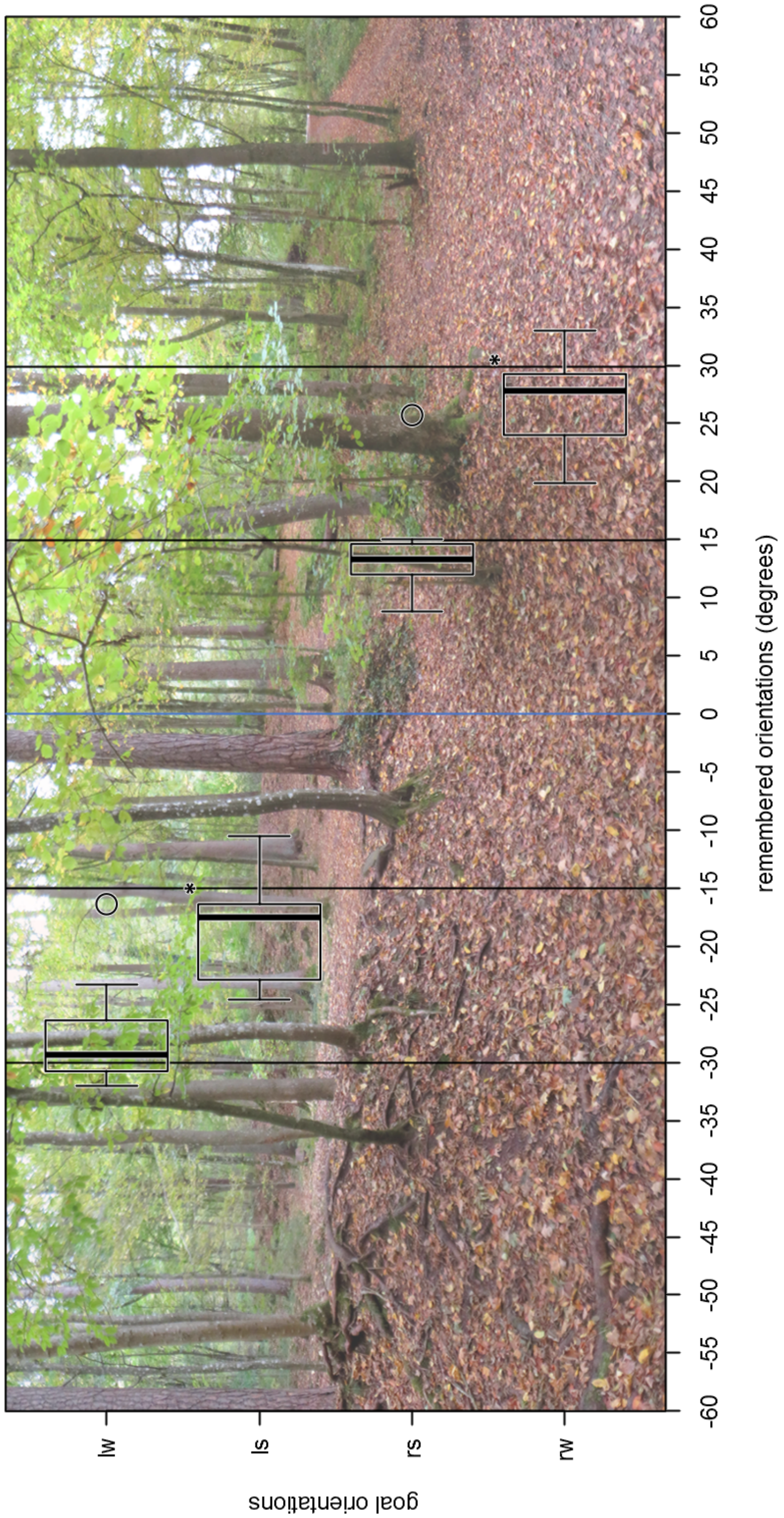

Figure 3.3: Boxplots of angle response distributions for each goal orientation (black lines) of the FO location (section of the image ±60° with respect to the canonical orientation (blue line)). Letter combinations on the y-axis stand for side (l = left, r = right) of the canonical orientation (blue line) and angle condition (w = wide, s = short). Significant deviations from the goal orientation are marked with one star for $p \leq .05$, two stars for $p \leq .01$ and three stars for $p \leq .001$.

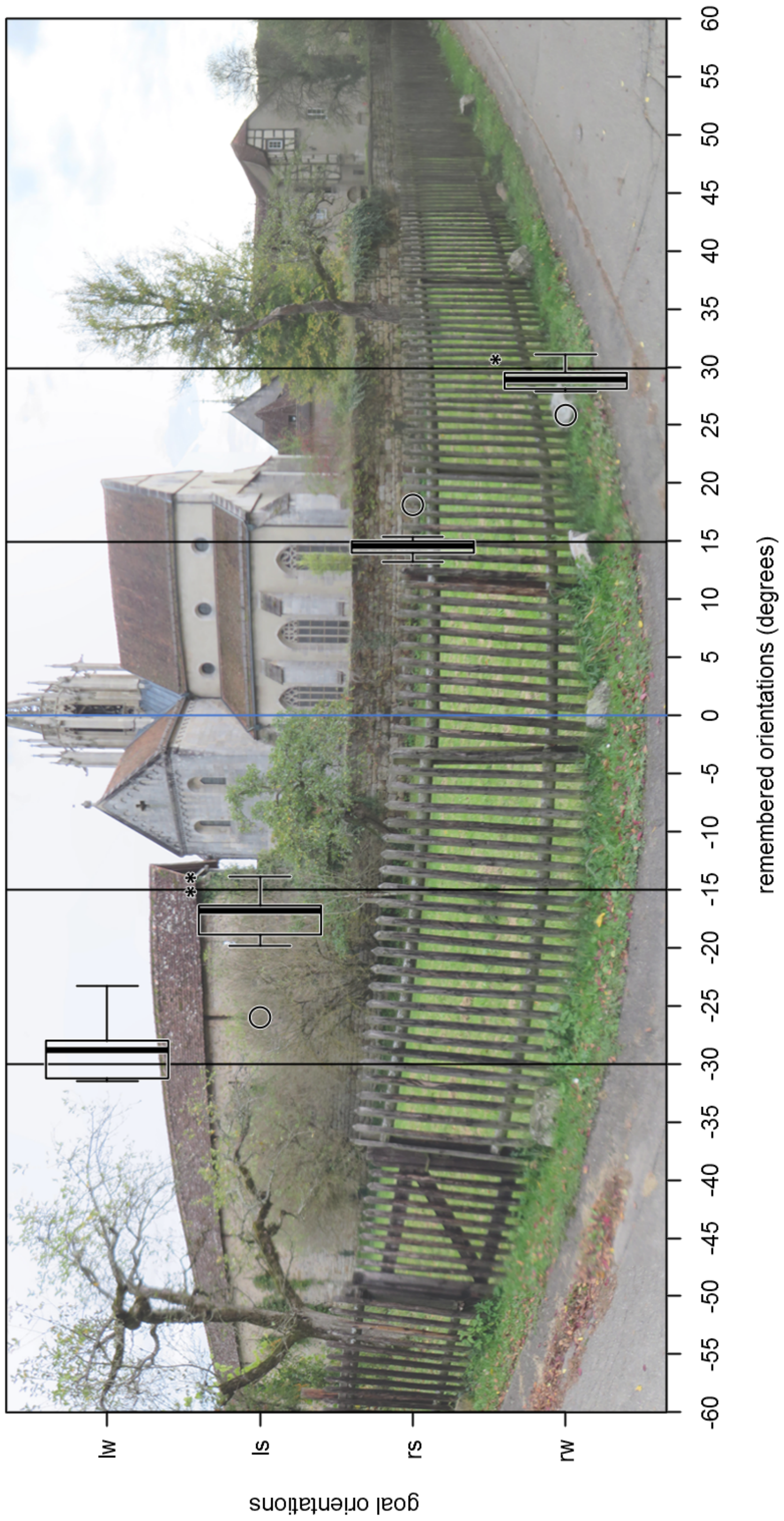

Figure 3.4: Boxplots of angle responses for each goal orientation of the BE location. See Figure 3.3 for further explanation.

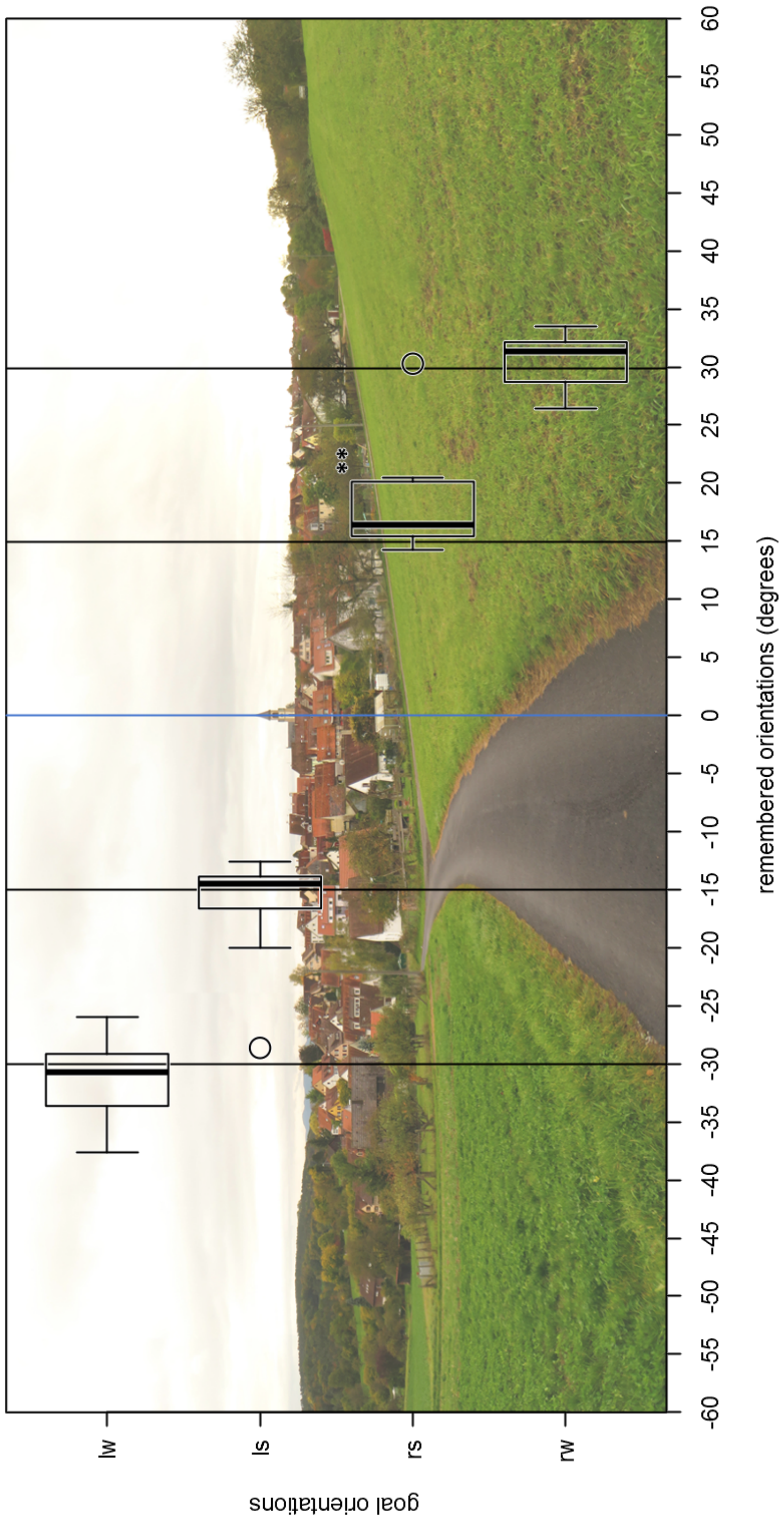

Figure 3.5: Boxplots of angle responses for each goal orientation of the HA location. See Figure 3.3 for further explanation.

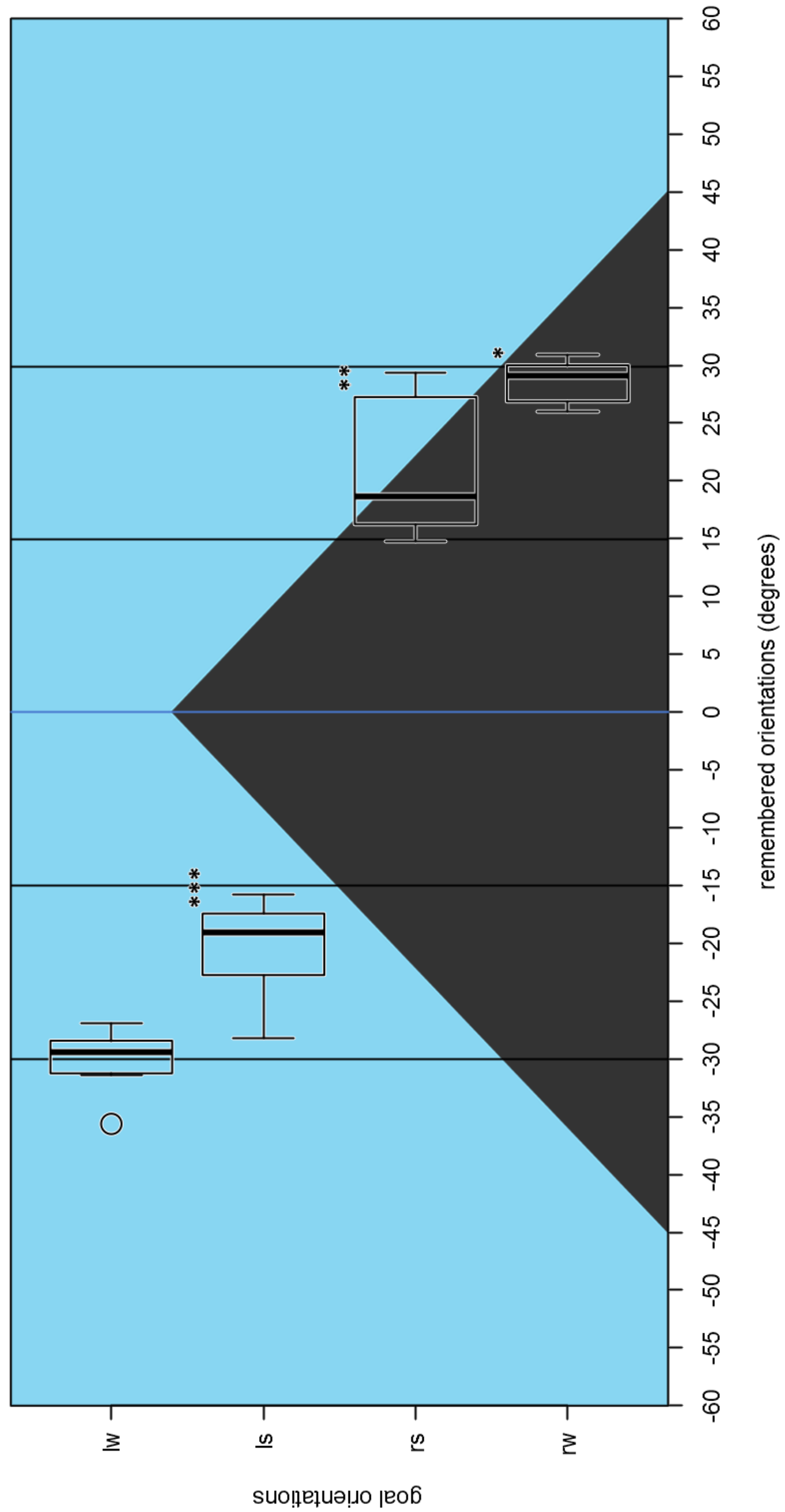

Figure 3.6: Boxplots of angle responses for each goal orientation of the PY location. See Figure 3.3 for further explanation.

4 Discussion

This study was aimed to bring us closer to answering the question of whether or not canonical orientations in scenes that are new to an observer can be identified. This would be evidence that canonicality affects the formation of allocentric representations in long–term memory in the way that intrinsic cues of a scene lead our perception to prioritize or prefer these orientations over others. In case of their existence, we predicted participants' memory of a goal orientation to be biased towards these preidentified canonical orientations. For this purpose, participants were asked to remember one of four orientations away from the canonical orientation, two on each side of it, as either a wide or short angle away from it. The dependent variable consisted of the corresponding number of the recalled orientation on a number scale that was verbally indicated by the participant.

4.1 Interpretation of the results

The analysis of the questionnaire confirmed that the scenes were indeed novel to the participants. Their accounts of how they consciously oriented themselves seemed to be in line with a hypothesized canonical orientation, because most subjects explicitly said to have used the church towers or tip of the pyramid for reference, while in the FO location trees in general, as opposed to one single feature, were used. In addition to landmarks, participants likewise reported using other prominent details in the BE, HA and PY location. This is not surprising, because if canonical orientations exist and play a role in the terms in which we represent our surroundings, it does not follow that other features cannot contribute to orientation purposes as well. The accuracy of letter responses were all significantly above chance level, which means that participants paid attention to the presented letters, so the performance in the recall of the goal orienation was not due to matching an afterimage of the shown orientation with orientations in the recall phase – this was one of the goals of adding a secondary task (see 2.3.2 for the other reasons).

The sought–after bias towards the canonical orientation is not reflected in the collected data. Mean deviations from the goal orientation were generally relatively small; they ranged only between ± 1–$2°$ with one exception. Only three of them pointed towards the canonical orientation and all of those in the wide angle condition of three different locations. Five conditions produced a significant bias, only two of which were in the direction of the canonical orientation – namely in the FO and BE locations. Our original expectation was that there would be no bias for the FO location, because no canonical orientation was identifiable. If our prediction had been correct, results of this location should have differed significantly from all others. At the very least, it should have differed from the PY location, because here potential confounding factors due to intrinsic image features that natural images might produce were excluded and we expected the largest bias in this scene. However, the Friedman test actually revealed no significant difference between any of the four locations. When examining the results which were further split into the left and right side (Figure 3.3), it looks as if people were not specifically biased towards the orienation we tested as the canonical orientation, but rather towards the nearest tree that could be used for orientation. The case of the BE location is interesting, because both biases of the short and wide angle conditions are significant, but they point in opposite directions. Because of this inconsistency, we think it unlikely that the biases were caused by the canonical orientation. In this location, many other feateres could have been used for orientation purposes or drawn attention, this might explain the lack of a bias towards the church tower. In the HA location, there was only a significant bias away from the goal orientation in the short angle condition. Although the church tower in this scene was also considered to be the most likely canonical orientation, it did not make up a very large portion of the picture, so it might not have been salient enough to cause a bias towards it. In the PY location, there was a comparatively large significant bias away from the canonical orientation in the short angle condition. There are at least two possible explanations: It might have been caused due to the relatively steep geographical slant of the pyramid, which might have guided the viewers gaze downwards. Another way to look at it is that, because there were only very few features available to use for orientation, people might have perceived the goal orientation as somewhere around the center of one of the sides of the pyramid, which is where the bias seems to point towards. On the background of the other results which are not consistent with our original prediction, we consider the latter to be the more probable scenario. We had also expected that the result pattern would be consistent for the short and wide angle condition which did not turn out to be the case; we take this as further evidence against our initial hypothesis.

There are some potential drawbacks in the design of our experiment or confounding factors which might have influenced our results. One was that we identified fixed canonical orientations to be tested prior to the experiment. There are two possible problems this might have caused: First, we might have wrongly judged an orientation to be the canonical one either if some of the criteria used were based on wrong assumptions or if we put the wrong emphasis on the relative importance of factors contributing to canonicality. The assumptions we made were based on our own evaluation of which orientation maximized information content in a scene – we picked orientations depicting landmarks. The other possible issue is that, in case canonical orientations do exist, they might not be universal for everyone and instead differ individually. We assumed the first of these options to be the case, because research on canonicality up until now seems to suggest canonical views arise due to intrinsic features like main axes, the degree of symmetry, functionality or the layout of a frame of reference (Blanz et al., 1999; Mou et al., 2007). A more explorative approach without preidentifying the canonical orientation would have been desirable, but then extensive sampling would have been necessary and the duration of the experiment might have become very long and tiresome for the participants, making their answers more error–prone. Also, if the experiment would have been too long, then it would not have been clear if the data resulted from canonicality as a characteristic of allocentric spatial representations. Another issue was, there might be a perceptual preference to process coarse before fine spatial scales in early perception for presentation durations of an image of 30 ms, but for longer presentation durations of about 150 ms this preference is reversed (Oliva and Schyns, 1997). The presentation duration of the target scene orientation in our experiment in the encoding phase was 500 ms, so it is probable that participants were able to recognize fine details of the scenes, although this was made more difficult because the scenes were degraded with a noise pattern. This might have enabled participants to memorize the scenes in terms of a few objects or prominent features in the center of their vision and to not perceive the gist of the scene, which could have prevented them from treating the scenes as if they were at the locations and trying to orient themselves as they would in a natural environment. One more possible issue is that although the shown scenes were new to the participants, as the analysis of the questionnaire established, the scenes (except for the PY location) depicted relatively generic places the likes of which are frequently encountered in everyday life. Therefore, allocentric representations of other well-known places might have had an influence on how people oriented themselves in the scenes.

The results of the current study are not in line with our prediction of a bias towards a canonical orientation and thus do not support our hypothesis of a canonical orientation in novel scenes, but neither do they necessarily contradict it. We have gatherd some possibilities from various perspectives how this result pattern could be explained in the context of our initial question. The negative results we have got are not as unambiguous in their interpretation as a positive result pattern would have been, because it is unclear if they might have been caused by troubles in the experimental setup or if there were confounding factors we did not eliminate, but we will try to formulate two scenarios we consider to be the most probable in light of the previous findings we considered at the beginning (in 1.1). The first scenario is that there are no canonical orientations in novel scenes. Beyond that, it is not possible on the basis of our findings to differentiate between the possibilites that canonical orientations do not exist at all or that canonical orientations in general do exist, as suggested by Ehinger and Oliva (2011), but that they only emerge on the basis of our experiences with places. These experiences lead to the formation of an allocentric spatial representation of the location inside our long–term memory. One finding by Blanz et al. (1999) was that no canonical orientation could be determined for nonsense objects and one possible explanation the authors gave was that the familiarity as a factor contributing to canonicality was missing in the cases of nonsense objects. If canonicality works in a similar fashion for both objects and scenes, the reason we found no bias towards a preidentified canonical orientation would have been the very fact that the scenes were novel to the observers. In this interpretation, canonicality is a property of allocentric spatial representations and might be manifested by storing more snapshot–like copies of prefered orientations than for others, as suggested by Röhrich et al. (2014).

The other possibility is that canonical orientations in novel scenes do exist, but that our prediction of a bias towards this orientation is erroneous. One argument for this possibility would be the accounts that participants gave in the questionnaire, where most said to have used the orientations we identified as the canonical ones for reference in their process of orienting themselves. Even if there is a preference of certain orientations in a scene over others, this might not lead our perception to be biased towards it when ones task is to find a specific orientation other than these most salient orientations. Therefore, maybe we should reconsider our conceptualization of canonicality. An alternative conceptualization of the perception of canonicality could be that canonical orientations are represented only in a nominal way – a certain orientation would then be classified as either the canonical orientation or not the canonical orientation. In the encoding phase, the goal orientation might have been encoded as being an orientation other than the canonical orientation and additionally what the relation to the canonical orientation was (e.g., the left/right side of it). In the recall phase participants might then have categorized the canonical orientation as not what they were looking for and been biased away from this orientation. This could be the reason we found biases away from the canonical orientation in the short angle conditions and no such tendencies in the wide angle condition, because the wide angle condition could clearly be classified as not being the canonical orientation and the short angle condition, which was closer to the canonical orientation, was more ambiguous in this respect. This would be in line with former findings that suggest that canonicality comes about based on intrinsic cues of a scene (Ehinger and Oliva, 2011; Mou et al., 2007; Röhrich et al., 2014) and such a nominal representation of canonicality would facilitate perceptual processing for a faster scene recognition.

4.2 Outlook

Now that we have discussed some possibilites to interpret our results, we want to introduce some follow-up experiments which might be able to distinguish between these and gain a better understanding of canonicality in spatial representations. If our original hypothesis and prediction are correct, the data was likely confounded by some external factor. To counteract some of the potential problems, we would propose the following improved setup. Instead of identifying the canonical orientation on the basis of our own evaluations, one could conduct a preliminary investigation of which is the most likely canonical orientation in a scene in a first experiment, perhaps with direct assessments by the participants like in Ehinger and Oliva (2011), and test these orientations. The basic procedure of the second experiment would then be similar to ours, with a few alterations. First, the image degradation with noise could be left out in the encoding phase and instead the presentation duration shortened to only around 30 ms, thereby priming visual input processing to operate at coarse spatial scales (Oliva and Schyns, 1997). To exclude influences of allocentric representations of other places, one could solely depend on the use of articifical scenes. A promising approach by Juliani, Bies, Boydston, Taylor and Sereno (2016) is to simulate natural environments virtually with the use of fractal geometry. The complexity level of scenes was operationalized with the fractal dimension in their study, so an additional advantage would be that connections between different fractal dimensions and the characteristic of canonicality in spatial representations could be drawn. A three-dimensional environment would also enhance participants' impression of being at the locations.

If there are no canonical orientations in novel scenes, the next question would be whether they exist in familiar scenes and if a bias towards them can be found in this case. One could simply conduct a similar experiment as this one, save for using familiar scenes as stimuli. The improvements proposed in the previous paragraph could still be applied, but instead of generating a virtual environment like Juliani et al. (2016), virtual environments of real places could be used. In case a bias would show, it could be taken as evidence that canonicality is a characteristic of allocentric representations.

If canonical orientations in novel scenes do exist and only our prediction of a bias towards them is wrong, we would have to think of another way to assess the presence of canonicality. The method of Ehinger and Oliva (2011) might be feasible, although the authors did not mention whether participants were familiar with the presented locations – this could be controlled for. Due to reasons mentioned in 1.2, an indirect measure would be preferable, though. If canonical orientations are represented nominally, there might be an advantage in the latency of recall of canonical orientations over others due to a faster recognition, so reaction times might be another feasible measure to use.

4.3 Conclusion

Canonical orientations have been proposed to exist by Ehinger and Oliva (2011), Mallot et al. (2014), McNamara et al. (2003), Mou and McNamara (2002), Mou et al. (2007) and Röhrich et al. (2014). We wanted to find out if canonical orientations exist in novel scenes. Our findings brought us closer to answering this question. No bias towards a preidentified canonical orientation was reflected in our data. As discussed in 4.1, this does not necessarily exclude the possibility that canonical orientations might be identifiable in novel scenes. If the original hypothesis and prediction is not to be rejected, one would have to find a plausible explanation for this result pattern, maybe a confounding factor in the design of the experiment for why no bias showed up. In case everything in the setup of our experiment was sound, this finding contradicts our prediction and might be taken as evidence against the hypothesis of the existence of canonical orientations in novel scenes. We consider the latter case more likely and opt for one of two alternative explanations of our results. One is that canonicality emerges only for familiar locations which have allocentric representations in long–term memory. Thus, canonical orientations exist, but not in novel scenes. Alternatively, the hypothesis of the existence of canonical orientations in novel scenes could be correct, but the prediction of a bias towards it is wrong. This could be the case if canonicality were represented purely on a nominal level. A combination of these two explanations is possible as well – maybe canonical orientations only exist for familiar scenes, but the prediction of a bias is wrong nonetheless. Altogether, the insights we gained do not clearly point in one single direction, but seem to go against our initial hypothesis of a canonical orientation in novel scenes. Further investigation – possibly in one of the ways discussed in 4.2 – might be able to differentiate between the presented possibilities.

Bibliography

Basten, K., Meilinger, T., & Mallot, H. A. (2012). Mental travel primes place orientation in spatial recall. In *International Conference on Spatial Cognition*, (pp. 378–385). Springer.

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception*, 28(5), 575–599.

Ehinger, K. A., & Oliva, A. (2011). Canonical views of scenes depend on the shape of the space. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Juliani, A. W., Bies, A. J., Boydston, C. R., Taylor, R. P., & Sereno, M. E. (2016). Navigation performance in virtual environments varies with fractal dimension of landscape. *Journal of environmental psychology*, 47, 155–165.

Mallot, H. A., Roehrich, W. G., & Hardiess, G. (2014). A view-based account of spatial working and long-term memories: Model and predictions. In *COGNITIVE PROCESSING* (Vol. 15, pp. S120–S123). Springer.

McNamara, T. P., Rump, B., & Werner, S. (2003). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonomic Bulletin & Review*, 10(3), 589–595.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 162.

Mou, W., Zhao, M., & McNamara, T. P. (2007). Layout geometry in the selection of intrinsic frames of reference from multiple viewpoints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(1), 145.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive psychology*, 9(3), 353–383.

Oliva, A. (2005). Gist of the scene. In *Neurobiology of attention* (pp. 251–256). Elsevier.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive psychology*, 34(1), 72–107.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in brain research*, 155, 23–36.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. *Attention and performance*, 135–151.

Röhrich, W. G., Hardiess, G., & Mallot, H. A. (2014). View-based organization and interplay of spatial working and long-term memories. *PloS one*, 9(11), e112793.

Schölkopf, B., & Mallot, H. A. (1995). View-based cognitive mapping and path planning. *Adaptive Behavior*, 3(3), 311–348.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science*, 5(4), 195–200.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97–136.

Appendix

Questionnaire

Fragebogen nach dem Experiment

Versuchspersonen–ID (wird vom Versuchsleiter ausgefüllt):

Wohnst du derzeit, oder hast du in der Vergangenheit in Tübingen gewohnt?

Wohnort (Stadtteil in Tübingen/Stadt außerhalb):

Dauer in Halbjahresschritten, die du schon in Tübingen lebst/ehemals gewohnt hast:

Bitte gib in der folgenden Tabelle an, wie vertraut du mit den gezeigten Orten aus dem Experiment vorher schon warst, 1 bedeutet dabei „Ich kannte diesen Ort vorher gar nicht“ und 7 „Ich kenne diesen Ort und weiß wo er sich befindet“ .

	1	2	3	4	5	6	7
Wald mit Brücke							
Stadt mit Kirche							
Ausblick auf ein Dorf							

Falls du die drei in der Tabelle genannten Orte schon kanntest/glaubst zu kennen, um welche Orte handelt es sich?

Hast du bewusst Strategien zum Lösen der Aufgaben verwendet? Falls ja, welche? Wie oder woran hast du dich in den gezeigten Szenen orientiert?

Tables with detailed results

Table A1: Accuracy rates of letter responses and the p values of the one–tailed binomial tests for $n = 128$ trials and an assumed probability of guessing the correct letter of 50%. All accuracy rates were significantly above chance level.

ID	p	accuracy rate (%)
1	<.001***	82.1
2	<.001***	87.4
3	<.001***	82.9
4	<.001***	85.5
5	<.001***	86.4
6	<.001***	95.0
7	<.001***	94.5
8	<.001***	93.4
9	<.001***	80.6
10	<.001***	96.8

Table A2: Results from Shapiro–Wilk tests. These were conducted to see if the assumption of normality was violated in each of the four combinations of angle condition (s = short/w = wide) and side (l = left/r = right) of the canonical orientation.

location	lw angle (W, p)	ls angle (W, p)	rs angle (W, p)	rw angle (W, p)
FO	0.82, .028*	0.94, .514	0.79, .010**	0.97, .845
BE	0.85, .064	0.80, .014*	0.82, .023*	0.94, .530
HA	0.98, .970	0.72, .001***	0.72, .001***	0.92, .331
PY	0.89, .184	0.89, .165	0.86, .076	0.90, .197

Table A3: Results of the Mann–Whitney tests to see whether deviations from the goal orientation were significant. Positive mean biases point towards the canonical orientation, negative ones point away from it. The letters in the angle condition stand for l = left, r = right, s = short and w = wide.

location	angle condition	U	p	M (°)	SD (°)
FO	lw	41	.092	2.4	4.7
	ls	10	.041*	-3.4	4.8
	rs	36	.061	1.1	4.7
	rw	46	.032*	2.9	4.2
BE	lw	42	.076	1.1	2.5
	ls	1	.004**	-2.8	3.3
	rs	33	.118	0.2	1.3
	rw	46	.033*	1.1	1.6
HA	lw	17	.161	-1.4	3.5
	ls	26.5	.479	-1.4	4.8
	rs	3	.004**	-3.2	4.7
	rw	17.5	.166	-0.8	2.2
PY	lw	31	.658	-0.1	2.4
	ls	0	<.001***	-5.5	4.1
	rs	1	.001**	-6.0	5.6
	rw	33	.021*	1.4	1.7