COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 40 (2016) 440–454

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12229

Seeing Like a Geologist: Bayesian Use of Expert Categories in Location Memory

Mark P. Holden,[a,b] Nora S. Newcombe,[a] Ilyse Resnick,[a] Thomas F. Shipley[a]

[a]*Department of Psychology, Temple University*
[b]*Department of Psychology, University of Western Ontario*

Received 29 July 2013; received in revised form 9 March 2014; accepted 22 May 2014

Abstract

Memory for spatial location is typically biased, with errors trending toward the center of a surrounding region. According to the category adjustment model (CAM), this bias reflects the optimal, Bayesian combination of fine-grained and categorical representations of a location. However, there is disagreement about whether categories are malleable. For instance, can categories be redefined based on expert-level conceptual knowledge? Furthermore, if expert knowledge is used, does it dominate other information sources, or is it used adaptively so as to minimize overall error, as predicted by a Bayesian framework? We address these questions using images of geological interest. The participants were experts in structural geology, organic chemistry, or English literature. Our data indicate that expertise-based categories influence estimates of location memory —particularly when these categories better constrain errors than alternative ("novice") categories. Results are discussed with respect to the CAM.

Keywords: Location memory; Expertise; Spatial cognition; Bayesian models; Categorization

1. Introduction

Memory for spatial locations is often biased in predictable ways. For example, adults recall irregular-shaped spaces or routes as being regular (e.g., Tversky, 1981), consistently estimate the distance from A to B as different than the distance from B to A (McNamara & Diwadkar, 1997; Newcombe, Huttenlocher, Sandberg, Lie, & Johnson, 1999), and systematically misremember locations as being more central to the surrounding region than they are (e.g., Holden, Curby, Newcombe, & Shipley, 2010; Holden, Newcombe, &

Correspondence should be sent to Mark P. Holden, Department of Psychology, University of Western Ontario, London, ON, Canada, N5A 5C2. E-mail: mholde3@uwo.ca

Shipley, 2013; Huttenlocher, Hedges, & Duncan, 1991). Some investigators have taken these errors to suggest that human spatial representations are distorted or schematized with respect to the physical space (e.g., Tversky, 1981), while others have proposed that such biases reflect processes of inference, based on hierarchical spatial representations (Stevens & Coupe, 1978). More recently, however, research has suggested that systematic errors in spatial memory can be accounted for in terms of the statistically optimal combination of multiple sources of information, which minimizes the mean squared error across multiple estimates (e.g., Huttenlocher et al., 1991; Newcombe & Huttenlocher, 2000).

For example, Huttenlocher et al.'s (1991) category adjustment model (CAM) postulates that locations are remembered hierarchically, at both a fine-grained (metric) level and a coarser, categorical level. That is, one may remember that one's keys were "on the table" (categorical), and, more precisely, that they were "5 inches from the edge" (fine-grained). According to the CAM, individuals combine these two representations, so that every estimate is the result of an optimal combination process that weights each representation by its relative certainty. This combination results in estimates that are biased toward the center of the categorical representation (sometimes called the category prototype), but it minimizes error variance across multiple estimates (Huttenlocher et al., 1991; see also Cheng, Shettleworth, Huttenlocher, & Rieser, 2007). The decreased variance optimizes performance by decreasing the average absolute error over multiple estimates—even though bias is introduced on individual trials. In this view, systematic errors in spatial judgments are not due to distorted representations of space (e.g., Tversky, 1981), but rather result from an optimal, Bayesian combination process across unbiased, but hierarchically ordered representations. Empirical evidence for the CAM has been found for numerous tasks, both spatial and non-spatial (e.g., Baud-Bovy & Gentaz, 2012; Corneille, Huart, Becquart, & Bredart, 2004; Duffy, Huttenlocher, & Crawford, 2006; Fitting, Wedell, & Allen, 2007; Holden et al., 2010, 2013; Huttenlocher, Hedges, & Vevea, 2000; Huttenlocher et al., 1991; Lee & Brown, 2003; Schutte & Spencer, 2002; Spencer & Hund, 2002; Spencer, Simmering, & Schutte, 2006; Wedell, Fitting, & Allen, 2007).

While there is abundant evidence that individuals encode locations at both a fine-grained and a categorical level, the types of information used to define categories is not clear. Complex environments can be organized in a variety of ways: a mountain may be perceptually distinct from the plateau beneath it and the sky above it, but *mountain* is also associated with specific concepts—such as the *peak* and *base* areas—that delineate more specific spatial categories. Indeed, spatial regions may be defined in many ways, including administrative procedures (e.g., political borders), thematic similarities (e.g., rainfall, presence of pine trees), functional interaction (e.g., migratory routes), or informal cognitive processes (e.g., "downtown"; Montello, 2003, 2008).

Whether these kinds of conceptual information are used in defining spatial categories is unclear. Spencer and colleagues' dynamic field theory (DFT; e.g., Spencer et al., 2006) is, in many ways, a complementary model to the CAM, making many of the same predictions, and offering a neurally plausible framework for the type of Bayesian combination proposed by the CAM. However, DFT explicitly states that *only* perceptual information is used to define spatial categories (e.g., Simmering & Spencer, 2007), implying that spatial categories are not based on conceptual information and are therefore non-malleable. Ignoring conceptual information in defining spatial categories seems unlikely under a Bayesian framework—in which estimates are adapted as more information becomes available. Holden et al. (2010) began to address this question, examining location memory errors in images manipulated to alter their low-level visual properties (color negatives) or the accessibility of their more conceptual structure (inverted images). Bias patterns differed for both conditions, suggesting that a variety of cues *can* be used in forming spatial categories. However, these results are based on scenes in which one type of information or the other was systematically disrupted (Holden et al., 2010). Thus, it remains unclear how categories are formed in canonical scenes in which various types of information are readily accessible.

This issue of the malleability of spatial categorization schemes has received some attention in recent years. In many studies of location memory, participants are asked to recall point locations within a blank circle, and estimates are often biased toward the center of the circle's quadrants, as defined by the horizontal and vertical axes. Huttenlocher, Hedges, Corrigan, and Crawford (2004) attempted to overcome this "default" categorization scheme by altering the distribution of points, and even requiring participants to categorize the points (during encoding) according to an alternative categorization scheme. Their data showed that participants continued to use the default scheme, leading the authors to state the use of the default scheme is "immutable" (p. 78). On the other hand, Simmering and Spencer (2007) determined that an alternative categorization scheme would be used, but *only* if it were perceptually supported throughout the trial. Finally, Sampaio and Wang (2010) demonstrated that the use of an alternative categorization scheme could be induced by visually cuing the alternative category membership at test or by using targets that were unique to each of the alternative categories (and while still providing perceptual cues as to the alternative categories at test). Each of these studies therefore converges on the idea that default categorization schemes are difficult to override, and that the use of an alternative categorization scheme requires fairly extensive perceptual support.

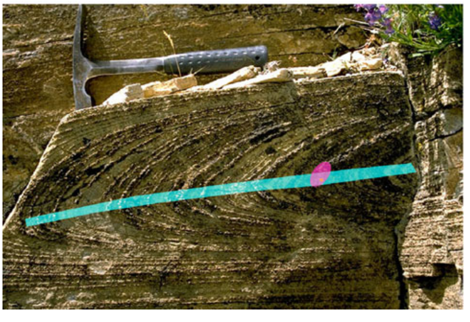
Here, we address these issues by analyzing the location memory errors made by experts in structural geology, organic chemistry, and English literature, using images of geological interest. Critically, the visual input for these images is identical across all participants, and the visual cues remain unaltered. However, we reasoned that experts in structural geology should be more likely to notice the presence of geologically relevant items—such as a subtle sedimentary structure—depicted in the images, and use that information to structure their memory for location. Indeed, in his seminal work on "professional vision," Goodwin (1994) states that domain expertise can affect how individuals process a given stimulus, noting that "an archaeologist and a farmer see quite different phenomena in the same patch of dirt" (p. 2). Similarly, research on perceptual learning has suggested that experts detect and distinguish features, differences, and relations not registered by novices in their field, and that they do so with increasing automaticity (see Kellman & Garrigan, 2009 for a review). If geology experts, therefore, process scenes of geological interest differently than novices, experts may segment these scenes into different spatial categories by using geologically relevant conceptual information. It would therefore be predicted that expert geologists' location memory estimates will be influenced by these geologically defined categories, implying that spatial categorization processes are both malleable and able to make use of conceptual information.

If structural geologists do use their expertise to define spatial categories in geologically relevant scenes, it is possible (though unlikely) that they might exclusively use geologically defined categories. Alternatively, when different categorization schemes are available, people might adaptively switch between them so as to minimize error on a trial-by-trial basis. For example, they might encode locations by using the smallest spatial category because this would best constrain their potential errors. If so, novices might define spatial categories based on low-level visual features or on informal conceptual information (e.g., the *base* of the mountain), while experts might adaptively choose between using these same "novice categories" or those based on domain-specific knowledge. Indeed, a third (though related) possibility is that the expert geologists may encode locations with respect to *both* the "novice" and the "expertise-based" categories. A fully Bayesian account would advocate this position, assuming that each source of information (metric and the two categorical representations) would *all* be weighted according to their relative reliability. In order to distinguish between the possibilities of relatively rigid versus flexible use of conceptual information in spatial categorization, we conducted our location memory task using stimuli for which the relative size of novice-defined and expert-defined categories differed.

The CAM makes two predictions for this study: First, if conceptual information is used to define spatial categories, then the location memory biases of structural geologists will differ from those of experts in unrelated fields. That is, the geology experts' errors will be influenced by the geologically relevant categories, while the novices' errors will be biased toward the center of the novice category. Second, and modifying the first prediction, if different reference frames can be used flexibly, then the frame that produces the smaller spatial category—which better constrains one's estimates, reducing overall error—will be more influential in guiding recall. Thus, the geology experts' biases may approximate or even match those of novices for images in which a novice-defined category is smaller than the geologically relevant category (as in the case of Fig. 1a and c) but will show distinct expertise effects when the opposite is true (as in the case of Fig. 1b and d). That is, the "expert" category in Fig. 1d would exert more influence in the estimates of structural geologists than the corresponding category in Fig. 1c because the expert category is smaller than the novice category in the former image, and larger in the latter. Thus, expert performance is predicted to more closely approximate novice performance when the novice categories better constrain errors than expert categories. Conversely, expert performance is predicted to differ from novice performance when the expert categories better constrain potential errors.

(a) Expert-Larger Image

(b) Expert-Smaller Image



(c) Expert and Novice Categories

(d) Expert and Novice Categories

Fig. 1. Sample Expert-larger and Expert-smaller images. (a) depicts the axial plane of a fold, while (b) is an example of "boudinage" (sausage links). The target location is indicated by a yellow dot. Images (c) and (d) mark sample spatial categories identified by the expert geologists (blue) and novices (pink). In image (c), the expert geologists have categorized the location as falling along the fold's axial plane (essentially the line of symmetry on the surface) while in image (d) the point lies along the line of boudinage, between two of the "sausage links." Novices appear to have categorized locations on the basis of low-level visual cues, such as "*the end of the lighter line*" (c) and "*on the flat region of rock*" (d).

2. Method

2.1. Participants

Three groups of experts were asked to recall point locations within scenes of geological interest (e.g., an outcrop of faulted rock). We examined bias in location memory estimates, comparing experts in structural geology (n = 9), organic chemistry (n = 11), and English literature (n = 9). Expertise was defined as having obtained a doctoral degree and having at least an additional 5 years of experience. All groups had therefore received the same level of education and had attained approximately equivalent levels of expertise in their respective fields. The stopping rule for data collection was to test all structural geologists who agreed to participate (and met the selection criteria) within a 100-mile radius of the tester. A similar rule, using a 40-mile radius, was used for the organic chemistry and English literature experts. A larger radius was used for the structural geologists because there were fewer in the immediate area.[1]

The groups of experts were chosen based on predicted differences both in terms of spatial abilities and in experience with images of geological content. That is, structural geology and organic chemistry are generally considered to be highly spatial domains, requiring skills such as spatial visualization, mental rotation, and visual penetrative thinking, among others; English literature, though, is not typically associated with these skills (e.g., Coleman & Gotch, 1998; Kastens et al., 2009; Resnick & Shipley, 2013). Furthermore, structural geology is a branch of geology that is highly associated with fieldwork and thus likely to be familiar with landscape-scale geological categories. Comparing location memory error patterns across these groups would therefore allow us to differentiate between effects due to domain-specific knowledge versus greater spatial reasoning skills.

2.2. Materials and procedure

The task used here was based on one described in detail by Holden et al. (2010). Briefly, participants were asked to recall point locations within high-resolution, color photographs of natural landscapes. Target locations in these images were indicated by a small, yellow dot, which stood out against the natural background colors. These dots were made to be elliptical, corresponding to the slope that they appeared on, and were blurred slightly at the edges in order to support the impression that it was a location within the scene, rather than a point on the photographic surface (Holden et al., 2010).

The landscape images used in this study were drawn from a larger pool of images, all of which contained some item of geological interest. Preliminary locations were chosen to be within these categories (or very near them, in the case of 2D categories such as a line indicating a geological fault or a fold's axial plane). In addition, locations were constrained such that they could not fall within 2 cm of the image edges, so as to prevent individuals from recalling the location based on distance from a border, rather than properties of the scene. All scenes were 20 cm in length, but varied in height.

Using these preliminary locations, "expert" and "novice" spatial categories were identified a priori, through consultation with two additional experts in structural geology, and by testing a group of novice undergraduate students.[2] The instructions for the category identification task explained that location memory "sometimes involves using broad levels of information—such as '*the keys are on the table*'—and sometimes involves more specific information—such as remembering that they were right *there*." The participants were then asked to identify the spatial category they would use in such a task by "identify[ing] the region that surrounds or encloses the [prospective] point's location." The spatial category could therefore be any regular or irregular shape.

Based on the results of the category identification task, 30 images were chosen from the pool based on high levels of agreement among the experts, and among the novices (see Scoring and Analysis), and such that equal numbers of images contained expert-defined categories that were larger or smaller in area than those identified by the novices (see Fig. 1). We will refer to these image types as “expert-larger” and “expert-smaller,” respectively.[3] Small adjustments were then made to the point locations within these images so as to approximately maximize the angle difference between the prototypes of the expert- and novice-defined categories (described below). For example, bias toward the center of the expert category in Fig. 1d would lead to approximately downward errors, while bias toward the center of the novice category would result in upward and leftward errors. No adjustment moved the location more than 5 mm, and the final target locations were still contained within the original categories.

For the location memory task, participants were asked to recall the target locations. On every trial, two scenes were presented serially, each containing one target location. These study images were presented for 2,500 ms each, separated by a 250 ms blank screen. Following another 250 ms delay, the identical scenes (but without the yellow points) were presented, in order. The participants’ task was to indicate the position of the dot, as accurately as possible, using the computer’s mouse. The test images each remained on-screen until a response was made. This method has been shown to increase the information collected while retaining the characteristic error patterns of bias toward category prototypes (e.g., Holden et al., 2010; Huttenlocher et al., 1991). In addition, previous work in our lab has found no difference in bias patterns between the first and second images in a pair (e.g., Holden, Duff-Canning, & Hampson, 2015; Holden et al., 2010). Stimuli were presented electronically using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA, USA), and responses were recorded by the program with pixel precision, allowing even subtle response patterns to be detected.

2.3. Scoring and analysis

The distributions of participants’ errors were examined for systematic bias toward either the expert- or novice-defined category prototypes. As outlined above, a separate group of structural geology experts and novices identified these categories a priori. The category identification task involved outlining the spatial category on a physical copy of the image, using a medium-tipped permanent marker. Each outlined category was then transformed into a digital copy, and the categories for a given location (within each group) were overlaid using photo-editing software. For the “expert” categories, we used a stringent criterion for agreement; only images for which the number of overlapping pixels was greater than 95% of the total identified pixels (across both experts) were used in this study. This often amounted to the difference between outlining along the edges of a rock (for example) and “coloring within the lines” of the same category. The center of mass of the category was then calculated by weighting all pixels by the degree to which they were agreed upon. Thus, portions of the category identified by both experts were weighted twice as heavily as those identified by only one. For the “novice” categories, digitized copies of the outlined categories were again overlaid. In this case, though, all regions that were identified by two or fewer participants were discarded, and the prototype of the composite category was again determined by weighting each pixel based on the level of agreement (see also Holden et al., 2010, Analysis 1B).

Location memory data were analyzed using the method described in detail by Holden and colleagues (Holden et al., 2010, 2013, 2015). Briefly, all responses were converted to vectors, originating at the correct location and ending at the response location. Similarly, the category prototype was also converted to a vector, also originating at the correct location and ending at the category prototype (see Fig. 2 for an example). These two vectors represent the observed and predicted error directions, respectively, for that location. The difference angle between these two vectors was then calculated. A difference angle of 0° would mean that the observed error was precisely in the direction of that category's prototype, while a difference angle of 180° would imply that the error was in precisely the opposite direction. All participants' responses were analyzed both with respect to the expert- and to the novice-defined categories.

The difference angles (and error magnitudes) for all responses by a given participant were then combined into an average "difference vector" for that participant (Holden et al., 2010, 2013, 2015). This point is important because minor deviations from correct are expected to be randomly distributed about the correct location (as they are more likely due to errors in physically responding than true errors of memory) and could inflate the average difference angle between the observed and predicted biases. To account for this issue, the difference angle for each response made by a given participant was converted into a vector with a length equal to the error magnitude (distance from correct) divided by the average absolute error across all responses for that participant. These "difference vectors" were then added, and the length of the summed vector is divided by the number of responses.

As a result, each participant is associated with a single, standardized average difference vector. The angle of this vector indicates the mean difference between observed errors

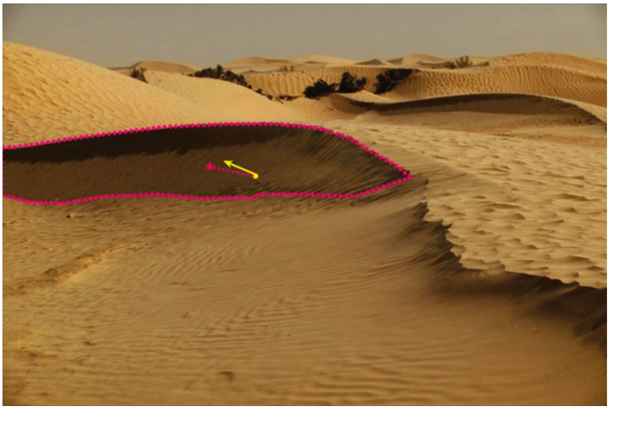


Fig. 2. An example of the data scoring procedure. Participants were asked to recall a given location, indicated by the small yellow dot on the figure. Here, the "novice" category is outlined, and the category prototype is indicated by a star. A sample participant's response is indicated by the arrow (i.e., the response was made at the arrow's head). The difference angle between the response vector and the dotted line (connecting the correct location with the category center) was measured.

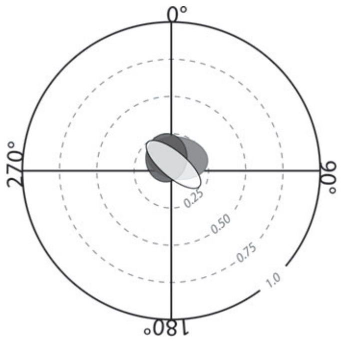
and the predicted direction of bias, while the length of the vector indicates how variable the responses were for that participant (i.e., vector length will range from 0 to 1, with longer lengths indicating that errors were more consistently in a given direction relative to predicted; Batschelet, 1981). Averaging vectors in this way ensures that minor deviations from correct exert much less of an effect on the average difference angle than larger errors, and standardizing the vectors controls for individual differences in response accuracy. Hotelling's one-sample test for circular data compares the distribution of these standardized average difference vectors against the null hypothesis that there is no mean direction of bias (i.e., random error, indicated by a small vector length; Zar, 1998), while Hotelling's two-sample test compares the distributions of these vectors between two groups (Zar, 1998).

3. Results

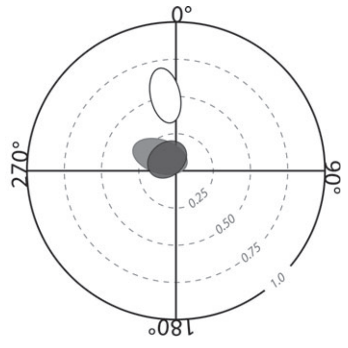
The 95% confidence ellipses of the mean (Batschelet, 1981) of the average difference vectors for each expert group, within each image type, and coded with respect to the expert- and novice-defined categories are shown in Fig. 3. The mean vectors for each group would connect the origin of the figure to the center of their respective ellipses. Recall that longer vectors indicate that errors were more consistently in a given direction relative to predicted (e.g., toward the category prototype—indicated by a difference angle near 0°), while shorter vectors indicate random errors. Therefore, a confidence ellipse that overlaps the origin of these figures indicates random error, while one that does not overlap the origin indicates that errors are significantly biased. Also recall that all responses were coded both with respect to the expert- and to the novice-defined categories, indicated by the two rows of the figure. Thus, if location memory errors for a given group are randomly distributed when coded with respect to one definition of a category, but are significantly biased when coded with respect to the other definition, then this would indicate that they used the latter definition to form their spatial categorical representation.

As expected, neither English nor chemistry experts' responses were biased toward the center of the geologically defined categories, whether these were larger or smaller than the novice-defined categories (all Fs < 1.07, *ns*; Fig. 3a and b) Their responses, however, were significantly biased toward the center of the novice-defined categories, in all conditions (all Fs > 9.17, all ps < 0.01; Fig. 3c and d). Furthermore, the bias patterns did not differ between these groups in any condition, (all Fs < 1.15, *ns*).

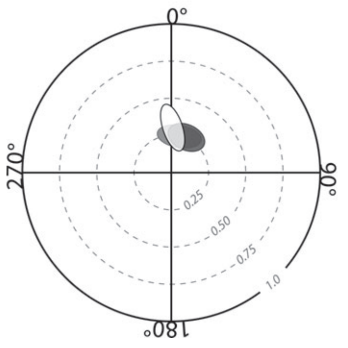
Geology experts' location memory errors were biased toward the center of both types of category, depending on the image type. Bias was primarily in the direction of the geologically defined category prototypes only for images in which these categories were smaller than the novice-defined categories (*expert-larger*: $F(2, 7) < 1$, *ns*; *expert-smaller*: $F(2, 7) = 11.09$, $p < .01$; Fig. 3a and b). Conversely, bias toward the novice-defined categories was found only for images in which the geologically defined categories were larger (*expert-larger*: $F(2, 7) = 6.15$, $p = .03$; *expert-smaller*: $F(2, 7) = 1.91$, *ns*; Fig. 3c and d). Critically, pairwise comparisons indicate that the geology experts' patterns of bias differ



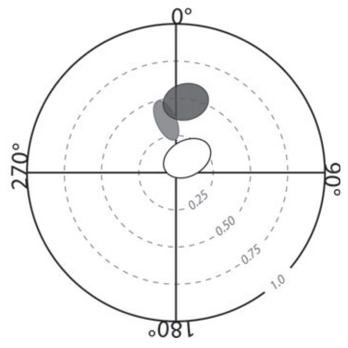
(a) Expert Categories, Expert-Larger Images



(b) Expert Categories, Expert-Smaller Images



(c) Novice Categories, Expert-Larger Images



(d) Novice Categories, Expert-Smaller Images

Fig. 3. Ninety-five percent confidence ellipses of the mean difference vectors (observed-predicted) for each group (white = Geology, medium-gray = Chemistry, light gray = English). Responses were coded with respect to the geologically defined categories (a and b) and to the novice-defined categories (c and d). Note that errors of 0° indicate bias toward the category prototype, and that ellipses that encompass the origin indicate no significant bias (i.e., random error).

significantly from those of the other groups only in the expert-smaller condition (geology-English: $F(2,15) = 3.77$, $p < .05$, $d = 1.00$; geology-chemistry: $F(2, 17) = 5.08$, $p = .02$, $d = 1.09$), but not in the expert-larger condition (all Fs < 1, *ns*).

We also examined location memory accuracy using the absolute displacement of the responses from the correct locations, in pixels. If using (or heavily weighting) the smallest potential spatial category was advantageous, as expected, then it would stand to reason that structural geologists' memory for spatial locations would be more accurate compared to the other expert populations, but only within the expert-smaller condition. The results support this prediction. When the geologically defined category was smaller, geology experts were significantly more accurate than English experts (geology: $M = 24.25$, $SD = 8.62$; English: $M = 35.17$, $SD = 9.98$), $t(16) = 2.49$, $p = .02$, $d = 1.25$, and than chemistry experts (chemistry: $M = 32.69$, $SD = 6.85$), $t(18) = 2.44$,

p = .03, d = 1.15). However, in the expert-larger condition, no significant differences in accuracy were found (geology: M = 23.50, SD = 8.81; English: M = 26.19, SD = 8.95; chemistry: M = 26.84, SD = 8.67), all ts ≤ 0.85, *ns*. No difference in accuracy was found between the English and chemistry experts in either condition, all ts < 0.66, *ns*.

4. Discussion

The primary goals of this study were to examine whether spatial categorization processes may be affected by relevant conceptual information, and—if both perceptual and conceptual cues are available—whether the spatial category that best constrains errors will be more influential in guiding recall (thereby minimizing error). Our results clearly indicate that, when recalling locations in scenes of geological relevance, geology experts' conceptual knowledge resulted in different patterns of bias compared to experts from unrelated fields. Critically, this difference was only significant for images in which the geologically defined categories were smaller in area than those identified by novices (either on the basis of low-level visual cues, or perhaps by using a different, informal conceptual framework). More specifically, expert geologists' errors were biased toward the center of the geologically defined categories in the expert-smaller images but were generally biased toward the center of the novice-defined category in the expert-larger images. In contrast, both English literature and organic chemistry experts' location memory errors were biased toward the center of the novice-defined categories in all conditions.

The CAM states that these systematic errors in location memory are the result of an optimal, Bayesian combination process across hierarchically ordered pieces of information (Huttenlocher et al., 1991). Accordingly, these results suggest not only that expert geologists were able to use acquired conceptual information to define spatial categories, but that they did so adaptively, using the source(s) of information that best constrained their estimates. The results in terms of absolute error support this conclusion: When geologically defined categories were smaller than the novice-defined ones (the expert-smaller condition), geologists were significantly more accurate in their recall than either of the other two expert groups. On the other hand, when the expert-defined categories were larger, no significant differences in absolute error were found between any of the groups.

It is important to point out that these results cannot be accounted for by any type of exposure effect or by differences between groups in the spatial content of their respective fields. That is, if geology experts' heightened accuracy in the expert-smaller condition were due to increased familiarity with geological images (and not to the use of relevant conceptual knowledge), then one would expect their accuracy to be higher across all conditions. However, neither the geology experts' accuracy nor their bias patterns differed significantly from those of the other groups in the expert-larger condition. Similarly, if the observed differences were due to the relatively spatial nature of structural geology, one would expect that organic chemists would perform similarly to geologists, given the highly spatial nature of this field as well. However, the organic chemistry and English literature experts did not differ in any of the comparisons.

It is interesting that the location memory task engaged the coding schemes relevant to the work of professional geologists. Identifying a fault is not a necessary step toward recalling a particular location on a rock face; indeed, the additional processing required might place further limits on the cognitive resources available to perform the location memory task. Nevertheless, despite the brief exposure time and (seemingly) unrelated nature of the task, geology experts did appear to extract this information in order to encode locations at a categorical level. This may imply an almost automatic extraction of conceptual information from the image, regardless of the nature of the task. Of course, this interpretation requires direct experimentation, but it is in line with research on perceptual learning suggesting that, with increasing expertise, individuals become more attuned to features relevant to that domain and extract them with increasing automaticity (e.g., see Kellman & Garrigan, 2009).

That structural geology experts' estimates of location were more strongly influenced by their specialized categories when these were smaller in area, and they were more strongly influenced by the novice categories when they were most helpful suggests that encoding a location entails adaptive use of available categories. Both the efficacy of acquired categories and the flexible use of both conceptual and visually defined categories would appear to argue against alternative models of spatial memory based on single perceptual categories. Although DFT (e.g., Spencer et al., 2006) may offer a neurally plausible framework for the type of Bayesian combination proposed by the CAM, it explicitly states that only perceptual information is used to define spatial categories. As such, it is not clear that DFT can accommodate the present data.[4]

Finally, although our results demonstrate that location memory estimates were most strongly influenced by the spatial category that best constrained potential errors, it remains an open question as to whether these different sources of information were used in isolation or in conjunction with one another. That is, any frame of reference that surrounds a given location will necessarily overlap with any other frame of reference that does the same. It is unclear, though, whether individuals simply use the (single) representation that best constrains errors, or whether they combine across the two representations, with categories akin to the overlapping region of a Venn diagram. A fully Bayesian account would advocate the latter position, assuming that all sources of information about a location (metric and the two categorical representations, in this case) would each be weighted according to its relative reliability. This would allow the individual to make use of *all* available sources of information to minimize errors. Unfortunately, our data do not allow us to distinguish between these alternatives because in some images, the smaller category was completely subsumed by the larger one. Future work should address this issue.

In summary, the study presented here is the first to directly assess the malleability of categorical spatial representations using new conceptual knowledge. Under the framework of the CAM, the results clearly indicate that domain-specific conceptual information was used by expert geologists to define spatial categories. Furthermore, different sources of information were used flexibly, so that spatial categories were defined by the reference frame that best constrained location memory estimates. These results suggest that our extraction and definition of spatial categories is malleable and changes adaptively to optimize particular tasks.

Acknowledgments

This research was supported by the National Science Foundation grants SBE-0541957 and SBE-1041707. Portions of this research were presented at the 53rd annual meeting of the Psychonomic Society, November, 2012. We thank Cathy Manduca, Carol Ormand and Basil Tikoff for help with the definition of geological categories in the images.

Notes

1. Because geology is generally a less populated field than chemistry, and because structural geology is a specific subfield within geology, there is an expected difficulty in obtaining a large sample of experts—even across a number of universities. However, power analyses in our lab using random sampling of groups with known differences (e.g., males and females; Holden & Hampson, 2014) had revealed a priori that sample sizes of approximately $n = 10$ per cell would be sufficient to detect significant differences in the data patterns.
2. In addition, "novice" categories were identified by two experts in English literature and two experts in organic chemistry. The categories identified by these individuals did not differ from those identified by the undergraduate students. However, because the undergraduate composite was based on 30 data points, rather than 4, we used that data set for subsequent analyses.
3. It is worth noting that the relatively large, expert-defined categories in the "*expert-larger*" images did not significantly differ in area from the relatively large, novice-defined categories in the "*expert-smaller*" images (large expert categories: $M = 29{,}703.40$ pixels, $SD = 19{,}783.11$; large novice categories: $M = 27{,}501.07$ pixels, $SD = 17{,}000.37$), $t(28) = 0.48$, *ns*. Likewise, the smaller of the expert- and novice-defined categories did not differ in area (small expert categories: $M = 5{,}430.27$ pixels, $SD = 2{,}569.58$; small novice categories: $M = 7{,}640.60$ pixels, $SD = 6{,}120.18$), $t(28) = 1.29$, *ns*.
4. It is possible that DFT, with some modifications, could accommodate this finding. Because perception includes both bottom–up and top–down processes, it could be that structural geologists' conceptual framework led them to literally perceive the scenes differently than the other expert groups. If so, the spatial categories formed by geology experts via conceptual knowledge could be considered the use of "perceptual" reference frames. However, DFT would need to include a mechanism by which top–down processes in spatial categorization proceed.

References

Batschelet, E. (1981). *Circular statistics in biology*. New York: Academic Press.

Baud-Bovy, G., & Gentaz, E. (2012). The perception and representation of orientations: A study in the haptic modality. *Acta Psychologica*, *141*, 24–30. doi:10.1016/j.actpsy.2012.06.002.

Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin*, *133*, 625–637. doi: 10.1037/0033-2909.133.4.625.

Coleman, S. L., & Gotch, A. J. (1998). Spatial perception skills of chemistry students. *Journal of Chemical Education*, *75*, 206–209. doi: 10.1021/ed075p206.

Corneille, O., Huart, J., Becquart, E., & Bredart, S. (2004). When memory shifts toward more typical category exemplars: Accentuation effects in the recollection of ethnically ambiguous faces. *Journal of Personality and Social Psychology*, *86*, 236–250. doi: 10.1037/0022-3514.86.2.236.

Duffy, S., Huttenlocher, J., & Crawford, L. E. (2006). Children use categories to maximize accuracy in estimation. *Developmental Science*, *9*, 597–603. doi: 10.1111/j.1467-7687.2006.00538.x.

Fitting, S., Wedell, D. H., & Allen, G. L. (2007). Memory for spatial location: Cue effects as a function of field rotation. *Memory & Cognition*, *35*, 1641–1658. doi: 10.3758/BF03193498.

Goodwin, C. (1994). Professional vision. *American Anthrolpologist*, *96*, 606–633. doi: 10.1525/aa.1994.96.3.02a00100.

Holden, M. P., Curby, K. M., Newcombe, N. S., & Shipley, T. F. (2010). A category adjustment approach to memory for spatial location in natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 590–604. doi: 10.1037/a0019293.

Holden, M. P., Duff-Canning, S. J., & Hampson, E. (2015). Sex differences in the weighting of metric and categorical information in spatial location memory. *Psychological Research*, *79*, 1–18. doi: 10.1007/s00426-013-0539-z

Holden, M. P., Newcombe, N. S., & Shipley, T. F. (2013). Location memory in the real world: Category adjustment effects in 3-dimensional space. *Cognition*, *128*, 45–55. doi: 10.1016/j.cognition.2013.02.016.

Huttenlocher, J., Hedges, L. V., Corrigan, B., & Crawford, L. E. (2004). Spatial categories and the estimation of location. *Cognition*, *93*, 75–97. doi: 10.1016/j.cognition.2003.10.006.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*, 352–376. doi: 10.1037/0033-295X.98.3.352.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgement? *Journal of Experimental Psychology: General*, *129*, 220–241. doi: 10.1037/0096-3445.129.2.220.

Kastens, K. A., Manduca, C. A., Cervato, C., Frodeman, R., Goodwin, C., Liben, L. S., Mogk, D. W., Spangler, T. C., Stillings, N. A., & Titus, S. (2009). How geoscientists think and learn. *EOS, Transactions of the American Geophysical Union*, *90*, 265–266. doi: 10.1029/2009EO310001.

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, *6*, 53–84. doi: 10.1016/j.plrev.2008.12.001.

Lee, P. J., & Brown, N. R. (2003). Delay related changes in personal memories for September 11, 2001. *Applied Cognitive Psychology*, *17*, 1007–1015. doi: 10.1002/acp.982.

McNamara, T. P., & Diwadkar, V. (1997). Symmetry and asymmetry in human spatial memory. *Cognitive Psychology*, *34*, 160–190. doi: 10.1006/cogp.1997.0669.

Montello, D. R. (2003). Regions in geography: Process and content. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), *Foundations of geographic information science* (pp. 173–189). London: Taylor & Francis.

Montello, D. R. (2008). Geographic regions as brute facts, social facts, and institutional facts. In B. Smith, D. M. Mark, & I. Ehrlich (Eds.), *The mystery of capital and the construction of social reality* (pp. 305–316). Chicago, IL: Open Court.

Newcombe, N. S., & Huttenlocher, J. (2000). *Making space: The development of spatial representation and spatial reasoning*. Cambridge, MA: MIT Press.

Newcombe, N., Huttenlocher, J., Sandberg, E., Lie, E., & Johnson, S. (1999). What do misestimations and asymmetries in spatial judgment indicate about spatial representation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 986–996. doi: 10.1037/0278-7393.25.4.986.

Resnick, I., & Shipley, T. F. (2013). Breaking new ground in the mind: An initial study of mental brittle transformation and mental rigid rotation in science experts. *Cognitive Processing*, *14*, 143–152. doi: 10.1007/s10339-013-0548-2.

Sampaio, C., & Wang, R. F. (2010). Overcoming default categorical bias in spatial memory. *Memory & Cognition*, *38*, 1041–1048. doi: 10.3758/MC.38.8.1041.

Schutte, A. R., & Spencer, J. P. (2002). Generalizing the dynamic field theory of the A-not-B error beyond infancy: Three-year-olds' delay- and experience-dependent location memory biases. *Child Development*, *73*, 377–404. doi: 10.1111/1467-8624.00413.

Simmering, V. R., & Spencer, J. P. (2007). Carving up space at imaginary joints: Can people mentally impose spatial category boundaries? *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 871–894. doi: 10.1037/0096-1523.33.4.871.

Spencer, J. P., & Hund, A. M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology: General*, *131*, 16–37. doi: 10.1037/0096-3445.131.1.16.

Spencer, J. P., Simmering, V. R., & Schutte, A. R. (2006). Toward a formal theory of flexible spatial behavior: Geometric category biases generalize across pointing and verbal response types. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 473–490. doi: 10.1037/0096-1523.32.2.473.

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, *10*, 422–437. doi: 10.1016/0010-0285(78)90006-3.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, *13*, 407–433. doi: 10.1016/0010-0285(81)90016-5.

Wedell, D. H., Fitting, S., & Allen, G. L. (2007). Shape effects on memory for location. *Psychonomic Bulletin & Review*, *14*, 681–686. doi: 10.3758/BF03196821.

Zar, J. H. (1998). *Biostatistical analysis* (4th ed). Upper Saddle River, NJ: Prentice Hall.