The evolution of spatial representations during complex visual data analysis: Knowing when and how to be exact

Christian D. Schunn (University of Pittsburgh)
Lelyn D. Saner (University of Pittsburgh)
J. Greg Trafton (NRL)
Susan B. Trickett (NRL)
Susan K. Kirschenbaum (NUWC)
Michael Knepp (University of Pittsburgh)
Melanie Shoup (University of Pittsburgh)

Contact author:
Christian Schunn
LRDC 821
3939 O'Hara St
Pittsburgh, PA 15260
schunn@pitt.edu

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
30-09-2005	Final	01-10-2003 — 30-09-2005

4. TITLE AND SUBTITLE
The evolution of spatial representations during complex visual data analysis: Knowing when and how to be exact

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-03-1-0061

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Schunn, C. D., Saner L. D., Trafton, J. G., Trickett, S. B., Kirschenbaum, S. K., Knepp, M., & Shoup, M.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

A — Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

How do problem solvers represent visual-spatial information in complex problem solving tasks? This paper explores the predictions of embodied problem solving and a neurocomputational theory for what factors influence internal representation choices. Data are collected from experts and novices in three different, complex visual-spatial problem-solving domains (weather forecasting, submarine target motion analysis, and fMRI data analysis). Internal spatial representations are coded from spontaneous gestures made during cued-recall summaries of problem solving activities. Analyses of domain differences, expertise differences, and changes over time with problem solving suggest that neurocomputational constraints play a larger role than the nature of the visual input or the nature of the underlying real world being examined through problem solving, especially for expert problem solvers. The particular neurocomptuational feature that was found to drive internal representation choice is the required spatial precision of the main goals of problem solving.

15. SUBJECT TERMS
spatial cognition; representation; complex problem solving

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*

Across all theoretical camps of information processing approaches to cognition, understanding the representation chosen by the problem solver is thought to be fundamental to understanding problem solving behavior (Markman, 1999); it is a key element of what distinguishes cognitive psychology/cognitive science from behaviorism. There may be some debate about the underlying nature of these representations, but all cognitive scientists endorse some form of underlying representation driving behavior (Dietrich & Markman, 2000).

Interestingly, at the same time, there are relatively few theories of internal representation choice (i.e., how problem solvers choose representations). This issue is especially problematic given how central representation is to the paradigm and how variable human internal representation is thought to be. One existing theory of representation choice (Kaplan & Simon, 1990; Lovett & Schunn, 1999) can be summarized under the heading of Search and Rational Choice: problem solvers consider different internal representations when they are unsuccessful and select the representations that turn out to lead to more successful problem solving. There is also an assumption that problem solvers start with certain salient features, but no theoretical specification of what features will be salient. Adding the expertise literature on representation (Chi, Feltovich, & Glaser, 1981; Kaplan & Simon, 1990; Klahr & Dunbar, 1988; Kotovsky, Hayes, & Simon, 1985; Larkin, McDermott, Simon, & Simon, 1980), one could assume that experts generally select representations that most directly captures functional aspects of the given problem (i.e., useful for solution).

But these symbolic approaches say relatively little about representations of very visual-spatial data. To bring the question of representation choice to the topic of visual-spatial problem solving, we develop two other approaches as logical extensions or applications of relevant cognitive theoretical frameworks that traditionally have made strong connections to visual-spatial problem solving. For example, a one approach could tie internal representation choice to the external world, building on the work of ecological psychology (Gibson, 1979; Neisser, 1976), situated cognition (Suchman, 1987), distributed cognition (Hutchins, 1995a, 1995b), embodied action (Fu & Gray, 2004; Gray, John, & Atwood, 1993), and perceptual symbol systems (Barsalou, 1999). Here, internal representation could be tightly connected to the input world (e.g., visual imagery similar to input computer screens) or the output world (mental representations tied to the real world the problem solver is reasoning about).

A second approach is neurocognitive in nature. Internal representations of visual-spatial data must have a neural home. Studies of visual imagery have found that visual imagery uses a large subset of brain regions used for vision itself and that visual imagery has many of the same psychological properties as vision (Kosslyn, 1994; Kosslyn *et al.*, 1999; Kosslyn, Thompson, Kim, & Alpert, 1995). One could extend those findings to suggest that 1) choices of internal representations of visual-spatial data are generally constrained by human neural hardware, 2) functional properties of internal representation will closely mirror functional properties of external input processing brain areas, and 3) people will tend to use internal representations whose functional capabilities best match the needs of the visual-spatial problem solving task at hand.

In both approaches that we have proposed here, like in the general symbolic to representation choice, there is an underlying assumption of optimality: people will tend to use representations that best support problem solving (at least best among the options available). Where the approaches vary is in terms of their views of what constitutes best support of problem solving: concrete match to the outside world or use of the most functionally relevant neural hardware.

No simple set of experiments can easily test between these four very different theoretical analyses of representation choice because the answer may depend upon the nature of the task being performed (e.g., more symbolic vs. more perceptual, or untrained vs. highly trained). Moreover, testing between such very different paradigms is problematic because of the existence of many additional assumptions required to be in place for the measurement of internal representations, and those assumptions themselves will differ across the paradigms.

However, we can ask how useful the different paradigms are for explaining internal representational choice in some interesting cases. In this paper, we present a study designed to look at internal representations of problem solvers in three complex domains. We focus on the case of domains with highly visual-spatial objects because the answer to our larger question is likely to relate to tasks being more perceptual vs. more symbolic in nature. But we use three very different domains within this general type of visual-spatial domains to help unpack what factors influence representation choice, teasing apart the theoretical approaches to predicting representation choice described above. Clearly one cannot generalize from this study to the utility of the different theoretical approaches overall. However, our study will provide a concrete example of how one can empirically test between the utility of the different approaches.

In all three settings, we will examine one particular but important aspect of internal representation: how people represent visual/spatial information. The world is 3-dimensional, but most information sources that experts in complex domains interact with are 2-dimensional (e.g., paper and computer screens). The world exists relative to the problem-solver in egocentric terms, but information sources often present visual/spatial data in exocentric terms. The world is life-sized (by definition), but expert information sources often present scaled versions, either much larger (e.g., via microscopes) or much smaller (e.g., satellite images). Given this diversity of reality and input, how will the problem solver represent their problem solving states internally?

The embodied problem solving approach suggests that representations will match either the form of the external input or the external reality of the problem. What about the neurocomputational problem solver? Here the devil is in the details—in order to develop predictions, we need to select an account (among several competing accounts) for how the brain represents visual/spatial information. We have selected the ACT-R/S theory, and explain it with just enough detail so that the predictions can be made for our current needs.

Brief Overview of ACT-R/S

ACT-R/S (Harrison & Schunn, 2001) is a neurocomputational theory of the visual/spatial representational and computational abilities of the human mind. It integrates current neuroscientific understanding of how the human brain represents visual/spatial information into the ACT-R 5.0 (Anderson, Bothell, Byrne, & LeBiere, 2002) view of how the mind achieves complex problem solving through a rich mixture environment encoding, memory retrievals, and skill applications through goal-directed behavior. In particular, ACT-R/S posits that there are three different visual/spatial representations (see Figure 1), which we call buffers. The three representations make use of different neural pathways, tend to get used for different kinds of basic perceptual/motor tasks, have fundamentally different ways of representing space, and have different strengths and weaknesses. Note that these buffers are multimodal in that they integrate spatial information coming from vision, audition, touch, locomotion, and joint sensors.

The first representation is the Visual Buffer. It is used for object identification and represents information primarily around the region that the eyes are attending to, and represents information in approximate shape terms and approximate size and location. Historically, this buffer has been called the "What" visual pathway. Its representation of the world is primarily a 2-dimensional

world, with objects occupying space in the fronto-parallel plane (i.e., like on a computer screen or chart on the wall in front of you). That is, there are approximate above/below and left/right relationships, but no strong distance and exact orientation information.

The second representation is the Manipulative Buffer. Historically, it has been called the "Where" visual pathway. It is used for grasping objects and tracking of moving of objects, representing information close to within reach, but also all the way around the person. It represents spatial information in highly accurate metric terms, which is required for object manipulation, and in a true 3-D fashion. It is not good at figuring out what objects are, but it knows exactly where they are and what there component shapes are.

The third representation is the Configural Buffer. It is used for navigation in small and large spaces, figuring out where you are, where you want to go, and how to get there. It represents information in terms of egocentric range vectors to blobs (e.g., the desk is approximately so far away, with the left and right side being at such and such angles from me). Locations are configurations of such vectors (e.g., I am at the location that is so far away from the door and such distance from the window, with a given angle between the two).

Note that ACT-R/S makes a 3-way distinction, whereas many spatial reasoning researchers have traditionally made only a 2-way distinction: either the what/where distinction (Ungerleider & Mishkin, 1982) or the small scale exocentric (sometimes called survey) representation versus a large scale egocentric (sometimes called route) representation (Hunt & Waller, 1999; Kozhevnikov & Hegarty, 2001; Tversky, Lee, & Mainwaring, 1999).

Complex-Problem Solving, Representation choice, and ACT-R/S

The strong assumption in ACT-R/S is that these three representations are the only visual/spatial representations that a novice or expert can use for problem solving. Obviously a problem solver may have verbal representations as well, but that type is not addressed by ACT-R/S. In other words, an expert cannot invent a new visual/spatial representation that does not use one (or more) of these three representations, and that their representations will be limited computationally in the same ways as novices based on the properties of these three visual/spatial representation systems. That is, people are assumed to be fundamentally limited by their neurobiology.

ACT-R/S assumes that people can translate between the three representations. In fact, for many tasks, translation and simultaneous activation of different representations is necessary. For example, in order to figure out one's location (a Configural task), one needs to identify what the landmarks are (a Visual task). This ability to translate between representations in general is what makes much of cognitive psychology so difficult because the internal representation can differ dramatically from the input form and can vary substantially across individuals, and the choice of internal representation fundamentally influences performance. For example, people can have visual representations of auditory stimuli, producing visual confusions rather than auditory confusions. In the case of ACT-R/S, a person can take arrangements of distant objects presumably only representable in the Configural space and translate it into a miniature 3D model in the manipulative space, or a flat visual map representation in the Visual space. The way that the person is internally representing the objects will then strongly determine how spatial features are encoded, and thus an important determiner of performance.

The choice of which representation is used will be influenced by input: things in flat displays will tend to start out as Visual; things within reach will tend to start out as Manipulative, and things out in the distance will tend to start out as Configural. However, the choice of representation will also be influenced by functional factors. ACT-R/S thus predicts that people

will tend to move towards representations that have been generally more functional for the goal task at hand. Because the three different representations have very different basic representational form and computational abilities, the match of representation to task should be a strong influence on representation choice.

Location Specificity Predictions from ACT-R/S

With all that theoretical background on ACT-R/S and how it might apply to complex problem solving, we can return to the issue of visual/spatial representations in complex problem solving. The three different spatial systems have varying degrees of match to spatial specificity. All things being equal, ACT-R/S then predicts that internal representation choice, especially in disciplines with complex visual displays, will vary as a function of spatial specificity levels of the scientist doing the data analysis. Manipulative representations will be used when spatial specificity requirements are the highest because the Manipulative space represents spatial location and features in very precise terms. Visual representations will be used when spatial specificity requirements are the lowest because the Visual space represents spatial location and features in very approximate terms, if it does at all. The Configural representation sits somewhere in between, with precise angles, but approximate distance and very approximate shape information.

It is important to note other factors beyond spatial specificity will impact representation choice even in the ACT-R/S framework. For example, the input form of the data will be processed in certain perceptual forms initially, and thus begin with representations tied to input. Expertise will play a role here, too, as experts may be more practiced at transforming from one representation into another. But we focus on the issue of spatial specificity because of the clear predictions from ACT-R/S and also because we have domains that clearly vary in this dimension.

In sum, ACT-R/S makes a variety of predictions for how experts will represent visual/spatial information during data analysis, and one of those predictions involves relative spatial specificity requirements. This spatial specificity requirements prediction is in clear contrast to the predictions of the embodied problem solving approach. The embodied problem-solving framework predicts a match of internal representations to either input or action external representations; it does not make a prediction about the relationship of internal representation choice and special specificity requirements, at least as a general predictor. We studied representation choice in complex problem solving in three different domains to see which perspective could successfully predict internal representation choices.

Measurement of Visuo-Spatial Representations with Gesture

How does one measure internal representations of visual-spatial information? All measures of mental representations is necessarily indirect. Verbal report, either retrospective or online verbal protocols is one general source of data regarding mental representation. However, for visual-spatial representations, it seems a suspect source, as verbal data are generally thought to capture the contents of verbal working memory, not spatial working memory (Ericsson & Simon, 1980). Retrospective or intermittent drawings could be another source of data, however one could imagine that such drawings would be strongly biased towards 2-D representations, and certainly difficult to use to distinguish between large-scale and small-scale internal 3-D representations.

A third approach is to use spontaneous gestures. In addition to serving a communicative act between speaker and listener, spontaneous gestures are thought to be an online measure of

mental representations much like verbal protocols (Alibali *et al.*, 1999; Alibali & Goldin-Meadow, 1993; McNeill, 1992). In our study, we develop an approach to coding gesture that closely maps to the distinctions made by ACT-R/S, namely between 2-D gestures (which we call Visual or Display-based gestures), 3-D small scale gestures (which we call Manipulative gestures), and 3-D large scale gestures (which we call Configural gestures). When our study participants make those kinds of gestures, we will assume that they are working from mental representations of space that have a corresponding structure (namely 2-D, 3-D small, or 3-D large). Of course, not all spontaneous gestures are indicators of a spatial representation—some are just fidgets or more emblematic of symbolic in nature (e.g., the ok sign or representing non-spatial features like time) rather than capturing visuo-spatial representations per se. Thus, our coding of gestures will also attempt to code the many different forms of gestures that occur, and then our analyses will focus in on the gestures that seem to be indicating visuo-spatial representational content.

Methods

Overview

The study examined participants at multiple levels of expertise in three different domains (submarine target motion analysis, meteorological forecasting (METOC), and fMRI data analysis), using a common data collection protocol across domains, which roughly consisted of videotaping a segment of problem solving, followed by a structured interview. We code the participants' representations of 3-Dimensional space from the spontaneous gestures provided by participants during the structured interview.

In the Submarine domain, problem solvers go through one complex scenario with a simulator that closely mirrors interfaces on modern US submarines. The participants' task is to locate an enemy submarine in an environment with a noisy merchant also moving around. In the fMRI domain, we observed participants analyzing their own data. In the weather domain, we observed participants making real forecasts. In all three domains, after 30-60 minutes of problem solving, we then stopped the data analysis activities, and showed the problem-solvers several one-minute videotape segments of their problem solving and asked them to explain what they knew and didn't know at that point in time, so that we could examine how they were representing their data spatially. We examined the gestures produced by problem solvers during those cued recall segments to measure the way they represented their data spatially.

Domain Descriptions

Submarine TMA Domain. While the basic task of finding other submarines using passive sonar (Target Motion Analysis or TMA) remains fundamentally the same very difficult task it has always been, modern computational algorithms and visual displays designed to help the submariner have improved significantly. Figure 3 presents the interface that was used. It runs on a high-end Windows© personal computer, and is an unclassified simulation environment used in engineering development and training situations. It closely mirrors the actual displays used in modern US Navy submarines. Explaining all the displays found in Figure 3 is beyond the scope of this paper. But the key points are that the display includes both egocentric and geosituational views, as well as alphanumeric best-guesses on target location. In general, this environment supports displayed-based problem solving, and thus we may see display-based representations of space in this domain.

fMRI Domain. The goal of fMRI is to discover both the location in the brain and the time course of processing underlying different cognitive processes. Imaging data are collected in

research fMRI scanners hooked to computers that display experimental stimuli to their human subjects. Generally, fMRI uses a subtractive logic technique, in which the magnetic activity observed in the brain during one task is subtracted from the magnetic activity observed in the brain during another task, with the assumption that the resulting difference can be attributed to whatever cognitive processes occur in the one task but not the other. Moreover, neuronal activity levels are not directly measured, but rather one measures the changes in magnetic fields associated with oxygen-rich blood relative to oxygen-depleted blood. The main measured change is not the depletion due to neuronal activity but rather the delayed over-response of new oxygen-rich blood moving to active brain areas, and the delay is on the order of 5 seconds, with the delay slightly variable by person and brain area. Data are analyzed visually by superimposing color-coded activity regions over a structural image of the brain (see Figure 2a), looking at graphs of mean activation level by region and/or over time (see Figure 2b) or across conditions (see Figure 2c), or looking at tables of mean activation levels by region across conditions (see Figure 2d). Elaborate, multi-stepped, semi-automated computational procedures are executed to produce these various visualizations, and given the size of the data (gigabytes per subject), many steps can take up to several minutes per subject. Inferential statistical procedures (e.g., t, ANOVA) are applied to confirm trends seen visually. Note that, as in the submarine domain, the input displays are very 2-dimensional, even though the underlying reality (activation in brain regions) is 3-dimensional. Unlike the submarine domain, however, the underlying reality takes place in a very small space (smaller than a breadbasket, relatively nearby) whereas in the submarine domain, the real space is many miles in every direction, with objects being the size of small to medium-sized buildings.

METOC Domain. Weather forecasters examine observations, summaries of those observations, and predictive forecast models that use those observations as input. While they do explicitly examine actual observations by examining satellite pictures or local wind-speed, the majority of their information comes from tools that summarize or use those observations. Figure 4 is a snapshot of a forecaster examining such a summary visualization. In general, the visualizations present exocentric top-down views of spatially distributed weather data. In our study, forecasters provided a mixture of local and remote weather forecasts for the near future.

Participants

In each of the three domains, we observed participants at different levels of expertise to see how representations change with expertise. As is typical of expertise studies in complex real-world domains, our Ns are not large. Moreover, we made use of populations that were available to us, which produced different distributions of participants along the expertise continuum in each of the three domains. In order to avoid representing expertise as a false dichotomy and to better facilitate alignment of our results between domains and with later research, we use the following labels. Novices are those participants who are minimally capable of doing the given tasks on their own but have only recently reached that level and still make considerable errors. Intermediates are those participants have progressed beyond the novice level but are not yet at the highest levels of performance. Experts are those participants who have progressed to the highest levels of performance in their field. Note that we do not study those participants completely unfamiliar with the given tasks; although commonly studied and given the label ‘novice’, those kinds of participants would not even be capable of approximating the complex real-world tasks that we examine.

Submarine. There were 7 submarine experts and 11 intermediates, all of whom were Officers who taught at Submarine School. The experts and intermediates were equally senior in rank, but

the experts regularly taught Target Motion Analysis whereas the intermediates taught other topic areas and had less practice with Target Motion Analysis.

fMRI. There were 10 fMRI participants, ranging from beginning graduate students to postdoctoral researchers. This study focused on naturalistic analysis of data, and faculty in this domain tend not to be directly involved in analysis of fMRI data; instead faculty work with students and postdocs after analyses have been carried out. We divided the participants into three expertise levels based on the number of studies they had carried out: 4 participants classified as Experts had carried out 4 or more fMRI studies, 4 participants classified as Intermediate has carried out between 2 and 3 studies, and 2 participants classified as Novices had carried out only 1 study.

METOC. There were 4 experts and 10 novices. The expert meteorologists had over 10 years experience working as Navy forecasters. The novice forecasters were junior and senior meteorology majors with an average of 2.75 years experience.

Procedure

The study took place at the participants' regular work location (or in a lab in the case of the meteorology students), and all participants used the tools, visualizations, and computer equipment that they usually employed. The one exception was that the submarine domain participants used the particular simulator we provided, primarily because videotaping is rather difficult in a real submarine. All participants agreed to be videotaped during the session. Participants were trained to give talk-aloud verbal protocols (Ericsson & Simon, 1993). All participants were instructed to carry out their work as though no camera were present and without explanation to the experimenter.

While the participants performed the task, the experimenter made note of "interesting events, such as major changes in the computer display, such as a new visualization or application or an event that spurred a burst of participant activity. This sampling approach allowed us to examine the evolution of spatial representations across the problem-solving episode without making the cued recall task monotonous for participants. After the task was completed, the experimenter showed the participant a one-minute segment of the video surrounding each of the interesting events. After reviewing each one-minute segment of videotape, the experimenter asked the participant "What did you know and what did you not know at this point?" Participants' responses to these cued-recall questions were also recorded on videotape.

Because participants were using computer interfaces during the actual problem solving with hand out mouse or keyboard, they made almost no gestures during that phase. Therefore, the spontaneous gestures that participants made during the cued recall are the focus of our analyses.

In the weather domain, we were able to test whether familiarity with the situation details could perhaps be a confound with (or less interesting source of) expertise differences. Therefore, we asked half of the novices to make a local forecast and half to make a forecast for a remote (and less familiar) location.

Predictions

Submarine domain. None of the input in target motion analysis tasks shows the equivalent of a view out of a window although there is a bird's-eye-view with ownship in the center and lines of bearing to other platforms, and current solution, if available. The visual/spatial displays are all 2-dimensional, complex displays. At the same time, the real world being reasoned about is a very large, 3-dimensional world.

The embodied problem solving perspective predicts that problem solvers will use either 2D display-based reasoning (the input) or large-scale 3D (configural) reasoning (the real world). By contrast, the neurocomputational perspective (in ACT-R/S) suggests that problem solvers will move from a display or configural representation to a manipulative (small 3D) representation because 1) configural or display representations are more appropriate for weak initial knowledge of location and distance, and 2) manipulative representations are more appropriate when location and distance are more accurately known. The neurocomputational perspective is the only one that very clearly predicts a change in internal representation choice for this task over time during problem solving.

fMRI domain. The embodied problem solving approach predicts that fMRI scientists should use manipulative (real-world) and display-based (input) representations. The neurocomputational perspective predicts that representations will go from manipulative representations to display-based representations, for the following reasons. When imaging data are first examined, determining precisely where the regions of activity are located in the brain is important. However, the end goal of fMRI analysis (at least as practiced by cognitive neuroscientists) is functional activity not precise location, so the problem solvers should move to less precise representations (e.g., display-based representations).

METOC domain. Weather forecasting from an embodied problem solving perspective is similar to submarine: the inputs are 2-dimensional and the real world is large 3-dimensional, and thus one would predict display-based and Configural representations. From a neurocomputational perspective, the weather forecasting domain is actually more like fMRI data analysis: problem solvers go from determining where weather patterns are located to developing a qualitative understanding of how weather is progressing over time. Thus, the neurocomputational perspective predicts a shift from manipulative to display-based representations.

The predictions across the 3 domains are summarized in Table 1.

Gesture Coding

Visual-spatial representations were coded from the spontaneous gestures made during the cued recall phase. Configural gestures were those made with the hand or arm such that the fingers are pointing in a direction without attempting to pick up or place or otherwise manipulate imaginary objects. These were usually one-handed gestures and one-dimensional, but some were two-handed when they have a quality of pointing into the distance. They could represent limited motion, for example in a single direction, but only if it seems the motion being capture was of an object in the distance rather than at the location of the hand itself. See Figure 5 for an example of a two-handed configural gesture in which the hands represent the angle at which the target is at relative to the heading of ownship.

Manipulative gestures placed objects and activity in a nearby space, such that the problem solver can actually manipulate or place the imaginary objects. Examples of manipulative gestures included one-handed gestures of a brain region and two-handed gestures showing two contacts and the relative motion involved or changes in bearing and curves in paths or course. Gestures in which the hand-shape suggests placing or holding as opposed to strictly pointing were also coded as manipulative. Figure 6 presents examples of manipulative gestures from two domains.

Display-based gestures were those gestures that place objects and activity on a flat surface in the fronto-parallel plane, mimicking a computer screen or map or diagram. Figure 7 presents an example display-based gesture in which the participant talks about brain activation of two

different spatial regions in terms of a flat bar-graph representation spatial region being represented one-dimensionally on the x-axis.

There were also several other kinds of gestures that were coded but do not bear on our analyses of spatial representations: 1) representationally ambiguous deictic gestures, which involved pointing to objects in the environment around the problem solver such as a piece of paper on the desk or the computer screen; 2) uncertainty-based gestures, in which participants shrugged or wiggled their hands indicating uncertainty about the situation; 3) iconic gestures, in which participants represented objects literally but the objects were not relevant to the problem solving domain (i.e., not part of the brain, a weather object, or an object in the surrounding water) such as gestures representing the computer mouse, a soda can, or a book; 4) metaphorical gestures, in which space was used to represent a non-spatial object or dimension (e.g., time); and 5) beat gestures, which were repetitive gestures thought either to be meaningless or perhaps ways of indicating emphasis in speech. Because these codes required some understanding of what was being represented, the gestures had to be coded with access to the spoken transcript.

Two coders independently coded all the data. They agreed upon whether there was a spatial gesture in a given segment 94% of the time (Kappa=.60). Of cases in which they independently noticed a spatial gesture, they agreed 74% of the time (Kappa=.46) on the form of the spatial gesture (configural, manipulative, or display). Because the coding task was relatively difficult, the coders both coded all the data and all disagreements were resolved through discussion.

Results

Domain Differences in Expert Representations

We begin with the representations chosen by experts in each of the domains. On the one hand, they are most likely to have come to understand the affordances of different representation choices for the tasks at hand. On the other hand, they may be the least bound by biological constraints, having had the opportunity to develop new, more abstract and functionally relevant internal representations of spatial information.

Figure 8 presents the proportion of spatial gestures of each type within each of the three domains using only the expert participants' data. We see that expert representations include a rich mixture of manipulative and display gestures in the fMRI and METOC domains, whereas in the Submarine domain, experts primarily use manipulative representations—the comparison of Submarine against fMRI and METOC were statistically significant, ($X^2(1)=17.6$, $p<.001$ and $X^2(1)=25.7$, $p<.001$ for the two pairwise comparisons respectively).

Comparing the three domains, we can suggest several conclusions about expert representations. First, the underlying reality appears to matter a little. There were no configural gestures in the fMRI domain (to a large or distant brain) but there were some (although relatively few) configural gestures in the submarine domain. But it cannot matter much, as there were no Configural gestures in the weather domain despite representing very large (and sometimes nearby) space. Second, the data from the submarine domain suggest that neurocomputational factors appear to matter a lot, because the most common representation (manipulative) corresponds to neither input nor external reality.

The diversity of representations within each group suggest that an account like ACT-R/S, in which there can be multiple spatial representations, is useful for highlighting representational variability. It is also the case that some participants used few spatial gestures overall. We suspect they were thinking spatially, but there are large individual differences in how much people gesture. The majority who used at least three gestures had both manipulative and display

gestures, suggesting the diversity does reside within individuals rather than reflecting individual choice of a single representation to use throughout problem solving.

Evolution of Representations

We were also interested in how internal representations evolved, both over time within a problem-solving episode, and over developmental time, with increasing expertise. To examine changes within a problem-solving episode, cued-recall minute segments were divided in to the first half vs. the second half of cued-recall minutes for each participant.

As there were interactions of domain by expertise by early/late, we present the full 3-way interaction graphs (see Figure 9). In the fMRI domain, Novices begin with primarily display gestures and move to manipulative gestures ($X^2(1)=10.4$, $p<.01$), whereas Intermediates and Experts show the opposite trend ($X^2(1)=4.1$, $p<.05$ for Intermediates, $X^2(1)<1$ for Experts). The proposed interpretation for that novice pattern is the following: 1) novices are initially tied to the input representation (2-dimensional displays); 2) novices then struggle to form a precise 3-D representation of the locations they have observed; and 3) novices are less like to make it to the more abstraction functional final understanding of the underlying brain activity. The proposed interpretation of the intermediate and expert pattern is the following: 1) intermediates and experts begin quickly with precise 3-D representations of the brain activity (despite being given the same 2-D display input as the novices); 2) intermediates and experts move to focusing on abstract functional activity, which is best-represented with 2-D visual mental representations; and 3) experts having made the transition to abstract representations earlier in problem solving than intermediates.

In the METOC domain, Novices move from an even balance of manipulative and display gestures to predominantly manipulative gestures; whereas Experts go from an even balance to more display gestures—the Expert/Novice difference at the Late period is statistically significant ($X^2(1)=6.5$, $p<.02$). Note that the representation that corresponds with the underlying reality (Configural) is very infrequent and only occurs in the early problem solving of the novices. The proposed interpretation of the novice data is: 1) novices are beginning with representations heavily tied to the input displays (2-D), and 2) are trying to build exact mental situation models which are 3-D. By contrast, we interpret the expert data in the following way: 1) the experts build their exact (3-D) situation models quickly, and 2) then move to more abstract, qualitative mental models of weather (Trafton *et al.*, 2000), which are best captured with 2-D mental representations.

In the Submarine domain, all participants appeared to have a high proportion of manipulative representations, but experts have the highest proportion, especially in the Late period ($X^2(1)=9.6$, $p<.01$). Experts have no display-based gestures, showing no dependence on the purely 2-dimensional input. The proposed interpretation is that the experts are best able to come to exact representations of the enemy submarine's location and thus will need to use manipulative representations, especially late in problem solving. By contrast, intermediates appeared to have struggled to develop exact location representations of the enemy submarine.

Testing the Impact of Familiarity

One might wonder whether familiarity with the particular problem-solving situation may account for some of the differences between representation choice of experts and novices. The data from the METOC domain can be brought to bear on this question, as half the novices were given a local forecast task and half the novices were given a remote (unfamiliar) forecast task.

Figure 10 shows the impact of forecast location on spatial gesture proportions. There is a fairly large effect, with local forecasts producing primarily manipulative gestures, and remote forecasts being more of a 50/50 mix ($X^2(1)=8.2$, $p<.01$). Consistent with these results, when asked in a debriefing, the local forecast participants typically described their mental images as being well within reach, whereas the remote forecast participants typically described their mental images as being beyond their reach.

In other words, the student mental representations were more similar to those of expert in their remote forecasts (i.e., more display, less manipulative gestures). If the expertise differences were actually due to familiarity, then we would have expected local forecasts (familiar to the students) to involve the more expert-like representations, whereas the reverse was found to be true.

Why would student mental representations be more similar to those of experts for the remote forecasts, rather than just showing no effect of remote vs. familiar? It seems unlikely that familiarity would reduce the quality of a student's representation. Instead, it is worth pointing out an ambiguity in the use of 2-dimensional representations: they could represent the earliest representation (very tied to input forms) or they could represent the most advanced representation (in the METOC domain), a qualitative mental model of the overall situation. It is most likely that familiarity does have some impact on representation choice and it is of allowing novices to move beyond the earliest representation (input-based) to a quantitative situation model.

General Discussion

Across the three domains, we found that the predictions of the neurocomputational account were generally met. The exact proposed interpretation of the observed spatial representation differences and changes over time are but one possible interpretation of the results; they were presented primarily to provide a working hypothesis that is plausible given the nature of the domains. However, it is clear that embodied problem solving predictions for internal representation choice were largely falsified, especially by expert representation choices. For example, it appears that reality primarily matters in novices and early in problem solving. Moreover, we saw that fMRI and METOC domains behave similarly to one another and quite differently from the submarine domain, whereas one would have expected METOC and submarine domains to behave more similarly if embodied problem solving were the better approach to predicting internal representations.

By contrasting representation choices across three very domains, we present evidence that suggests that spatial informational specificity is related to the selection of internal visual/spatial representations. In particular, we found that the domain with high spatial specificity goals and low initial spatial precision of input (submarine TMA) shows movement over time towards manipulative representations, which we argue are best suited for representing precise spatial locations and shapes. Consistent with this view, experts, who are most able to develop precise final solutions showed the strongest movement to manipulative representations. The two domains with intermediate initial spatial specificity goals and low final spatial specificity goals (METOC and fMRI) showed movement over time away from manipulative representations. Again consistent with this view, experts, who are most able to move beyond the first steps of data analysis, showed the strongest movement away from manipulative representations.

Moving to the next level of abstraction, this paper presents some of the first results suggesting that expert representations are better predicted by the match of task goals to

neurocomputational constraints—experts appear to use existing visual/spatial systems for problem solving on the basis of how well the computational abilities of those systems support the current needs/features of the given task.

The Absence of Configural Representations

An interesting result of this study was that there were relatively few configural gestures in any of the domains. We assume in this work that configural gestures are the indicators of large-space egocentric representations, and display and manipulative gestures are indicators of smaller, exocentric representations. Where do configural representations occur? We found them in experts, intermediates, and novices. We found them in the METOC and submarine domains, but not the fMRI domains. Although not a significant effect in any one case, in all participant types in both domains, we see them occur more often earlier in problem solving than later in problem solving. The results, brought together, suggest that configural spatial representations (or route representations in other researchers' language) occur only in reasoning about situations that involve large spaces (or at least larger than a brain), and occur primarily early in problem solving as the problem solver tries to contextualize results in a instantiated model of the real situation. Note, however, that familiarity with the situation is not key because the local vs. remote forecasts manipulation in the METOC domain had no strong impact on configural representations, or at least not in the direction of familiarity producing more configural representations.

The most interesting feature of this general absence or very low levels of configural representations is that it suggests the focus of past researchers on small vs. large or exocentric vs. egocentric representations in studies of visuo-spatial reasoning may be misplaced, especially when brought to the context of studying real world problem solving with complex visuo-spatial data. In our three domains, it appears that problem solvers generally work with small scale, exocentric representations of visuo-spatial data and rarely use large scale, exocentric representations. Yet, they appear to be making predictable and, in some cases, large changes with time and expertise in their spatial representations: between 2-D fronto-parallel representations (display-based or visual in our terminology) and 3-D representations (manipulative in our terminology). If these results correctly characterize visuo-spatial representations in complex visuo-spatial problem solving domains, then it suggests a change in focus of spatial reasoning research. At the very least, it suggests that researchers in general should distinguish between the two different forms of small-scale, exocentric representations.

Caveats

It is important to note several caveats. First, there were not a large number of participants within each of the three domains, especially when divided into different expertise levels. Many of the observed results were rather large effects and thus statistically significant, but, nonetheless, the results should be replicated with additional participants.

Second, the inference from spontaneous gestures to internal representation choice is somewhat indirect. Given the large differences over time and by domain and expertise group, the spontaneous gestures are clearly not random. Prior work on gestures suggests that gestures do provide good insights into internal representations generally (Alibali et al., 1999; Alibali & Goldin-Meadow, 1993) and spatial representations in particular (Emmorey, Tversky, & Taylor, 2000; McNeill, 1992). However, our approach to coding spatial representations uses new divisions that have not been validated and future studies should use a variety of measures of representation choice.

References

Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E., & Goldin-Meadow, S. (1999). Illuminating mental representations through speech and gesture. *Psychological Science, 10*(4), 327-333.

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*(4), 468-523.

Anderson, J. R., Bothell, D., Byrne, M., & LeBiere, C. (2002). An integrated theory of the mind. from http://act-r.psy.cmu.edu/papers/403/IntegratedTheory.pdf

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences, 22*(4), 577-660.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Dietrich, E., & Markman, A. B. (2000). Cognitive dynamics: Computation and representation regained. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics* (pp. 5-30). Mahwah, NJ: Erlbaum.

Emmorey, K., Tversky, B., & Taylor, H. A. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Journal of Spatial Cognition and Computation, 2*, 157-180.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.

Fu, W.-T., & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science, 28*(6).

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project ernestine: Validating goms for predicting and explaining real-world task performance. *Human Computer Interaction, 8*(3), 237-309.

Hunt, E., & Waller, D. (1999). *Orientation and wayfinding: A review. Onr technical report n00014-96-0380*. Arlington, VA: Office of Naval Research.

Hutchins, E. (1995a). *Cognition in the wild*. Cambridge, MA, USA: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science, 19*(3), 265-288.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*, 374-419.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1-48.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge. MA: MIT Press.

Kosslyn, S. M., Pascual-Leone, A., Felican, O., Camposano, S., Keenan, J. P., Thompson, W. L., et al. (1999). The role of area 17 in visual imagery: Convergent evidence from pet and rtms. *Science, 284*(April 2), 167-170.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual conrtex. *Nature, 378*(Nov 30), 496-498.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from tower of hanoi. *Cognitive Psychology, 17*(2), 248-294.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29*, 745-756.

Larkin, J. H., McDermott, J., Simon, D., & Simon, H. (1980). Expert and novice performance in solving physics problems. *Science, 208*, 140-156.

Lovett, M. C., & Schunn, C. D. (1999). Task representations, strategy variability and base-rate neglect. *Journal of Experimental Psychology: General, 128*(2), 107-130.

Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Erlbaum.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL, USA: University of Chicago Press.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman.

Suchman, L. A. (1987). *Plans and situated action: The problem of human-machine communication*. New York: Cambridge University Press.

Trafton, J. G., Kirschenbaum, S., S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human Computer Studies, 53*(5), 827-850.

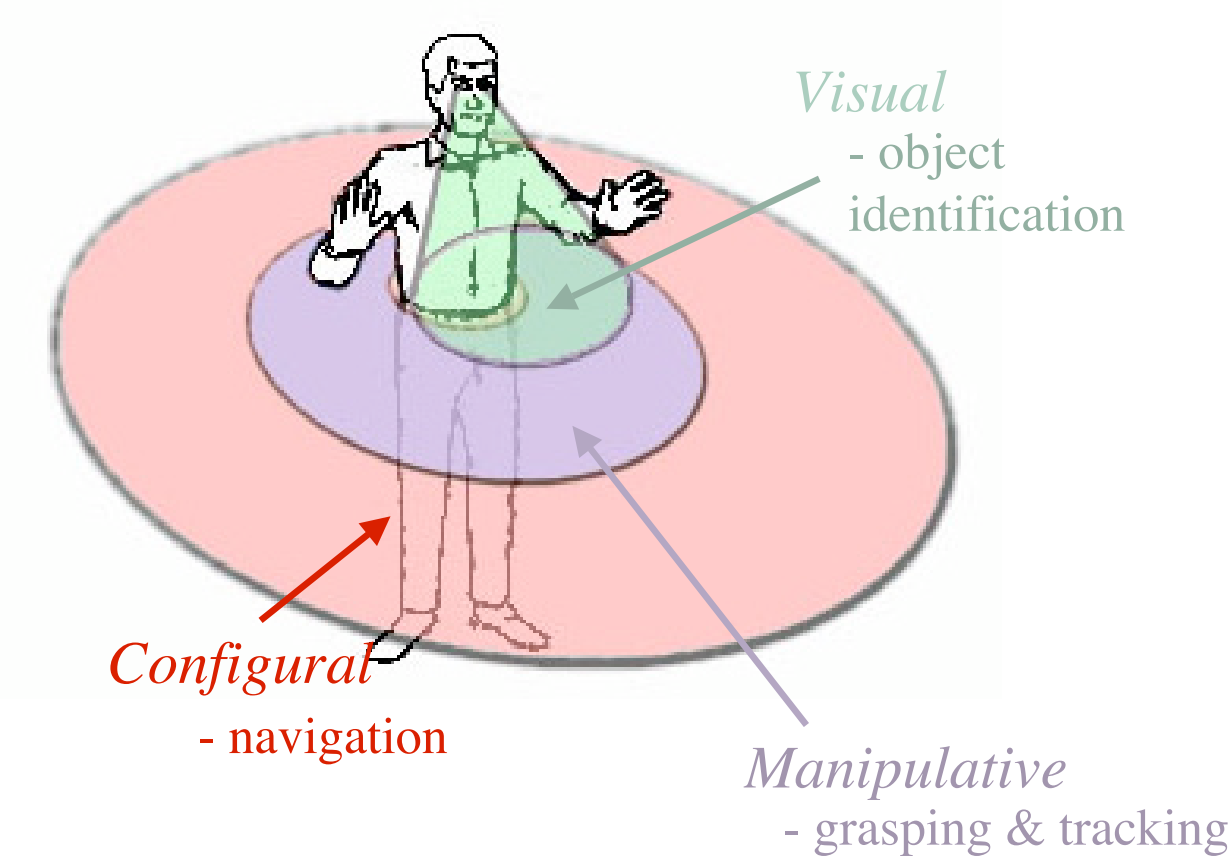
Tversky, B., Lee, P. U., & Mainwaring, S. (1999). Why speakers mix perspectives. *Journal of Spatial Cognition and Computation, 1*, 399-412.

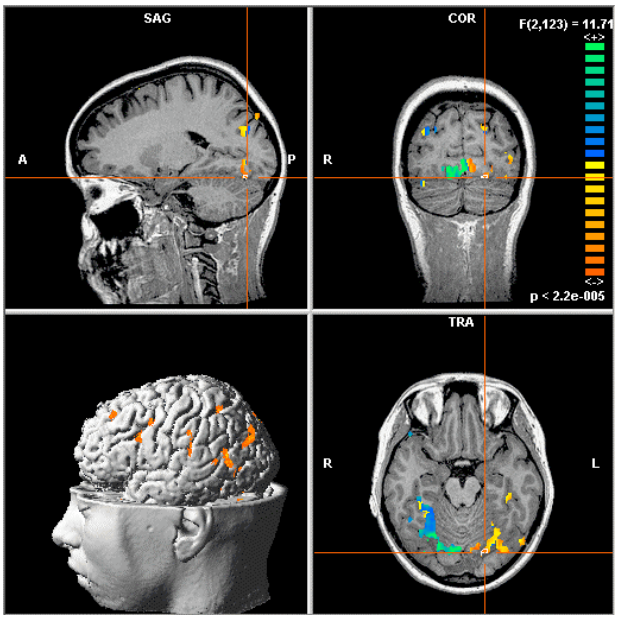
Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior*. Cambridge, MA: MIT Press.

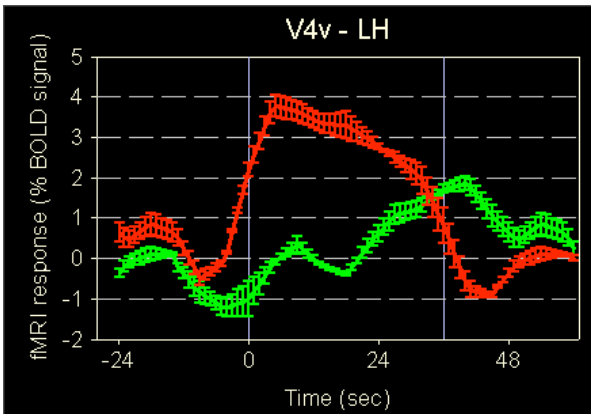
Table 1. Predictions for visual-spatial representation choice (and thus observed gesture type) in the 3 domains under the two different prediction frameworks.

Domain	Embodied Framework (Input/Output)	Neurocomputational Framework
Submarine TMA	Display or Configural	Manipulative
fMRI	Display or Manipulative	Display
METOC	Display	Display

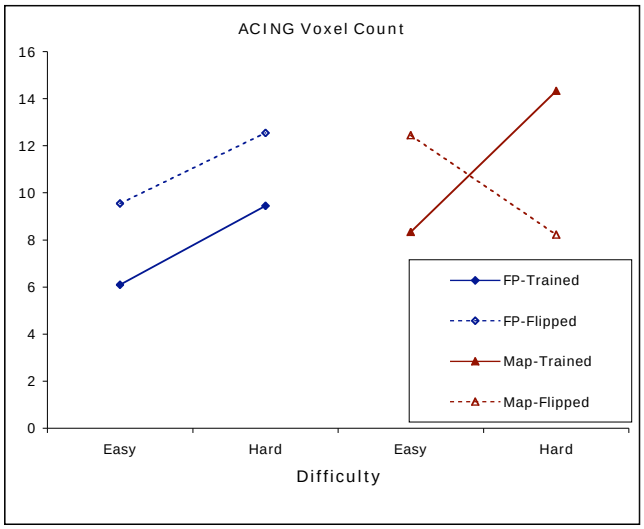
Figure 1. Three visual/spatial representation systems posited in ACT-R/S, the size and location of space they cover, and the basic tasks they typically support.







c)



d)

ROI	TrainedEasy Count	TrainedHar Count	FlippedEasy Count	FlippedHard Count
1	2	3	4	5
ACING	8.33	14.33	12.44	8.22
CALC	72.33	87.56	75.89	103.22
LCBELL	21.11	33	27.22	28.89
RCBELL	19.44	38.44	28.33	33.67
LCBELL+RCBELL	40.56	71.44	55.56	62.56
LDLPFC	**32.78**	**51**	**38.22**	**40.56**
RDLPFC	**37.56**	**61.67**	**43.44**	**57.33**
DLPFC	**70.33**	**112.67**	**81.67**	**97.89**
LFEF	5.33	8.22	4.67	7.33
RFEF	3.44	5.56	3.78	6.56
LFEF+RFEF	8.78	13.78	8.44	13.89
LHG	12.67	1.44	5.56	1.56
RHG	14	3	6	3.78
LHG+RHG	26.67	4.44	11.56	5.33
LIES	77.56	104.89	84.78	109.44
RIES	59.11	76.56	59.56	81.67
LIES+RIES	136.67	181.44	144.33	191.11
LIPL	26.44	22.22	27.33	20.67
RIPL	19.11	21.89	20.33	20.56
LIPL+RIPL	45.56	44.11	47.67	41.22
LIPS	**54.67**	**76.56**	**61**	**75.22**
RIPS	**37.67**	**64.11**	**48.11**	**72.56**
IPS	**92.33**	**140.67**	**109.11**	**147.78**

Figure 2. Kinds of visualizations examined in analysis of fMRI data: a) degree of activation indicated with a color scale superimposed over a gray-scale structural brain image in three different planar slices and a surface cortex map; b) graph of percent signal change in a brain region as a function of time relative to a stimulus presentation in two different conditions (red and green); c) graph of number of activated voxels in an area as a function of various condition manipulations; and d) table of number of activated voxels in different brain areas (Regions of Interest) as a function of different conditions.

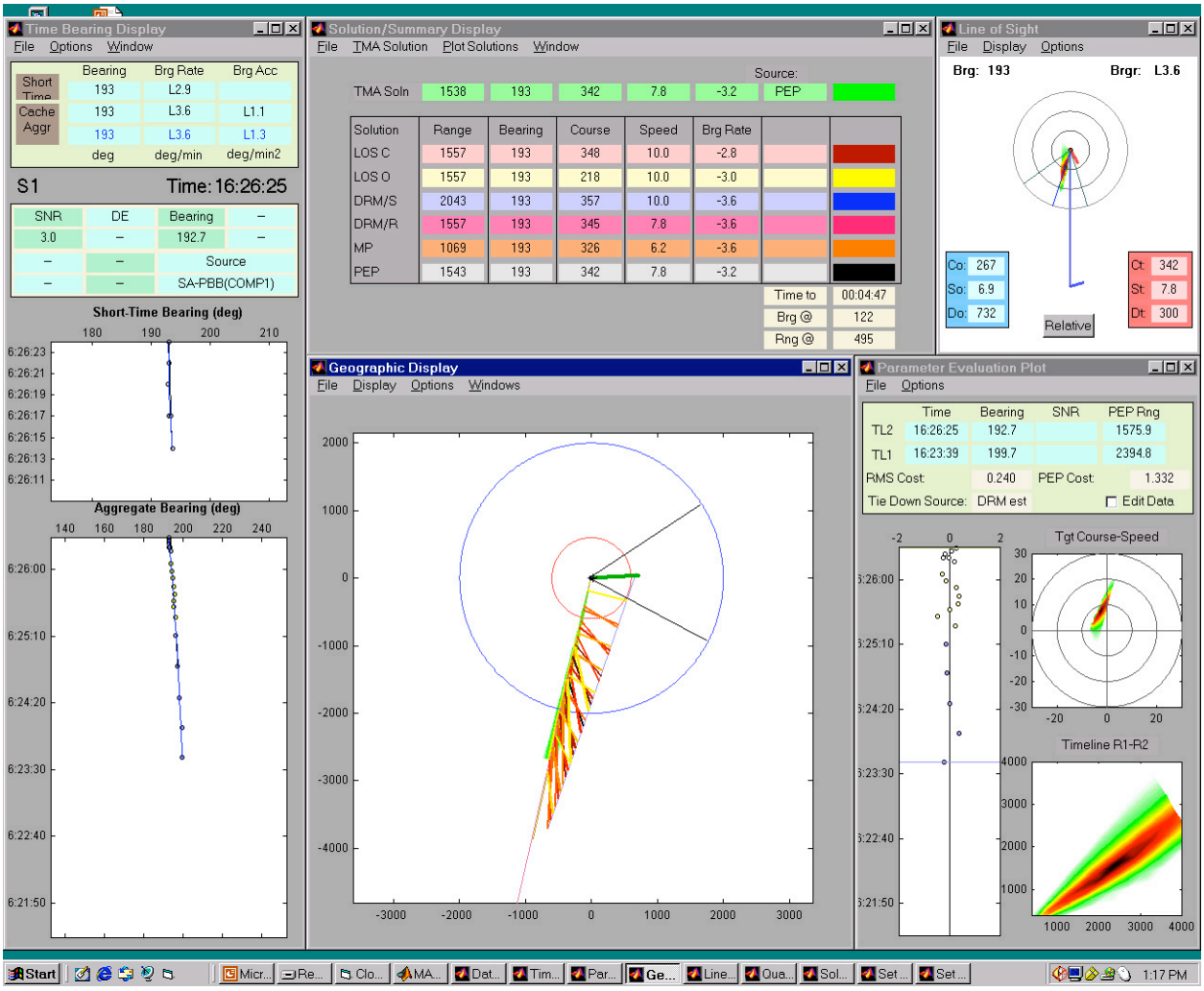


Figure 3. Display used in the Submarine Target Motion Analysis task. The table in the middle shows possible solutions calculated through 6 different algorithms. The graph in the middle is a geosituational plot with ownship in the center, the dark green line indicating ownship motion direction, and the red lines indicating possible distance/location/direction of the target. The graphs on the right and bottom left are various complex egocentric perspectives on the target.

Figure 4. A typical display used by weather forecasters in this study.

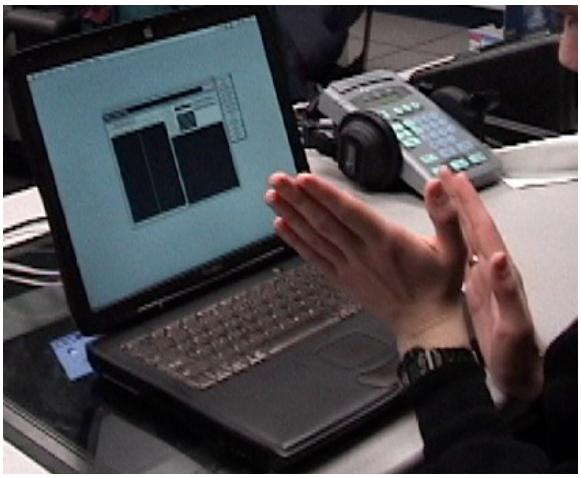


Figure 5**.** A participant's configural gesture while saying "…bearing around course oh three five, our ownship course is about three five seven, we'll be about…here".

a) b)

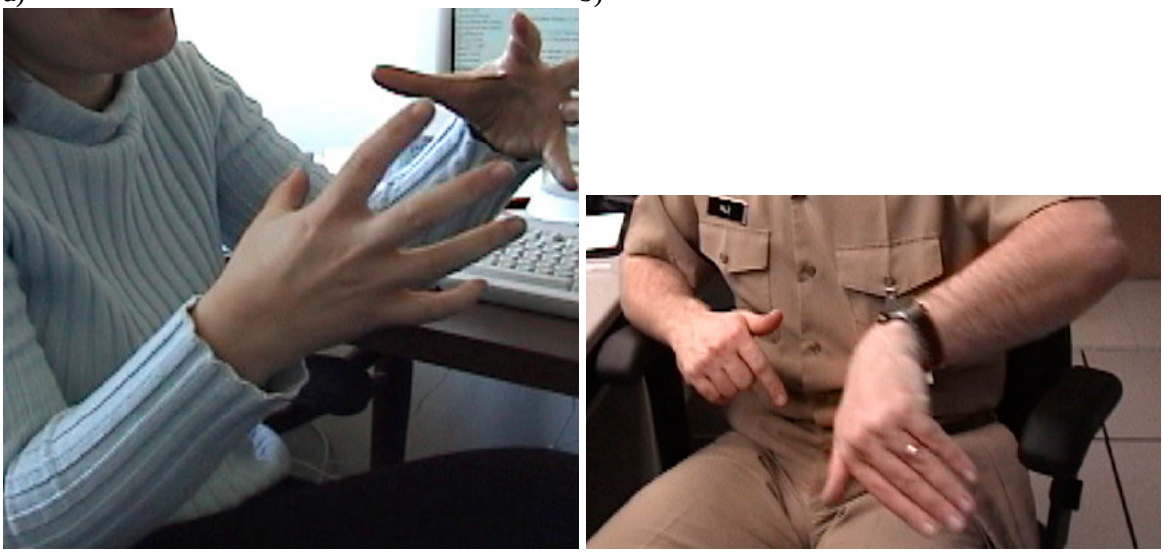


Figure 6. Examples of manipulative gestures. a) A fMRI participant's manipulative gesture, while saying, "… if you have, like, this massive thing, the peak is really in there…", b) A submarine participant's manipulative gesture, while saying *"I should've gone left…come left and gone be**hind** him…"*.

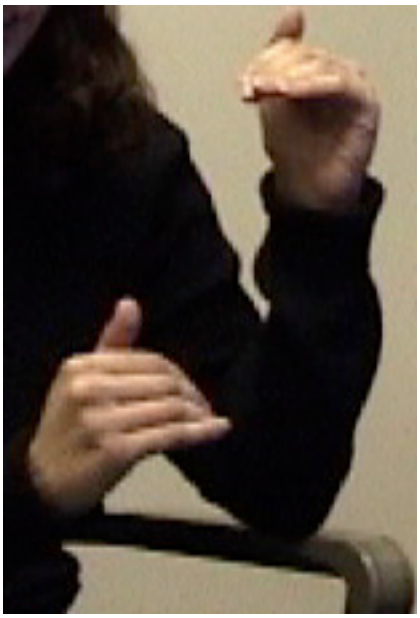


Figure 7. An example display-based gesture, "...I found out that, it looked like there's a difference between frontal and hippocampal activation..."

Figure 8. For experts only, the proportion of spatial gestures of each type within each domain (with SE bars).

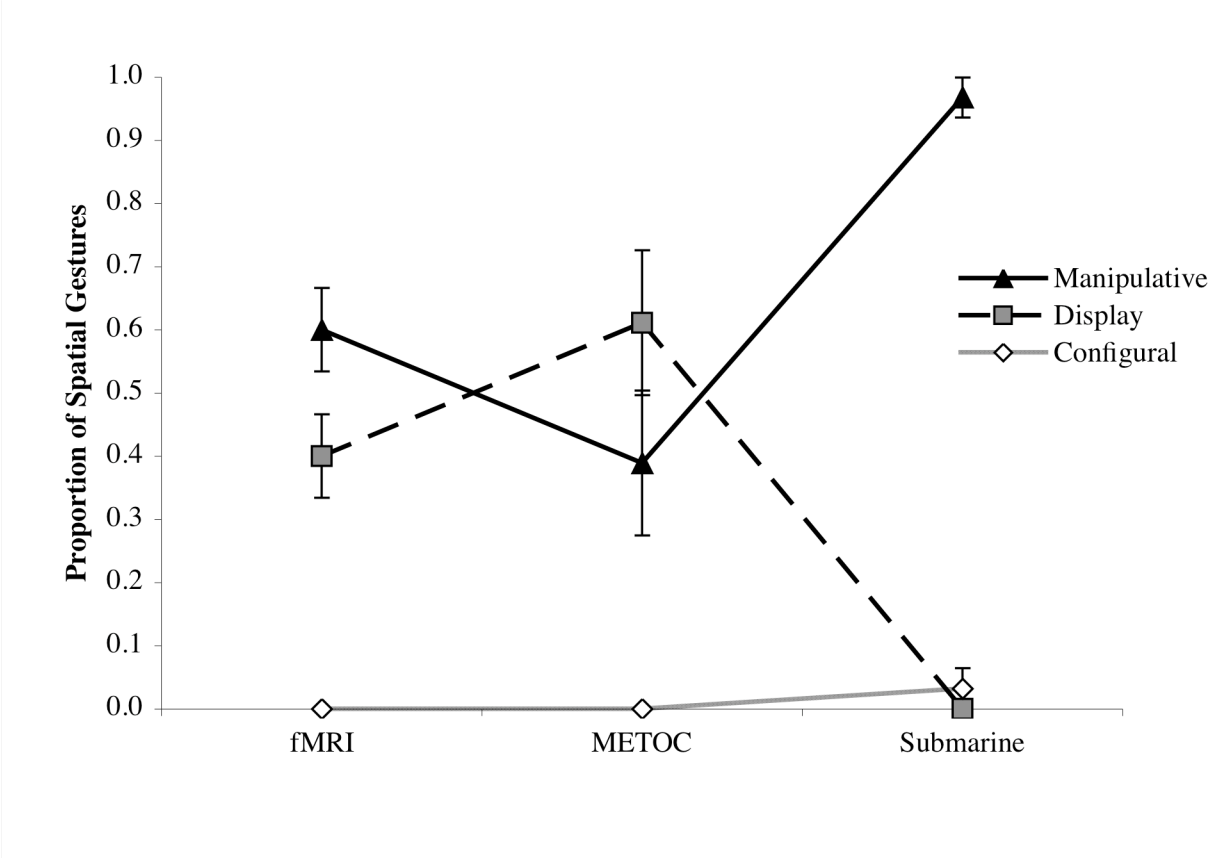


Figure 9. The distribution of spatial gestures early and late in each problem-solving episode for each expertise group in each domain (with SE bars).

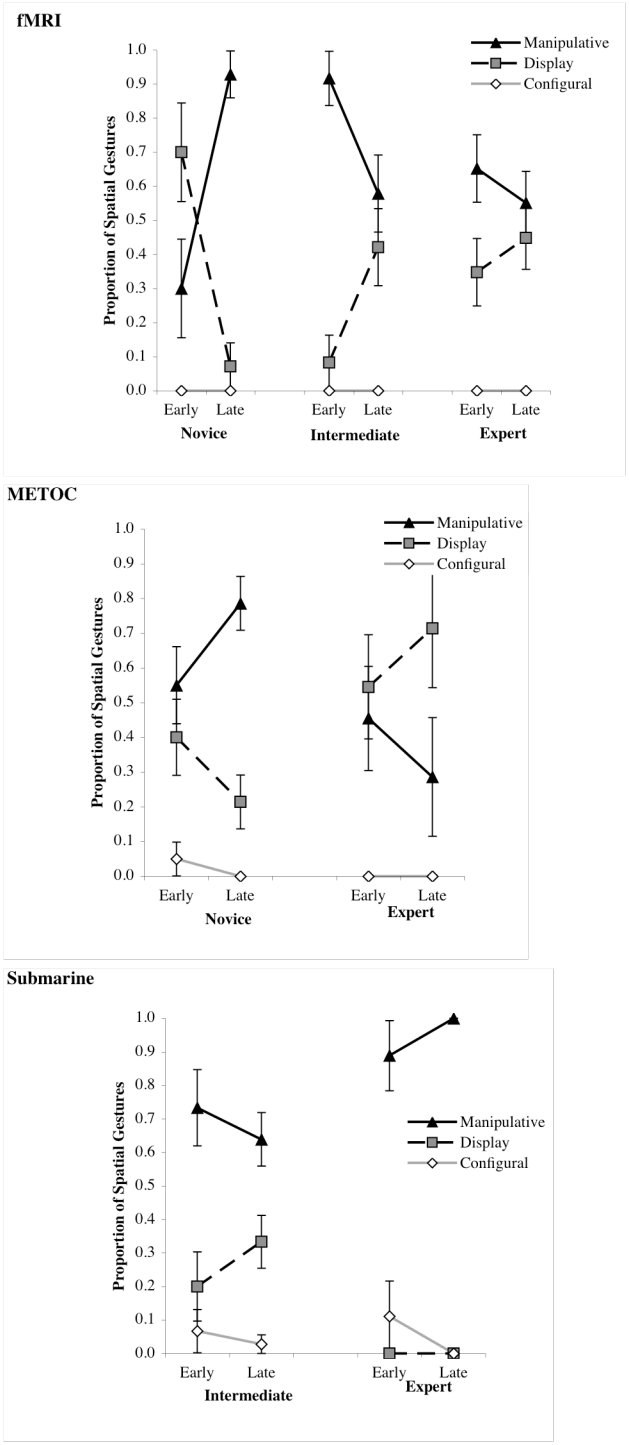
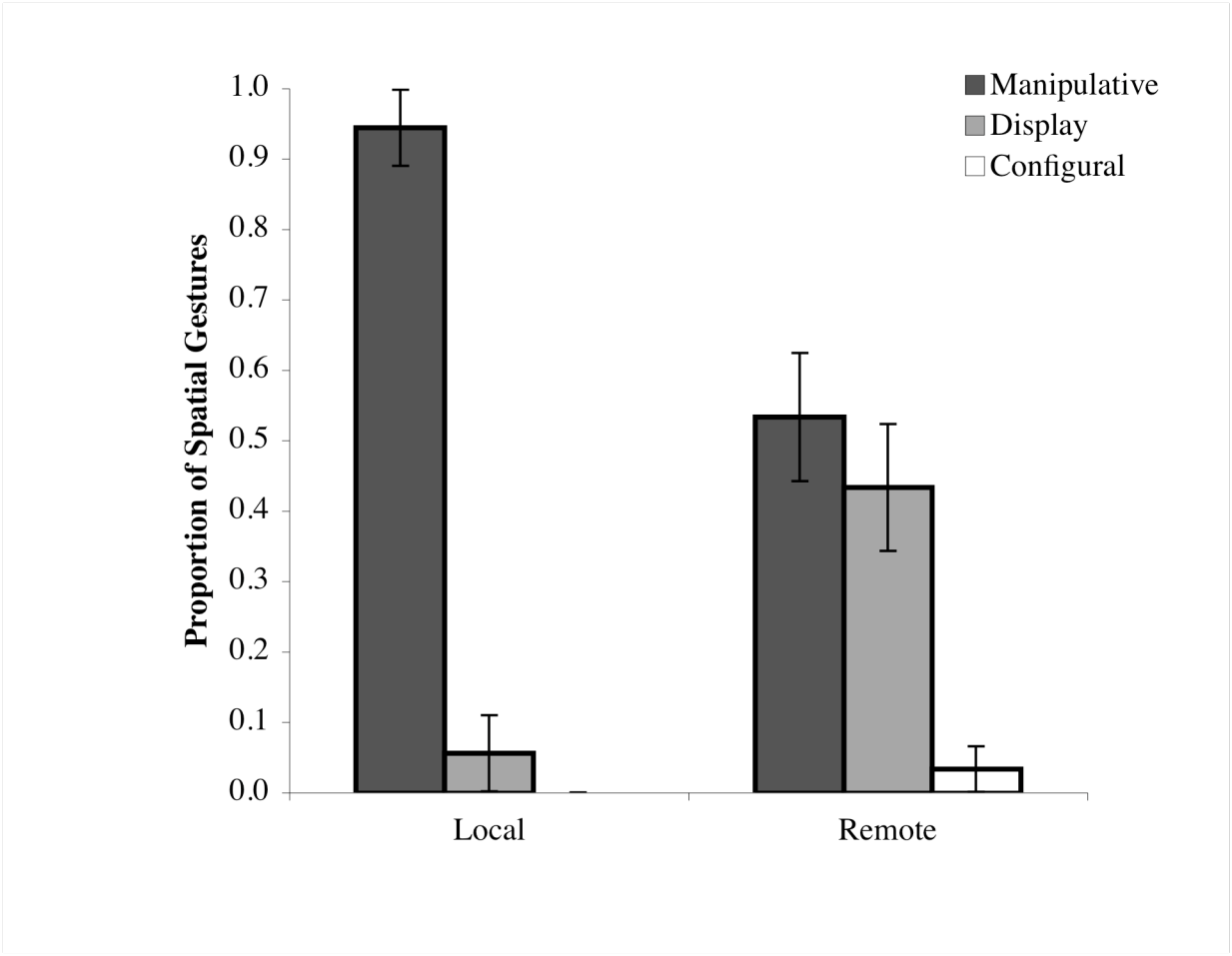


Figure 10. The impact of forecast location (local vs. remote) on spatial gesture types, for the Novice METOC participants (with SE bars).


Manipulative
Display
Configural
Proportion of Spatial Gestures
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Local
Remote