SPATIAL COGNITION AND COMPUTATION, *6*(1), 1–29

The Relationship Between Spatial Transformations and Iconic Gestures

J. Gregory Trafton and Susan Bell Trickett
Naval Research Laboratory

Cara A. Stitzlein
George Mason University

Lelyn Saner and Christian D. Schunn
University of Pittsburgh

Susan S. Kirschenbaum
Naval Undersea Warfare Center

Current theories of gesture production all suggest that spatial working memory is a critical component of iconic gesture production. However, none of the models has a selection mechanism for what aspect of spatial working memory is gestured. We explored how expert and journeyman scientists gestured while discussing their work. Participants were most likely to make iconic gestures about change over time (spatial transformations), less likely to gesture about spatial relations and locations (geometric relations), and far less likely to gesture about the magnitude of spatial entities. We also found that experts were especially likely to have a high degree of association between iconic gestures and spatial transformations. These results show that different features of spatial language are gestured about at different rates. We suggest that current gesture production models need to be expanded to include selection mechanisms to account for these differences.

Keywords: Gestures, expression, inference, spatial reasoning.

Correspondence concerning this article should be addressed to Greg Trafton, NRL, Code 5515, Washington, DC 20375-5337; email: trafton@itd.nrl.navy.mil.

INTRODUCTION

One of the fundamental findings within the gesture research community is that people gesture when they are thinking about something spatial. For example, it is common to see people thinking, talking, and gesturing about spatial items or actions at the same time: a meteorologist thinking about a particular weather pattern may say "Cyclogenesis[1] is forming" while making a spinning motion with her hand. These types of gestures that "act out" what has been said have been labeled *iconic* or *representational* gestures (McNeill, 1992), and are of particular interest to many researchers. In fact, researchers have found that gestures occur more frequently during spatial tasks (Lavergne & Kimura, 1987; Morsella & Krauss, 2004; Wesp, Hesse, Keutmann, & Wheaton, 2001) and while talking about something spatial (Alibali, Heath, & Myers, 2001; Kita & Özyürek, 2003; Rauscher, Krauss, & Chen, 1996) than during non-spatial tasks.

Where and how do spatial cognition and iconic gesture production overlap? Current theories and models suggest that gestures and spatial cognition intersect at spatial working memory or imagery. For example, Krauss and his colleagues suggest that spatial working memory plays a key component in gesture production (Krauss, Chen, & Gottesman, 2000; Morsella & Krauss, 2004; Rauscher et al., 1996); Kita and Özyürek (2003) suggest that spatio-motoric processes are key, and other researchers suggest that there is a strong imagistic component to gesture production (De Ruiter, 2000; McNeill, 1997; Wesp et al., 2001). Researchers use different terms, but most theorists believe that gestures are tied to some form of spatial imagery or working memory, typically non-propositional (though c.f. Wagner, Nusbaum, & Goldin-Meadow, 2004).

This overlap is probably most evident in the three existing process models of gesture production (De Ruiter, 2000; Kita & Özyürek, 2003; Krauss et al., 2000). All three gesture production models are based on Levelt's (1989) speech processor and have several similarities, including having a three part working memory (propositional, spatial/dynamic, and other). Levelt's (1989) speech production model has three stages: conceptualizing (constructing a communicative intention); formulating (the intention is converted into an abstract symbol with syntactic structure); and articulating (the physical action of speaking). Working memory is critical for all three models: propositional working memory is important to standard utterance formation, and spatial working memory is critical for both gesture production and speech production when the topic has a spatial component.

Spatial working memory plays a major role in each of the theories. Krauss et al.'s (2000) model, for example, has a spatial / dynamic feature selector that

[1]Cyclogenesis is the development or intensification of a low-pressure center (a cyclone) showing up as circulation or rotation in the same direction as the Earth's rotation (i.e., counterclockwise in the northern hemisphere). Almost all storms contain cyclogenesis.

chooses which features within spatial working memory to gesture about. De Ruiter's (2000) model uses an imagistic representation that feeds into the conceptualizer that makes a determination about what and when to gesture. Kita and Özyürek's (2003) model has an action generator for gestures that uses spatial imagery and spatial working memory as the basis for what to gesture. For all three models, spatial working memory is one of the least specified components. Our view of spatial working memory is that there are several different visual/spatial representations which 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 (Harrison & Schunn, 2002, 2003). For the purposes of this discussion, spatial working memory can be considered a multiple-component memory system with a domain-general executive (Anderson & Lebiere, 1998; Baddeley & Logie, 1999).

It should be noted that all three models are quite complex and have multiple constraints on when and what to gesture about; Kita and Özyürek's (2003) model decides to gesture about objects in spatial working memory, but those gestures are constrained by communicative intention, action schemata (a mix of person and environmental possibilities), and linguistic encoding of the referent. The important part of this brief review of the different models is that whereas all of the models specify that spatial working memory is the source of gestures, none of the models have any selection mechanism for what *aspect* of spatial working memory or imagistic memory is gestured. In other words, there is no current theory or data about what type of spatial information within working memory co-occurs with gesture.

This paper explores how different aspects of spatial working memory influence iconic gesture production as well as how expertise affects iconic gesture production.

What spatial features could people gesture about? That is, what features could be extracted from a spatial representation held in spatial working memory? One possibility is that people could extract spatial primitives from the spatial representation they hold, and then perform some iconic gesture about those extracted features. There are several spatial ontologies (e.g., DAML-Space) and spatial primitives (Golledge, 1995; Nystuen, 1963) that could be used for this purpose; we have chosen to focus on Golledge's spatial primitives, though most other accounts would yield a similar analysis.

Golledge (1995) suggests that there are four types of spatial primitives: identity, location/relation, magnitude (size or amount of the object), and time (transformation of an object). Identity determines what an occurrence is, or equates an occurrence with a name. Location/Relation provides information about where an occurrence exists in either absolute terms (world-based) or relative terms (with respect to another object). Magnitude is the amount of an occurrence's feature; for example, the size or weight of an occurrence are magnitude issues. Finally, time concerns not only when an occurrence takes place, but also how it changes, when it is created or moved or removed, and so forth. From these primitives, Golledge proposes a number of derived concepts for identity (class and category), location/relation (distance, angle and direction, etc.), magnitude (frequency and hierarchy), and time (growth, change, periodicity, etc.).

Golledge (1995) suggests that these spatial primitives tie directly to human cognition and spatial language. If we assume that these primitives and associated derived concepts are basic building blocks of spatial cognition, then it follows that these primitives will be represented in spatial working memory; iconic gestures will then be a reflection of at least one type of spatial primitive. Our goal in this paper is to explore what types of spatial primitives are accompanied with iconic gestures.

In order to identify the relationship between gesture and spatial language, it will be important to identify spatial primitives from the linguistic stream, independently of gesture. Fortunately, researchers have proposed different measures that map very well onto each of the spatial primitives. We assume that the spatial primitive of identity will occur primarily at the visual/perceptual level and will have a minimal effect on spatial working memory, so we will restrict the rest of our discussion to the remaining three spatial primitives: location/relation, magnitude, and time. We will identify the occurrence of spatial primitives by coding three types of linguistic behavior: descriptions of geometric relations for location/relation, magnitude utterances for magnitude, and spatial transformations for time.

Descriptions of geometric relations have been hypothesized to account for a great deal of spatial cognition, especially as shown through spatial language (Coventry & Garrod, 2004; Herskovits, 1986; Landau & Jackendoff, 1993; Talmy, 1983b). Geometric relations deal with the spatial relationships between different objects (and to a lesser extent the characteristics of the objects themselves). The geometric relation framework is thus an excellent candidate for identifying the spatial primitive of location/relation.

When the size or amount of a spatial entity is used, the spatial primitive of magnitude is assumed to be part of spatial working memory. Magnitude can be described in many different ways, but it is frequently used as a descriptor of a spatial object.

Identifying the spatial primitive of time linguistically is more difficult than identifying either geometric relations or magnitude because of the variety of utterances that can describe change over time. Mental spatial transformations correspond to exactly the same concepts as the spatial primitive of time, including change, transformation, movement, creation, removal, and so on, so descriptions of spatial transformations will be used to identify the spatial primitive of time. A mental spatial transformation occurs when a spatial object is transformed from one mental state or location into another mental state or location. Mental spatial transformations occur in a mental representation that is an analog of physical space and are frequently part of a problem solving process. Further, they can be performed purely mentally (e.g., purely within spatial working memory or a mental image) or “on top of” an existing visualization (e.g., a computer-generated image). Mental spatial transformations may be used in all types of visual-spatial tasks, and thus represent a general problem-solving strategy in this area. Note that for the spatial primitive of time, we are focusing on the object-change aspect of time rather than just the passage of time itself.

People can and do verbally describe the mental spatial transformations they engage in; these utterances are frequently seen as a window upon the mental operations they are performing (Ericsson & Simon, 1993). These verbal mental spatial transformations are frequently seen in direction giving (Franklin, Tversky, & Coon, 1992; Taylor & Tversky, 1992; Tversky, Lee, & Mainwaring, 1999) as well as in spontaneous descriptions of technical and scientific material (Trafton, Marshall, Mintz, & Trickett, 2002; Trafton & Trickett, 2001; Trafton, Trickett & Mintz, 2005; Trickett & Trafton, under review; Trickett, Trafton, Saner, & Schunn, under review). For the remainder of the article, we will call mental spatial transformations simply "spatial transformations."

Not all primitives are captured by the methods we have selected. There could be, for example, situations where a spatial primitive occurs, but it would not be identified by our methods: the utterance "It is raining right now" is clearly about time, but it is not a spatial transformation. However, most of these situations deal primarily with perception, and have a very small (if any) spatial component, and presumably would not enter into spatial working memory. Note also that each of these spatial primitives could occur singly or in combination. It is common for people to think, gesture, or talk about all three spatial primitives, or any one by itself. For example, if a weather forecaster said, "This low's coming over here" it would refer to both the geometric representation of location and relation as well as being a spatial transformation. A later analysis will take this possibility into account.

In summary, we will examine the types of spatial features associated with gestures by focusing on the spatial primitives of location/relation, magnitude, and time.

Gesture in Context

We are interested in understanding how people generate and use gestures in complex, real-world situations. Thus, unlike previous gesture studies, which have for the most part been performed in a laboratory setting, with carefully crafted, artificial materials, we decided to explore how people gesture in their own work setting. We focus on the gesturing behavior of participants who not only were able to perform challenging tasks, but who understood how to use the complex tools, theories, and issues within different, complex domains. We selected two different domains, neuroscience and meteorological forecasting, both of which are highly spatial and involve the interpretation of spatio-visual representations of complex data. We do not expect any large differences between domains with respect to gesture production, but the different domains should provide generalizability.

The focus on real-world situations raises the issue of the effect of expertise on behavior. In the laboratory studies discussed above, participants neither possessed nor needed to possess any specific domain knowledge in order to perform the task. In our real-world domains, however, it is impossible to perform the task without some domain knowledge. Given that experts and novices have different knowledge structures (Chase & Simon, 1974; Chi, Feltovich, & Glaser, 1981; Larkin, 1983; Schunn & Anderson, 1999), we expect that they will display some differences in their use of spatial reasoning.

In order to investigate which type of spatial primitive is associated with greater expertise, we studied participants who had differing levels of expertise (Experts and Journeymen). Journeymen are still students of the domain, but have some understanding of the theory, data, and tools needed to succeed in the domain. We collected data on journeymen instead of true novices because the tasks that they were to perform were, in general, too complex for a true novice to complete with any level of accuracy. By examining differing levels of expertise, we were also able to investigate gesture rate and type differences across levels of expertise, an issue that has received little empirical investigation (c.f. Schunn, Saner, Kirschenbaum, Trafton, & Littleton, under review).

The purpose of this study, then, is two-fold. First, we examine the different influences of spatial primitives (location/relation, magnitude, and time) on iconic gesture production. Second, we explore how differing levels of expertise influence gesture production. We explore these issues in the context of the real-world task of data analysis in two complex, visuo-spatial domains, neuroscience and meteorological forecasting.

Predictions

Our predictions focus on a common assumption made in all current gesture production models. All three models (De Ruiter, 2000; Kita & Özyürek, 2003; Krauss et al., 2000) predict that when a spatial entity enters spatial working memory, it is likely to be associated with iconic gesture. Because spatial language is expected to use spatial working memory (Hayward & Tarr, 1995; Shah & Miyake, 1996), iconic gestures should occur more frequently when talking about spatial occurrences. However, the models are underspecified about the types of spatial cognition that are likely to be gestured about—according to the models, all types of spatial cognition are equally likely to co-occur with an iconic gesture.

This paper thus attempts to refine the gesture production models: if people do gesture equally about different types of spatial cognition or spatial primitives, then the current models can be assumed to be correct in their selection process. However, if people gesture more about one spatial primitive (time, for example) than the others, the implication would be that the type of spatial working memory selected for gesture is sampled differentially. Such differences should be made explicit in the models.

Finally, one of the features of expertise is increased spatial skill within the area of expertise, though experts in a highly spatial domain do not necessarily have high general spatial ability (Doll & Mayr, 1987; Sims & Mayer, 2002). This expertise certainly extends through different spatial primitives, including identifying spatial objects (Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988), determining the location/relation of spatial objects (Chase & Simon, 1973; Chase & Simon, 1974), and transforming spatial objects (Sims & Mayer, 2002). It follows, then, that experts will be more adept than journeymen at performing spatial actions and more fluent in describing those actions within their area of expertise. Since spatial working memory is implicated in all these spatial activities, we expect experts to perform more spatial activities, and perhaps more iconic gestures associated with those spatial actions.

METHOD

Participants

Participants were eight neuroscience fMRI researchers (four experts and four journeymen) and fourteen meteorologists (four experts and ten journeymen). The more expert neuroscience researchers had conducted an average of 6.25 studies and had an average of 4.6 years experience in fMRI research.[2] The neuroscience journeymen had completed 0 studies and had 2.5 years experience (see Schunn et al., under review for additional descriptions of these researchers). The expert meteorologists each had over 10 years experience working as Navy forecasters, and were thus experts in this domain (Ericsson, Krampe, & Tesch-Roemer, 1993; Hayes, 1985). The journeyman forecasters were junior and senior undergraduate meteorology majors with an average of 2.75 years experience.

Procedure

The experiment took place at the participant's regular work location, and all participants had access to all the tools, visualizations and computer equipment that they usually employed. All participants agreed to be videotaped during the session. Participants first performed their "normal" working activity—data analysis for the neuroscientists, and creating a weather forecast for the meteorologists. During this activity, the experimenter made note of "interesting events," such as a major change in the computer display or something that spurred a burst of participant activity. Then participants were asked two questions about the activity they had just performed. We used their answers to these questions as the basis for our analyses of spatial cognition and gesture.

The questions were designed to allow our participants to talk freely about something they had done that had a large spatial component and something they had done that had a smaller spatial component. Our spatial question asked about

[2]Expert neuroscience researchers had less experience than the traditional definition of expertise primarily because fMRI analysis is a relatively young field. The more experienced neuroscientists were definitely more expert than their less experienced counterparts were. Their behavior in all analyses discussed below is comparable with the meteorologist experts.

the 3-dimensionality of the data, and our less-spatial task asked for knowledge recall. Both questions are described below. Note that because of the highly spatial nature of the domains, even the recall question may have had *some* spatial component; however, we expect substantial differences in the level of spatial reasoning involved in answering the two questions.

After participants had completed their task (analyzing their data or making a forecast), the experimenter showed the participant a one-minute segment of the video surrounding the "interesting event." For each of these minutes, after reviewing the videotape, the experimenter asked the participant "What did you know and what did you not know at this point?" For the purposes of this paper, only the first and last interesting events (the "Recall" Question) will be analyzed and described.[3] The first Recall Question always occurred at the very beginning of the participant's task, and the last Recall Question occurred during the last minute the participant worked.

Next, the participant was asked a "3D" Question that was designed to elicit explicit spatial reasoning: "How do you deal with the three dimensionality of your data?" Participants' responses (including gestures) to both Recall and 3D questions were recorded on videotape.

Coding

There were very few gestures during the *in vivo* problem solving session, so gesture rates, utterances, and other problem solving measures from the online session will not be discussed further. All utterances from both Recall and 3D minutes were transcribed and segmented such that each segment contained a single conceptual idea (typically a subject and verb); we called this coding by "complete thought" (Trafton, Kirschenbaum, Tsui, Miyamoto, Ballas & Raymond, 2000; Trafton et al., 2005; Trickett, Schunn, & Trafton, 2004; Trickett & Trafton, under review; Trickett et al., under review). For example, if a forecaster said, "It's going to slow down, start to develop," there are two complete thoughts, the first about slowing down and the second about the development. Each utterance was then coded explicitly (yes/no) as having each type of spatial primitive and each gesture type.

Spatial primitive coding: Location/relation. The spatial primitive of location/relation was coded using the geometric framework. The geometric framework assumes that the geometric relations (and to a lesser extent the characteristics about the objects themselves) can be captured very well through the use of spatial prepositions. Thus, when people use spatial prepositions they are representing geometric components (especially spatial relations) to a large extent. Multiple researchers have suggested that when people use utterances with spatial prepositions (e.g., "The weather system is above California"), they

[3]We chose the first interesting minute as a control condition because participants had different numbers of interesting events and the first one was the most comparable across domains. We chose the last interesting event for a subset of analyses to show that our effects were not due to time on task or learning.

are focusing on geometric locations and relations, which is the same as the spatial primitive of location/relation. We will use this geometric framework as a proxy for the spatial primitive of location/relation.

A pure spatial preposition coding scheme is not appropriate since some spatial prepositions are used in metaphorical manners. Thus, all spatial prepositions (under, across, into, over, etc.) were coded as either spatial or metaphorical (Coventry & Garrod, 2004; O'Keefe, 1996). For example, utterances like “if that low was really going to move out” would have been coded as being a “spatial” spatial preposition (“out”), but the utterance “…and figure out what effect that’s going to have” would have been coded as having a metaphorical spatial preposition (“out”). Only “spatial” spatial primitives will be used in the following analyses.

Spatial primitive coding: Magnitude. There are many ways of describing spatial magnitude, and there are no a priori linguistic or other types of coding schemes for spatial magnitude. Therefore, all utterances that modified a spatial object’s magnitude (size, amount, etc.) were coded as containing a magnitude spatial primitive. For example, an utterance like “if you’re just going to get spotty showers…” would be coded as a magnitude primitive because “spotty” modifies the amount of “showers.”

Spatial primitive coding: Time. In order to identify the spatial primitive of time, spatial transformation coding was used. There are many types of spatial transformations: creating a mental image, modifying that mental image by adding or deleting features, mental rotation (Shepard & Metzler, 1971), mentally moving an object, animating a static image (Bogacz & Trafton, 2005; Hegarty, 1992), making comparisons between different views (Kosslyn, Sukel & Bly, 1999; Trafton et al., 2005), and any other mental operation which transforms a spatial object from one state or location into another.

When spatial transformations are used to solve a problem, they can occur quite quickly (e.g., a simple mental rotation can take less than four seconds; Shepard & Metzler, 1971), but spatial transformations in natural and scientific contexts frequently occur in a series. For example, a scientist attempting to match empirical data to a computational model may perform a series of spatial transformations (extension of a line, movement of that line, mental comparison of the two images) in order to come to a tentative conclusion about the fit of the model to the data. These different spatial transformations occur in identifiable steps (Trafton et al., 2005).

We operationally defined a spatial transformation as any time a coder judged whether a complete thought expressed a mental transformation from one spatial object or location into another spatial object or location. Additionally, two types of context were used to identify spatial transformations: current and previous goals, and what participants could see. Thus, if the person could see something on a computer screen and directly extracted that information, or simply remembered something from a previous activity, it would not be coded as a spatial transformation. Note that because spatial transformations are quite diverse (creation of an imaginary object, moving an existing object, etc.), and that context is absolutely critical, a simple linguistic analysis is not appropriate for identifying when a spatial transformation occurs: linguistic analyses (Levin, 1993; Talmy, 1975, 1983a, 1988) over-or under- classify true mental spatial transformations

Table 1 shows examples of each type of spatial primitive along with domain, expertise, and condition. We used the rate of occurrence per minute for each spatial primitive for all analyses except the logistic regression.

Gesture Coding

All gestures from both Recall and 3D minutes were coded. Gesture rates per minute were calculated throughout for all gesture types. A two-step gesture-coding scheme was used. All gestures were initially coded, with all sound turned off, as either simple (beats or deictic gestures) or complex gestures (all

Table 1
Examples of Utterances That Contain Spatial Primitives of Location/Relation, Magnitude, and Time

Utterance	Location / Relations	Magnitude	Time	Domain	Expertise	Condition
When you're in a location, // the vertical dimension's probably the most important	Yes	No	No	Meteorology	Journeyman	3D
If the longwaves were deepening or the ridge was building	No	No	Yes	Meteorology	Expert	Recall
You can see a legitimate time-course of what you reasonably would think the blood flow would be	No	No	Yes	Neuroscience	Expert	3D
If that low was really going to move out	Yes	No	Yes	Meteorology	Journeyman	Recall
The low is coming down through the surface	Yes	No	Yes	Meteorology	Expert	3D
Because often you can get **some** activity in a region	Yes	Yes	No	Neuroscience	Novice	3D

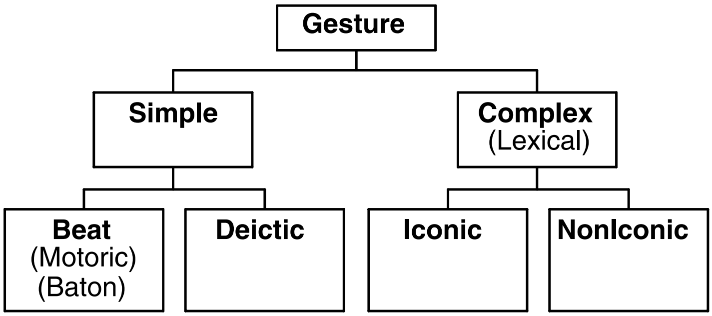

Figure 1. The gesture coding scheme we used, with common names that other researchers have used for similar gesture types.

other gestures). Complex gestures were then further coded with the sound turned on. We did not code personal adjustments (e.g., scratches, clothing adjustments, etc.). For the most part, McNeill's (1992) coding scheme was followed.

Simple gestures. Simple gestures were typically brief, motorically simple gestures (e.g., simple gestures had only two component movements like up/down). Gestures were coded as beats if they were rhythmic. For example, hand flicks and waves were coded as beats. Gestures were coded as deictic if there was a directed, explicit pointing action, usually involving a finger or forearm, and a purposeful direction toward a display or item in the environment. For example, a neuroscientist may point directly to the video screen where she just saw a visualization of the brain.

Complex gestures. Complex gestures were further coded, with access to what the participant said, as either iconic or non-iconic gestures.[4] Gestures that had a strong relationship to the semantics of the utterance, or "acted out" what was said, were labeled iconic gestures. For example, a neuroscientist may gesture to an imaginary brain in her hands and refer to its different regions in space with her hands. Complex, non-iconic gestures were a mix of metaphoric gestures (McNeill, 1992) and non-codable gestures (they were not iconics, beats or deictic gestures). A gesture was non-iconic if it did not have a clear connection to what was being said. Figure 1 depicts our coding scheme, and Table 2 shows example utterances and a brief description of different gesture types.

[4] Other researchers have described complex gestures as lexical gestures (Krauss, 1998; Krauss et al., 2000; Morsella & Krauss, 2004; Rauscher et al., 1996). We prefer the term complex because the term "lexical" presupposes a theory which is still under investigation—that people gesture to facilitate lexical access (Krauss, 1998; Krauss et al., 2000; Rauscher et al., 1996). "Complex" is a much more agnostic term, which requires fewer assumptions.

Table 2
Examples of Utterances That Contained Different Gesture Types With Our Coding of the Gesture and a Brief Description of the Physical Movements; Square Brackets [ ] *Show Gesture Occurrence*

Utterance	Gesture		Domain	Expertise	Condition
	Type	Description			
[this was the first thing I looked at]	Simple: Deictic	Points at monitor	Meteorology	Journeyman	Recall
Basically, I knew there [were frontal] regions that were more significant [for our] semantic condition than [for a] detection task	Simple: Beat	Rhythmic gesture at each gesture point	Neuroscience	Journeyman	Recall
OK, [you kind of think of it, OK, this low's coming over here,]	Complex: iconic	Raises left forearm, left hand, fingers open and slightly curled in, up and sweeps around to left, down and up again in circular motion	Meteorology	Expert	3D
You can go from, like, [left to right],	Complex: iconic	Raises right arm, elbow at side and bent, forearm raised, hand extended and flat, facing right ear; symmetrical gesture with left hand	Neuroscience	Expert	3D
So for instance, the wind's blowing from the [west], I'm picturing whatever's out [west] like [west of where the wind was]	Complex: iconic	Forearms on desk, hands are with palms together; both hands point to 11 o'clock; right hand crosses over left, pointing to left, this gesture is then repeated	Meteorology	Journeyman	3D
Then [you look at], whatever it is that's going on…	Complex: noniconic	Right arm raised, hand holding pen between 2nd and 3rd fingers, sweeps hand back and then forward in single circular motion	Meteorology	Expert	3D
It's a [very time intensive thing]	Complex: noniconic	Elbows resting on knees, left and right hands raised and circle around each other several times	Neuroscience	Journeyman	Recall

Results and Discussion

First, we deal with inter-rater-reliability, time on task, etc. Then we address baseline performance for both spatial language and gestures. Finally, we address the primary analysis of interest: the relationship between spatial language and iconic gesture formation.

Inter-Rater-Reliability

One coder coded all verbal utterances; a second independent coder coded 18% of the verbal utterances. Initial agreement for the spatial transformation coding was 88%, $\kappa = .73$, $p < .0001$. Initial agreement for the "spatial" spatial preposition versus metaphorical spatial preposition coding was 91%, $\kappa = .81$, $p < .0001$. One coder coded all gestures; a second coder coded 11% of the gestures. Initial agreement for the simple/complex coding was 93%, $\kappa = .85$, $p < .0001$. Initial agreement for the type of gesture (beat, deictic, iconic, non-iconic) was 94%, $\kappa = .88$, $p < .0001$. One of the coders performing spatial language analysis was unaware of the hypotheses under investigation and did not have access to the gestures. Similarly, one of the coders examining gestures was unaware of the hypotheses. Keeping this information compartmentalized prevented the hypotheses from influencing the coding. Thus, agreement between all coders was quite strong. Disagreements were resolved by discussion.

There were very few statistical differences between domains (meteorology vs. fMRI) and domain did not interact significantly with any other variable (see Table 3), so all analyses will be collapsed across domain. Similarly, participants spent the same amount of time on each question type (an average of 1 minute) regardless of condition or expertise; additionally there was no interaction, all *p's* > .10.

Spatial Language

Two assumptions drove the analyses, first, that our two questions would elicit different amounts of spatial reasoning, measured by the three spatial primitives location/relations (by geometric relations), magnitude (by magnitude language), and time (by spatial transformations), and second that these measures of spatial reasoning indicate different (though partially overlapping) aspects of spatial cognition. In order to verify these assumptions, we first analyzed the use of spatial primitives for the different questions by computing an ANOVA with condition as a within-subjects factor and expertise as a between-subjects factor.

Participants talked more about geometric relations when answering the 3D question than the Recall question, $F(1, 20) = 21.0$, MSE = 3.1, $p < .001$ (see Figure 2a). They also spoke more about time (via spatial transformations) in answering the 3D question than the recall question (see Figure 2b), $F(1, 20) = 17.3$, MSE = 11.5, $p < .001$. Interestingly, however, participants did not use more magnitude language when answering the 3D question (M = 0.33) than

Table 3
Means and Standard Deviations (in Parentheses) for Domain and Expertise

	Neuroscience		Meteorology	
	Journeyman	Expert	Journeyman	Expert
Duration (mins)	1.2 (0.5)	1.3 (0.7)	0.8 (0.3)	0.8 (0.8)
Rate of Geometric Relations (per minute)	1.7 (2)	1.7 (1.7)	4.5 (2.9)	1.3 (1.6)
Rate of spatial transformations (per minute)	1.4 (2.6)	3.1 (3.8)	3.3 (4.4)	3.0 (3.6)
Rate of magnitude (per minute)	0.5 (0.7)	0.2 (0.5)	0.8 (1.2)	2 (4.4)
Rate of beat gestures (per minute)	2.9 (2.9)	6.2 (3.6)	3.8 (3.9)	9.7 (6.2)
Rate of deictic gestures (per minute)	0.2 (.3)	0.5 (.7)	1 (2.4)	2.3 (2.9)
Rate of iconic gestures (per minute)	1.2 (2.7)	4.2 (3.9)	2.5 (3.5)	3.4 (3.7)

when recalling what they had done in the first minute (M = 1.3), $F(1, 20) = 2.9$, MSE = 4.6, $p > .10$. The overall number of magnitude utterances was quite low overall, which perhaps contributed to this null effect.

This result suggests that our two question types did indeed elicit different amounts of spatial reasoning, at least for locations/relation and time. Figures 2a and b also show that the primitives of location/relation (geometric relations) and

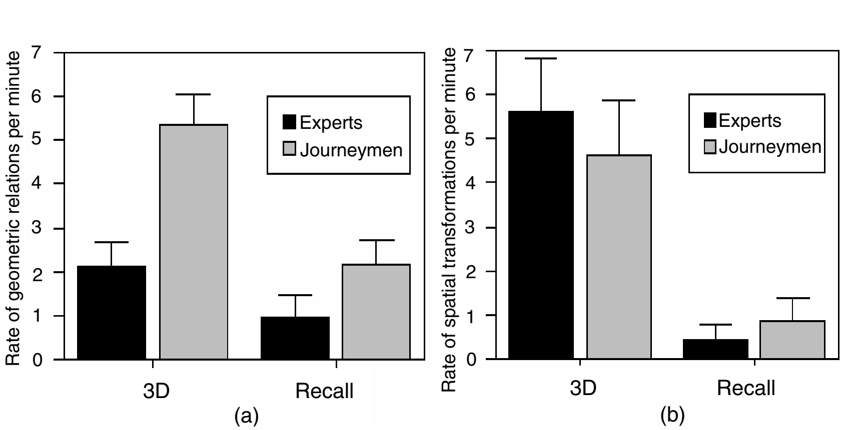

Figure 2. Rate per minute of geometric relations (a) and spatial transformations (b). Error bars are standard errors of the mean.

Table 4
The Rate of Geometric Relations, Spatial Transformations, and Iconic Gestures; Standard Deviations are in Parentheses

	First Recall	Last Recall	3D
Geometric Relations	1.7 (2.0)	1.4 (1.3)	4.2 (2.8)
Spatial Transformations	0.7 (1.6)	0.7 (0.7)	5.0 (4.3)
Iconic Gestures	0.9 (2.0)	1.0 (1.2)	4.5 (3.8)

time (spatial transformation) were used at different rates in the more spatially-oriented 3D question, thus confirming our assumption that these represent different measures of spatial reasoning. We assume that if the questions asked had been more about the magnitude of particular spatial entities, we would have found a comparable result with magnitude.

Because of experts' greater spatial skill within their own area of expertise, we expected experts to produce more spatial utterances than journeymen. However, a consistent effect of expertise across spatial primitives did not emerge from the data. There was no difference between experts' (M = 1.1) and journeymen's (M = 0.7) rate of magnitude utterances, $F(1, 20) < 1$, MSE = 3.9, nor any interaction, $F(1, 20) < 1$, MSE = 4.6. Overall, journeymen had a higher rate of geometric relations than experts, $F(1, 20) = 7.9$, MSE = 6.3, $p < .05$. This higher rate of geometric relations was especially pronounced during the 3D question as shown by a marginal interaction, $F(1, 20) = 3.4$, MSE = 3.1, $p = .08$. However, there was no effect of expertise on the rate at which the spatial primitive of time (spatial transformations) was used, $F(1, 20) < 1$, MSE = 10.1 nor an interaction between condition and expertise, $F(1, 20) < 1$, MSE = 11.5.

Note that one possible confound in this analysis is order: because of the manner in which the data was collected (the Recall question was always before the 3D question), participants could have changed their spatial language over time becoming more spatial over time. To investigate this possibility, we coded the rate of geometric relations and spatial transformations in the Recall question about the last interesting minute (typically the very last minute of a participant's session). If learning were driving the overall effect, we would expect a general increase in geometric relations and spatial transformations over time. However, this was not the case, as Table 4 suggests. For both variables, we collapsed across expertise and found an overall difference between conditions for both geometric relations, $F(2, 42) = 13.0$, MSE = 3.9, $p < .001$, and spatial transformations, $F(2, 42) = 18.2$, MSE = 7.4, $p < .001$. Post-hoc Tukey tests suggested that the 3D condition differed from each of the two Recall conditions, $p < .05$, but the two Recall conditions did not differ from each other, $p > .4$. Thus, people do not seem to be learning to perform more spatial transformations

Table 5
The Rate per Minute of Different Gestures Types; Standard Deviations are in Parentheses

		Journeymen	Experts
Beat Gestures	Recall	3.6 (3.8)	7.8 (5.4)
	3D	3.5 (3.6)	8.1 (5.4)
Deictic Gestures	Recall	1.2 (2.6)	2.0 (2.9)
	3D	0.4 (1.4)	0.7 (0.9)
Complex Gestures	Recall	1.8 (2.8)	3.1 (2.9)
	3D	6.8 (4.9)	7.5 (4.3)
Noniconic Gestures	Recall	1.3 (2.3)	1.4 (1.6)
	3D	1.8 (2.0)	1.6 (1.2)

or talking about more geometric relations. The differences between conditions seem to be because of the spatiality of question type.

Gesture Analysis (Simple Gestures)

Next, we investigated the overall use of gestures by all participants. There were 518 gestures in the entire corpus, 193 (37%) complex, and 325 (63%) simple. We first examined the simple gestures, which included beats and deictics (pointing gestures). Since beat gestures are associated with narrative content and discourse (Alibali et al., 2001; McNeill, 1992) and experts are typically more fluent than journeymen, and gesture rate has been associated with fluency (Rauscher et al., 1996), we expected experts to have a higher rate of beat gestures. As Table 5 suggests, experts did make more beat gestures than journeymen, $F(1, 20) = 5.9$, MSE = 32.6, $p < .05$, though there was no effect of condition nor any interaction, all Fs < 1. It appears that experts' greater knowledge and recollection of what they were doing (Chase & Simon, 1973; Chase & Simon, 1974; Ericsson et al., 1993) allowed them to be more fluent and make more beat gestures.

We next examined deictic or pointing gestures. As Table 5 suggests, there were more deictic gestures during the Recall minute than during the 3D minute, $F(1, 20) = 5.6$, MSE = 1.8, $p < .05$. Neither expertise nor the interaction between expertise and condition approached significance, all Fs < 1. Participants probably made more deictic gestures during the Recall minute because they could point to relevant items in the environment. For example, a meteorologist may have pointed to a specific printout of the weather and said, "This was the first thing I looked at." Specific items in the environment were less relevant to answering the 3D question.

Gesture Analysis (Complex Gestures)

Our main interest in this study is the complex gestures, specifically the iconic gestures. As Table 5 suggests, participants made many more complex gestures while answering the 3D question than the Recall question, $F(1, 20) = 21.4$, MSE = 11.8, $p < .0005$. However, there was no difference between Experts or Journeymen, and no interaction, all Fs < 1. This result replicates Krauss' finding that people make more lexical gestures while speaking about spatial material than when speaking about non-spatial material (Krauss, 1998; Morsella & Krauss, 2004), since our complex gestures correspond directly to his lexical gestures.

Of the 193 complex gestures, 127 (66%) were iconic and 66 (33%) were non-iconic. As Table 5 suggests, the rate of non-iconic gestures did not differ by condition or expertise, nor was there a significant interaction, all Fs < 1. The fact that there was no difference between experts' and journeymen's use of non-iconic gestures suggests that non-iconic gestures are linked to a more generic representation, available to journeyman and expert alike.

In contrast, as Figure 3 suggests, participants made more iconic gestures when answering the 3D question than when answering the Recall question, $F(1, 20) = 5.1$, MSE = 9.2, $p < .05$. Interestingly, there is a trend that experts used more iconic gestures than journeymen, $F(1, 20) = 3.5$, MSE = 9.2, $p = .08$, and no interaction, $F < 1$. Note that, as with geometric relations and spatial transformations, it is possible that people learned to perform more iconic gestures as time went on. However, as Table 4 suggests, the difference is not practice or learning-based. We again collapsed across expertise and found an overall difference between conditions for iconic gesture rate, $F(2, 42) = 14.1$, MSE = 6.6, $p < .001$. Post-hoc Tukey tests suggested that the 3D condition differed from each of the two Recall conditions, $p < .05$) but the two Recall

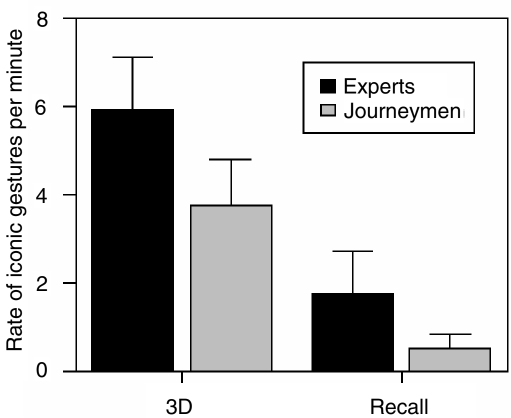

Figure 3. Rate per minute of iconic gestures. Error bars are standard errors of the mean.

conditions did not differ from each other, $p > .5$. Thus, people were not learning to perform more iconic gestures over the course of the session.

In summary, participants performed more iconic gestures while answering the more spatial 3D question than they did while answering the less spatial Recall question. However, there was no such difference in the rate of non-iconic gestures. This difference suggests that focusing on the relationship between iconic gestures and spatial cognition is appropriate.

Spatial Cognition and Iconic Gestures

In order to investigate the relative strengths of the relationship between iconic gesture and the different spatial primitives, we examined the frequency of co-occurrence of the spatial primitives geometric relation, time, and magnitude with iconic gesture within each utterance. Table 6 shows the numbers of utterances that contain location/relations, magnitude, and time used in this analysis.

Out of 614 utterances, 114 contained geometric relations (about 13%), 39 contained magnitude (about 6%), and 127 (about 19%) contained spatial transformations. As noted above, the majority of spatial language occurred during the 3D minute—there were 307 total utterances in the 3D minute, of which 79 (about 26%) were geometric relations, 12 were magnitude (about 4%), and 112 (about 36%) were spatial transformations.

Magnitude was not correlated across utterances with either geometric relations, $r_\phi = 0.047$ or spatial transformations, $r_\phi = 0.001$. Spatial transformations and geometric relations were only moderately correlated across utterances, $r_\phi = 0.25$, $p < .05$, providing further support that they represent distinct types of spatial cognition or spatial language (see Table 6). There was no correlation between iconic gestures and magnitude utterances, $r_\phi = 0.06$, $p > .10$. However, consistent with the results of other researchers (Alibali et al., 2001; Rauscher et al., 1996), there was a positive correlation between geometric relations and iconic gestures, $r_\phi = .20$, $p < .0001$. Finally, we found a strong, positive correlation between iconic gestures and spatial transformations, $r_\phi = .52$, $p < .0001$.

Table 6
Number of Utterances Containing Unique and Overlapping Instances of Each Type of Spatial Primitive

Location/Relation	Magnitude	Time	# Utterances
No	No	No	397
Yes	No	No	59
No	Yes	No	24
No	No	Yes	74
No	Yes	Yes	5
Yes	Yes	No	7
Yes	No	Yes	45
Yes	Yes	Yes	3

These results suggest that the spatial primitive of magnitude is gestured about only infrequently, whereas both location/relations and spatial transformations are frequently gestured about. Note, though, that the relationship between iconic gestures and spatial transformations is more than twice as strong as the correlation between iconic gestures and geometric relations. These results suggest a hierarchy of gesture frequency with spatial primitives: spatial transformations are most likely to be gestured about, then geometric relations, and magnitude least of all.

These analyses, however, only show simple pairwise relationships. As a more direct test of the contribution of the different spatial primitives, we performed a logistic regression analysis to predict when an iconic gesture might occur. A logistic regression analysis was used because the outcome variable (existence of iconic gesture) was a dichotomous variable, which violates many of the assumptions of standard linear regression (Tabachnick & Fidell, 2001). An oversimplified description of logistic regression is that it is a multiple linear regression model with a dichotomous variable as an outcome variable. An excellent description of logistic regression can be found in Peng, Lee, & Ingersoll (2002).

Two logistic regression analyses were performed to investigate the relative predictive contribution of the different spatial primitives. The first model was a complete model, with all relevant variables put into the equation. The second model contained only the significant regressors from the first model for predictive purposes. All logistic regression analyses were carried out by the Logistic Regression Model within the Design library (Harrell, 2004) of R (R-Development-Core-Team, 2004).

The outcome variable of both logistic regression analyses was the existence of an iconic gesture (1 = yes, 0 = no). The full model had six predictors: the expertise of the speaker (1 = Expert, 0 = Journeyman), the domain (1 = meteorology, 0 = neuroscience), the condition (1 = 3D, 0 = Recall), the existence of a geometric relation in that utterance (1 = yes, 0 = no), the existence of magnitude in the utterance (1 = yes, 0 = no), and the existence of a time based relation in that utterance (1 = yes, 0 = no). Equations 1 and 2 show the result of these logistic regression analyses.

Equation 1: Predicted logit of Iconic gesture = -3.5 + (geometric relation × .71) – (magnitude × .98) + (spatial transformations × 2.08) + (Condition × 1.04) + (Expertise × .92) + (Domain × .35)

Equation 2: Predicted logit of Iconic gesture = -3.3 + (geometric relation × .699) + (spatial transformations × 2.06) + (Condition × 1.01) + (Expertise × .81)

The overall logistic regression equation is significant for the six variable model, $\chi^2(6) = 127.9$, $p < .0001$. The log odds of an iconic gesture occurring was positively related to geometric relations ($p < .05$), time ($p < .0001$), condition ($p < .001$), and expertise ($p < .005$); neither magnitude ($p > .10$) nor domain ($p > .10$) were significant predictors. Thus, magnitude is not a strong predictor of iconic gesture production, whereas both geometric relations and spatial transformations are.

The second model used only the significant predictors from the first model. Thus, geometric relations, time, condition, and expertise were the four predictors used. The overall logistic regression equation was significant for the four variable model, $\chi^2(4) = 127.9$, $p < .0001$ model. The log odds of an iconic gesture occurring was positively related to geometric relations ($p < .05$), time ($p < .0001$), condition ($p < .001$), and expertise ($p < .005$). The two models differed in their coefficient weights very little for the significant predictors and, in fact, the model with all variables predicted the log odds of an iconic gesture no better than the reduced model, $p > .05$; a comparison of AIC (Akaike Information Criterion) values also showed no substantial difference between the models. Since the coefficient weights were very similar and neither model was significantly better, all remaining discussion will focus on the second model with only four significant predictors. Details of the model are shown in Table 7.

It is difficult to determine relative contributions of different predictor variables within a logistic regression (squared semi-partial's can not be calculated, for example). However, it is possible to examine the proportion of the total likelihood ratio accounted for by each variable in the corrected model. As Table 7 suggests, the biggest contributor to the model by far is spatial transformations.

Since logistic regression equations predict log odds of a particular event occurring, it is sometimes more comprehensible to convert specific data to probabilities (see Peng et al., 2002 for a discussion on these issues). In order to convert these numbers to probabilities, specific values for each variable can be

Table 7
Results of the Logistic Regression

Predictor	β	SE β	Wald's χ^2	df	*p*	Proportion Total Likelihood	e^{β} (odds ratio)
Constant	-3.3	0.28	-11.64	1	< .0001		0.04
Spatial Transformation	2.1	0.27	7.6	1	< .0001	36%	8.2
Geometric Relations	.7	0.30	2.3	1	.02	3%	2.0
Condition	1.0	0.29	3.5	1	.0005	8%	2.7
Expertise	.8	0.26	3.1	1	.002	6%	2.2
Test			χ^2	df	*p*		
Score Test			127.9	4	< .0001		

substituted into the equation to get a log odds value. This log odds number can then be substituted into the following equation to calculate a probability $\frac{e^{value}}{(1+e^{value})}$.

According to the full model, an iconic gesture is not very likely (about 4%) to occur during an utterance that has no spatial transformation, no geometric relation, occurred in the Recall condition, and was uttered by a Journeyman (e.g., zeroes are substituted into Equation 2 above). Conversely, an utterance that has a spatial transformation, a geometric relation, occurred in the 3D condition, and was uttered by an expert (i.e., all factors are used) has approximately a 78% chance of using an iconic gesture.

How well does the full model fit the current data? The *c* statistic represents the proportion of gesture pairs with different observed outcomes for which the

Table 8

Examples of Utterance with Different Characteristics, the Predicted Probability of an Iconic Gesture, and Whether or Not an Iconic Gesture Actually Occurred; Condition = 3D or Recall; Expertise = Expert or Journeyman; Square Brackets [ ] Show Iconic Gesture Occurrence

Utterance	Spatial Trans.	Geometric Relation	Condition	Expertise	Predicted Probability	Iconic Gesture? 1=yes, 0=No
OK, [you kind of think of it, OK, this low's coming over here,]	Y	Y	3D	Expert	.78	1
So for instance, the wind's blowing from the [west]	Y	Y	3D	Journeyman	.61	1
[Cyclogenesis is forming]	Y	N	3D	Expert	.64	1
You know they were moving [west to east]	Y	Y	Recall	Journeyman	.36	1
So, I mean it gives us an easy way of looking at certain levels in the atmosphere	N	Y	3D	Journeyman	.17	0
Basically I knew there were frontal regions that were significant for our semantic condition	N	N	Recall	Journeyman	.04	0

model correctly predicts a higher probability for observations with the event outcome (e.g., an iconic gesture) than the probability for nonevent observations (e.g., no iconic gesture). The *c* value for this logistic regression is 0.826, which means that for 82.6% of all possible pairs of utterances, the model correctly assigned a higher probability to utterances that contained iconic gestures than to utterances that did not contain an iconic gesture.

It is also possible to examine how the model predicts iconic gesture co-occurring with specific utterances. We examined specific utterances that were used as examples throughout this paper and made model predictions for each utterance. As Table 8 suggests, the model did quite a good job predicting the existence of iconic gestures, though of course it is not perfect (see the low predicted probability of the utterance "You know they were moving [west to east]" which contains both a spatial transformation and a geometric relation). However, in general, the model does do an excellent job of predicting when an iconic gesture will occur. Further support for the model is found when we compare the model to all the actual data and assign any model value greater than .5 as an iconic gesture prediction and any model value less than or equal to .5 as a lack of iconic gesture prediction. According to this analysis, the model fit the data on 86% of the cases.

Another way to look at the individual contribution of each spatial primitive is to perform a logistic regression on each participant to calculate each individual's set of coefficients, and then perform statistical tests on those coefficients. This method does not work in this case because (a) logistic regression requires a great deal of data to return stable coefficient weights, and (b) some participants were completely missing some variables (e.g., some participants did not have any iconic gestures in the Recall condition), making individual logistic regressions non-computable. Additionally, it could be argued that the logistic regression analysis suffers from non-independent data. The logistic regression analysis does, however, allow us to examine individual contributions of predictors at the utterance level; other analyses would not allow such fine-grained analysis. We can bolster the logistic regression analysis by coming up a level of analysis and examining the co-occurrence of iconic gestures with spatial transformations, geometric relations, and magnitude.

Co-occurrence was calculated by counting the number of times each variable co-occurred with an iconic gesture for each participant. As Figure 4 suggests, experts had a higher rate of co-occurrence of spatial activity with iconic gesture than journeymen, $F(1, 20) = 7.3$, MSE = 10.2, $p < .05$. More importantly, there was an overall difference between types of spatial activity, $F(2, 40) = 15.7$, MSE = 4.8, $p < .001$. Finally the differences between spatial activity was much bigger for experts than it was for journeymen, interaction, $F(2, 40) = 5.8$, MSE = 4.8, $p < .01$. A significant linear contrast, $F(1, 20) = 22.1$, MSE = 6.7, $p < .001$, and a Bonferonni adjustment suggests that iconic gestures co-occurred least with magnitude, an intermediate amount with geometric relations, and the most with spatial transformations.

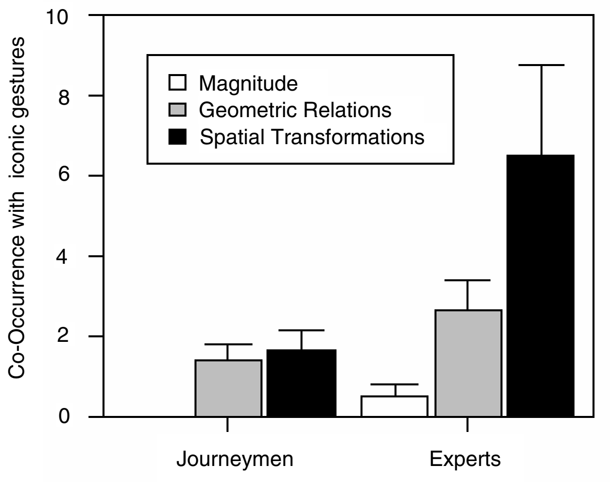

Figure 4. Co-occurrence of magnitude, geometric relations, and spatial transformations with iconic gestures. Error bars are standard errors of the mean.

This analysis shows the same pattern of results as the logistic regression; the number of iconic gestures that a person performs is dependent on the number of spatial transformations and geometric relations. Again, magnitude is the spatial primitive that is least likely to co-occur with an iconic gesture. Geometric relations do co-occur with iconic gestures, but not as much as spatial transformations.

GENERAL DISCUSSION

What do people gesture about? Current theories of gesture production suggest that when something enters into spatial working memory, an iconic gesture is likely to co-occur with it. In fact, all current gesture production models suggest that, all else being equal, anything that enters spatial working memory is equally likely to co-occur with a gesture as anything else within spatial working memory. This study suggests that, as far as iconic gesture is concerned, all elements in spatial working memory are not created equally. People are most likely to gesture about the spatial primitive of time (spatial transformations), less likely to gesture about spatial relations and locations (geometric relations), and far less likely to gesture about magnitude.

These conclusions are supported in several ways. First, iconic gestures co-occur with both spatial transformations and geometric relations, though the relationship between spatial transformations and iconic gestures was much stronger than the relationship between geometric relations and iconic gestures.

Perhaps just as importantly, we did not find a co-occurrence between magnitude and iconic gestures. We also performed a logistic regression analysis, showing that spatial transformations were a better predictor of iconic gestures at the utterance level. In no case did we find support for the spatial primitive of magnitude corresponding with iconic gestures. It is this finer-grained level of analysis that has allowed us to investigate the distinct roles of the spatial primitives of location/relation (represented by geometric relations) and time (represented by spatial transformations) in the production of iconic gestures.

This study has implications for current methodology within the gesture community. Most researchers who deal with gestures and spatial cognition have used one of two methods of identifying spatial language. The first is to provide spatial and non-spatial materials and examine gesture production within those cases (Krauss, 1998; Lavergne & Kimura, 1987); the second method is to identify spatial language within the linguistic stream by marking spatial prepositions (which we used to mark the spatial component of geometric relations; Alibali et al., 2001; Kita & Özyürek, 2003; Rauscher et al., 1996). This study suggests that both within an utterance and within spatial conditions, iconic gestures can differ depending on what is entering spatial working memory. Furthermore, since spatial transformations seem to be a stronger predictor of iconic gesture production than geometric relations, researchers should take into account both spatial prepositions and spatial transformations when examining iconic gesture production. One reason that previous theories have been so successful is, perhaps, that spatial transformations and geometric relations (spatial prepositions) overlap a great deal in the domains that have been studied before. By focusing on how gestures co-occur with different aspects of spatial cognition, stronger, more generalizable theories can be built.

These findings also have implications for gesture production models. As currently formulated, gesture production models do not specify what types of information from spatial working memory is gestured about; once something enters into spatial working memory, it is just as likely to be associated with an iconic gesture as anything else in spatial working memory. This study suggests that this view is overly simplistic: spatial transformations are most likely to be gestured about, followed by geometric relations, with magnitude unlikely to be gestured about at all. This information can be incorporated into the different models in model-specific manners. Kita and Özyürek's (2003) model, for example, could simply add the type of spatial cognition as an additional constraint, along with communicative intention, action schemata, and linguistic encoding of the referent. Krauss et al.,'s (2000) model has an explicit "Feature selector" that could prefer spatial transformations that enter into spatial working memory. Similarly, the conceptualizer's output to the gesture planner in De Ruiter's (2000) model could be adapted to have a predisposition for spatial transformations. Another way to deal with this finding within the context of gesture production models would be to add a spatial working memory "Sampler" component that sampled spatial entities within spatial working memory. Any spatial transformation would have a large (approximately 70%)

chance of co-occurring with a gesture, and any geometric relation would have a small to medium (approximately a 25% chance) of co-occurring with an iconic gesture. Additional constraints (e.g., temporal proximity, co-occurrence of different types of spatial cognition, etc.) would need to be worked out, but this study provides a starting point for deciding what type of information within spatial working memory will be gestured about.

One important point that should be noted is that in many gesture studies, researchers attempt to understand and explain iconic gestures. In fact, our emphasis was no different. However, in our study, there were far more simple gestures than complex gestures. It is clear that, in many ways, the iconic gestures seem to be a window into what people are thinking (Goldin-Meadow, Alibali & Church, 1993; Goldin-Meadow, Wein, & Chang, 1992; McNeill, 1992). However, the fact that in our study and in Wagner et al., (2004) there were far more simple than complex gestures suggests that simple gestures may represent a largely untapped source of information.

We believe that focusing on gestures in more naturalistic and complex environments will allow researchers to understand how gestures are related to other complex behavior, such as problem-solving. Finally, we believe that focusing on knowledge-rich domains will allow a richer exploration of the relationship between gestures and underlying knowledge representations, and that this will lead researchers to develop a better understanding of how, when, and why people use gestures of every type.

ACKNOWLEDGMENTS

This article is based on a presentation in a symposium on gesture and spatial cognition at the Annual Meeting of the Cognitive Science Society in August 2004. Mary Hegarty acted as action editor for the article. This work was partially supported by the Office of Naval Research to Greg Trafton under work order numbers N0001403WX30001 and N0001402WX20374. The views and conclusions contained in this document should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U. S. Navy. Thanks to Debbie Boehm-Davis, Raj Ratwani, Peter Squire, Melanie Diez, and Michelle Harper for comments on an earlier version.

REFERENCES

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169–188.

Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In P. Shah & A. Miyake (Eds.), *Models of working* *memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge: Cambridge University Press.

Bogacz, S., & Trafton, J. G. (2005). Using images to reason dynamically. *Cognitive Systems Research*, *6,* 312–319.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing.* (pp. 215–281). New York: Academic Press.

Chase, W. G., & Simon, H. A. (1974). Perception in chess. *Cognitive Psychology, 4*, 55–81.

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.

Coventry, K. R., & Garrod, S. C. (2004). *Saying, seeing, and acting: The psychological semantics of spatial prepositions*. New York: Psychology Press.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge, UK: Cambridge University Press.

Doll, J., & Mayr, U. (1987). Intelligenz und schachleistung—eine untersuchung an schachexperten. [Intelligence and achievement in chess—a study of chess masters.] *Psychologische Beitrage, 29*, 270–289.

Ericsson, K. A., Krampe, R., & Tesch-Roemer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* (2nd ed.). Cambridge, MA: MIT Press.

Franklin, N., Tversky, B., & Coon, V. (1992). Switching points of view in spatial mental models. *Memory and Cognition, 20*, 507–518.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review, 100*, 279–297.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition & Instruction, 9*, 201–219.

Golledge, R. G. (1995). Primitives of spatial knowledge. In T. L. Nyerges, D. M. Mark, R. Laurini & M. J. Egenhofer (Eds.), *Cognitive aspects of human-computer interaction for geographic information systems* (pp. 29–44). Dordrecht: Kluwer.

Harrell, F. E. (2004). *Design: Design package. R package version 2.0-9.*

Harrison, A. M., & Schunn, C. D. (2002). ACT-R/S: A computational and neurologically inspired model of spatial reasoning. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the twenty fourth annual meeting of the cognitive science society* (pp. 1008). Fairfax, VA: Lawrence Erlbaum Associates, Inc.

Harrison, A. M., & Schunn, C. D. (2003). ACT-R/S: Look Ma, no "cognitive-map"! In F. Detje, D. Doerner, & H. Schaub (Eds.), In *Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 129–134). Bamberg, Germany: Universitats-Verlag Bamberg.

Hayes, J. R. (1985). Three problems in teaching general skills. In S. Chipman, J. W. Segal & R. Glaser (Eds.), *Thinking and learning skills, vol. 2* (pp. 391–406). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hayward, W. G., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition, 55*, 39–84.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1084–1102.

Herskovits, A. (1986). *Language and spatial cognition*. Cambridge, UK: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gestures reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16–32.

Kosslyn, S. M., Sukel, K. E., & Bly, B. M. (1999). Squinting with the mind's eye: Effects of stimulus resolution on imaginal and perceptual comparisons. *Memory and Cognition, 27*, 276–287.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54–59.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge: Cambridge University Press.

Landau, B., & Jackendoff, R. (1993). What and where in spatial language and spatial cognition. *Behavioral and Brain Sciences, 16*, 217–265.

Larkin, J. H. (1983). The role of problem representation in physics. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 75–98). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Lavergne, J., & Kimura, D. (1987). Hand movement asymmetry during speech: No effect of speaking topic. *Neuropsychologia, 25*, 689–693.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. Chi, R. Glaser & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

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

McNeill, D. (1997). Growth points cross-linguistically. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization. Language, culture and cognition* (Vol. 1, pp. 190–212). New York: Cambridge University Press.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology, 117*, 411–424.

Nystuen, J. D. (1963). Identification of some fundamental spatialconcepts. *Papers of the Michigan Academy of Science, Arts, Letters, XLVIII*, 373–384.

O'Keefe, J. (1996). The spatial prepositions in english, vector grammar and the cognitive map theory. In P. Bloom, M. A. Peterson, L. Nadel & M. F. Garrett (Eds.), *Language and space* (pp. 277–316). Cambridge, MA: MIT Press.

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The journal of Educational Research, 96*, 3–14.

R-Development-Core-Team. (2004). *R: A langugage and environment for statistical computing*: R Foundation for Staistical Computing.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226–231.

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science, 23*, 337–370.

Schunn, C. D., Saner, L. D., Kirschenbaum, S. S., Trafton, J. G., & Littleton, E. B. (under review). Complex visual data analysis, uncertainty, and representation. In M. Lovett & P. Shah (Eds.), *Thinking with data.* Mahway, NJ: Lawrence Erlbaum Associates, Inc.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*, 4–27.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology, 16*, 97–115.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics (4th ed.).* Needham Heights, MA: Allyn & Bacon.

Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), *Syntax and semantics* (Vol. 4, pp. 181–238). New York: Academic Press.

Talmy, L. (1983a). How language structures space. In H. L. Pick & L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225–282). New York: Plenum.

Talmy, L. (1983b). How language structures space. In H. Pick & L. Acredolo (Eds.), *Spatial orientation: Theory, research and application* (pp. 225–282). New York: Plenum Press.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49–100.

Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language, 31*, 261–292.

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*, 827–850.

Trafton, J. G., Marshall, S., Mintz, F. E., & Trickett, S. B. (2002). Extracting explicit and implicit information from complex visualizations. In M. Hegarty, B. Meyer & H. Narayanan (Eds.), *Diagramatic representation and inference* (pp. 206–220). Heidelberg: Springer-Verlag.

Trafton, J. G., & Trickett, S. B. (2001). A new model of graph and visualization usage. In J. D. Moore & K. Stenning (Eds.), *The proceedings of the twenty third annual conference of the cognitive science society* (pp. 1048–1053). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Trafton, J. G., Trickett, S. B., & Mintz, F. E. (2005). Connecting internal and external representations: Spatial transformations of scientific visualizations. *Foundations of Science, 10,* 89–106.

Trickett, S. B., Schunn, C. D., & Trafton, J. G. (2004). Puzzles and peculiarities: How scientists attend to and process anomalies during data analysis. In M. Gorman, A. Kincannon, D. Gooding & R. D. Tweney (Eds.), *Spherical horses and shared toothbrushes: Recent developments in scientific and technological thinking*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Trickett, S. B., & Trafton, J. G. (under review). Movies-in-the-mind: The instantiation and use of conceptual simulations in scientific reasoning.

Trickett, S. B., Trafton, J. G., Saner, L. D., & Schunn, C. D. (under review). I don't know what's going on there: The use of spatial transformations to deal with and resolve uncertainty in complex visualizations. In M. Lovett & P. Shah (Eds.), *Thinking with data*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Tversky, B., Lee, P., & Mainwaring, S. (1999). Why do speakers mix perspectives? *Spatial Cognition and Computation, 1*, 399–312.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language, 50*, 395–407.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology, 114*, 591–600.