Coordinating Gesture, Word, and Diagram: Explanations for Experts and Novices

Abstract

Successful explanations are a symphony of gesture, language, and props. Here, we show how they are orchestrated in an experiment in which students explained complex systems to imagined novices and experts. Visual-spatial communication--diagram and gesture--was key; it represents thought more directly than language. The props, real or virtual diagrams created from gestures, served as the stage for explanations, enriched by language and enlivened by deictic gestures to convey structure and iconic gestures to enact the behavior and functionality of the systems. Explanations to novices packed in more information than explanations to experts, emphasizing the information about action that is difficult for novices, and expressing information in multiple ways, using both virtual models created by gestures and visible ones.

Keywords: Gesture; Explanation; Diagram; Adult Expert; Young Novice; Complex systems

Coordinating Gesture, Word, and Diagram: Explanations for Experts and Novices

Watch people explaining how to get somewhere or how something works to someone else. Words are not enough. Speakers integrate their words with gestures, sketches, and props (e. g., Clark, 1996; Engle, 1998; Goldin-Meadow, 2003; McNeill, 1992; Tversky, Heiser, Lee, & Daniel, 2009). Both gestures and sketches are spatial forms of communication; they use elements or actions in space and the spatial relations among them to represent meanings that can be visuo-spatial, as in maps or architectural plans, or abstract, as in schedules, corporate charts, decision trees, and scientific diagrams. One advantage of visuo-spatial forms of communication is that they apply people's highly-practiced spatial skills to making abstract inferences. Another advantage of forms of visuo-spatial communication, gestures, sketches, and props, is that they can represent content more directly than words, using elements, relations, and actions in space. They can show structure, that is, relative locations of people or parts, and they can demonstrate actions (e. g., Tversky, 2011; Tversky and Kessell, in press). Together, word, gesture, sketch, and prop can create and annotate models that represent the situation speakers are trying to convey (e. g., Enfield, 2003; Engle, 1998; Emmorey, Tversky, & Taylor, 2000). These multi-modal models draw from speakers' mental models and can serve to create mental models in the minds of their interlocutors. These layers of communication, word, gesture, props (and more) complement and supplement one another, and are coordinated to create meaning (e. g., Clark, 1996; Goldin-Meadow, 2003; McNeill, 1992). Here, we explore how gesture, word, and diagram are used and coordinated in explanations of complex scientific processes to novices and to experts.

Countless human interactions include explanations to novices or to experts. Everyday conversation typically entails both: when I relate something to you that I know but you do not, and when you check your understanding by explaining it back to me. We are all experts in our own experiences, and our audiences are necessarily novices. Collaborations provide many more examples. They usually consist of individuals with different expertise, so that they often entail many explanations to experts and to novices. Educational settings also provide such situations. Explaining to a student is explaining to a novice; explaining to a teacher is explaining to an expert. Under scrutiny, it seems that the norm in communication is between experts and novices.

Clearly, explanations to novices and explanations to experts do and should differ, but how? Although there is considerable research on what experts and novices know, there has been scant analysis of how people explain to experts and novices. One way such explanations differ is in terms of reference. When New Yorkers need to distinguish photos of New York landmarks for other New Yorkers, they share common ground and use the specific names of the buildings, for example, CitiCorp; when they refer to the same landmark to novices, they use discriminative descriptions, such as “the building with the slanted roof” (Isaacs & Clark, 1987). More generally, speakers make attempts to take the perspectives and knowledge of addressees into account, when they are able to (e. g., Bavelas, Chovil, Coates, & Roe, 1995; Fussell & Krauss, 1992; Mainwaring, Tversky, Ohgishi, & Schiano, 2003; Schober, 1993, 1995; Tversky & Hard, 2009). But these findings only scratch the surface, partly because they have been directed at simple references to objects, rather than explanations, and partly because explainers use far more than words in their explanations.

The case of explaining complex scientific systems is richer, going far beyond terms of reference, and more characteristic of explanations to experts and novices in the wild. Like everyday events, scientific systems have *structure*, that is a configuration of parts, they have *relations* among the parts, and they have properties of the parts, notably, *action*, the behavior and interactions of the components, the causes of the behavior and the consequences of the processes (e. g., Heiser & Tversky, 2006; Hmelo-Silver, Marathe, & Liu, 2007; Tversky, Heiser, & Morrison, 2013). Real world explanations have beginnings, middles, and ends (e. g., Tversky, et al., 2009; Tversky & Zacks, 2013). Philosophers and psychologists alike acknowledge that the form and structure of explanations are far from uniform, and vary with domain (e.g., Lombrozo, 2012). Many, such as the scientific systems investigated here, rely on causal mechanism, that is, the organization of parts and processes. In practice, explanations are grounded in pertinent prior beliefs, and appeal to simplicity and the unification of different phenomena in an integrated mechanism that can account for structure, behavior, and outcomes.

Explanations of complex systems and events are additionally of interest because they can use the full panoply of human communicative expression, words, gestures, props. One goal of the present study is to characterize how people orchestrate those communicative channels to create explanations of complex systems, notably of their structure and their function. In order to do so, we compared explanations to imagined experts and novices. To magnify the effects, we chose the extremes on this continuum: our (imagined) experts were knowledgeable adults and our (imagined) novices were children, though we are aware that there are bound to be differences in explanations to adults and children independent of expertise. The analyses reported here focus on the coordination of gesture, speech, and artifacts that are likely to be part of any explanation to experts or novices.

What do experts know that novices need to know? The short answer is: many things (e. g., Chi, 2006; Feltovich, Prietula, & Ericsson, 2006; Hmelo-Silver, et al., 2007; Tversky, et al., 2013; Tversky & Suwa, 2009). Specific to explanations of scientific systems, experts can be expected to have the pertinent prior beliefs as common ground, but these must be made explicit to novices. Another fundamental and critical difference between experts and novices is that novices may know or readily grasp structure, but have difficulties making inferences from structure to action, process, cause, and function (noted by those just cited, among others). If explainers are aware of that crucial difference, implicitly or explicitly, then their explanations to novices should emphasize the actions of complex systems.

Explanations of everyday or scientific events entail creating a model of a structure or a process, a situation or a system. Models can be created quite directly by diagrams (e. g., Glenberg & Langston, 1992; Tversky, 2011) or gestures (e. g., Enfield, 2003; Emmorey et al., 2000; Tversky, Jamalian, Giardino, Kang, & Kessell, 2013), using elements in space and spatial relations to map situations that are spatial or metaphorically spatial. They can be created by words (e. g., Franklin & Tversky, 1990; Glenberg, Meyer, & Lindem, 1987; Kintsch, 1998; Taylor & Tversky, 1992), but words bear only arbitrary relations to what they are representing and creating mental models from them is more effortful. They can also be created by combinations of word, gesture, and diagram (e. g., Engle, 1996). Diagrams are especially effective for conveying mental models because they map the actors, objects, and places to be communicated isomorphically to an external space. Diagrams can also map actions and functional or causal relations using non-depictive elements like lines and arrows (e.g., Tversky, 2001; 2011). Like diagrams, gestures can map elements and relations of a complex system directly, but to virtual spaces created in the air. A coherent connected sequence of representational gestures can be used to create a model of a situation, for example, of an environment (Emmorey et al., 2000) or a family tree (Enfield, 2003). For environments, explainers created a virtual space using as many as 15 integrated gestures in a row, placing landmarks and paths appropriately and to scale within it, as if drawing a sketch map. In addition, diagrams can and do serve as a stage for gestures that animate the diagrams by pointing to the relevant parts and especially by enacting the behaviors of a system. Language can accomplish these things only through an arbitrary symbol system.

Gestures vary widely and convey many different kinds of information, serving both those who make them and those who see them (e. g., Goldin-Meadow, 2003; Kendon, 1994; McNeill, 1985, 1992). Gestures accompanying phrases like “first,” “second,” are called *beats* and serve to mark the structure of the discourse. More central to the present topic, explanations, are gestures that deliver semantic content, notably *deictic, iconic,* and *metaphoric* gestures (e.g., McNeill, 1985, 1992). Together, iconic, deictic, and metaphoric gestures are called *representational* gestures as they serve as external representations or depictions of the semantic content of the information that speakers are trying to convey. Deictic gestures point, typically to something visible in the world, but also to places or things in virtual worlds, worlds that were visible or that are imaginary, constructed worlds. Deictic gestures are for the most part static and derive their meaning from what they point to. Iconic gestures can create meaning by their actions; they can illustrate or depict object qualities, actions, relations among objects, and more. Metaphoric gestures do the same, metaphorically. Iconic and metaphoric gestures are typically more complex than deictic, which typically do little more than point to referent or virtual referent (e.g., Goldin-Meadow, 2003; Kendon, 1994; McNeill, 1992).

Here, we analyze explanations given to young novices or adult experts. Participants learned two scientific systems, the rock cycle and the circulatory system, from diagrams and were allowed to use those diagrams in their explanations. These systems differ in many ways; the rock cycle is a slow physical system loosely related to objects dispersed in the world, and the circulatory system is a rapid biological system tightly related to an interconnected and enclosed set of parts. All explanations were given to a video camera; that is, they were not interactive. Interactions with novices and experts are undoubtedly interesting but our goal here was to characterize the ways explanations integrate multiple modes of expression, not to study interaction. Interactions would have interrupted and changed the explanations. Communicating an entire system is more characteristic of teaching situations and of real-world collaboration where the explainer explains without or with minimal interruption.

Explanations to both young novices and to adult experts should use the given diagrams as foundations because the diagrams show the essential structure of the systems. For both audiences, explainers can point to the parts of the diagrams as they explain the behavior of each of them and of the systems as wholes. When explaining to experts, the participant's job is to convince the expert that they understand the information the experts already have about the system and its workings. Experts and novices share common ground (e.g., Clark, 1996). Thus, the explainer only needs to provide the essential information, information that is minimal, coherent, and complete. When explaining to novices, participants need to teach, to make sure that someone who knows little or nothing will be able to understand the causal mechanisms, what the basic parts are, what they do, how the system as a whole operates. They need to explicitly ground the explanation in prior beliefs. Child novices are unlikely to have knowledge of the specific systems, although they do understand causal systems in general (e.g., Keil, 2006). To insure that novices understand, explanations to them need to establish common ground, to provide redundant information, and to emphasize information about action, as action is less evident in the diagrams and harder to grasp. Thus, explanations to novices should open with an introduction grounding the explanation in prior beliefs, they should rely on multiple models of the system, both the given diagrams and virtual diagrams created by gestures, they should use relatively more iconic gestures that show action than explanations to experts, and they should use larger gestures, for emphasis. Because eye contact is important for learning and teaching, explainers should make more eye contact with novices (e. g., Fry & Smith, 1975; Fullwood & Doherty-Sneddon, 2006; Otteson & Otteson, 1980).

Method

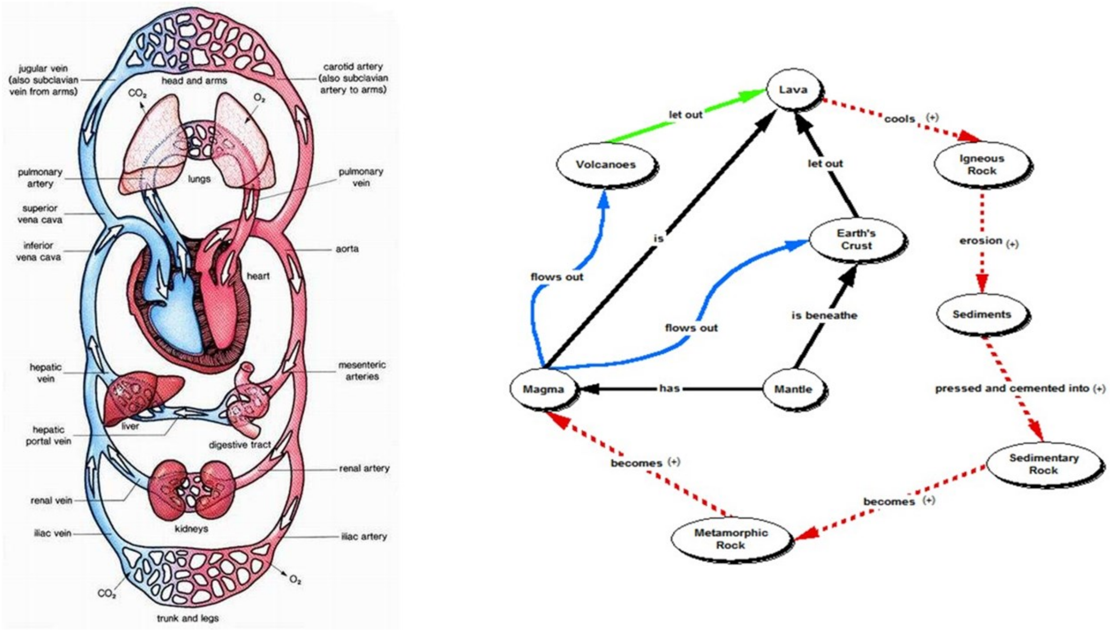
Participants. Thirty (21 women, 9 men) native English speakers (aged 23-33 years old), students at Columbia Teachers College, participated in the experiment in exchange for pay or course credit.

Stimuli. The stimuli were diagrams of the circulatory system (left panel on Fig. 1) and the rock cycle (right panel on Fig. 1).

Design. Half the participants explained both systems to “young novices,” half to “adult experts.” These extreme target audiences were chosen to maximize the differences between the explanations. The order of systems was counter-balanced across participants.

Procedure. Each participant first learned the workings of the circulatory system or the rock formation by studying a diagram displayed on a computer. After learning the system, participants turned to face a video camera to explain the system to either adult experts or young novices, free to the use the diagram in their explanations.

For the circulatory system, participants in the "explanations for adult experts" group were told: “Your job today is making a video explaining how human circulatory system works to a cardiovascular specialist who works in a Columbia University medical school.” The participants were told that the specialist would be watching the video they made. They were then told: “You will have time to prepare for the explanation before explaining the concept and you can use as much time as you want for the preparation. When you are ready to explain, please let me know by knocking or opening the door.” Then the experimenter left the room. When the participant was ready, the experimenter entered in the room and set the video camera. The instructions for the "explanations for young novices" group were similar except that they were asked to explain the human circulatory system to a second grade elementary school student who did not know anything about the concept. They were also told that the student would be watching the video that they made.



Human Circulatory System (Curry, 2002) **Rock Formation System (Tsuei, 2004, p. 28)**

Fig. 1. Diagrams used in the study.

For the rock cycle, participants in the "explanations for adult experts" group were told: “Your job today is making a video explaining how certain types of rock are formed to an earth science specialist who teaches at a university." The participants were then told that the specialist would be watching the video they made. Participants in the "explanations for young novices" group were asked to explain rock formation to a second grade elementary school student who did not know anything about the concept. As before, the two groups were given time until they felt ready to explain the concepts in the diagram.

Participants were reminded of their audiences both before preparing their explanations and before explaining the system.

Results

Gesture Coding. Following McNeill (1992), a gesture unit was defined as "the period of time between successive rests of the limbs (p. 83)." A movement of the hand or hands that started and returned to a resting position was regarded as a single gesture. If the hands did not return to a resting position between two gestures, the boundary was defined by a pause in motion and an obvious change in shape or trajectory. When both hands were used simultaneously to describe one object, concept, or part, it was regarded as a single gesture. If the two hands had different roles, for example, one hand used to describe an object, a concept, or a part and the other hand a different concept, the gestures were coded as two different gestures. Emblematic gestures such as the "OK" sign, counting on the fingers, and thumbs-up were excluded from data analysis.

Based on McNeill (1992), gestures were categorized as: iconic, deictic, or metaphoric. Describing a path or an object by tracing its outline, or using a closed fist to represent a rock are examples of iconic gestures. Few metaphoric gestures occurred, probably because the participants explained concrete systems; these were too few to be reported. Gesture size was coded; because large gestures are more emphatic, participants were expected to use larger gestures when addressing novices. Gestures were also analyzed according to the stage on which they were delivered, on the diagram (left panel) or in the air (right panel) as in Figure 2.

Fig. 2. Gesture delivered on the diagram (left panel) or in the air (right panel)

Inter-rater reliability was assessed on randomly selected subsets (490, about 15%) of the data by a second coder who was trained but blind to the experimental design. Agreement for identifying gestures was 91.8% and for categorizing gestures was 94.7%.

The results will be reported separately for each system, circulatory and rock formation. The same patterns emerged for both systems, so they serve as replications.

Gesture Stage. As apparent from Fig. 3, participants generally opened their explanations to adult experts by directly gesturing on the diagram, whereas they opened their explanations to young novices typically with some general introductory statements before using the diagram (for the circulatory system, $\chi^2(1, N = 30) = 3.97$, $p < .05$, $V = .36$; for the rock cycle, $\chi^2(1, N = 30) = 13.89$, $p < .001$, $V = .68$).

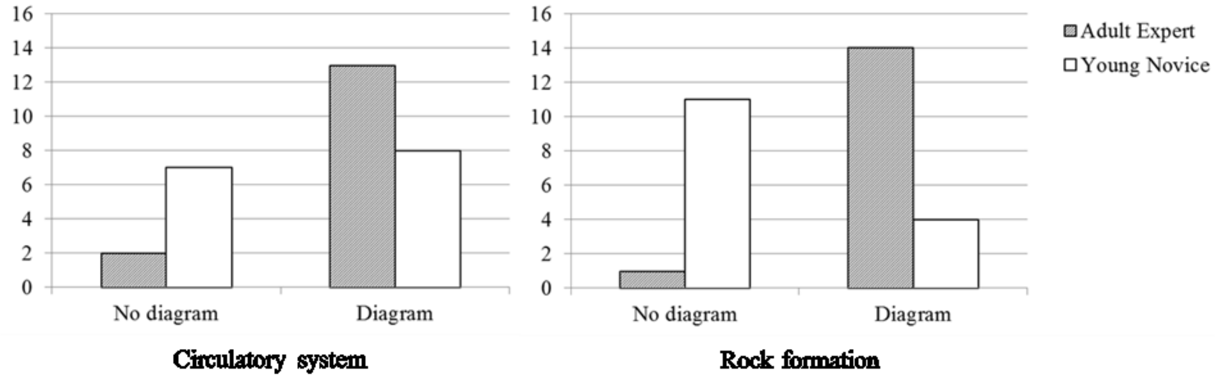


Fig. 3. Number of participants opening explanations with and without diagram to experts and novices

For explanations to novices, all 15 participants created a virtual diagram stage in front of their bodies for both the circulatory system and the rock cycle. By contrast, in explanations to experts, only 8 out of 15 participants gestured at least once in the air for the circulatory system, and 10 out of 15 for the rock cycle.

Speech Coding. As is characteristic of spontaneous speech, participants' explanations were full of incomplete sentences, interjections, self-interruptions, and the like. To simplify the analyses, their explanations were segmented into phrases, small if not always minimal units of information. Units were coded as *structural, functional*, or *other*. Phrases containing 'is-a' or 'has-a' or 'contains' were coded as *structural* information (following Heiser & Tversky, 2006). For the human circulatory system, structural information included locations and properties of the organs. For example, *"left ventricle... this is very important…the left ventricle is key part of your heart…ok?"* and *"…inferior vena cava, which is, I think, the largest vein in the human body…"* were coded as structural. For the rock formation system, structural information included locations and properties of types of rock. For example, *"…when it becomes solid, it becomes* *rock again, it's not liquid anymore."* was regarded as structural because it refers to properties rather than process. Similarly, *"…igneous rock is basically ahh...dark color rock…"* was coded as structural.

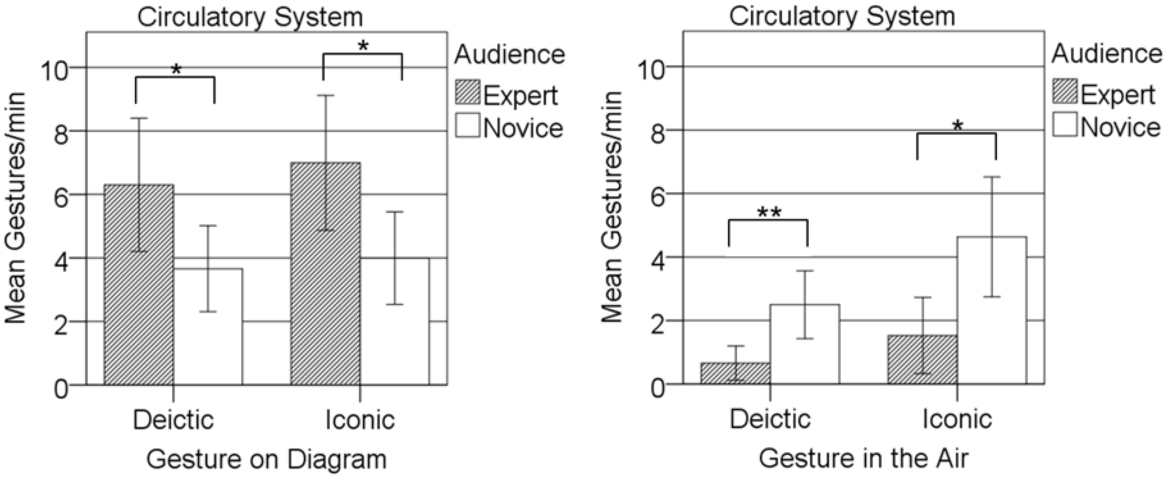
Phrases using action verbs were coded as *functional* information. Previous research had shown that distinguishing causes, processes, and behavior was difficult (Heiser and Tversky, 2006), so all these were grouped as "functional." For the circulatory system, functional information included behaviors of the organs as well as causal relations among them. For example, *"Oxygen is going to your body, going to your lungs... and going to your heart"* was coded as functional. Similarly, *"…it (air) comes in through the lungs, and then right away it makes couple of different paths…to give of course our brain a lot of blood and oxygen…"*, was coded as three functional information units. For the rock formation, functional information included changes of state, movements of entities, and causal relations. For example, *"once the lava gets out of the earth…"* was coded as functional as was *"... when the lava cools off, it turns into this type of rock…"*

The *other* category included introductory information, meta-comments, and analogies. The introductory information and the meta-comments served to provide a narrative structure for the explanations.

Explanations of the circulatory system to adult experts and young novices did not differ in length: when directed to young novices (M = 212 sec, SD = 119); when directed to adult experts (M = 167 sec, SD = 109), $F(1, 28) = 1.13$, $p = 0.30$, ns. However, explanations of rock formation were marginally longer when directed to young novices (M = 213 sec, SD = 122) than to adult experts (M = 141 sec, SD = 66), $F(1, 28) = 4.03$, $p = .054$). In addition, many qualitative differences between explanations to young novices and adult experts emerged.

Gesture Rate. Although there were no significant differences in the time and in overall numbers of gestures in explanations to two types of audience, explanations to young novices took longer on the whole. Consequently, subsequent analyses will report rate of gesturing rather than frequency of gestures, though the same patterns emerged in gesture frequency analyses. Almost all explanations made gestures on both the given diagrams and on virtual diagrams created in the air; however, there were significantly more gestures on the provided diagrams for adult experts than for young novices and significantly more gestures on the virtual diagrams created in the air for young novices than for adult experts for both the circulatory system ($F(1, 236) = 14.40$, $p < .01$) and the rock formation ($F(1, 236) = 15.80$, $p < .01$). These differences also held for each type of gesture.

As evident in Fig. 4, explanations of the circulatory system to young novices used more deictic gestures/min in the air ($M = 2.50$, $SD = 2.07$) than explanations to adult experts ($M = 0.66$, $SD = 1.04$), $F(1, 28) = 9.47$, $p < .01$, $d = 1.12$. Similarly, explanations to young novices used more iconic gestures/min in the air ($M = 4.63$, $SD = 3.65$) than explanations to adult experts ($M = 1.53$, $SD = 2.33$), $F(1, 28) = 7.71$, $p < .05$, $d = 1.01$. In contrast, explanations to adult experts used more deictic gestures/min staged on the diagram ($M = 6.30$, $SD = 4.06$) than explanations to young novices ($M = 3.66$, $SD = 2.62$), $F(1, 28) = 4.47$, $p < .05$, $d = 0.77$, and used more iconic gestures/min staged on the diagram to adult experts ($M = 6.99$, $SD = 4.12$) than to young novices ($M = 3.99$, $SD = 2.83$), $F(1, 28) = 5.42$, $p < .05$, $d = 0.85$.



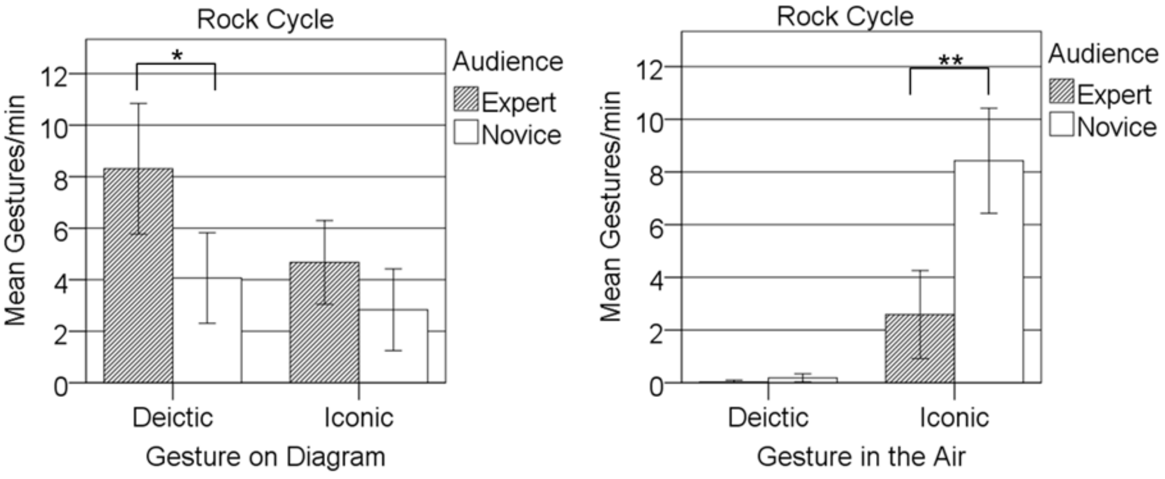
* $p < .05$, ** $p < .01$

Fig. 4. Gesture rate on given diagrams (left panel) and diagrams in the air (right panel) for explanations of the circulatory system by gesture type and audience. Error bars represent standard errors of the means.

A similar pattern emerged for the rock cycle: more gestures of both types on the diagram for experts (left panels of Figures 4 and 5) and more gestures of both types in the air (right panels) for novices. Overall, explanations of the rock formation to adult experts were accompanied by more deictic gestures/min (*M* = 8.34, *SD* = 4.92) than explanations to young novices (*M* = 4.25, *SD* = 3.43), *F*(1, 28) = 6.97, $p < .05$, *d* = 0.96. In particular, explanations to adult experts (*M* = 8.31, *SD* = 4.92) used more diagram-based deictic gestures/min than explanations to young novices (*M* = 4.07, *SD* = 3.40), *F*(1, 28) = 7.54, $p < .05$, *d* = 1.00.

In contrast, explanations to young novices (*M* = 11.26, *SD* = 3.23) used more iconic gestures/min than explanations to adult experts (*M* = 7.27, *SD* = 2.45), *F*(1, 28) = 14.49, $p < .01$, *d* = 1.39. In particular, explanations to young novices (*M* = 8.43, *SD* = 3.86) used more iconic gestures/min in the air than explanations to adult experts (*M* = 2.59, *SD* = 3.24), *F*(1, 28) = 20.15,

$p < .01$, $d = 1.64$. Fig. 5 shows the rates of deictic and iconic gestures to young novices and adult experts in explanations of rock formation.



* $p < .05$, ** $p < .01$

Fig. 5. Gesture rate on the given diagrams (left panel) and in the air (right panel) for explanations of the rock cycle by gesture type and audience. Error bars represent standard errors of the means.

Gesture Stage and Audience. Univariate statistical analyses were used to examine the rate of gestures directed to young novices or adult experts by the gesture stage, diagram or air. Regardless of gesture type, there was an interaction between gesture stage and audience in explanations of both systems (Fig. 6). Those addressing young novices gestured more in the air and those addressing adult experts gestured more on the diagram for both the circulatory system ($F(1, 236) = 14.40$, $p < .01$) and rock formation ($F(1, 236) = 15.80$, $p < .01$).

For the circulatory system, those addressing adult experts used an average of 3.47 ($M = 3.47$, $SD = 4.30$) diagram-based gestures and an average of 0.82 ($M = 0.82$, $SD = 1.46$) gestures in the air. Those explaining to young novices used an average of 1.96 (*M* = 1.96, *SD* = 2.67) diagram-based gestures and an average of 2.22 (*M* = 2.22, *SD* = 2.74) gestures in the air.

For explanations of the rock formation, those addressing adult experts used an average of 3.50 (*M* = 3.50, *SD* = 4.41) diagram-based gestures and an average of 1.03 (*M* = 1.03, *SD* = 2.00) gestures in the air. Those addressing young novices used an average of 1.78 (*M* = 1.78, *SD* = 2.84) diagram-based gestures and an average of 2.82 (*M* = 2.82, *SD* = 3.95) gestures in the air.

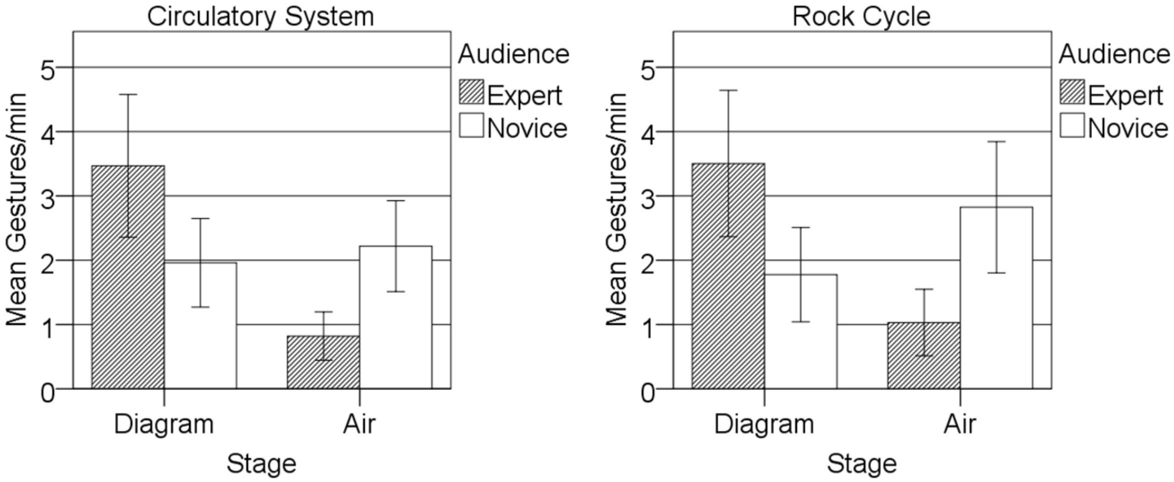


Fig. 6. Gesture rate by stage and audience for explanations of the circulatory system (left panel) and the rock cycle (right panel). Error bars represent standard errors of the means.

For explanations of the circulatory system, gesture rate significantly interacted with gesture stage and audience for both deictic (*F*(1, 56) = 10.48, *MSE* = 7.19, *p* < .01) and iconic gestures use (*F*(1, 56) = 12.80, *MSE* = 10.93, *p* < .01). For rock formation explanations, gesture rate significantly interacted with gesture stage and audience for deictic (*F*(1, 56) = 8.07, *MSE* = 8.97, *p* < .01), iconic (*F*(1, 56) = 19.79, *MSE* = 11.17, *p* < .01), and metaphoric gestures (*F*(1, 56) = 5.50, *MSE* = 0.46, *p* < .05). The patterns of kinds of gestures and gesture stage in explanations to young novices and adult experts were strikingly different, yet consistent across systems. Gestures to adult experts were based primarily on the diagram, and gestures to young novices were more balanced, on both the diagram and in the air, with more in the air. This means that explanations to young novices were more redundant, that is, they explained the systems both using the diagram as a model of the system and also using a virtual model in the air. Later analyses of the language used in explanations will corroborate and extend these findings.

Gesture Size. Gestures in the air were coded as large when they went beyond the speaker's trunk, that is, the area from the neck to the waist, vertically, and from the right side of the body to the left, horizontally. Gestures on the diagram were coded as large when they went beyond the space of the diagram. For the circulatory system, 13 out of 15 participants who explained to young novices used large gestures, but only 5 out of 15 participants who explained to adult experts used large gestures, χ^2 (1, N = 30) = 8.89, $p < .01$. Perhaps because the space of the diagram clearly delineates the boundaries of a stage for gestures, the gestures of only one participant were larger than the diagram, in an explanation to a novice.

For gestures in the air in explanations of the rock formation system, 13 out of 15 participants who explained to young novices used large gestures, and 5 out of 15 participants who explained to adult experts used large gestures, χ^2 (1, N = 30) = 8.89, $p < .01$. Again, because the diagram provides a constrained stage for gestures, only one participant explaining to an adult expert group and one participant explaining to a young novice made a gesture beyond the space of the diagram.

Gesture Size, Stage, and Audience. As noted, large gestures were rarely used on diagrams. For explanations of the circulatory system (left panel in Fig. 7), more large gestures were directed to young novices than to adult experts, $F(1, 56) = 5.90$, $MSE = 18.54$, $p < .05$; but the interaction was not significant, *F*(1, 56) = 3.56, *MSE* = 15.20, *p* = .064. For explanations of rock (right panel in Fig. 7) formation, the interaction between gesture size, gesture stage, and gesture audience was significant, *F*(1, 56) = 4.23, *MSE* = 54.00, *p* < .05). Altogether, the vast majority of large gestures were made in virtual spaces created in the air and directed to young novices.

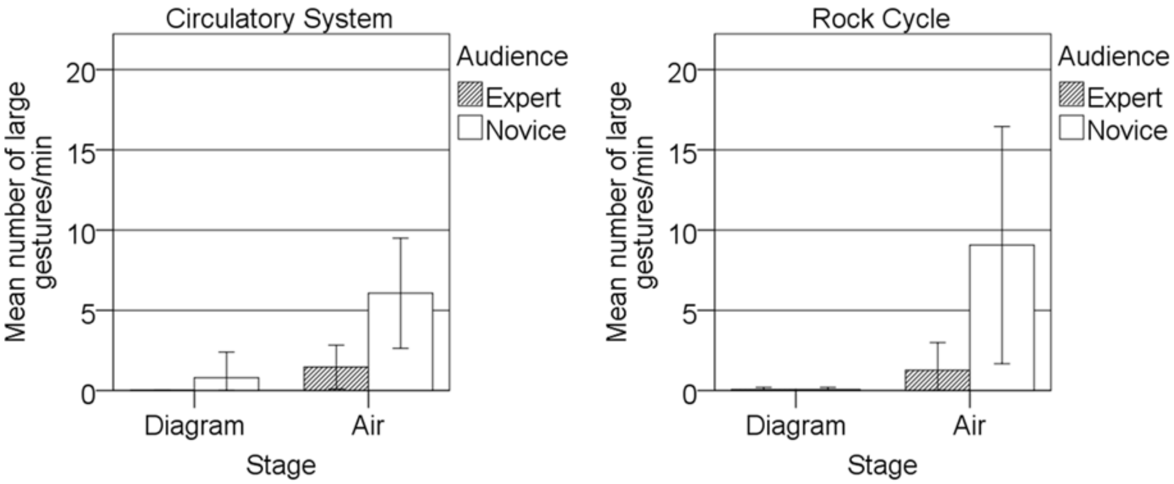


Fig. 7. Mean number of large gestures in explanations of the circulatory system by audience and stage. Error bars represent standard errors of the means.

Direction of Gaze. We computed the proportion of time that participants looked at the imagined listener, that is, the camera, irrespective of speaking and gesturing. For both systems, participants looked at the imagined listener longer when explaining to young novices than when explaining to adult experts (Fig. 8). This is likely to be due in part to the fact that while gesturing in the air in front of the body, it is possible to look at the camera, but when gesturing on the diagram, explainers need to look at the diagram to be sure they are pointing correctly.

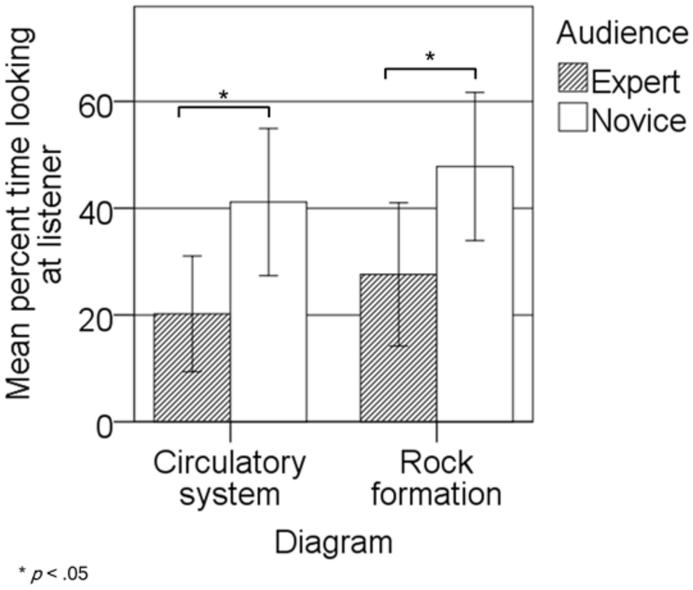


Fig. 8. Mean percentage of time explainers looked at imagined listener by audience and diagram. Error bars represent standard errors of the means.

As evident from Fig. 8, for the circulatory system, explainers to young novices (*M* = 41.13, *SD* = 24.91) gazed at the imagined listener longer than explainers to adult experts (*M* = 20.20, *SD* = 19.55), *F*(1, 28) = 6.56, *p* < .05, *d* = 0.93. Similarly, those explaining the rock formation to young novices (*M* = 47.80, *SD* = 25.08) gazed at the imagined listener longer than those explaining to adult experts (*M* = 27.60, *SD* = 24.25), *F*(1, 28) = 5.03, *p* < .05, *d* = 0.82.

Speech Coding. As is characteristic of spontaneous speech, participants' explanations were full of incomplete sentences, interjections, self-interruptions, and the like. To simplify the analyses, their explanations were segmented into phrases, small if not always minimal units of information. Units were coded as *structural, functional*, or *other*. Phrases containing 'is-a' or 'has-a' or 'contains' were coded as *structural* information (following Heiser & Tversky, 2006). For the human circulatory system, structural information included locations and properties of the organs. For example, *"left ventricle... this is very important…the left ventricle is key part of your* *heart…ok?"* and *"…inferior vena cava, which is, I think, the largest vein in the human body…"* were coded as structural. For the rock formation system, structural information included locations and properties of types of rock. For example, *"…when it becomes solid, it becomes rock again, it's not liquid anymore."* was regarded as structural because it refers to properties rather than process. Similarly, *"…igneous rock is basically ahh...dark color rock…"* was coded as structural.

Phrases using action verbs were coded as *functional* information. Previous research had shown that distinguishing causes, processes, and behavior was difficult (Heiser and Tversky, 2006), so all these were grouped as "functional." For the circulatory system, functional information included behaviors of the organs as well as causal relations among them. For example, *"Oxygen is going to your body, going to your lungs... and going to your heart"* was coded as functional. Similarly, *"…it (air) comes in through the lungs, and then right away it makes couple of different paths…to give of course our brain a lot of blood and oxygen…"*, was coded as three functional information units. For the rock formation, functional information included changes of state, movements of entities, and causal relations. For example, *"once the lava gets out of the earth…"* was coded as functional as was *"... when the lava cools off, it turns into this type of rock…"*

The *other* category included introductory information, meta-comments, and analogies. The introductory information and the meta-comments served to provide a narrative structure for the explanations.

Explanations of the circulatory system to adult experts and young novices did not differ in length: when directed to young novices ($M = 212$ sec, $SD = 119$); when directed to adult experts ($M = 167$ sec, $SD = 109$), $F(1, 28) = 1.13$, $p = 0.30$, ns. However, explanations of rock formation were marginally longer when directed to young novices (*M* = 213 sec, *SD* = 122) than to adult experts (*M* = 141 sec, *SD* = 66), *F*(1, 28) = 4.03, *p* = .054).

Information Transmitted to Experts and Novices. Recall that the language of participants was coded into structural and functional information units. Do explainers alter the amount and kind of information they communicate depending on their audiences, young novice or adult expert?

More information transmitted to novices. Explanations of the circulatory system to adult experts contained 551 information units, 83 structural units, 377 functional units, and 91 *other* units. Explanations of the circulatory system for young novices contained 1067 information units, 181 structural units, 692 functional units, and 194 *other* units. As shown in Fig. 9 (left panel), explanations of the circulatory system directed to young novices (*M* = 12.07, *SD* = 7.86) contained more structural units than explanations to adult experts (*M* = 5.53, *SD* = 5.55), *F*(1, 28) = 6.91, *p* < .05, *d* = 0.96. Similarly, explanations to young novices (*M* = 46.13, *SD* = 28.45) contained more functional units than explanations to adult experts (*M* = 25.13, *SD* = 14.15), *F*(1, 28) = 6.55, *p* < .05, *d* = 0.93.

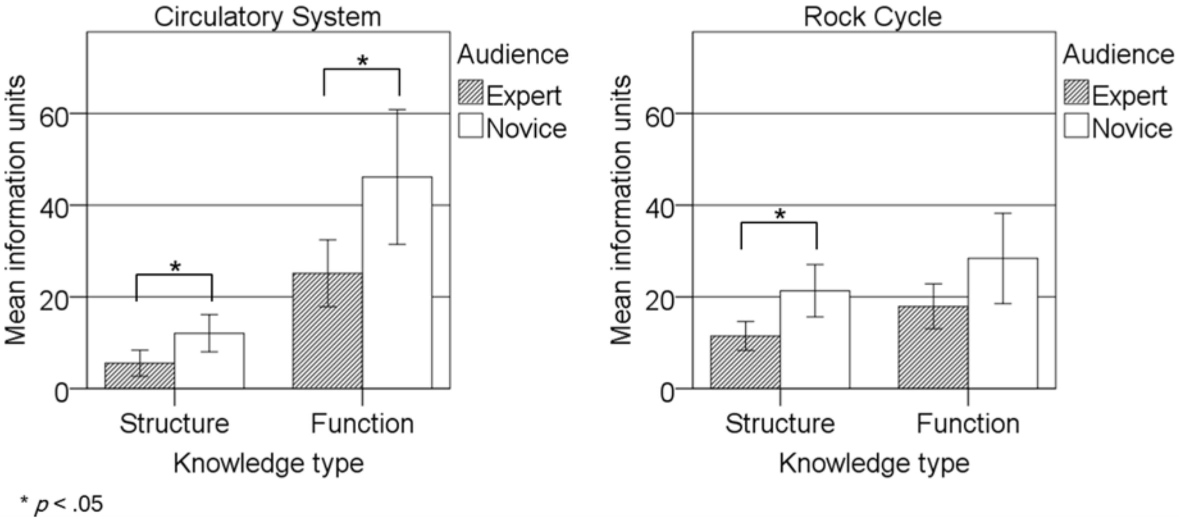
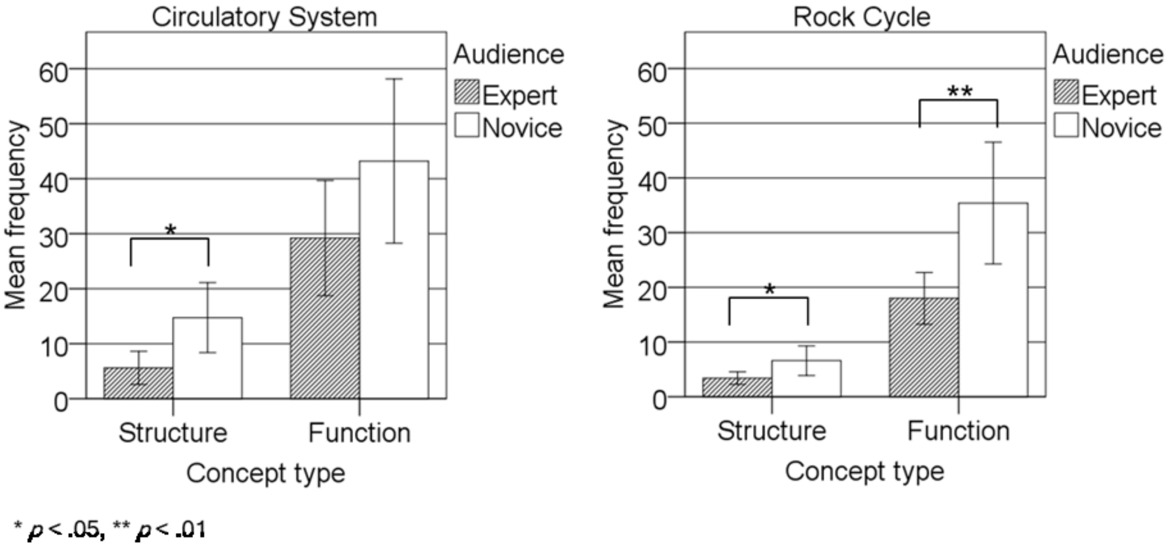


Fig. 9. Mean number of information units directed at adult experts and young novices in explanations of the circulatory system (left panel) and rock cycle (right panel). Error bars represent standard errors of the means.

Explanations of the rock cycle to adult experts contained 577 information units, 172 structural units, 269 functional units, and 136 *other* units. Explanations of the rock cycle to young novices contained 1023 information units, 320 structural units, 426 functional units, and 277 *other* units. As shown in Fig. 9 (right panel), explanations directed to young novices (*M* = 21.33, *SD* = 11.05) contained more structural information than explanations to adult experts (*M* = 11.47, *SD* = 6.13), *F*(1, 28) = 9.15, *p* < .05, *d* = 1.10. A similar pattern emerged for functional information, more to young novices (*M* = 28.40, *SD* = 19.08) than to adult experts (*M* = 17.93, *SD* = 9.47), but it was not statistically significant, *F*(1, 28) = 3.62, *p* = .07.

Rate of information transmission. Recall that there were no statistically significant differences in time to explain to adult experts and young novices. Yet, more information units were directed to young novices than to adult experts. First, we checked the rate of information conveyed to adult experts and young novices. For the circulatory system, explanations to adult experts contained on average 0.23 information units per second (*SD* = 0.07) and explanations to young novices contained on average 0.33 information units per second (*SD* = 0.07), *F*(1, 28) = 14.84, *MSE* = 0.07, $p < .01$, $d = 1.43$. For the rock cycle, explanations to adult experts contained an average of 0.28 information units per second and explanations to young novices contained an average of 0.32 information units per second, *F*(1, 28) = 2.50, $p = .13$. Thus, speakers delivered more information units per second to young novices than to adult experts in explaining the circulatory system, but not in explaining the rock cycle.

Key information. To determine whether participants' explanations were richer for young novices than for adult experts, we made a master list of key concepts in each diagram, where each information unit about each part or process was counted as a key concept. For both systems, participants were expected to explain the core parts and processes. Fig. 10 shows the numbers of concepts explained by two groups for each system.

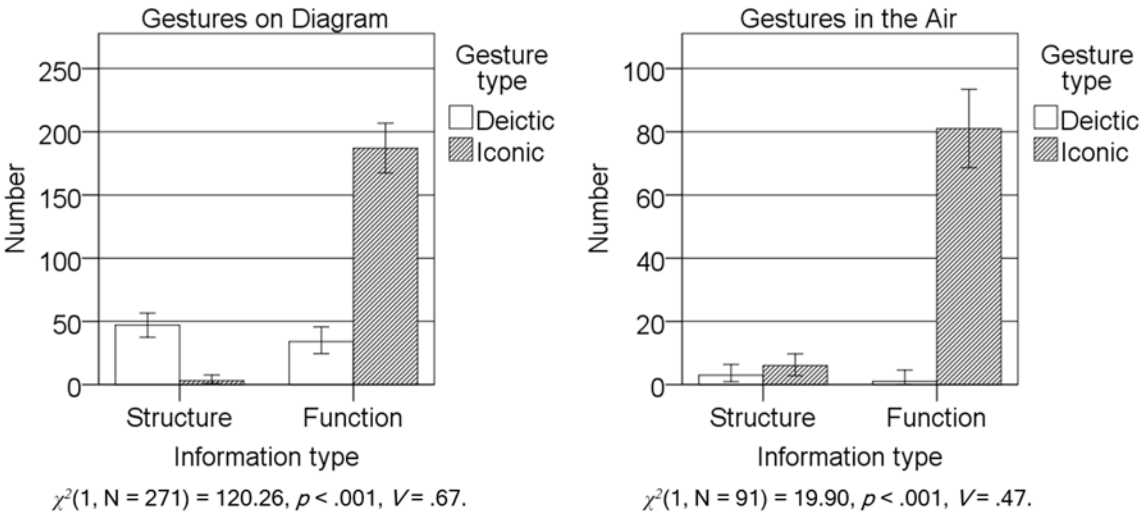


* $p < .05$, ** $p < .01$

Fig. 10. Mean frequency of concepts explained to each audience for the circulatory system (left panel) and the rock cycle (right panel). Error bars represent standard errors of the means.

For the circulatory system, there were marginally more key concepts explained to young novices than to adult experts, $F(1, 28) = 3.80$, $p = .06$. In addition, explanations to young novices included more key structural concepts than explanations to adult experts, $F(1, 28) = 6.75$, $p < .05$, $d = 0.95$. For the rock cycle, explanations to young novices contained more key concepts than explanations to adult experts, $F(1, 28) = 8.36$, $p < .01$, $d = 1.06$. In addition, explanations to young novices contained both more key structural concepts ($F(1, 28) = 4.77$, $p < .05$, $d = 0.80$) and more key process concepts ($F(1, 28) = 8.27$, $p < .01$, $d = 1.05$) than explanations to adult experts. Overall, explanations to young novices contained more information about parts of the systems and to some extent more information about processes than explanations to adult experts.

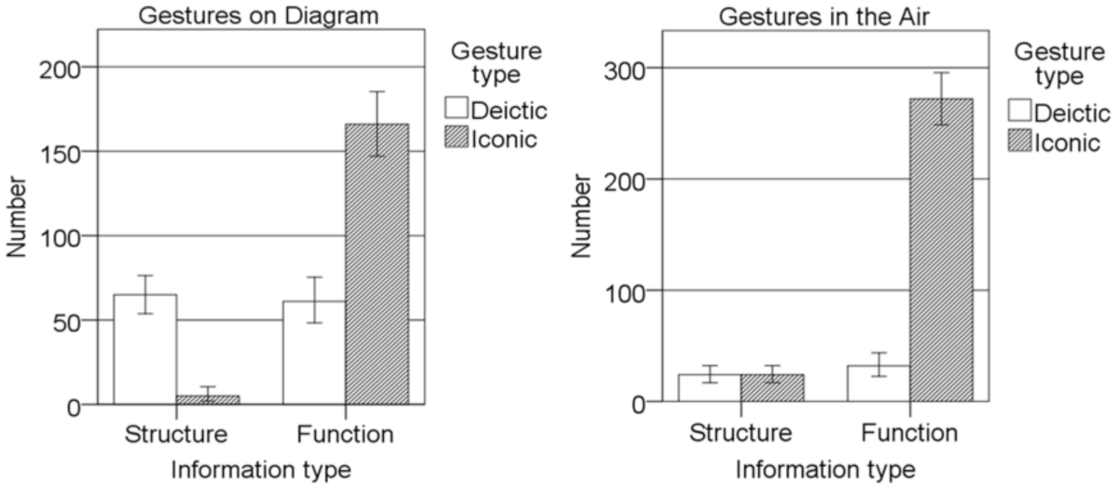
Correspondences between information type and gesture type. Overall, we expected that structural information would be accompanied by deictic gestures because deictic gestures point, indicating the mentioned part. Similarly, we expected that functional information would be accompanied by iconic gestures, as iconic gestures can enact the appropriate action. To assess whether structural information tended to elicit deictic gestures and functional information iconic gestures, we analyzed the cases where information statements were accompanied by gestures, impressively, 66% of the statements about the circulatory system and 56% about the rock cycle. Fig. 11-14 present the correspondences between structural and functional information and deictic and iconic gestures for each system, audience, and gesture platform separately. The figures report the numbers of deictic and iconic gestures for structural and functional statements separately by audience, gesture stage, and system. Because the numbers of gestures and statements were highly variable and in some cases, very small, the data in the figures are for total gestures over participants, not averages. As is apparent from the data and the Chi-square values in Fig. 11-14, the data strongly support the correspondence of structural information with deictic gestures and functional information with iconic gestures. Note the consistency across the two systems. Remember also the previously reported findings that explanations to adult experts staged gestures primarily on the diagrams, but explanations to young novices staged gestures both on the diagrams and on virtual diagrams created in the air.



$\chi^2(1, N = 271) = 120.26, p < .001, V = .67.$

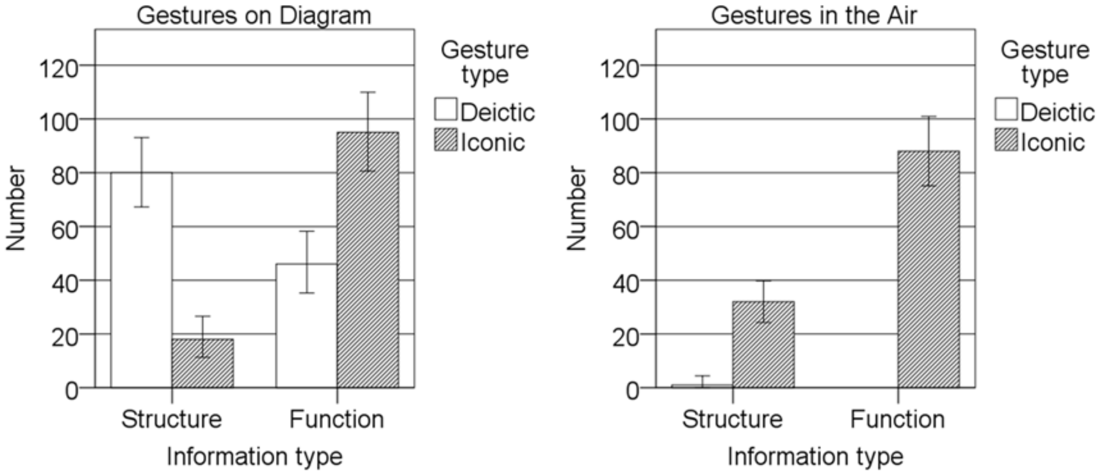
$\chi^2(1, N = 91) = 19.90, p < .001, V = .47.$

Fig. 11. Numbers of types of gestures for adult experts on diagram (left panel) and in the air (right panel) by statement information type for circulatory system. Error bars represent a 95% confidence interval of the counts.



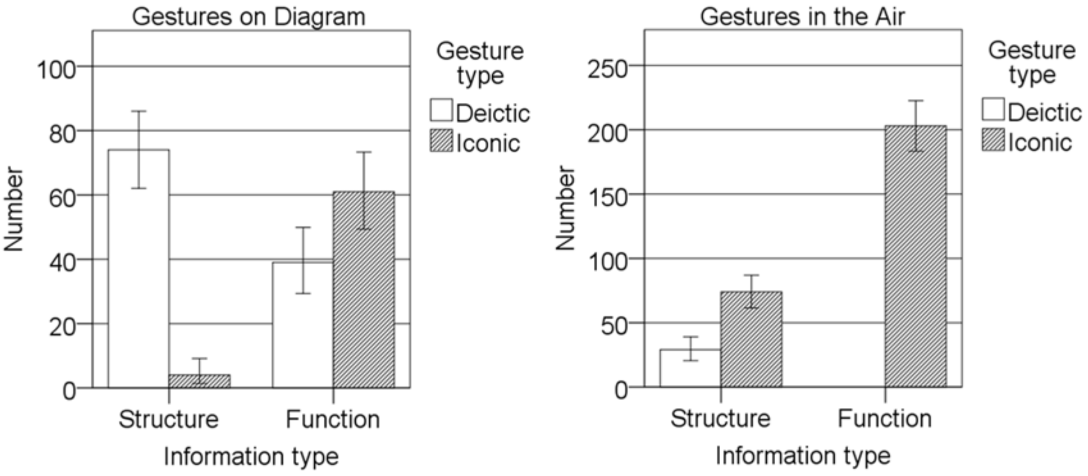
$\chi^2(1, N = 297) = 95.37, p < .001, V = .57.$ $\chi^2(1, N = 352) = 48.28, p < .001, V = .37.$

Fig. 12. Numbers of types of gestures for young novices on diagram (left panel) and in the air (right panel) by statement information type for circulatory system. Error bars represent a 95% confidence interval of the counts.



$\chi^2(1, N = 239) = 55.71, p < .001, V = .48.$ $\chi^2(1, N = 121) = 2.69, p = .10, V = .15.$

Fig. 13. Numbers of types of gestures for adult experts on diagram (left panel) and in the air (right panel) by information statement type for rock formation. Error bars represent a 95% confidence interval of the counts.



$\chi^2(1, N = 178) = 59.00, p < .001, V = .58.$ $\chi^2(1, N = 306) = 63.14, p < .001, V = .45.$

Fig. 14. Numbers of types of gestures for young novices on diagram (left panel) and in the air (right panel) by information statement type for rock formation. Error bars represent a 95% confidence interval of the counts.

Summary of the Findings. The predictions were upheld, along with ancillary corroborative findings. Explanations to novices opened with an introduction that established common ground in prior beliefs, whereas explanations to experts began directly with the diagrams. Explanations to adult experts primarily relied on the provided diagrams, whereas explanations to young novices relied on both the provided diagrams and virtual diagrams created by gestures in the air. The virtual diagrams were larger than the given ones and elicited more large gestures. Gestures on actual diagrams were primarily deictic gestures pointing to parts, whereas gestures on virtual diagrams were primarily iconic gestures demonstrating action. Consequently, relatively more iconic gestures showing action were directed at novices. Although explanations to adult experts and young novices did not differ in time, far more information per unit time was delivered to novices than to experts. Explanations to both young novices and adult experts contained far more information about function than information about structure, perhaps because the structure was apparent from the diagrams. Verbal functional information tended to be accompanied by iconic gestures, whereas verbal structural information tended to be accompanied by deictic gestures. Thus, as predicted, explanations to novices contained more information, contained more information about action in both language and gesture, and used multiple means of explanation. Explanations to novices also made eye contact a larger percent of the time.

Discussion

Explanations in the wild are a harmonious symphony of language, gesture, gaze, and props, real or virtual, partly overlapping, partly solo, sometimes in unison, sometimes in counterpoint, playing roles that are sometimes interlocking, sometimes independent. We captured some of that dynamic interaction in this study of explanations of complex systems, one biological, the circulatory system, and one physical, the rock cycle. Each system has a structure, parts, properties, and interrelations, and each system has an associated set of mechanisms or causal chain of behavior, action or process, referred to as its function.

Here, participants first studied the systems from diagrams, and explained each system to an imaginary adult expert or child novice. Despite the large qualitative differences between the biological and the physical systems, the properties of the explanations were quite similar. The diagrams were available for explanations. Even so, most participants not only used the given diagrams as props, but also used gestures to create virtual diagrams of the systems in the air, especially when addressing young novices, and referred to them by gesturing as they spoke. Nearly two-thirds of the information statements were accompanied by gestures. Two kinds of gestures dominated, deictic gestures that gave structural information, pointing to the relevant parts and structures, and iconic gestures that showed the relevant actions of the parts and structures. By pointing to parts and enacting actions, explainers avoided cumbersome verbal descriptions of the appearance and properties of the parts, the structural and causal relations among the parts, and the physical qualities of the processes. Iconic gestures enacting the behaviors of the systems outnumbered deictic ones that pointed to parts, perhaps because actions are typically harder to understand, and because the structure and parts were displayed in the diagrams but the actions were not. Notably, the diagrams, real or virtual, were the backbone of the information; the language and the gestures served to annotate and animate the diagrams.

Although the explanations of the physical and biological systems were quite similar, explanations directed to young novices were quite different from those directed to adult experts. Some of the differences may be due to the age of the addressees. However, the notable differences appear to be related to the relative states of knowledge of experts, novices, and explainers. Experts, by definition, know a lot, and novices, by definition, know little, but expert and novice knowledge is not simply in quantity; there are also significant qualitative differences between expert and novice knowledge. Novices usually can readily understand structure. What is typically difficult for novices is understanding change, behavior, action, process, and causality (e. g. Hmelo-Silver, et al., 2007; Tversky & Suwa, 2009). Because explanations to experts only need to demonstrate knowledge that the experts already have, explanations to experts should be straightforward and direct, with little embellishment. Explanations to novices, by contrast, should be redundant, and should emphasize action and change. Explanations to novices are like those of teachers to students, whereas explanations to adult experts are like those of students to teachers. Presumably, explainers share knowledge or common ground with experts, but have less shared knowledge or common ground with novices (Clark, 1996). If explainers know and understand the differences between adult expert and young novice knowledge and understanding, implicitly or explicitly, and if they can spontaneously design explanations to adapt to those differences, their explanations to experts and novices should differ, and they did.

Explainers knew the diagram correctly represented the important parts and connectivity of the system. Their explanations to experts were direct; they began their explanations on the given diagram, explicating it by pointing to successive parts of the system and elaborating the processes that corresponded to each place on the diagram.

Explanations to young novices were dramatically different. Explainers appeared to regard their task as one of teaching the young novices the basics of the systems. They looked more at the young novices than at the adult experts. They began their explanations with an overall introduction before turning to the diagrams. They gave more information and they gave redundant information. Importantly, they supplemented the provided diagrams with large virtual diagrams that they created in the air, they used larger gestures for emphasis, and they went back and forth between the provided diagram and the virtual diagram.

If an explanation is a symphony, the theme, the melody, is the representation of the system, primarily carried by the props, the real or virtual diagrams of the system. The gestures that point and the gestures that enact bring the props to life and the language provides commentary and continuity.

Each mode, explaining to young novices or to adult experts, is a different variation on the same theme, the structure and workings of a complex system. Each variation uses the instruments, language, gesture, gaze, and props, in coordinated, integrated ways that are customized for their audiences. There are recommendations implicit in the theme and in the variations. Create a model of the system, coordinating gestures with language to props, notably diagrams. Using props, gesture and language, build a representation of the complex system, deictic gestures for constructing its framework of components in relation to one another, iconic gestures for enacting the behavior of each component and component-to-component. To demonstrate knowledge to adult experts, use a familiar visible prop that shows the components and their relations. To transmit knowledge to young novices, use a familiar visible prop and also build the framework component by component, creating a large virtual prop. To young novices, everything large, everything repeated, in different ways. Create a harmonious variation for each.

Spatial cognition is more typically studied in understanding and navigating real or virtual environments; these are important, even crucial skills. However, what is singularly impressive about the human mind is that it can use the fundamentals of environmental cognition, such as landmarks, paths, and actions, as a foundation for conceptualizing abstract, even arcane systems, real and imaginary, and to communicate them, indirectly with language, just as for navigation, and, again as for navigation, directly in gesture and diagram.

Acknowledgments. Preparation of the manuscript was partially supported by National Science Foundation grants; HHC 0905417, IIS-0725223, IIS-0855995, and REC 0440103.

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.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin, 21*, 394-405.

Chi, M. (2006). Two approaches to the study of experts' characteristics. In Feltovich, P. J., Prietula, M. J, & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 7-76). Cambridge, UK: Cambridge University Press.

Clark, H. H**.** (1996). *Using language*. Cambridge: Cambridge University Press.

Curry, D. (2002). Circulatory system. Retrieved May 2, 2007, from http://faculty.plattsburgh.edu/david.curry/NUR464/Circul2.gif

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

Enfield, N. (2003). Producing and editing diagrams using co-speech gesture: Spatializing non-spatial relations in explanations of kinship in Laos. *Journal of Linguistic Anthropology, 13*, 17-50.

Engle, R. A. (1998). Not channels but composite signals: Speech, gesture, diagrams and object demonstrations are integrated in multimodal explanations. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Feltovich, P. J., Prietula, M. J, & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 39-68). Cambridge, UK: Cambridge University Press.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General, 119*, 63-76.

Frick-Horbury, D., & Guttentag, R. (1998). The effect of restricting hand gesture production on lexical retrieval and free recall. *American Journal of Psychology, 111*, 43-62.

Fry, R., & Smith, G. F. (1975). The effects of feedback and eye contact on performance of a digit-coding task. *Journal of Social Psychology, 96*, 145-146.

Fullwood, C., & Doherty-Sneddon, G. (2006). Effect of gazing at the camera during a video link on recall. *Applied Ergonomics, 37*, 167-175.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62,* 378-391.

Glenberg, A., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31*, 129-151.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26*, 69-83.

Goldin-Meadow, S. (2003). *How our hands help us think.* Cambridge, MA: Harvard University Press.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science, 30*, 581-592.

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-Novice understanding of complex systems. *Journal of the Learning Sciences, 16,* 307-331.

Isaacs, E. A., & Clark. H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology*, *116*, 26-37.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, 57: 227–254.

Kendon, A. (1994). Do gestures communicate?: A review. *Research on Language and Social Interaction, 27,* 175-200.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Lombrozo, T. (2012) Explanation and adductive inference. In K.J. Holyoak and R. G. Morrison (Eds.) *The Oxford handbook of thinking and reasoning* (pp. 260-297). Oxford: Oxford University Press.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Science, 10,* 464-470.

Mainwaring, S. D. Tversky, B., Ohgishi, M., & Schiano, D. J. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spatial Cognition and Computation, 3*, 3-42.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92,* 350-371.

McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.

Otteson, J. D., & Otteson, C. R. (1980). Effect of teacher's gaze on children's story recall. *Perceptual and Motor Skills. 50*, 35–42.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*, 1–24.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes, 20,* 219–247.

Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language, 31*, 261-282.

Tsuei, L (2004). *Using simulation diagrams to facilitate reasoning about mechanisms and systems.* Retrieved from Dissertations and Theses database. (AAT 3129039)

Tversky, B. (2001). Spatial schemas in depictions. In M. Gattis (Ed.), *Spatial schemas and abstract thought* (pp. 79-111). Cambridge: MIT Press.

Tversky, B. (2011). Visualizations of thought. *Topics in Cognitive Science, 3*, 499-535.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective taking. *Cognition, 110*, 124-129.

Tversky, B., Heiser, J., Lee, P., & Daniel, M. P. (2009). Explanations in gesture, diagram, and word. In K. R. Coventry, T. Tenbrink, & J. A. Bateman (Eds.), *Spatial language and dialogue* (pp. 119-131). Oxford: Oxford University Press.

Tversky, B, Heiser, J., & Morrison, J. B. (2013). Space, time, and story. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (pp.47–76). Academic Press, Elsevier Inc.

Tversky, B., Jamalian, A., Giardino, V., Kang, S., & Kessell, A. M. (2013). Comparing gestures and diagrams. *Proceedings of the Tilburg Gesture Research Meeting*.

Tversky, B. and Kessell, A. M. (In press). Thinking in action. *TopiCS in Cognitive Science*.

Tversky, B., & Suwa, M. (2009). Thinking with sketches. In A. B. Markman and K. L. Wood (Editors). *Tools for innovation*. Oxford: Oxford University Press.

Tversky, B., & Zacks, J. M. (2013). Event perception. In D. Riesberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 83-94). Oxford: Oxford.