RESEARCH REPORT

How to put things together

Marie-Paule Daniel · Barbara Tversky

Abstract Instructions for putting things together or understanding how things work are notoriously frustrating. Performance relies on constructing mental models of the object and the actions of the object from text or diagrams or both. Here, we show that instructions can be improved by turning users into designers and deriving design principles from their designs. People first assembled an object and then crafted assembly instructions, using text alone or text and diagrams. Some were required to be brief and to include only the most essential information. Users' instructions had a narrative structure with an introduction, a middle, and an end. The essential middle described or depicted the step-by-step sequence of actions on parts. Diagrams were regarded as fundamental, and redundancy of depictions and descriptions desirable. These design principles have wide applicability to many kinds of explanations.

Keywords Assembly instructions · Diagrams · Spatial ability

Introduction

So often people bring home a bright box with a new toy or gadget or piece of furniture, enticed by the colorful photo on the box. That excitement is often dampened by the emergence from the box of a pile of mysterious parts and an opaque set of instructions, written in what seems to be a translation of a translation. Problems with instructions are not limited to consumer products; they are rampant, in explanations of complex systems, in instructions to operate equipment, and in navigation (e.g., Mijksenaar and Westendorp 1999; Norman 1998). No wonder consumers so frequently ignore instructions (e.g., Ganier 2004; Carroll et al. 1994) or find themselves with extra parts after they think they are done.

Understanding instructions or explanations entails forming a mental model of the object and of its actions from language or diagrams or a combination of both. Forming mental models of complex systems is known to be challenging (e.g., Gentner and Stevens 1983). The language of the instructions can facilitate or exacerbate formation of mental models (e.g., Dixon 1987a, b; Mani and Johnson-Laird 1982). Little is known about instruction design. Because plans are organized hierarchically, instructions are read faster when general organizational information precedes specific component information. Statements in which the action precedes the object are easier to understand because they provide the information in the order in which it is to be implemented (Dixon 1987a, b). The design of assembly instructions needs more guidance than that.

Understanding assembly instructions requires constructing a series of mental representations of structure and of action. The structural representation includes information about the object to be assembled, its parts, their form, their structure, and their spatial relations with respect to the other parts and the whole. Representations of structure may also contain information about identifying features of the parts, such as color, shape, texture, and weight. The action component includes the actions required to proceed from

M.-P. Daniel (✉)
Limsi-CNRS, Orsay, France
e-mail: mpd@limsi.fr

B. Tversky
Columbia Teachers College,
Stanford University, Stanford, CA, USA

step to step and the changes in structure accomplished by each step (Tversky et al. 2007). Ideally, instructions should allow users to form and implement this set of interrelated mental models.

Constructing mental models can be facilitated by well-crafted diagrams (Bauer and Johnson-Laird 1993; Glenberg and Langston 1992; Hegarty et al. 1990; Novick 2001, 2006; Larkin and Simon 1987; Scaife and Rogers 1996; Schnotz 2002; Tversky 2001, 2005, 2011). Among the reasons that diagrams are effective are the direct corre-spondence of the elements and spatial relations in the diagram to the parts and spatial and action organization of assembly (Tversky 2001; Krull et al. 2004) as well as the spatial convergence of the relevant information (Larkin and Simon 1987). Nevertheless, even when instructions are accompanied by diagrams, following instructions is not as simple or efficient as it should be (Cheng 1996; Novick and Morse 2000; Marcus et al. 1996). In addition, diagrams may not be sufficient (e.g., Stenning and Oberlander 1995) or may be difficult for some users, so well-designed verbal instructions can be essential.

Given the complexity of the tasks and the difficulty of constructing effective instructions, either descriptive or depictive, designing effective instructions presents chal-lenges. Typically, instructions are composed by technical writers but technical writers are not trained in human cognition, and may find it difficult to take the perspective of inexperienced users. Here, we try a different approach to instructional design, applying what has proved to be an efficient way to design effective instructions and interfaces, namely, relying on users as designers (Kessell and Tversky 2011; Tversky et al. 2007, 2012). Those studies have shown that instructions and interfaces designed by users, especially by users of high spatial ability, are effective, that is, they improve performance in the tasks they were designed to facilitate. The present series of experiments extends the previous research program to the design of verbal as well as visual instructions.

We chose a common object, a TV cart, which is typical of the kinds of consumer products that require assembly. In order not to bias our amateur designers, they assembled using only the photograph on the box as a guide. After-ward, they were asked to design instructions for others to use as an aid in assembling the TV cart.

The goal of the first experiment was to produce desiderata for the kinds of information to include in verbal instructions. Requiring producers of instructions to be concise forces them to select the most essential information. In previous research, way-finding instructions constrained to be brief were even more effective than unconstrained ones (Daniel and Denis 2003). Half the participants were asked to pro-duce concise instructions while the other half was not con-strained. What kinds of information do users regard as

essential and retain in the concise instructions? The second experiment examined spatial ability in the design of both verbal and diagrammatic components of instructions. Spatial ability is related to producing and understanding diagrams (e.g., Tversky et al. 2007); would spatial ability also be related to quality of verbal instructions? How do designers combine text and diagrams in their instructions? The third experiment investigated the role of the brevity restriction for both verbal and diagrammatic components. Together, the experiments will provide insight into the design of instruc-tions and explanations, for text and for text and diagrams in combination.

Experiment 1: Designing verbal instructions, concise or unconstrained

The first experiment was designed to elicit and code the kinds of information users find important for assembly instructions, and especially, the critical information they regard as essential when instructions must be brief. In previous work on route directions, encouraging the par-ticipants to include the absolute minimal information elicited directions that were not only concise and efficient but also more accurate (Daniel et al. 2003).

Method

Participants

Forty-one Stanford undergraduates, all native English speakers, were tested individually. They participated in this experiment for course credit.

Materials

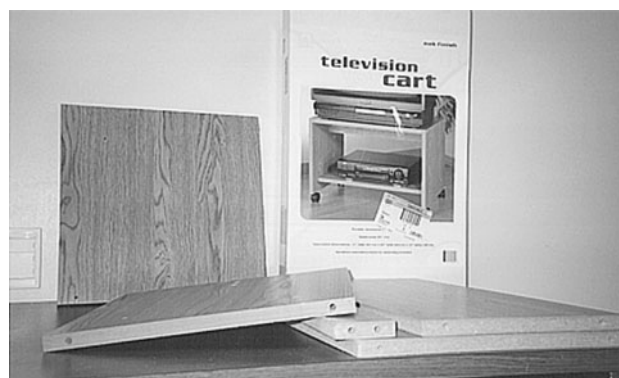The parts of the disassembled TV cart, the tools needed for assembly, and the box with the photograph of the

Fig. 1 Photograph of the TV stand

completed TV cart, as shown in Fig. 1, were provided to participants.

Procedure

The participants' first task was to assemble the TV cart, without instructions but with the help of the photograph, which was available during the task. After assembling the cart, the participants were asked to produce a set of simple, straightforward instructions for assembling it. They were told that the instructions should be aimed at consumers like themselves. Specifically, the participants were instructed:

> Suppose you have to explain how to assemble this TV stand. Please write instructions to assemble the TV stand so that someone else can use your instructions to easily and efficiently to assemble the stand (no drawings at all, only verbal instructions). Please, write legibly.

Twenty-one participants were given 2 sheets of paper and did not receive any further instruction (unconstrained condition), while the remaining twenty participants were given a sheet of paper on which a rectangle enclosed a space of ten lines. They received additional instructions as follows: "One concern is that instruction manuals are too long. Please include only the absolute minimal amount of information for a person to assemble the TV stand easily and efficiently. Your descriptions must be written only within this rectangle" (concise condition).

Results

All participants were able to assemble the TV stand. The concise and unconstrained instructions were coded for content, length, and narrative structure by propositions.
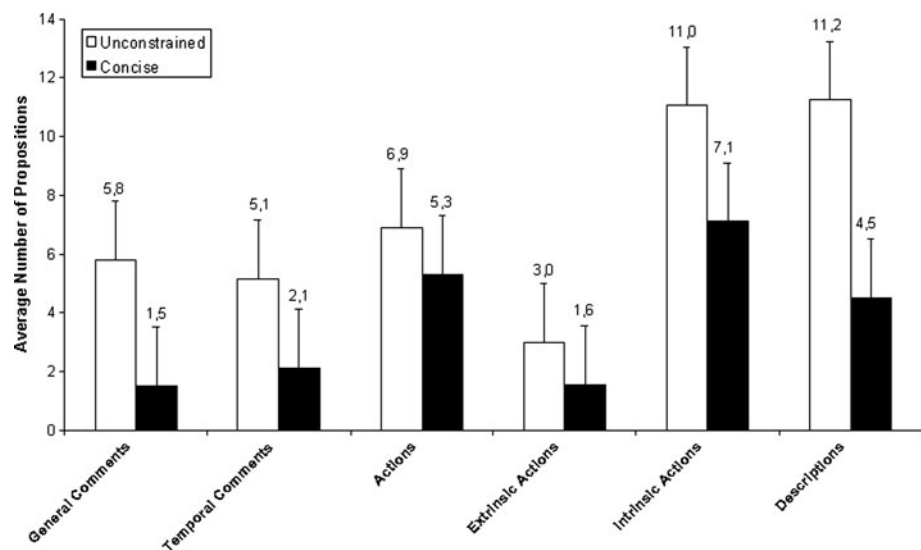
Content and length of assembly instructions

What kinds of information do people include in their instructions and what kinds do they discard when constrained to brevity? We addressed these questions by dividing the instructions into propositions, that is, minimal units of information combining a predicate and one or two arguments, a method previously used to decompose route instructions into informational units (Denis 1997). The content of each individual description was recoded into a proposition-like format. By using equivalent categories, this procedure established uniformity across individual contributions. There were six classes of propositions. General comments: such as "this might be convenient" or "you're ready to use it." Temporal comments: propositions containing temporal references such as "first of all" or "the next step is." Action: propositions prescribing a general action without referring to a spatial orientation or a specific assembly step, such as "take the smaller board," "use the screwdriver," or "insert the pegs in the holes." Extrinsic Action: propositions containing an action with respect to spatial reference system of the body or world, such as "insert them horizontally," "the rough finish side must face you," or "rotate stand 90?" Intrinsic Action: propositions containing prescriptions of actions with respect to an object, such as "attach other side panel to the open side," "slide the piece between end and support sides," or "the small panel should be perpendicular to the square sides." Description: propositions containing descriptions of objects such as "the pegs are white," "there is a rough side," or "one side has 6 holes."

As for the route directions, the instructions varied widely in their length. Among the protocols collected in the unconstrained condition, the number of propositions ranged from 29 to 90 ($M = 43.1$, $SD = 14.2$); in the Concise

Fig. 2 Average number of propositions of each category per protocol in free and constrained conditions

Fig. 3 Type of introductions
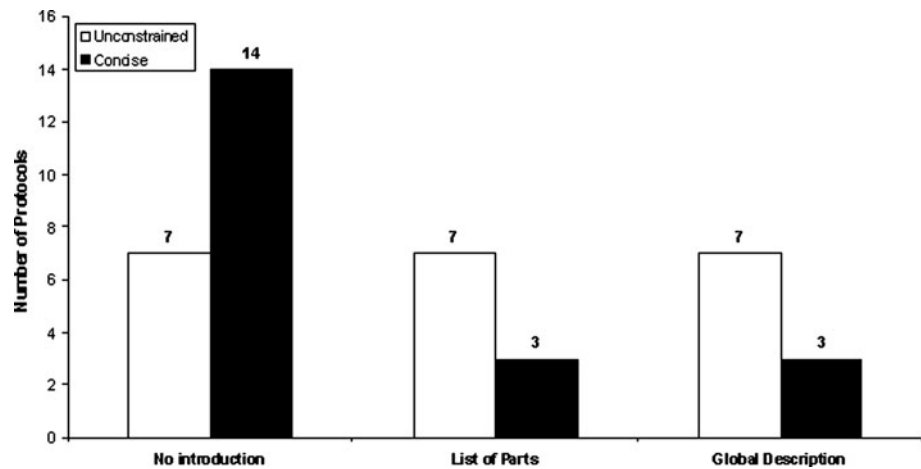
one, it ranged from 12 to 35 ($M = 22.1$, $SD = 5.7$). Unsurprisingly, the constraint of brevity led to a significant reduction in the number of propositions. The difference between the standard deviation values indicates that the brevity instructions not only reduced the number of statements, but also resulted in individual descriptions that were more uniform in size. Overall, the average number of propositions decreased from 43.1 to 22.1 (−48.8 %): $t(39) = 6.20$, $p < .0001$. In particular, there were dramatic decreases in descriptions, temporal comments, and general comments under the constraint of brevity: the number of descriptions dropped nearly 60 % [11.24 vs. 4.5, $t(39) = 4.54$, $p < .001$], the number of temporal comments decreased from 5.1 to 2.1 [$t(39) = 3.84$, $p < .001$], and the number of general comments from 5.81 to 1.5, $t(39) = 5.38$, $p < .001$.

In contrast to the prescriptive parts of the instructions, the reduction was more selective and quite subtle. The total number of Actions categories ("Actions", "Extrinsic Actions", and "Intrinsic Actions" altogether) decreased significantly from unconstrained to concise conditions: 20.95 ($SD = 6.9$) vs. 13.95 ($SD = 3.51$) [$t(39) = 4.05$, $p < .001$]. However, as evident from Fig. 2, this reduction is mainly due to the dramatic decrease in Intrinsic Actions, 11 versus 7.1, $t(39) = 3.47$, $p < .001$. The reduction in Extrinsic Actions was smaller, 3 versus 1.55 [$t(39) = 2.40$, $p < .05$], the decrease in the number of Action propositions was smaller still and failed to reach significance [6.90 vs. 5.30 (−23.2 %); $t(39) = 1.56$, $p = .12$, ns].

The results of the present experiment clearly show that when participants are required to be brief, they produce shorter descriptions, but in a highly selective manner. They reduce the less crucial general, descriptive, and temporal information but do not reduce the information critical to performing the task, the action information. A more microscopic analysis of the narrative structure of the instructions reveals further refinement in concise instructions.

Narrative structure of instructions

In addition to containing the step-by-step information needed for assembly, the instructions also had a narrative structure, that is, a beginning, middle, and end. The narrative structure was most evident in the ways the instructions opened and in the way the middle was organized, step-by-step. These devices were used differently when the instructions were constrained to be brief.

Introductions There were three main ways participants began their instructions. One method was to skip an introduction entirely and begin directly with the assembly steps, for example, "first use the white pegs to attach skinny board to the two square boards." Another method, similar to a recipe, was to list the needed parts and tools, for example: "there are two long boards, two short boards, and a board that is long." Yet, a third method was a general introduction to the task, such as explaining the global shape of the completed TV stand or giving a global description. Examples of the last method include "the two square-shaped pieces of wood are the sides of the stand" and "look at all the materials you have to work with and match up parts of the same size." As can be seen in Fig. 3, participants in the unconstrained condition chose equally among the three kinds of beginnings, a third (7 out of 21) began directly with the assembly, 7 out of 21 with a list of parts, and 7 out of 21 with a global description. Those constrained to brevity overwhelming chose to begin the step-by-step instructions directly (70 %, 14 out of 20 participants).

Temporal markers The participants described the assembly task step by step. To mark each step, constrained or not, the majority of the participants used adverbs such as "first," "then," "begin by," "once this is done," or "lastly". In the control condition, a third of the subjects

added numbers or letters (a, b, c…) to accentuate the segmentation and the sequence.

Hierarchical structure of actions Assembly actions are perceived to be hierarchically organized, around goals and sub-goals (Zacks et al. 2001). A closer examination of the action propositions used in the instructions revealed that they were organized hierarchically, by goals and sub-goals. The higher-level action statements provided the "what-to-do" information (the Actions category statements), while the sub-goal statements provided the "how-to-do" information (the Extrinsic and Intrinsic Actions categories). Those constrained to brevity primarily reduced the lower-level information, sparing the higher-level information. The actions were categorized into these two hierarchical levels, based on the verbs used. At the higher level were the verbs that directly expressed the end goals of each step of the task, namely, "to assemble," "to attach," "to join," "to connect," or "to screw." Similarly, the higher level also included expressions such as "insert the wheels" and "flip over the stand." The lower-level category included the other verbs, conveying the sub-goals such as "to put," "to orient," "to slide in," or "to align." The expectation was that under instructions to be concise, more of the higher-level actions would be retained than the lower-level ones.

Overall, the average number of action verbs dropped from 21.10 (SD = 6.55) in the unconstrained condition to 13.85 (SD = 3.47) (t (39) = 4.31, $p < .0001$) in the concise condition (see Fig. 4). This reduction was accounted for by a decrease in sub-goal actions referring to manner of action [$M = 16.43$ (SD = 5.93) vs. $M = 9.15$ (SD = 3.34), t (39) = 4.71, $p < .0001$]. In contrast, the average number of higher-level goal actions did not significantly decline (4.67, SD = 1.96, vs. 4.70, SD = 1.49, ns).
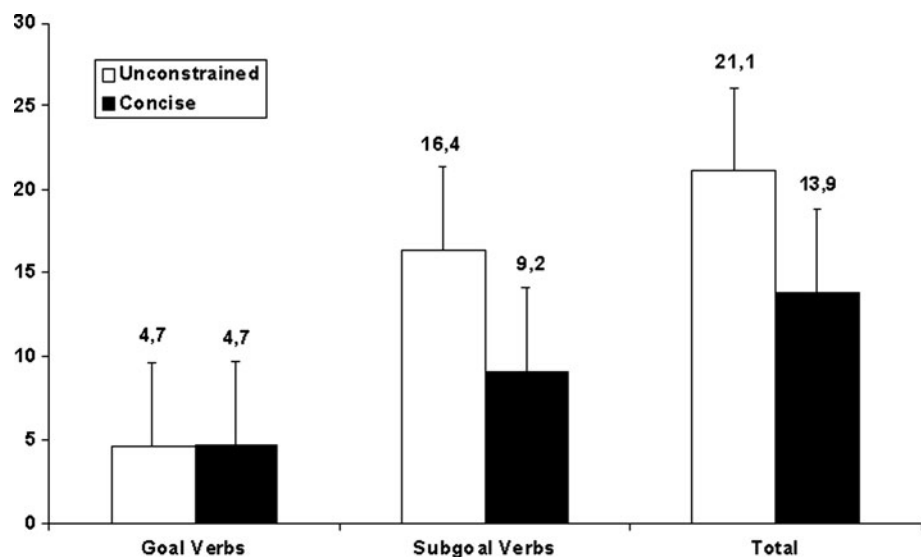
Discussion

Can human users serve as effective designers? The present experiment adds to the evidence that they can. In this experiment, participants assembled a TV cart and then wrote verbal instructions for others to perform that task, either constrained to be concise or unconstrained. The content of the instructions was coded into kinds of propositions. Appropriately, the dominant form of information in the instructions was actions, general ones, or actions described relative to an extrinsic or intrinsic spatial reference system. The instructions also included general comments, temporal markers, and descriptions, typically of parts. Together, the unconstrained instructions had a narrative structure, a beginning introducing the task, a step-by-step middle, and an ending. The concise instructions tended to drop the beginnings and endings, but not the step-by-step middle. What else did the concise instructions omit?

When required to be concise, participants complied by reducing the amount of information. That reduction was by no means uniform; it selectively reduced all categories except for general actions and actions with respect to an extrinsic reference system. Some of the information reduced could be inferred from other information or from the situation. For the introductory material, both the overall goal and the list of parts were evident from the physical parts on the table. The temporal order of steps could be inferred from the temporal order of the instructions, so adverbs like *first* and *next* could be omitted. The higher-level action information was retained but some of the more detailed action information about the exact manner and placement of action was eliminated. The higher-level action information and the constraints of the parts would allow the recipient recover that information.

Fig. 4 Average number of action verbs of each category per protocol in free and constrained conditions

Users who create instructions, then, regard general information about actions to be performed in the service of higher-level goals to be the critical information in instructions, information that should not be omitted from them. Is information about action in fact the information that users want and that helps them to perform the task? We did not address those questions explicitly here but we did in prior research. In earlier work on designing diagrams for assembly, diagrams that conveyed action were rated the highest and were also the most effective in actual assembly (Tversky et al. 2007). Thus, the same action information is regarded as essential for both diagrammatic and verbal instructions and that information promotes performance. It stands to reason that verbal instructions containing action information would also be rated highly and would be effective for assembly.

Users, then, can serve as effective designers of verbal instructions. How will user-designers distribute information across text and diagrams? Are some users better instruction designers than others? The second experiment turns to those questions.

Experiment 2: Spatial ability in designing diagrammatic and verbal instructions

Producing a coherent and efficient set of assembly instructions entails imagining the object, imagining the step-by-step changes, and transforming the successive images into clear prose. Is spatial ability needed to create good verbal instructions? The second experiment addresses that question. Spatial ability is not a unitary ability. Because the task requires imagining spatial transformations, we chose a spatial ability measure that reflects mental spatial transformations, the Vandenberg-Kuse Mental Rotation Task (Vandenberg and Kuse 1978; see also Linn and Petersen 1986). Initial unaided assembly of an object also requires mental spatial transformations, so that those high in spatial ability were also expected to perform the assembly task more efficiently.

What kinds of diagrams will users design, and how will the inclusion of diagrams affect design of text? In this experiment, participants were allowed to create instructions using both text and diagrams. Explanations that use both diagrams and text allow parallel and redundant modes of explanation. Would users regard the redundancy as distracting or helpful? One prominent theorist of visual explanations has argued against redundant information (Tufte 1983); do users agree? Finally, because producing diagrams also requires mental spatial transformations, spatial ability is expected to be associated with quality of diagrams. In earlier work, spatial ability correlated with both quality of diagrams and assembly performance (Heiser et al. 2004).

In short, this experiment will reveal the kinds of information users believe should be included in verbal and diagrammatic instructions. It also allows assessment of the role of spatial ability in producing both verbal and pictorial instructions, as well as the assembly task. In addition, it will reveal whether users prefer instructions that express information redundantly in both language and diagrams or prefer to reduce redundancy.

Method

Participants

Twenty-one undergraduate students (eleven male, ten female) from Stanford University participated in this experiment for course credit. All participants were native English speakers and tested individually.

Procedure

The experiment had three phases. First, participants completed the Vandenberg Mental Rotation Task, which assesses one kind of spatial ability, the ability to perform mental spatial transformations (Vandenberg and Kuse 1978). Next, they assembled the TV stand using the photograph on the package as a guide, as in the first experiment. Finally, they produced assembly instructions for assembling the TV stand they had just assembled. They were given 2 sheets of paper to write their instructions (verbal and diagrammatic). They were told that the instructions should be aimed at consumers like themselves. Specifically, the participants were given the following instructions:

> Suppose you have to explain how to assemble this TV stand. Please write instructions to assemble the TV stand using a combination of pictures and words so that someone else can use your instructions to easily and efficiently assemble the stand. The pictures can be sketches; there is no need to worry about the way they look, as an artist will do the actual drawings. Please, write legibly.

Results

Length and content of assembly instructions

Among the protocols collected, the number of propositions ranged from 21 to 65, with an average length of 38.3 (SD = 13.5). The average length did not differ significantly from the average value obtained in the first experiment (unconstrained condition): 38.3 versus 43.1 (SD = 14.2), $t(40) = 1.14$, ns. The comparison with the results obtained in the first experiment, where only verbal

Fig. 5 Average number of propositions of each category per protocol in Text-Only versus Text-plus-Diagrams conditions
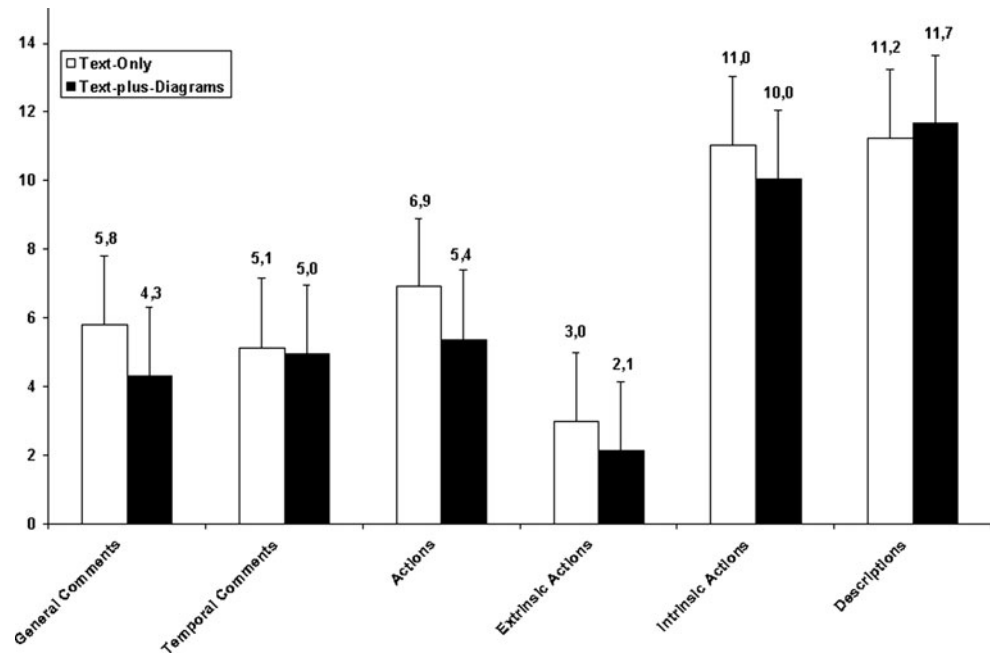
propositions were possible, reveals that the inclusion of diagrams in this new experiment did not substantially change the structure of the descriptions. As can be seen in Fig. 5, the inclusion of diagrams did not significantly reduce the number of Extrinsic and Intrinsic Actions nor the average number of descriptions or temporal comments. The general information decreased somewhat with the inclusion of diagrams but did not reach significance 5.81 versus 4.30, $t (1, 40) = 2.68$, $p < .07$.

Narrative structure of the descriptions

As before, the instructions of participants allowed to use both diagrams and text had a distinct narrative structure, a beginning, middle, and end.

Introductions The distribution of the types of introduction was equivalent in Text-Only group and Text-plus-Diagrams group, respectively, 7/21 versus 8/21 out of the participants omitted any introduction and proceeded directly to step-by-step assembly, 7/21 versus 8/21 opened with a list of parts, and 7/21 versus 5/21 began with a global information about the task.

Temporal markers Participants in both studies used temporal markers equally: average number of 5.1 (SD = 3) in the Text-Only condition versus 5.0 (SD = 2.2) in the Text and Diagrams one.

Hierarchical structure The inclusion of diagrams in the instructions reduced the total number of action verbs slightly, but this difference did not reach significance.

There were 21.10 (SD = 6.55) higher-level action verbs in the Text-Only condition versus 17.57 (SD = 6.92) in the Text-plus-Diagrams one, $t (40) = 1.67$, $p < .10$. However, including diagrams did reduce the number of sub-goals action verbs, as shown in Fig. 6, and that difference nearly reached significance, 16.43 (SD = 5.93) versus 12.86 (SD = 5.82), $t (40) = 1.94$, $p < .06$.

Overall, adding diagrams did not substantially change the structure of the descriptions regarding the general organization of the texts as well as their intrinsic contents: explaining using both diagrams and verbal instructions created parallel modes of explanation. The results show that, globally, the participants regard this kind of cross-modal redundancy as helpful. Would spatial ability affect the structure and content of the instructions?

Role of spatial ability

The mean score on the Vandenberg Mental Rotation Task (MRT), which assesses the ability to perform mental rotation, was 10.4. The median (10) was used to split participants into 10 low spatial participants and 11 high spatial participants.

Assembly performance All the participants were able to assemble the TV stand. Mean assembly time was 10.5 min. (SD = 4.1). A Mann–Whitney U test was conducted. The results showed that participants low in spatial ability took significantly longer to assemble the TV stand: $M = 12.9$ min (SD = 3.7) versus $M = 8.3$ min (SD = 3.0), $z = 2.67$, $p < .01$ (rank, 148 vs. 83).

Fig. 6 Average number of
action verbs of each category
per protocol in Text-Only
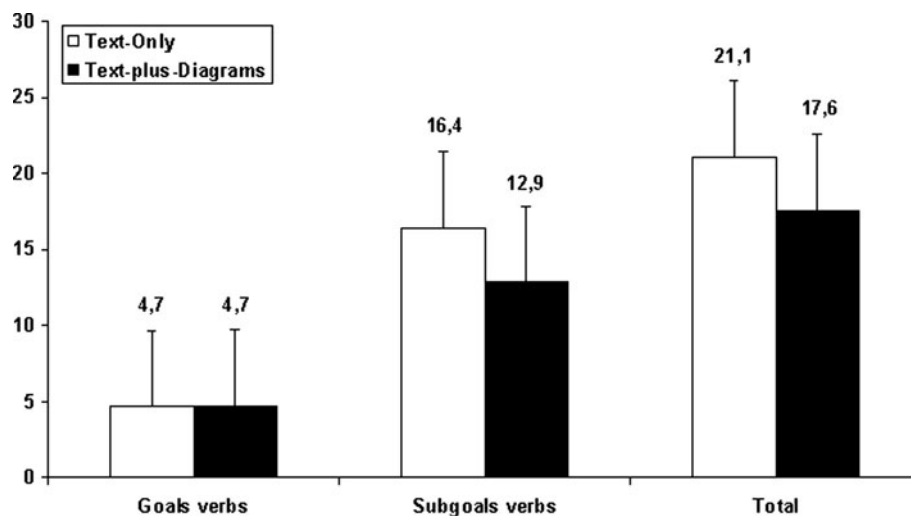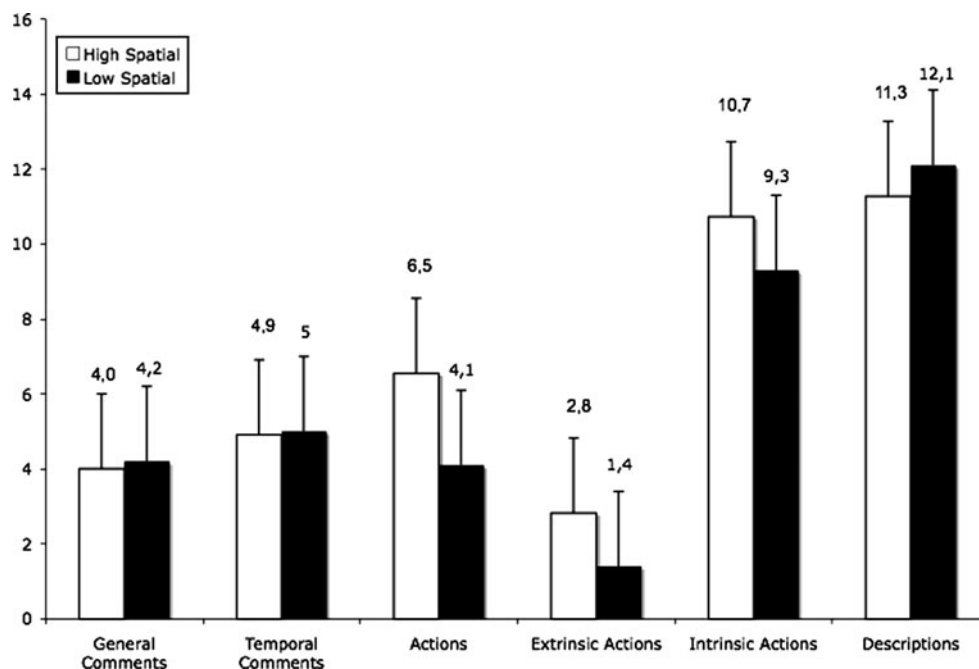versus Text-plus-Diagrams
conditions

Fig. 7 Average number of
propositions of each category
per protocol for low spatial
versus high spatial participants

Characteristics of instructions The descriptions of the high spatial participants were not longer than those of the low spatial ones: the average number of propositions in instructions by high spatial participants was 40.27 (SD = 13.53) versus 36.1 (SD = 13.12) for participants of low spatial ability, $z = 0.81$, ns. However, high spatial participants produced more dynamic texts, reflected in greater prevalence of two language categories: "Actions" and "Extrinsic Actions" (see Fig. 7). High spatial participants produced more actions (M = 6.55, SD = 2.9) than low spatial participants (M = 4.1, SD = 2.7). $z = 2.19$, $p < .05$. High spatial participants had an average rank of 151, while low spatial ones had an average rank of 79.50. High spatial participants also produced more Extrinsic Action statements (M = 2.82, SD = 1.25) than low spatial

participants (M = 1.4, SD = 1.6), $z = 2.04$, $p < .05$. High spatial participants had an average rank of 150, while low spatial ones had an average rank of 81.

The descriptions of the high spatial participants were also more refined as well as more dynamic: they produced more action verbs (19.9 vs. 15, $z = -1.93$, $p < .05$; rank, 148.50 vs. 82.50) in their descriptions, an effect primarily due to including more sub-goal actions, 15.63 versus 9.80 sub-goals verbs ($z = 2.37$, $p < .05$; rank, 154.50 vs. 76.50).

Errors in instructions Instruction protocols were coded for errors and missing assembly steps. For example, because of the constraints of parts attachments, the TV stand must be assembled in one of two orientations: upright

(upside-down) or on the side. Omitting this information was considered as an error. Seven out of ten low spatial participants had this error in their instructions, compared with only two out of eleven high spatial participants [Pearson χ^2 (1, $N = 21$) = 5.7, $p < .05$]. High spatial participants also mentioned more of the assembly steps, 5.88 (SD = 1.2) than low spatial participants versus 4.8 (SD = 1.2) ($z = 2.21$, $p < .05$; rank, 116 vs. 74).

Prescribed direction of rotation Although the TV cart requires rotation to assemble, the majority of participants' instructions described the assembly without mentioning any rotation, as if they were keeping in mind the upright (vertical) orientation of the completed TV stand. Out of 21 participants, 14 described the assembly keeping an upright orientation. Instead of prescribing the successive rotations of the different panels requisite for assembly, they preferred to instruct their imagined addressees first to maintain the sides up, vertically oriented, and second to screw in the three perpendicular panels to these sides in succession. This way of proceeding would make attaching these panels difficult and result in awkward performance. Although avoiding rotation would make the physical task more effortful, it makes the imaginary task easier, as it maintains the orientation of assembly as close as possible to the orientation of the stand when completed (Fig. 8).

In practice, assembly is steadier and easier when it begins by first placing the top panel upside-down or sideways, and then attaching the sides to it. However, explaining this method requires good spatial ability. The necessary spatial transformations appear to be too demanding for most participants. On total, only a third (7 out of 21) participants prescribed a 90° or 180° rotation in order to achieve the assembly. Of the 7, 6 were high in spatial ability.

The necessity to assemble upside-down or sideways was most likely omitted because the pragmatics of assembly require it, and in fact, all participants assembled by putting the top upside-down or sideways. The top of the cart must be supported while other parts are added, and they can be added only if the top is upside-down or sideways. Instructions are always used in a context, and the context places constraints. Route instructions, for example, do not usually specify exact angles of turns or distances because the context constrains that information (e.g., Tversky 2011).

Analysis of diagrams in instructions The diagrams produced by participants were coded as one of three types (see Fig. 9): parts menus, structural diagrams, and action diagrams. Parts menus consisted of a list of the appearances of the parts. Structural diagrams depicted two or more parts in configuration. Structural diagrams were used to show a
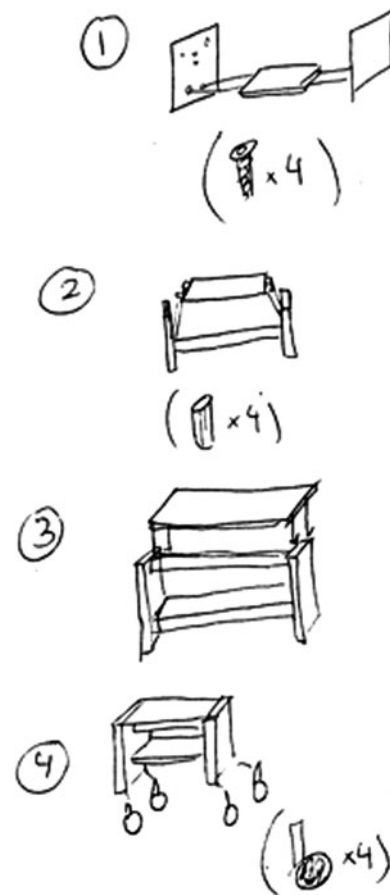

Fig. 8 Example of "upright" assembly, without any prescribed rotation

completed step or to demonstrate what an object should look like at a given point. Action diagrams showed the actions required to attach a part, that is, diagrams that depict one part joining another, demonstrating the necessary assembly procedures. The latter are preferred by users of instructions (Tversky et al. 2007).

High spatial ability participants did not produce a larger number of diagrams than those of low spatial ability (summing part, structural, and action diagrams), high ($M = 5.9$, SD = 2.3), and low spatial participants ($M = 6.5$, SD = 2.8), $p > .05$. However, high spatial participants drew better diagrams. In particular, high ability participants produced significantly more action diagrams (2.81, SD = 2.3) than low spatial (0.60, SD = 1.4) participants, $z = 2.25$, $p < .05$ (rank, 153 vs. 78). They also produced significantly more diagrams with depth information than low spatial participants 2.18 (SD = 1.2) versus 0.80 (SD = 0.69), z = 2.11, $p < .05$, rank, 151 vs. 80. In addition, high spatial participants were more likely to use a 3/4 perspective, which best shows the action, in their instructions than low spatial participants (1/10), χ^2 (1, $N = 21$) = 6.34, $p < .05$ (see Fig. 10).

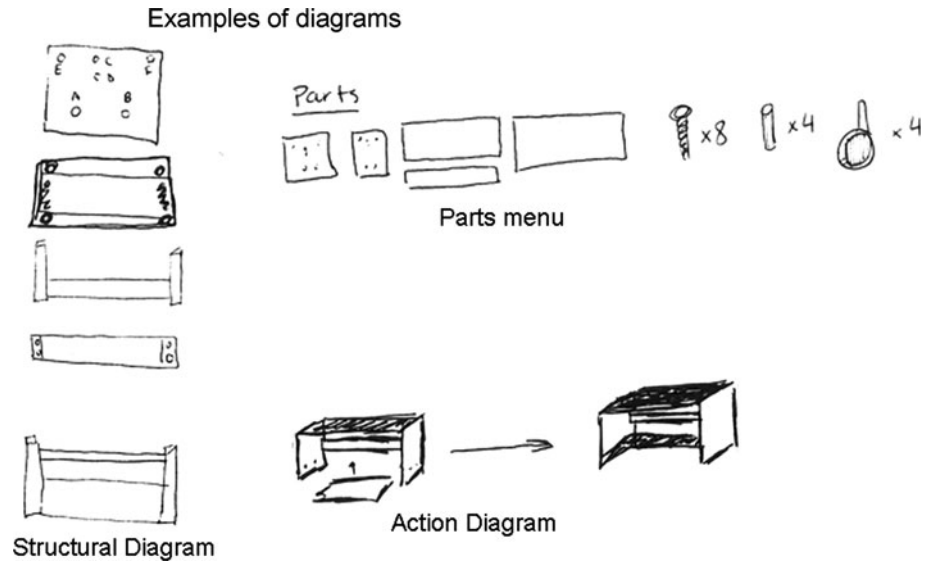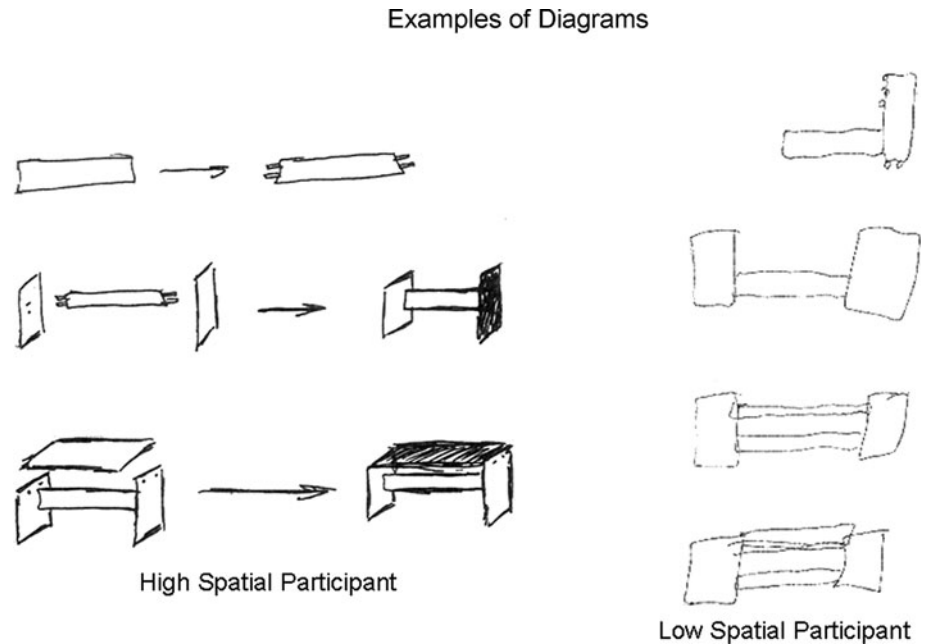**Fig. 9** Examples of part menu, action, and structural diagrams

Examples of diagrams

Parts menu

Structural Diagram

Action Diagram

Fig. 10 Examples of diagrams from high versus low spatial participants

Examples of Diagrams

High Spatial Participant

Low Spatial Participant

Discussion

As in the first experiment, participants in this experiment first assembled a TV cart using a photograph of the completed cart as a guide. Then, they produced instructions to aid another in assembling the TV cart. In contrast to the first experiment, participants were allowed to use diagrams as well as language to convey assembly.

One question was whether the addition of diagrams would reduce or alter the verbal portion of instructions. That is, would users choose to convey certain types of information in language and other in diagrams, or would

they simply use both? Including diagrams did not change either the quantity or the quality of the language portion of instructions. This suggests that users regard both text and diagrams as potentially useful to conveying assembly, even if they are redundant.

A second question was the effects of spatial ability on assembly of the TV cart as well as the quality of instructions for assembly. In fact, high spatial participants both assembled the TV cart more efficiently and produced more effective diagrams. Diagrams were of three types: part menus, diagrams showing structure, and diagrams showing action. Diagrams conveying the step-by-step actions are the

most effective for assembly (Tversky et al. 2007). The high spatial participants in fact produced more action diagrams than the low spatial participants.

More surprising was the effect of spatial ability on the quality of the verbal instructions. High spatial participants included more action information in their verbal instructions and made fewer errors. Thus, spatial ability confers an advantage even in a verbal task when the task is to describe spatial transformations.

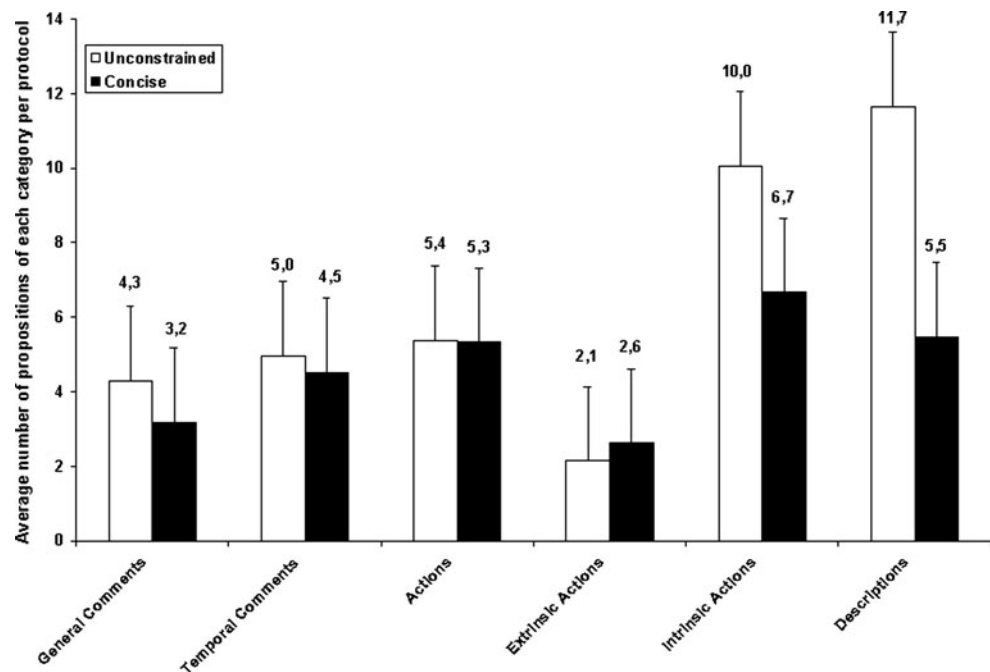## Experiment 3: Producing concise instructions with diagrams and text

The third experiment combines and replicates the first two experiments. Here again, participants were asked to produce concise instructions but allowed to use both text and diagrams. Would the requirement to be concise reduce the redundancy of the instructions? If so, how will redundancy be reduced, by reducing the verbal instructions or the number of diagrams or both? As in the second experiment, spatial ability was assessed to determine whether high spatial individuals will again produce better verbal instructions than low spatial individuals.

Method

Participants

Twenty-one undergraduate students (eight male, thirteen female) from Stanford University participated in the experiment for course credit. All participants were native English speakers. They were tested individually.

Procedure

As in the previous studies, the procedure had three phases: first, the participants completed the Vandenberg Mental Rotation Test after which they assembled the TV stand. In the third phase, they were asked to write assembly instructions for the TV stand that they had just assembled. They were told that the instructions should be aimed at consumers as themselves. They received instructions as follows: "Suppose you have to explain how to assemble this TV stand. Please write instructions to assemble the TV stand using a combination of pictures and words so that someone else can use your instructions to easily and efficiently assemble the stand. One concern is that instruction manuals are too long. Please use a combination of pictures and words including only the absolute minimal amount of information for a person to assemble the TV stand easily and efficiently. The pictures can be sketches; there is no need to worry about the way they look, as an artist will do the actual drawings. Please, write legibly." The participants were given a half sheet of paper on which to write their assembly instructions instead of the two sheets of paper provided in the previous experiment.

Results

Content and length of the constrained descriptions (Text-plus-Diagrams)

Comparing the results of the second experiment to this one allows assessing the effects of restricting the space for instructions. Restriction led to small decreases in language

Fig. 11 Effect of the constraint of conciseness in the Text-plus-Diagrams condition

Fig. 12 Comparison of the decreasing effect of the constraint of conciseness according to absence versus presence of diagrams
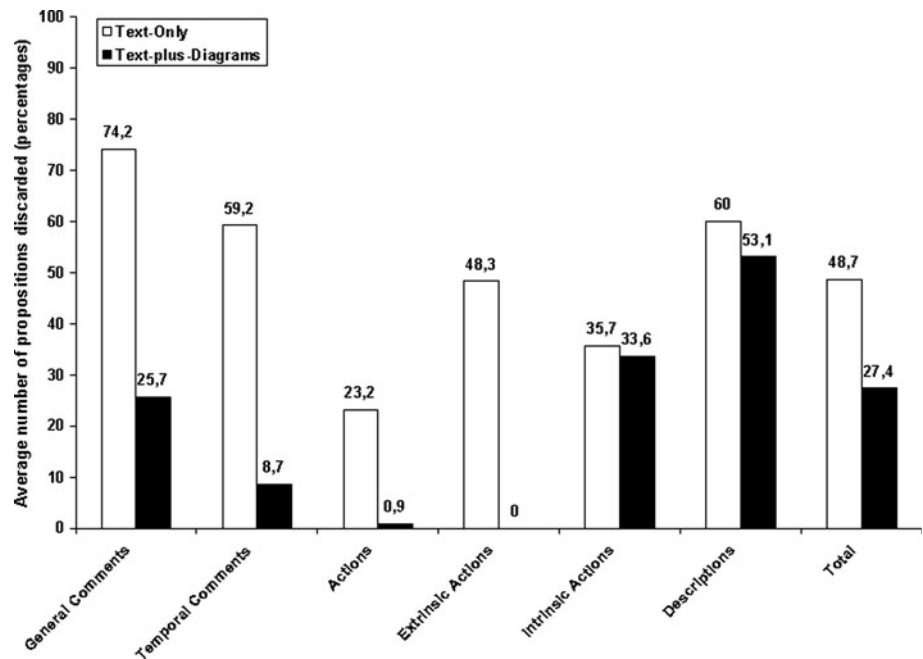
similar to those observed in the first experiment (see Fig. 11): the descriptions' length ranged from 15 to 42 propositions, with an average value of 27.8 (SD = 9.1) compared with 21 to 65 propositions in the Text-plus-Diagrams condition of the second experiment, with an average value of 38.3 (SD = 13.5) in the unconstrained condition [38.3 vs. 27.8, $t(40) = 2.94$, $p < .005$].

Intriguingly, the constraint to be brief had larger effects in the Text-Only condition, with a 48.6 % reduction, than in the Text-plus-Diagrams condition, where the reduction was only 27.8 %. Whereas in the language alone condition, all categories but two (Extrinsic Actions and Actions) were significantly reduced, in the Text-plus-Diagrams condition, the reduction significantly affected only two categories: "Intrinsic Actions" [$M = 10$, SD = 4.73 vs. $M = 6.7$, SD = 2.15, $t(40) = 2.98$, $p < .005$] and "Descriptions" [11.7, SD = 6.22 vs. 5.5, SD = 4.58, $t(40) = 3.67$, $p < .001$]. Figure 12 compares the extent of the reduction due to the constraint in Text-Only versus Text-plus-Diagrams conditions.

Narrative structure of the descriptions

The constraint of brevity had effects on the narrative structure of the descriptions and especially the introductions to the instructions similar to those observed in Experiment 1. When constrained, the majority of the participants omitted introductions and began directly with step-by-step assembly (Fig. 13).

The average number of action verbs was not substantially reduced either ($M = 17.57$ in the Text-Only condition, SD = 6.91; vs. $M = 14.85$, SD = 4.89 or in the

Text-plus-Diagrams one), $t(40) = 1.46$, ns. The average number of goal verbs was 4.71 (SD = 2.61) versus 4.28 (SD = 1.15), whereas the average number of sub-goal verbs was 12.85 (SD = 5.82) versus 10.57 (SD = 5.00).

The constraint to be concise had no effect on the diagrams, suggesting that users regarded the diagrammatic information as already streamlined and more important than the verbal. The diagrams were schematic, so there was no clear way to reduce them while retaining the critical information. Because there was typically one diagram per assembly step, the number of diagrams per instruction set did not significantly decrease when the length of the instructions was limited: the mean number of diagrams participants included in their instructions was 5.9 (SD = 2.5) in this experiment versus 6.2 (SD = 2.5) in the second experiment.

Effects of the presence of diagrams on the language of the descriptions

Surprisingly, even under the constraint of brevity, the average length of the verbal part of the descriptions did not decrease when diagrams were included, but instead, increased [$M = 22.05$ (SD = 5.65); vs. $M = 27.81$ (SD = 9.20), $t(39) = 2.41$, $p < .05$]. The data (see Fig. 14) show that the participants preferred to present comparable information in each mode, text and diagrams, in parallel. Moreover, including diagrams increased the number of comments: $M = 3.60$ (SD = 2.30) versus $M = 7.71$ (SD = 3.65), $t(39) = 4.29$, $p < .0001$.

Rather than allowing a decrease in textual information, then, including diagrams appeared to serve as a guide to

Fig. 13 Distribution of the types of introductions according to the constraint of brevity in the Text-plus-Diagrams conditions
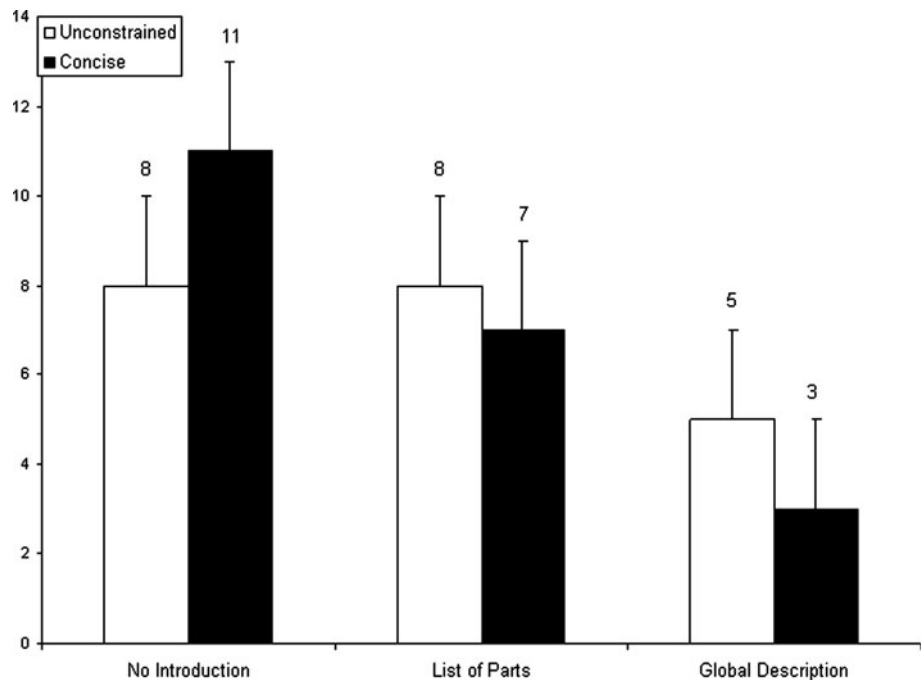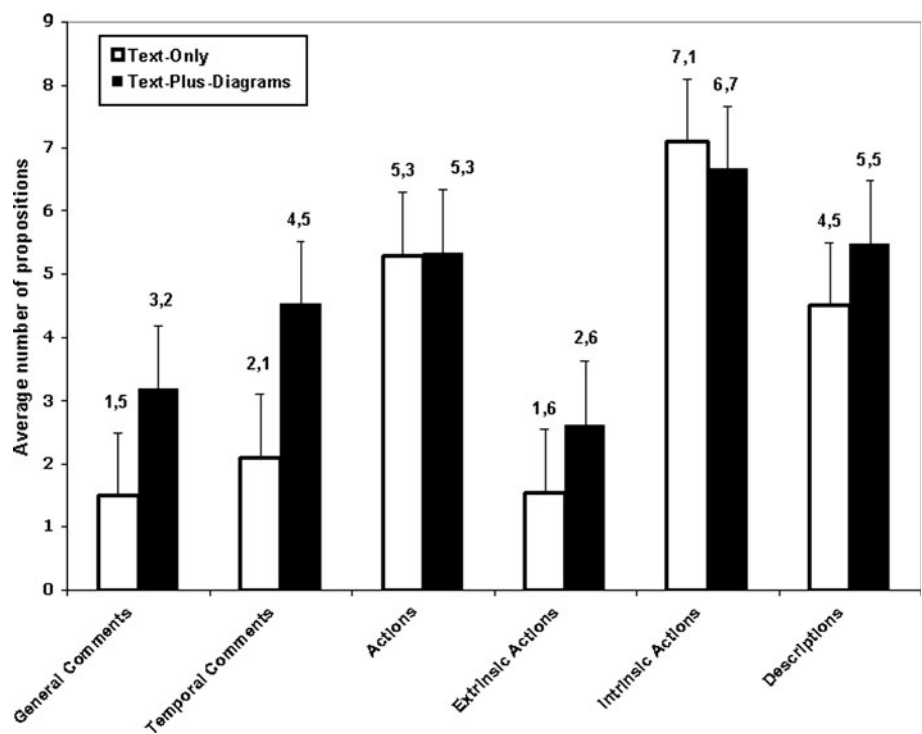
Fig. 14 Effect of the presence of diagrams on both concise conditions

participants for constructing a more complete and better organized set of procedures. By illustrating each assembly step in order, the diagrams reinforced the main actions entailed by assembly and guided participants to produce a refined step-by-step structure. The result is that text and diagrams proceed in independently and in parallel.

Role of spatial ability

The mean score on the Vandenberg Mental rotation test was 10.23 (SD = 5.2), with a median of 11. A median split yielded 11 low spatial and 10 high spatial participants.

Assembly performance

The mean assembly time was 10.1, SD = 4.1. As in the previous experiment, low spatial participants took significantly longer to assemble the TV stand ($M = 12.73$, SD = 3.7) than high spatial participants ($M = 6.64$, SD = 1.03) Mann–Whitney U test, $z = 3.55$, $p < .001$. High spatial participants also produced more dynamic texts, with more actions than low spatial participants. High spatial participants had an average rank of 59.5, while low spatial ones had an average rank of 171.5.

Text As in the second experiment, the length of the descriptions of the high and the low spatial participants did not differ, 25.5 versus 29.91, respectively. Because the brevity constraint reduced the average number of propositions, spatial ability was not reflected in the language categories. Thus, there were no differences in frequency of action verbs due to the spatial ability (14.36, SD = 5.55 vs. 14.9, SD = 3.78, ns).

Replicating Experiment 2, more low spatial participants had an error in their instructions (9/11) than high spatial participants (2/11), χ^2 (1, $n = 22$) = 8.90, $p < .001$. High spatial participants also distinguished more steps, 5.7 (SD = 1.5) than low spatial participants 4.8 (SD = 0.6) (rank, 131.5 vs. 99.5, $z = 1.61$, $p < .11$).

Quality of the diagrams The effect of the spatial ability previously observed on the quality of diagrams was confirmed in this experiment. There was no difference in total number of diagrams produced by high ($M = 5.7$, SD = 3.2) and low ($M = 6.1$, SD = 2.5) participants, $p > .05$, but high spatial participants drew significantly more action diagrams ($M = 2.60$, SD = 1.8) than low spatial participants ($M = 0.63$, SD = 1.2), $z = 2.74$, $p < .005$, rank, 149 vs. 82). High spatial participants also produced significantly more diagrams with depth information ($M = 2.5$, SD = 1.8) than low spatial participants ($M = 0.55$, SD = 0.69), $z = 2.29$, $p < .05$. In addition, more high spatial participants used the 3/4 perspective that showed the actions in their diagrams (6/10) than low spatial participants (0/11), χ^2 (1, $N = 21$) = 6.12, $p < .05$.

Discussion

As in the previous two experiments, participants first assembled a TV cart using a photograph of the completed cart as a guide. They then produced instructions to guide another to assemble the TV cart. As in the second experiment, participants were allowed to use diagrams as well as language to convey assembly. But as in the first one, they were asked to produce concise descriptions. The spatial ability of participants was measured to determine whether

it would affect assembly as well as quality of instructions. Replicating the second experiment, high spatial participants assembled the TV cart more efficiently and produced more effective diagrams. The high spatial participants produced more action diagrams than the low spatial participants and made fewer errors in their instructions. In earlier work (Heiser et al. 2004), the instructions produced by the high spatial participants, those that contained step-by-step procedures emphasizing the actions to be performed, were rated higher by both high and low spatial participants. The preferred instructions contrast with those that came in the box, which were an exploded diagram rather than step-by-step action instructions. The kind of instructions produced by high spatial participants was more effective than the instructions in the box, especially for low spatial participants.

One new question examined in the third experiment was how the constraint to be concise would affect the diagrammatic and verbal portions of the instructions. That is, would users reduce certain types of information in language and others in diagrams, would they simply reduce both and, would they reduce one and not the other? In the first experiment, where the instructions were only in text, users required to be concise reduced the text by 48.6 %. Surprisingly, when participants were allowed to use both text and diagrams, instructions to be concise reduced the text by only 27.8 %. Furthermore, the constraint to be brief did not reduce the number of diagrams, perhaps because there was only one diagram per step, so eliminating a diagram would eliminate a critical step. Participants unequivocally voted for redundancy: even under a constraint of brevity, participants preferred to present complete information in both media.

It seems paradoxical that when instructions included a complete set of diagrams user-designers left more text when they were required to be concise than when there were no diagrams. It seems likely that the diagrams drove the text. The step-by-step diagrams concisely and completely conveyed the step-by-step assembly. It seems that users then added language parallel to the diagrams, describing the actions and consequences of the steps shown in the diagrams. Thus, when diagrams were present, the text was guided not by internal representations of the objects and actions but rather by external representations, the sequence of diagrams of the objects and actions. The diagrams did not illustrate the text; rather, the text described what was in the diagrams.

General discussion

Explanatory instructions, such as those to assemble something or to operate something, can be enormously frustrating (e.g., Norman 1998). In fairness, designing good

instructions is not easy, in large part because the instructions need to convey mental models of both structure and action, and each may be subtle and intricate. In addition, instructions are often written by technical writers who use technical language and for whom the task may be trivial or obvious. Instruction writers are experts who may not be able to enter the minds of novices to comprehend their perspectives, experience, and knowledge in order to know what information would be most helpful to them and how to convey it. Indeed, recent learners might have the best balance of expertise and naivete to serve as designers. With this in mind, we asked naïve students to become new experts, by having them first assemble a simple object, a TV cart, and then design instructions that would guide novices, like their former selves, to efficiently assemble the cart. Over three experiments, participants were asked to design purely verbal instructions or instructions using both diagrams and text. Some were asked to make their instructions brief and others were not constrained. Spatial ability, as assessed by mental rotation, was measured as previous research had shown that spatial ability contributes to understanding and producing diagrams (e.g., Heiser et al. 2004; Tversky et al. 2007).

Rather than asking user-designers directly for guidelines, probably a futile task, design guidelines were inferred from the designs of user-designers. The kinds of information that participants included in their instructions, especially the information kept under the constraint of brevity, are likely to be the information that would most benefit novices.

Can we rely on users as designers? Are their designs effective? Previous research has shown that the designs created by new users are indeed effective for performance (Kessell and Tversky 2011; Heiser et al. 2004, 2007; Tversky 2011). Here, we turned users into designers of verbal and diagrammatic components of instructions: who are the best designers and what are their implicit recommendations?

The best designers of assembly instructions turned out to be those high in spatial ability. It was perhaps not surprising to find that those high in spatial ability were proficient at assembling quickly and accurately. Assembly of parts to form a functioning 3-D object entails spatial thinking, including thinking about actions on parts and wholes in different orientations. It is also not surprising to find that those high in spatial ability design more powerful and effective diagrams for assembly (Heiser et al. 2004). In particular, the instructions designed by high spatial participants for assembly were rated higher by both high and low spatial participants. They were also more effective than the exploded diagram instructions that came in the box, notably for low spatial participants (Heiser et al. 2004). More surprising was the finding here that those high

in spatial ability produced better and more accurate *verbal* instructions, including more comprehensive descriptions of the actions to be performed and fewer errors. Creating assembly instructions depends on imagining the step-by-step actions on objects required for assembly. Those high in spatial ability are better able to imagine those actions and then to transform them into prose.

Assembly instructions are a narrative. Interestingly, both verbal and diagrammatic instructions had a narrative structure. The unconstrained textual instructions had a beginning, middle, and end. Most instructions began with an introduction, for the verbal instructions typically a general statement such as the goal of the task and for the diagrammatic instructions typically a "menu" of the parts to be assembled. For both verbal and diagrammatic instructions, the middle was a step-by-step sequence of actions, where each new step corresponded to a new object part the instruction was to show or say how to attach. For the verbal instructions, the end was often a summary statement indicating that the task was finished or suggesting what could now be done with the object, how to enjoy it. For the diagrammatic instructions, the end was often a depiction of the completed TV cart with sparkly lines radiating from it.

The major portion of the instructions was the step-by-step sequence of actions. For the textual condition, the middle portion consisted of a sequence of higher-level actions, such as attaching a major part, and lower-level statements giving more details on the orientation of the part and the manner of attachment. Notably, under the constraint to be brief, the high-level actions and the diagrams were retained but the low-level details, the beginning and end, the introduction and the conclusion were eliminated. Users can infer this information from the affordances of the situation.

User-designers produced several kinds of diagrams, notably "menus" of parts, structural diagrams, and action diagrams. The action diagrams used the best perspective to show the attachment process for each step and added guidelines to indicate the orientation of the part with respect to the whole and arrows to indicate the attachment action. The structural diagrams showed only the relations of the parts to each other. This structural information comes for free in the action diagrams. High spatial ability participants created more action diagrams than low spatial ability participants. Not only were the diagrams of the high spatial participants more dynamic, they were also more complete and had fewer errors. Just as for describing actions, depicting actions completely and accurately reflects the ability of high spatial participants to imagine dynamic changes over and above imagining static structural relations.

This narrative structure with clear segmentation of the steps and clear descriptions or depictions of the actions to

be performed at each step stands in stark contrast to the sets of instructions that commonly come in the bright boxes of consumer products, which are often merely a single exploded diagram (Mijksenaar and Westendorp 1999).

Action is fundamental. The critical information to include in instructions, then, is the sequence of actions on objects that users need to perform to correctly assemble the object. This information constituted the majority of the verbal instructions, even more so for the brief instructions. It was hierarchically organized, appropriately, as perception and execution of ongoing assembly is hierarchical (Zacks et al. 2001). At the higher level in the text were the action goals of each step; at the lower level, details about the orientation of the part and the manner of action. To a great extent, the lower-level information could be inferred from the situation, combining the specification of the higher-level action with the affordances of the tools, parts, and the whole. Indeed, under the constraint of brevity, it was the lower-level information that was eliminated, and the higher-level information retained. Similarly, details of the object's appearance that were not relevant to assembly were also eliminated under instructions to be concise.

Diagrams are key. User-designers regarded the diagrams as conveying information critical for assembly. When asked to be brief, they cut extraneous information in the prose but not the diagrams. Because there was on the whole one diagram per assembly action, there was no sensible way to cut the diagrams. User-designers apparently had the correct intuition that diagrams are more effective than words for these kinds of tasks (e.g., Glenberg and Langston 1992; Larkin and Simon 1987; Tversky 2011).

Redundancy is desirable. Although the text and diagrams were redundant, user-designers preferred to keep that redundancy even when directed to be concise. They neither reduced the text to bare-bones text nor the diagrams. This contrasts with the advice of a notable expert (Tufte 1983). Again, user-designers had the correct intuitive understanding that two quite different modes of explanation can complement each other, reduce error, and help novice learners. If information is not clear from one mode, it might be clear from the other quite different mode.

Users, notably new experts, can be effective designers. They know the information that is essential for performing the task, and how to convey it to others like themselves. Their designs revealed a number of design guidelines for assembly instructions. Instructions should be a narrative, with a beginning, middle, and end. Instructions should make the sequence of actions explicit and complete. Instructions should rely on diagrams. Instruction should be redundant, with a parallel textual track. The principles uncovered here have wider applicability than assembly instructions. The components of assembly are structure, parts and wholes, and the step-by-step sequence of actions

that transform the parts and wholes. These are the elements of instructions for performing almost any task. They are also the components of explanations for how things work and how to get from here to there. It's all about action.

References

Bauer MI, Johnson-Laird PN (1993) How diagrams can improve reasoning. Psychol Sci 4(6):372–378

Carroll JM, Mack RL, Robertson SP, Rosson MB (1994) Binding objects to scenarios of use. Int J Hum Comput Stud 41:243–276

Cheng PC-H (1996) Functional roles for the cognitive analysis of diagrams in problem solving. In: Cottrell GW (ed) Proceeding of the eighteenth annual conference of the cognitive science society. Lawrence Erlbaum, Hillsdale, pp 207–212

Daniel M-P, Denis M (2003) The production of route directions: investigating conditions that favor concise spatial discourse. Appl Cogn Psychol 18(1):57–75

Daniel MP, Manghi E, Tom A, Denis M (2003) Testing the value of route directions through navigational performance. Spat Cogn Comput 3(4):269–289

Denis M (1997) The description of routes: a cognitive approach to the production of spatial discourse. Curr Psychol Cogn 16:409–458

Dixon P (1987a) The processing of mental plans for following directions. J Exp Psychol Learn Mem Cogn 13(1):18–26

Dixon P (1987b) The processing of organizational and component step information in written directions. J Mem Lang 26:24–35

Ganier F (2004) Factors affecting the processing of procedural instructions: implications for document design. IEEE Trans Prof Commun 47(1):15–26

Gentner D, Stevens AL (1983) Mental models. Erlbaum, Hillsdale

Glenberg AM, Langston WE (1992) Comprehension of illustrated text: pictures help to build mental models. J Mem Lang 31:129–151

Hegarty M, Carpenter PA, Just MA (1990) Diagrams in the comprehension of scientific text. In: Barr R, Kamil ML, Mosenthal P, Pearson PD (eds) Handbook of reading research. Longman, New York

Heiser J, Phan D, Agrawala M, Tversky B, Hanrahan P (2004) Identification and validation of cognitive design principles for automated generation of assembly instructions. In: Proceedings of advanced visual interfaces'04, pp 311–319, ACM

Kessell AM, Tversky B (2011) Visualizing space, time, and agents: production, performance, and preference. Cogn Process 12:43–52. doi:10.1007/s10339-010-0379-3

Krull R, D'Souza S, Roy D, Sharp DM, Roy D (2004) Designing procedural illustrations: IEEE transactions on professional communication. Acquir Proced Knowl Technol Interface 47(1):27–33

Larkin JH, Simon HA (1987) Why a diagram is (sometimes) worth ten thousand words. Cogn Sci 11(1):65–99

Linn MC, Petersen AC (1986) A meta-anaysis of gender differences in spatial ability: Implications for mathematics and science achievement. In: Hyde JS, Linn MC (eds) The psychology of

gender: advances through metaanalysis. Johns Hopkins University Press, Baltimore, pp 67–101

Mani K, Johnson-Laird PN (1982) The mental representation of spatial descriptions. Memory Cogn 10:181–187

Marcus N, Cooper M, Sweller J (1996) Understanding instructions. J Educ Psychol 88:49–63

Mijksenaar P, Westendorp P (1999) Open here: the art of instructional design. Thames and Hudson, London

Norman DA (1998) The design of everyday things. Doubleday, New York

Novick LR (2001) Spatial diagrams: key instruments in the toolbox for thought. In: Medin DL (ed) The psychology of learning and motivation, vol 40. Academic Press, San Diego, pp 279–325

Novick LR (2006) Understanding spatial diagram structure. Q J Exp Psychol 59:1826–1856

Novick LR, Morse DL (2000) Folding a fish, making a mushroom: the role of diagrams in executing assembly procedures. Memory Cogn 28(7):1242–1256

Scaife M, Rogers Y (1996) External cognition: how do graphical representations work? Int J Hum Comput Stud 45:185–213

Schnotz W (2002) Towards an integrated view of learning from text and visual displays. Educ Psychol Rev 14(1):101–120

Stenning K, Oberlander J (1995) A cognitive theory of graphical and linguistic reasoning: logic and implementation. Cogn Sci 19:97–140

Tufte ER (1983) The visual display of quantitative information. Graphics Press, Cheshire

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

Tversky B (2005) Visualspatial reasoning. In: Holyoak K, Morrison R (eds) Handbook of reasoning. Cambridge University Press, Cambridge, pp 209–249

Tversky B (2011) Visualizations of thought. Top Cogn Sci 3:499–535

Tversky B, Agrawala M, Heiser J, Lee PU, Hanrahan P, Phan D, Stolte C, Daniel M-P (2007) Cognitive design principles for generating visualizations. In: Allen G (ed) Applied spatial cognition: from research to cognitive technology. Erlbaum, Mahwah

Tversky B, Corter JE, Yu L, Mason DL, Nickerson JV (2012) Representing category and continuum: visualizing thought. In: Rodgers P, Cox P, Plimmer B (eds) Diagrammatic representation and inference. Springer, Berlin, pp 23–34

Vandenberg SG, Kuse AR (1978) Mental rotations, A group test of three-dimensional spatial visualization. Perceptual Motor Skills 47:599–604

Zacks JM, Tversky B, Iyer G (2001) Perceiving, remembering, and communicating structure in events. J Exp Psychol: Gen 130:29–58