People, Place, and Time: Inferences from Diagrams

Barbara Tversky (btversky@stanford.edu)
James E. Corter (jec34@columbia.edu)
Jie Gao (jg2902@columbia.edu)
Teachers College, Columbia University
New York, NY 10027 USA
Yuko Tanaka (yuko.tanaka@stevens.edu)
Jeffrey Nickerson (jnickerson@stevens.edu)
Stevens Institute of Technology
Hoboken, NJ 07030 USA

Abstract

Keeping track of things as they move in space and time is a task common to scientists, marketers, spies, coaches, and more. Visualizations of complex information aid drawing inferences and conclusions but there are many ways to represent data. Here we show that the kinds of inferences people draw depend on the kind of visualization, boxes in tables or lines in graphs. Lines link and boxes contain; they both direct attention and create meaning.

Keywords: diagrams; information visualization; inference; data displays.

Introduction

People are always on the move. So are other living things, and even inanimate things, not just tangibles like packages, airplanes, and lava but also slang, fashion, music, rumors. Tracking and understanding movements of things in space and time is a task shared by scientists, historians, football coaches, paparazzi, marketers, physicians, spies, Facebook, event planners, Foursquare, police, culture mavens, advertisers, gossip columnists, friends, and more. The movements of beings and things through space over time are valuable data to be explained by theories. Why do people or things cluster in one place or avoid another? Why did person X see Y and then Z? Why did they meet there? Why is this place popular at one time and not at another? Speculating about the movements of people or things over time is endlessly fascinating, and the number of queries, hypotheses, and explanations that can be generated enormous.

Making sense of complex data like the movements of things in space and time is made easier by organizing it spatially into diagrams. Diagrams are composed of simple geometric forms, dots, lines, boxes, and more that both carry meaning and direct attention (e.g., Tversky, 2011; Tversky, Zacks, Lee, & Heiser, 2000). Lines direct attention by drawing the eye from place to place, point to point, connecting the dots. Lines create meaning by conveying relationships, connections from one place or point to another, as in route maps or networks or line graphs. Boxes also direct attention, by bringing the eye to the contents of the boxes. Boxes are containers, they enclose one set of elements and separate them from elements in other boxes. Boxes create meaning by creating categories. They indicate that everything within the box is similar, sharing features, and different from everything outside the box. Lines and boxes, like other simple geometric marks, are replete with meaning. They alter conclusions, inferences, and interpretations. The same data, height of 8 and 10 year olds or height of women and men, are interpreted as trends when displayed as lines and as discrete comparisons when displayed as bars (Zacks & Tversky, 1999). For example, when lines connected the height of men and women, some people said, “As you get more male, you get taller.”

Lines and boxes should also bias data exploration and inferences from displays of people, place, and time. Previous research evaluated production, preference, and performance of displays of people, place, and time (Kessell & Tversky, 2010). When asked to create ways to keep track of movements of people across space and time, most participants created matrices or tables; a minority connected people over time with lines. Preference by other participants followed the same pattern. Overall, matrices with people as cell entries and time and place in rows and columns respectively were most commonly produced and preferred. This format has good foundations. Place and time are fixed, immutable, but people can move from cell to cell. Performance was assessed by the time to verify many kinds of inferences from the data. Lines facilitated inferences about time, but all other kinds of inferences were faster from tables.

Displays of people, place, and time are frequently used for data exploration, to generate conclusions from the data and inferences about the underlying processes. Here, we investigate the roles of lines and boxes in the spontaneous generation of inferences from data displays. Because lines connect people over time, lines should bias conclusions and inferences about people, and secondarily about time. Boxes emphasize their contents, the confluence of people, place, and time, and should support a greater variety of conclusions and inferences.

Method

Participants

Eighty-one people, 39 of them men, participated through Amazon's Mechanical Turk website. Their ages ranged from 18 to 75, with a mean of 30.9. Forty-six percent had a Bachelor's degree or higher, 39.4% had some college education, 12.2% went to high school, and 2.5% did not specify education level. Most (93.4%) were native English speakers.

Stimuli

The stimuli (Figures 1 and 2) were taken from Kessell & Tversky (2010). Both showed the locations of four students at four times of day with time horizontal, place vertical, and people as cell entries. For the boxes condition (Figure 1) people were color-coded dots. For the lines condition (Figure 2), people were coded as colored lines going from cell to cell. Note that both conditions have boxes, but in the *box* condition, they are filled with the individuals. In the *line* condition, the boxes are empty, in the background, acting as points that are connected by lines.

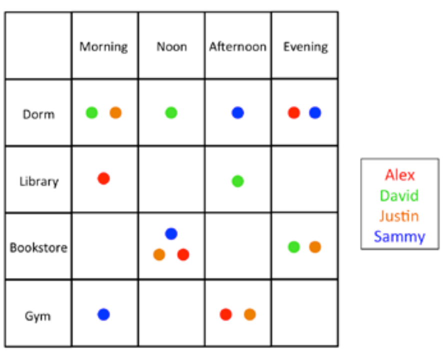

Figure 1: The *box* stimulus display

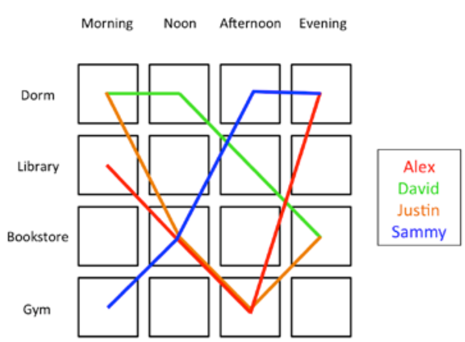

Figure 2: The *line* stimulus display

Procedure

Participants were randomly assigned to the *box* or *line* condition. For both, the first screen, seen in Figure 3, showed an example of a data display, a bar graph, along with several possible conclusions and inferences that could be drawn from the display.

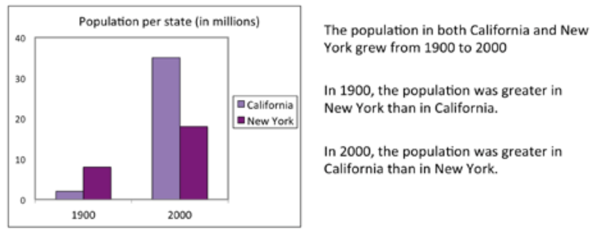

Figure 3. Example used in the instructions.

The sample inferences given were: "The population in both California and New York grew from 1900 to 2000. In 1900, the population was greater in New York than in California. In 2000, the population was greater in California than in New York."

Then either the *box* (Figure 1) or *line* (Figure 2) diagram was presented and participants were directed: "Please study the following graph and use the space below to draw as many inferences as possible." After this task, participants were asked for demographic information.

Results

Participants typically generated many inferences, often several in a single phrase, complicating the coding and the counting. Consequently, inferences and interpretations were coded and analyzed in two ways: the primary and secondary organizer used; and the number of different types of statements/inferences produced. Two people coded; in the few cases where they disagreed, they discussed the cases and came to agreement.

Primary and secondary organizers

In order to capture the overall structure of the organization of the interpretations and to compare the organization produced for each diagrams, we coded the primary organizer and secondary organizer for each participant. The inferences could be organized by *Time People*, or *Location*. Here is an example with *People* as primary organizer and *Time* as secondary organizer:

*"**David** went to the dorm in the morning, stayed at the dorm until noon, went to the library at the afternoon, and ended up at the bookstore at evening. **Justin** went to the dorm in the morning, the bookstore at noon, the gym in the afternoon and back to the bookstore at evening. **Alex** went to the library in the morning, the bookstore at noon, the gym at the afternoon and to the dorm at evening. **Sammy** went to the gym in the morning, to the bookstore and noon, to the dorm at the afternoon, and stayed at the dorm until the evening."*

The results of this coding are shown in Figures 4 and 5. The distribution of primary organizers differs between the two conditions, $\chi 2$ (2; n=81) = 5.815, p=.043; However, the secondary organizers did not differ between conditions: $\chi 2$ (2; n=81) = 2.489, p=.288.

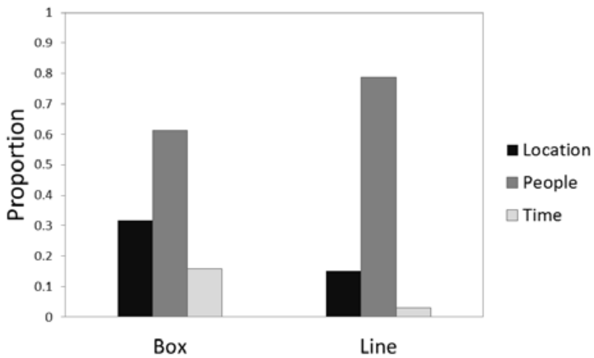

Figure 4. Distribution of primary organizer by condition.

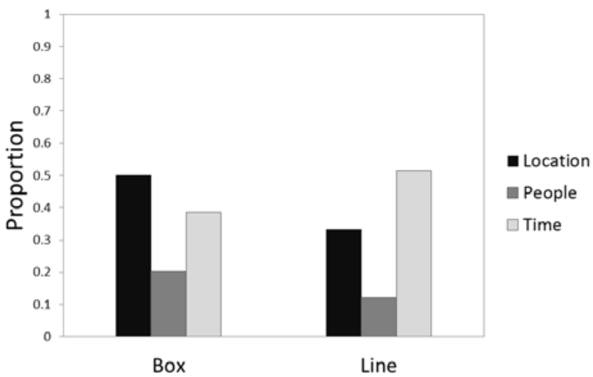

Figure 5. Distribution of secondary organizer by condition

For both box and line displays, the default primary organizer was *People*, followed by *Location* and then *Time*. However, people dominated far more for lines than for boxes. The dominant secondary organizer for lines was *Time*, whereas the dominant secondary organizer for boxes was *Place*.

Number of statements/inferences

As the previous example illustrates, the inference statements were organized and structured. Careful examination of the protocols revealed that most statements could be categorized as follows.

Single statement: a statement that referred to a single cell of the matrix, i.e., one person, one time and one place. For example, *"David is in the dorm in the morning."*

Parallel: a set of related statements in the same format, that is, organized by the same features in the same way. Parallel statements contain many inferences, that is, they refer to information in many cells. For example, the following statement is counted as one parallel statement: *"Justin went to the dorm in the morning, the bookstore at noon, the gym in the afternoon and back to the bookstore in the evening."* Parallel statements invite repetition, and were often repeated.

Generality: any statement that involves more than one person, time, or place (but is not a parallel statement). For example, *"The bookstore is the most consistently visited places for the guys"* and *"David and Sammy spent more time in the Dorms that the others."* Generalities also include many inferences.

Leap: any interpretation that went beyond the information given. For examples, *"David and Sammy are friends," "David is probably unfit," "Alex manages his time well and gets everything done,"* and *"Since Justin does not return to the dorm in the evening I would infer that he is probably dating a student who works at the bookstore and spends evenings at her place."*

Negation: a negative statement from information given in the diagrams. For examples, *"David never goes to the gym," "No one goes to the bookstore in the morning,"* and *"Students are not required to use the library or gym."*

The mean numbers of statements in each category are given in Figures 6 (error bars indicate standard error).

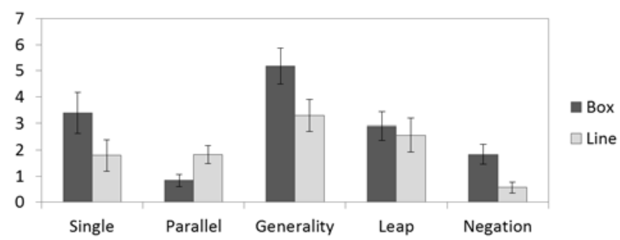

Figure 6. Frequencies of kinds of inferences given to box and line displays.

Differences in the frequencies of the statement categories between the two displays were examined by a generalized linear model with a Poisson distribution for each dependent variable. The results ($\chi 2$ statistics and p-values) are shown below in Table 1.

Table 1. Tests of differences between conditions

Inference type	Wald $\chi 2$ (1)	p-value
Single	17.828	< .001
Parallel	14.608	< .001
Generality	15.156	< .001
Leaps	0.771	.380
Negation	20.86	<.001
Word Count	30.413	<.001

The analyses confirm that the box displays yielded more single statements, generalities, and negations than the line displays, and that the line displays yielded more parallel statements than the box displays. There was no significant difference in number of leaps.

Because of the diversity of statements, there is no sensible way to count and compare the total number of inferences drawn from each display. Some brief statements summarized many information cells, and other statements conveyed none. However, the word count was higher for *lines* (mean=90) than *boxes* (mean=79); this may be due to the large quantity of parallel statements, which are relatively long, for *lines.*

Discussion

Diagrams of complex information use marks and place on the page to convey information effectively (e.g., Tversky, 2011; Nickerson, Corter, Tversky, Rho, Zahner & Yu, 2013). Such diagrams are meant to spur a wide range of conclusions, inferences, and hypotheses. Designers of displays are faced with many decisions for portraying the data, and those choices affect the kinds of inferences that viewers make. In particular, data points can be connected by lines or enclosed in boxes. Lines suggest relationships and links whereas boxes contain and suggest contrasts with other boxes.

Here, information about movements of people in place and time were organized with lines or boxes, corresponding to two common diagrammatic formats, line graphs and tables. Participants were asked to make as many inferences as they could from one of the displays. Overall, participants produced a large number of generalities that linked information that was separated in the data, showing that they did attempt to integrate the information. In general, *People* was the dominant organizer of inferences. As predicted, the two spatial organizations of data, lines and boxes, had dramatic effects on the kinds of inferences drawn from the data, movements of people in space and time. Lines connected people over space and time. Although *People* was the dominant organizer in both cases, *People* was far more dominant when lines connected each person's movements over time, and *Time* was the dominant secondary organizer. Lines also encouraged more *parallel* inferences, inferences with the same structure and format. These are sets of inferences structured in the same way: *X went to A at time 1, to B at time 2*, etc. With boxes, people dominated as first organizer, but *Place* rather than *Time* dominated as secondary organizer. Boxes also encouraged more statements about single features of the information, more generalities involving many features, more leaps that went far beyond the information given, and more negations, that is, statements about empty cells.

Displays of this information are used for exploration and understanding of the underlying phenomena driving the movements as well as conveying them to others. Visuospatial characteristics of information displays affect the kinds of inferences drawn from the information, factors like position in space, marks such as lines and boxes, and content of the dimensions. People, place, and time are three-dimensional data, and three-dimensional displays are famously difficult to comprehend, biased toward the variables on the axes (e. g., Carpenter & Shah, 1998). Based on previous research (Kessell & Tversky, 2010), we chose the consensus arrangement of the three variables, time on the Y axis, place on the X axis, and people as cell entries. Time and space are fixed dimensions (place was not located dimensionally here, but commonly is, in maps). Only people are movable, perhaps the reason they were selected for the cell entries.

People was by far the most popular organizer for inferences. This is most likely due to that fact that people are agents, for the most part, they decide where to go and when. *People* is also preferred to *Place* or *Time* for organizing both episodic (e. g., Taylor & Tversky, 1997) and autobiographical memory (e. g., Wagenaar, 1986). In both cases, organization of memory is multiple and flexible, but organization by *People* is privileged. Location and time, like people, can be good predictors of activities, but people are agentive, and for that and a variety of other reasons, are better and preferred as organizers of memory.

The display format, line graph or table, affected both quantity and quality of inferences. The different patterns of inferences suggest that tables and line graphs induce different strategies for exploring the data. Those presented with tables seemed to focus on the cells, producing more single statements that described single cells. They noticed when cells had many entries, producing relatively more generalities, such as the crowd at the bookstore at noon. They also noted empty cells, producing negations that observed the absence of people in the bookstore in the morning or the gym at night. By contrast, those presented with lines used the lines to explore the data, focusing on each person's movements in turn across cells. Lines led the eye and the mind from cell to cell; matrices led the eye and the mind to the cells.

Which is better? Like almost everything, it depends. If you are tracking parcels or thieves or spies or consumers or celebrities, then lines will focus you on the important information. On the other hand, if you're entertaining many hypotheses, then use tables. Just be aware that what you choose makes a difference.

Acknowledgments

The authors gratefully acknowledge awards IIS-0725223, HHC-0905417, IIS-0855995 and IIS-0968561 from the National Science Foundation, and the Stanford Regional Visualization and Analysis Center for partial support of the research or preparation of the manuscript.

References

Carpenter, P.A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*, 75-100.

Kessell, A.M. & Tversky, B. (2010). Visualizing space, time, and agents: Production, performance, and preference. *Cognitive Processing, 12*, 43-52.

Nickerson, J. V., Corter, J. E., Tversky, B., Rho, Y-J., Zahner, D., & Yu, L. (2013). Cognitive tools shape thought: diagrams in design. *Cognitive Processing*, published online Feb. 15, 2013.

Taylor, H. A., & Tversky, B. (1997). Indexing events in memory: Evidence for index preferences. *Memory*, *5*, 509-542.

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

Tversky, B, Zacks, J., Lee, P. U., & Heiser, J. (2000). Lines, blobs, crosses, and arrows: Diagrammatic communication with schematic figures. In M. Anderson, P. Cheng, & V. Haarslev (Eds.), *Theory and application of diagrams*. Pp. 221-230. Berlin: Springer.

Wagenaar, W.A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology, 18*, 225-252.

Zacks J & Tversky B. (1999) Bars and lines: a study of graphic communication. *Memory and Cognition*, *27*, 1073–1079.