Evaluating 2D and 3D Visualizations of Spatiotemporal Information

ANDREAS KJELLIN and LARS WINKLER PETTERSSON
Uppsala University
STEFAN SEIPEL
University of Gävle and Uppsala University
and
MATS LIND
Uppsala University

Time-varying geospatial data presents some specific challenges for visualization. Here, we report the results of three experiments aiming at evaluating the relative efficiency of three existing visualization techniques for a class of such data. The class chosen was that of object movement, especially the movements of vehicles in a fictitious landscape. Two different tasks were also chosen. One was to predict where three vehicles will meet in the future given a visualization of their past movement history. The second task was to estimate the order in which four vehicles arrived at a specific place. Our results reveal that previous findings had generalized human perception in these situations and that large differences in user efficiency exist for a given task between different types of visualizations depicting the same data. Furthermore, our results are in line with earlier general findings on the nature of human perception of both object shape and scene changes. Finally, the need for new taxonomies of data and tasks based on results from perception research is discussed.

19

Categories and Subject Descriptors: H.1.2 [**Models and principles**]: User/Machine Systems—*Human factors*; H.5.1 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/methodology*; J.4 [**Social and Behavioral Sciences**]: Psychology

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: 2D, 3D, animation, space–time cube, spatiotemporal, user studies

ACM Reference Format:

Kjellin, A., Pettersson, L. W., Seipel, S., and Lind, M. 2010. Evaluating 2D and 3D visualizations of spatiotemporal information. ACM Trans. Appl. Percept. 7, 3, Article 19 (June 2010), 23 pages.
DOI = 10.1145/1773965.1773970 http://doi.acm.org/10.1145/1773965.1773970

The work presented here is the result of a subproject funded by Swedish National Defence College (SNDC) in the framework of project AQUA.

Authors' addresses: A. Kjellin and M. Lind: Uppsala University, Department of Informatics and Media/Human-Computer Interaction, Box 513, SE-751 20 Uppsala, Sweden; email: kjellin.andreas@gmail.com, mats.lind@dis.uu.se; L. W. Pettersson and S. Seipel: Uppsala University, Department of Information Technology, Box 337, SE-751 05 Uppsala, Sweden; email: mail@larspettersson.se, ss@hig.se.

DOI 10.1145//1773965.1773970 http://doi.acm.org/10.1145//1773965.1773970

1. INTRODUCTION

Although always utilized as tools for navigation, the use of maps in later years has expanded to a variety of other uses, for example, as an aid in the understanding of a complex chain of events not primarily of a geographical nature. An often-cited early example of this is when the English physician John Snow, in 1850, used a map to better understand the nature of a cholera epidemic in London. What he did was to mark on a map the homes of infected people. In viewing these marks, he realized that water pumps could be central to the spreading of the disease. For a more detailed discussion on Snow and this episode, see Tufte [1983]. Another famous use of map-based visualizations is, of course, Minard's map of Napoleon's Russian campaign in 1815. In this visualization, several different data dimensions are simultaneously shown such as time, number of troops, and temperature [Tufte 2006]. The field of geovisualization is thus not new, but the foundations for it have radically changed because of the development of computers. A visualization that used to take hours to produce can now be made in real time, and there seems to be few limits to its complexity. Thus, we now have the potential of visualizing spatiotemporal data in many different forms and the graphical power of today's computer makes all of them perfectly viable. The question then arises if there is a difference in usability between these techniques and, if this is the case, which one is optimal for a decision maker given a set of circumstances?

The time dimension is usually important in geovisualization, and as such, it needs to be represented in the map. Two obvious choices when visualizing time-dependent events is either to label important moving objects with time stamps (this being a common technique in command and control applications) or to use animation. Another idea is to map data in three dimensions. If the direction orthogonal to the map is considered to be time [Hägerstrand 1970], something called the *space–time cube* arises (see e.g., Andrienko et al. [2003a]).

The combined effect of spatial and temporal dimensions is especially important in command and control situations, for instance military decision making, fire fighting and air traffic control. The tasks for a decision maker in these kinds of command and control situations vary, but in this article, we will look at two different kinds of spatiotemporal tasks that a decision maker may face: One is to predict where vehicles will meet in the future, the other is to decide in which order four vehicles arrive at a meeting point.

In this article, we report three different experiments as a start of an investigation of the usability of different visualizations of spatiotemporal data and tasks. In Section 2, we give some theoretical background on recent findings in human visual perception, and on different approaches on how to classify a visualization. In Section 3, the three experiments are presented. Finally, there is a general discussion of the findings from these experiments.

2. THEORETICAL BACKGROUND

2.1 Visualization Types

In spatiotemporal (geo-) visualization, three main techniques have evolved over the years.

2.1.1 *2D Maps.* In the area of military command and control, the use of maps is extensive. Until now, the primary artifacts were ordinary paper maps with clear plastic overlays where additional information could be drawn. Time on these maps is often shown by explicitly noting time stamps. By the introduction of computer-aided support to the decision makers, the possibility to evolve new and possibly more effective types of visualizations to annotate time has arisen. However, even today, traditionally used ways of marking time are often used. To visualize the movement of a vehicle during a time period, for example, two common ways of depicting time exists. One is to write down the actual time when the vehicle passed different landmarks. The other is to use cross-lines, orthogonal to the

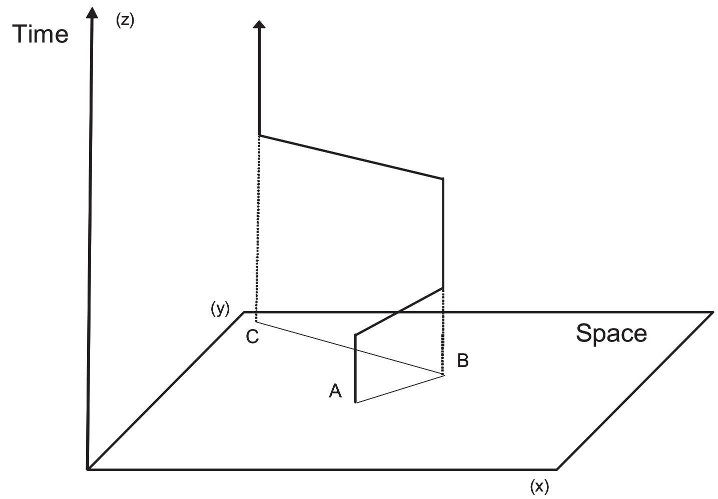

Fig. 1. In a space–time cube, time is visualized orthogonally to the map surface. In this figure a vehicle has been moving from A to B and then to C. At A, the vehicle stood still for some time and then moved to B. At B, the vehicle again stood still before moving to point C. The slope of the line during movement is an indication of the speed of which the vehicle moved. The steeper the slope of the line, the slower the vehicle moved and vice versa.

trajectory that the vehicle has followed. The distance between these cross-lines represents a fixed time interval. Such cross-lines allow a decision maker to quickly judge the speed of motion. The closer the cross-lines are to each other, the slower the vehicle is moving and vice versa.

2.1.2 *Animations.* Another way to show a movement is to use animation, a technique used in a variety of fields, for instance in geovisualization (e.g., Andrienko et al. [2000], Harrower [2002], Lightner [2001] and Wiegmann et al. [2005]). Often, the primary goal of these animations is not to show the motion of an individual or a group of objects, but rather to visualize other aspects of a changing phenomenon. An exception is the work reported by Andrienko et al. [2000], where animation is used to visualize movements of individual migrating storks.

What is largely missing in the work using animations, so far, are evaluations of their effectiveness and efficiency in relation to human decision makers. Some basic aspects for simple tasks have of course been studied by experimental psychologists (e.g., Schipper and Versace [1956] and Tresilian [1995]), and recently Lind and Kjellin [2005] have reported an experiment where the goal was to investigate the optimal speed of an animation, but much remains to be done. A related area of interest is the work on multiple object tracking (e.g., Alvarez et al. [2005], Scholl [2001] and Yantis [1992]). A general conclusion is that there exists an upper limit to the number of moving objects a human observer can track with accuracy. Evidence suggests that five moving objects is close to that limit [Pylyshyn and Storm 1988].

2.1.3 *Space–Time Cubes.* A third way of visualizing space and time in geovisualizations is the use of a space–time cube. In a space–time cube, the time dimension is orthogonal to the surface of the map, along an axis that points up from the map surface. One of the origins of this idea can be found in an article by Hägerstrand [1970] (Figure 1). In his article, Hägerstrand discusses how to track people's movements during their lives and how they move to work and to other locations. In recent years, other scientists have also used this technique (see ITC-Minard [2007] for an example of a space–time cube visualization of Napoleon's Russian campaign). Further examples of the use of the space–time cube can be found in Andrienko et al. [2003, 2003a] and Kapler and Wright [2004]. The space–time cube idea has an intuitive appeal. The time of an event can be directly shown by spatial position without resorting to the use of graphical expressions such as labels, color, line width, and so on. This means that such graphical expressions are left to be used for other important characteristics of the visualized events, and this is a fundamental difference between the space–time cube and any 2D visualization. Thus, the space–time cube offers the potential of showing more data directly coded to unique graphical expressions than any 2D visualization. This, of course, presupposes that humans can reliably perceive all aspects of 3D structure, something that unfortunately must be questioned based on the existing results from studies on human perception.

2.2 Human Visual Perception of a 3D Environment

Klein [1939], a German mathematician, proposed a hierarchy among geometries. He suggested that as long as structures remain invariant (are preserved) over different types of transformations, the structure belongs to a particular geometry. This inspired Todd and coworkers [Todd and Norman 2003; Todd et al. 2001] to use this hierarchy to investigate human perception of 3D structure. Although some controversy exists, the vast majority of studies clearly point to a common conclusion: Tasks that human observers can perform well are tasks that can be solved by estimating properties of a viewed scene that are invariant under affine transformations, or lower, in the Klein hierarchy. Tasks that require human observers to estimate properties of a viewed scene that are invariant only under Euclidean or similarity transformations cannot be performed without large errors (e.g., see Forsell [2007], Lind et al. [2003], Norman et al. [1996], Suppes [1977], Tittle et al. [1995], Todd and Norman [2003], Todd et al. [2001], Todd and Bressan [1990] and Todd et al. [1998]).

Derived from these results, a prediction about the performance in different tasks in 3D visualization environments can be drawn. Tasks that require the user to perceive aspects of a visualization that are invariant only under Euclidean or similarity transformations should be substantially more difficult to solve in a 3D visualization compared to in a task-relevant frontoparallel 2D visualization.

In the case of a space–time cube, this means that properties such as absolute speed of a moving object should be difficult to perceive because speed is reflected by the angle of the object's trace line. Also, relative speed between objects should be difficult to perceive apart from the special case where objects are traveling in parallel paths in two spatial dimensions. On the other hand, there are a number of properties that should be easy to perceive, such as the order of arrival of objects at a certain point in space or time for instance.

2.3 Constraints on a Visualization Due to Tasks and Data

As illustrated by the earlier discussion on metric and affine aspects of 3D structure, it is important to know the task faced by a user when designing a visualization; if the task requires the viewer to use quantitative information, for instance, it seems reasonable to assume that, in the general case, only 2D visualizations can be used. Like other researchers, for instance, Andrienko et al. [2003b], Koua et al. [2006], and Peuquet [1994], but for slightly different reasons, we, therefore, find it clear that tasks are constraining the set of visualizations that will be effective and efficient in any usage situation. Likewise, the data available will constrain the set of visualizations that will be effective and efficient.

Obviously, the data type will do this [Mackinlay 1999], but a slightly more intriguing aspect is that the spatial and temporal characteristics of the data-generating process most probably will affect this as well [Yattaw 1999].

2.3.1 *Data Classification.* For any geospatial system, be it a GIS-system for road planning or a command and control center for the military, the underlying need is the same: to show geographic movement. According to Yattaw [1999], geographic movement can be subdivided into 12 different categories. Two dimensions, temporal and spatial, form the basis of this categorization.

The temporal dimension has three descriptors: continuous, cyclical, and intermittent. Any movement is a part of these three categories; for example, if motion is uninterrupted throughout time, it can be categorized as continuous. Otherwise, if the motion is regular and it is possible to predict it, it can be categorized as a cyclic movement. The last category is called intermittent; if motion is sporadic or unpredictable.

The spatial characteristics of geographical motions are point, linear, areal, and volumetric. The point characteristics of a movement can be understood as sampling of the movement. Yattaw [1999] gives the example of the tracking of an animal, where you do not need to have a continuous tracking of it. It is sufficient to sample the movement at intervals to be able to grasp the situation. Linear motion is following a continuous track, like traveling along a road. Areal and volumetric motion takes place when the objects either spread across an area, like the spreading of a disease over a geographic area, or when spreading in a volume, like spreading in water or in air. By combining these spatial and geographical characteristics, a matrix is formed containing 12 different classes of movements in geospatial systems.

This taxonomy offered by Yattaw [1999] has some problems, we believe. For instance, as we interpret them, the different categories are not mutually exclusive. Furthermore, a phenomenon like an earthquake can be described in many different ways. For example, you can consider an earthquake an intermittent point if you are interested in where earthquakes erupted during the 20th century. Or, you can see it as a continuous volume, a shock wave spreading in the earth's crust. The classification of the data thus does not seem to be independent of how the data will be used or interpreted.

However, the taxonomy offered by Yattaw [1999] is sufficient for describing the data used in the experiments reported here in that these experiments investigate the case of vehicle movement for a command and control situation. This kind of data can simply be categorized as continuous point motion in Yattaw's taxonomy. In the following, we, therefore, adopt this description.

2.3.2 *Task Classification.* When the characteristics of the underlying geographical data are defined, the next step in the process of understanding geospatial visualization is to apply a taxonomy describing the different tasks a user can perform on the data. A number of different approaches have been suggested in the literature (see e.g., Andrienko et al. [2003a], Koua et al. [2006], Ogao and Kraak [2002], Peuquet [1994] and Wehrend and Lewis [1990]).

Koua et al. [2006] have, based on Wehrend and Lewis [1990], made a list of the different tasks and operations a user of a geographical system can perform. These operations are as follows.

Identify. Characteristics of an object.

Locate. Absolute or relative position.

Distinguish. Recognize as the same or different.

Categorize. Classify according to some property (e.g., color, position, or shape).

Cluster. Group same or related objects together.

Distribution. Describe the overall pattern.

Rank. Order objects of like types.

Compare. Evaluate different objects with each other.

Associate. Join in a relationship.

Correlate. A direct connection.

Another proposition of task classification comes from Peuquet [1994], where she introduces a typology for spatiotemporal data. In this typology, three questions can be asked: *where* (space), *when* (time) and *what* (objects). This has later been further developed by Andrienko et al. [2003a], where Peuquet's triad of questions about spatiotemporal data is extended.

—*when* + *where* ⇒ *what*: State the properties of an object or objects at a certain time, or set of times, and a certain place, or set of places.

—*when* + *what* ⇒ *where*: State the location or set of locations.

—*where* + *what* ⇒ *when*: State the time or set of times.

This also corresponds to Bertin's [1983] approach on how data can be analyzed. He also suggests that there are two concepts, *question types* and *reading levels*. Question types refer to the values of variables in the data. For example, data about fever in a patient has two questions, temperature and time: "How high was the fever at a particular time?" and "When did the patient recover?" The reading level for each question is whether it refers to a single data element, a group of elements, or to the whole phenomenon.

In the following, we will adopt this approach stemming from Peuquet [1994], Bertin [1983], and Andrienko et al. [2003a], although we generally believe it, like Yattaw's taxonomy of data, needs to be refined. For instance, the findings on human perception of 3D structures need to be incorporated. This, in turn, means that a task analysis needs to start with the question of whether or not the answer to a task can be based on visualized information that is invariant under Euclidean or similarity transformations. If it is, the visualization most probably needs to be different.

However, in the following experiments, we use the classifications suggested earlier in the text. In all three experiments, the spatial data are points and the temporal dimension is a continuous dimension, according to Peuquet's classification. The tasks the participants in the experiments need to solve are, according to Andrienko et al., the question of "where" in the first two experiments. In the third experiment, the task is "when."

2.4 User Studies

In the world of visualization, many interesting new techniques frequently emerge and there is an increasing need to evaluate the usability of these visualizations. For discussions on this, see Andrienko and Andrienko [2003], Chung et al. [2005], Forsell [2007], House et al. [2006], Kosara et al. [2003] and Kulyk et al. [2007].

When conducting a user study, the goal for the study is to measure the suitability of the visualization in some sense. What is actually measured is a fundamental question that we believe can be handled by using the concepts of effectiveness, efficiency, and satisfaction. These three concepts are derived from the ISO standard of usability 9241-11, [ISO 9241-11 1998, 2].

> Extent to which a product can be used by specified users to achieve specified goals with *effectiveness*, *efficiency*, and *satisfaction* in a specified context of use.

In the ISO standard, the first of the three concepts, effectiveness, relates to the possibility of a user accurately solving the task. The efficiency refers to the cost involved in solving the task. The cost can be measured in different ways, but it is usually thought of in terms of the time required to solve the task. The last concept, satisfaction, is the user's own subjective experience of the task.

2.5 Research Questions and Hypotheses

Based on the previously theories mentioned the research questions and hypotheses investigated in these three experiments are as follows.

Fig. 2. The experimental environment. On the right is the view of the screen, which measures 0.8 × 0.8m. The screen is rear projected with two projectors each having a polarized filter attached. By wearing polarized eyewear, a disparity between the pictures in the two eyes occurs, and together with the head-tracking system, a correct stereoscopic and perspective view is shown to a viewer.

H1. Given that the task requires the perceiving of structure in the visualization that is invariant only under Euclidean or similarity transformations, traditional 2D visualizations, with cross-lines representing time, will lead to a performance that is either more effective, more efficient or both, compared to when the same data is presented in the form of a space–time cube.

—This follows from the discussions in the theoretical background described earlier.

H2. Given that the task only requires the perceiving of affine structures in the visualization, the space–time cube representation will lead to a performance that is at least as effective and efficient as compared to when the same data are presented in the form of a traditional 2D visualization, with cross-lines representing time.

—The simultaneous presentation of spatial and temporal information in the space–time cube could even be beneficial to a decision maker in comparison to the more indirect representation with cross-lines in this case.

H3. Regardless of the requirements of the task, 2D animations and the traditional 2D visualization will lead to a comparable performance given that the number of tracked objects is less than five.

—There is simply no hard data to suggest a difference.

3. EMPIRICAL STUDIES

3.1 General Method

3.1.1 *Hardware and Software*. The display environment used in the experiments has previously been described in Pettersson et al. [2004]. In short, the display environment uses rear projection capable of displaying eight images on top of each other in the same display area. To accomplish this, four pairs of DLP projectors are positioned below each side of a quadratic display surface, each projector pair pointing at approximately 45 degrees toward the surface. To each projector pair, polarized filters are applied to match those of standard polarizing eyewear. The display surface is designed to allow light to penetrate primarily along the line of sight without losing polarization. A viewer wearing standard polarized eyewear located to one side of the table, therefore, perceives a correct stereoscopic view from two of the eight projection images (Figure 2).

The display surface is a horizontal square screen area of 0.8 × 0.8m. The visible pixel resolution of each projection image is 768 × 768 pixels. The eight projectors can provide four head-coupled stereoscopic views using position information from four Flock-of-Bird sensors. Four commercial off-the-shelf computers equipped with NVIDIA Quadro FX 500 graphics cards are connected using a Gigabit Ethernet local area network. Each computer renders one of the four stereoscopic views of the display environment. The software environment used to render the visualization is based on OpenGL and is described in Pettersson et al. [2004].

In the experiments, one of the four projector pairs was used and the other three were switched off. The projector pair provided a stereoscopic head-coupled view using a Flock-of-Birds sensor and a pair of polarizing eyewear. The approximate distance and angle from the observer's eyes to the center of the display surface was 1.5m at 45 degrees.

3.1.2 *Observers.* In these three experiments, all of the participants were either students or staff at the Royal Institute of Technology (Kungliga Tekniska Högskolan, KTH), or at the Swedish National Defence College, SNDC, (Försvarshögskolan, FHS). All of the observers declared that they had normal, or corrected to normal, vision and normal color vision.

They all received a movie ticket as a small compensation for participating in the experiment. None of the observers had any prior knowledge of the purpose of the experiments or of the specific hypotheses employed.

3.1.3 *Procedure.* Upon arrival at the lab, the participants were first asked to read and sign an informed consent, and then they received their compensation. Some background information, such as age, color vision, and if they had normal or corrected to normal vision, was then collected. The observers were then seated in front of the screen and were given both written and oral instructions for the experiment in question. The observers were told that accuracy was the most important but time was also crucial. After the experiment, the observers were able to ask questions about the experiment and its purpose.

In all three experiments, the trials were self-paced, meaning that the observers could decide for themselves the tempo between the trials. This was done to relieve some stress on the observers.

3.2 Experiment 1

In the first experiment, hypotheses H1 and H3 were tested. Hypothesis H1, which states that if the task requires perceiving structures that are invariant under Euclidean or similarity transformations, traditional 2D visualizations should lead to a better performance than a 3D visualization in the form of a space–time cube. H3 states that animation should lead to the same performance as that of a 2D visualization. The layout of this first experiment was to investigate how different ways of visualizing movement on a map affects the ability to predict a future state.

3.2.1 *Task and Stimuli.* What the subjects saw in each trial was a stylized map. In the map, or in the space above the map, the movements of three vehicles were visualized. The lines for each vehicle was colored (blue, green, and orange) to distinguish them from each other. Each visualization condition illustrated the vehicles' movements during 5.5 hours. The participants' task were to predict where the vehicles would meet 4.5 hours into the future, given that they would move with exactly the same speed characteristics as in the previous 5.5 hours. In the instructions to the participants, they were told that they should extrapolate from the history of the movements to the future point where the vehicles would meet.

The three visualizations used are as follows (Figures 3 through 6).

2D. Here, the movements of the vehicles were drawn as lines on the map surface and on every 20th minute a cross-line was drawn. This resulted in 25 visible cross-lines. By just looking at the shape of the trail of the object, the subjects could see what speed the different vehicles were doing at different times. The closer the cross-lines were to each other, the slower the speed of the vehicle and vice versa.

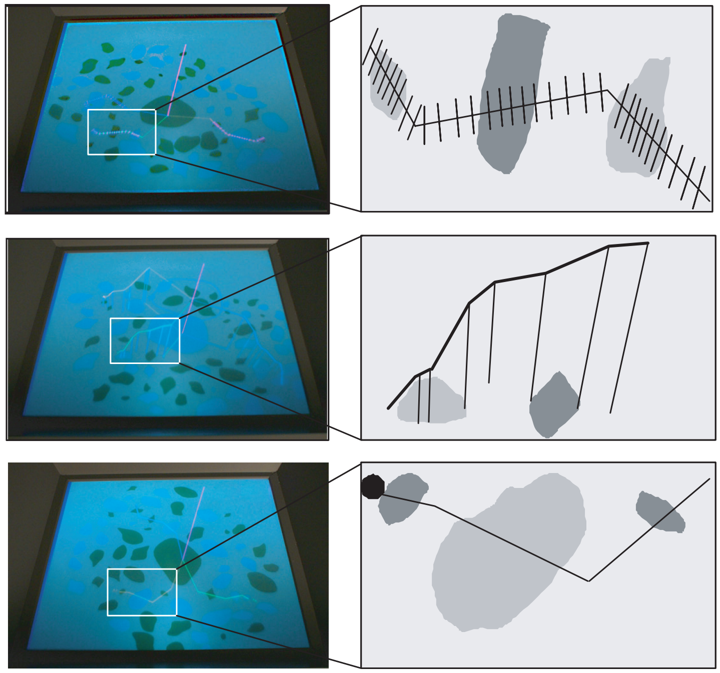

Fig. 3. The three different visualization conditions in experiment one. On top is the 2D static visualization, in the middle is the 3D, and below, the animation. On the right, a blow-up shows a simplified drawing of the path of one of the vehicles in each visualization. Each of the paths of the three vehicles was colored in a different color (green, blue, and orange). The three different colored areas on the map represent different surfaces in the terrain, with each of these surfaces affecting the three vehicles differently. The surface pattern was randomized for each task.

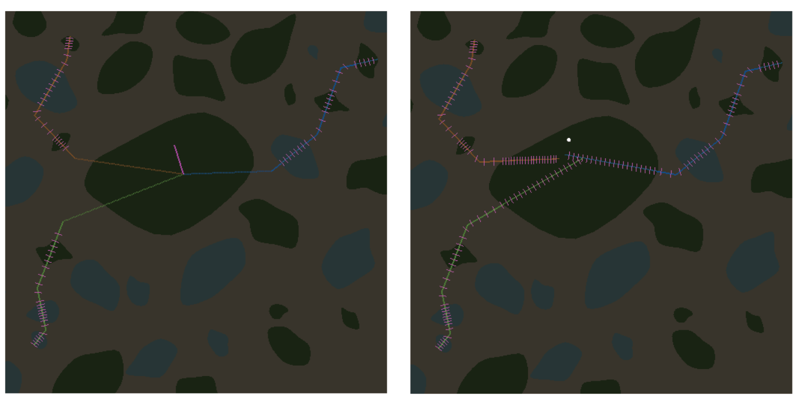

Fig. 4. The static 2D visualization from experiment 1. The left picture shows the stimuli as the participant saw it before he/she gave an answer regarding the meeting point. The vertical line was controlled by the participant. The right picture shows the feedback he/she received, the white point is the correct meeting point, with lines drawn in the direction of the point the participant thought was the correct meeting point.

3D. Here, we used the third dimension, orthogonal to the table surface, as a time dimension. Speed is, therefore, indicated by the slope of the line, a steeper slope shows a slower speed and vice versa (see Figure 1).

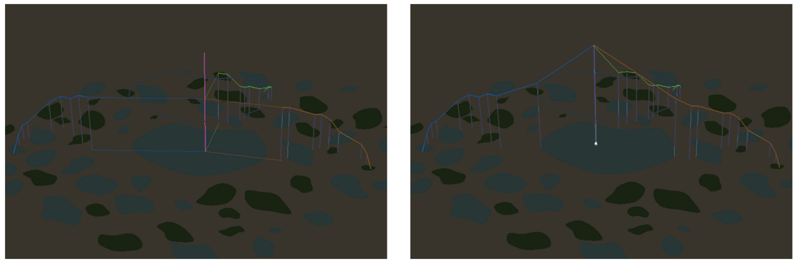

Fig. 5. The space–time cube visualization from Experiment 1. The left picture shows the stimuli as the participant saw it before he/she gave an answer regarding the meeting point. The vertical line was controlled by the participant. The right picture shows the feedback he/she received, the white point is the correct meeting point, with lines drawn in the direction of the point the participant thought was the correct meeting point.

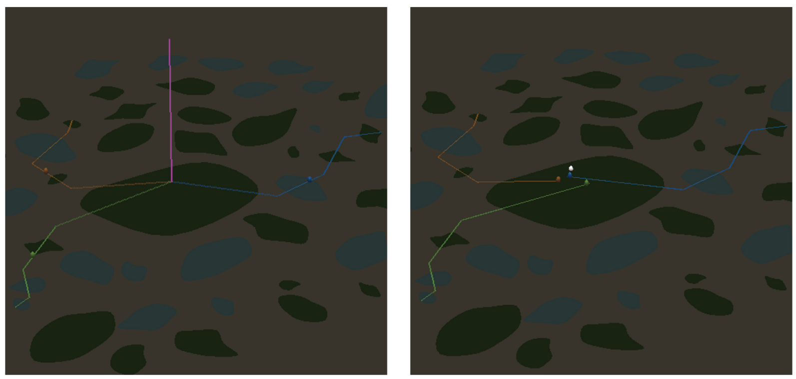

Fig. 6. The animation visualization from Experiment 1. The left picture shows the stimuli as the participant saw it before he/she gave an answer regarding the meeting point. The vertical line was controlled by the participant. The right picture shows the feedback he/she received, the white point is the correct meeting point, with the movements of the objects drawn in the direction of the point the participant thought was the correct meeting point.

Animation. In this visualization, the different speeds of the vehicles were shown directly by a moving sphere along the path in the map along which the vehicles moved. Based on the findings in Lind and Kjellin [2005], the mean animation speed was set to approximately 13 degrees of visual angle per second. The actual mean speed, in degrees of visual angle per second, of the animation varied because the viewpoint of the participants was not constrained.

Care was taken while constructing the stimuli not to produce any easily detectable motion path lengths or starting position that could be used to solve the task.

Task. The goal given to the subjects in this experiment was to determine the only possible simultaneous meeting point of the vehicles, given that they would use the same speed characteristics as previously. This meeting point was always 4.5 hours into the future from the instant the reply period of the experiment commenced. The map contained three different surface types indicated by three different colors. These different kinds of surfaces influenced the vehicles in different ways; however, each surface type always influenced each vehicle in the same manner throughout the experiment. These properties of the task are not as artificial as might seem initially. For example, a big truck cannot move as fast as a tracked vehicle can on marshland. On the other hand, the truck moves faster than

Table I. The Different Speeds for the Different Vehicles on the Different Surfaces of the Map

Vehicle type	Normal surface	Surface one	Surface two
1	55	16.5	33
2	60	36	18
3	65	29.25	29.25

The speed is in kilometers per hour.

a tracked vehicle on other surfaces. The vehicles' default speeds were 55, 60, and 65km per hour, respectively. Each of the three vehicles was affected differently by the different surfaces (Table I).

Stimuli. The motions of the vehicles up to the "present" time was shown according to the condition in question. The vehicles were at this "present" time positioned close to the edge of a large, central surface patch somewhere in which the optimal meeting point was located. The color of the central surface patch followed the convention of the whole experiment, that is, it was indicative of the speed each type of vehicle would have when traveling over it. This means that the observers knew that the vehicles would have the same speed during the whole period of motion prediction. In the instructions, the participants were also told that there was no change of direction allowed for a vehicle once it had started its motion over this central surface patch. This meant that the vehicles always would travel in straight lines, but not necessarily in the same direction as the most recent line segment. The total size of the map was 800 $\times$ 800km and the size and position of the different surface segments were randomly determined for each trial.

All the participants were shown the same 72 stimuli (6 pre- and 6 posttrials and 60 learning trials). The order of trials within each of the three blocks (pre, learning, and post) was completely randomized.

Dependent measure. The dependent variable used was the absolute distance between the position indicated by the subject and the correct optimal meeting point calculated by the program.

3.2.2 *Experimental Design.* The experiment used a mixed design with one between-subject factor "visualization type" with three levels (2D, 3D, and animation) and one within-subject factor "learning" with two levels (pretest and posttest).

The layout of this experiment was a learning experiment with two blocks of trials (six pretest and six posttest) administered without feedback and a block of 60 trials with feedback in between. The pretest and posttest trials were identical but administered in a randomized order in each of block of trials, respectively.

3.2.3 *Procedure.* The experiment was self-paced, that is, the participants could themselves decide when a trial would begin by pressing the space bar. By pressing it, the trial began and the visualization appeared. The type of the visualization was determined by condition of the subject. To give an answer, the participant clicked on the map by using the mouse. No practice trials were administered. In the pre- and posttrials, the screen went blank after each mouse click and the system was in a waiting state, ready to start a new trial. In the learning trials, the correct answer was indicated by a white marker in the map. The participant's response was used to generate the motion paths that would have occurred should the vehicles have passed though the position indicated by the participant's response. These paths were subsequently visualized in the same manner as the stimuli, that is, according to the condition the participant was in. In this manner, each participant had a chance to learn the relationship between the different surfaces and the vehicles. For each task, the answer time was limited to 20 seconds, after this, a magenta circle appeared, and the participant was forced to answer.

3.2.4 *Observers.* In this experiment, 30 participants took part, 15 male and 15 female. Their age varied between 21 and 37 years, with a median age of 24.

Table II. Descriptive Statistics from Experiment 1

Visualization	Mean	Std.Err.	−95.00%	+95.00%	N
2D pre	42.43	4.12	33.97	50.90	10
2D post	23.91	4.07	15.56	32.27	10
3D pre	57.93	4.12	49.46	66.39	10
3D post	42.47	4.07	34.11	50.83	10
Anim pre	54.70	4.12	46.24	63.17	10
Anim post	50.51	4.07	42.16	58.87	10

The numbers represent distances in millimeters on the table surface between the marked meeting point and the correct meeting point. "Pre" measures are the first six trials the participants performed, and "post" are the same trials, performed last in the experiment.

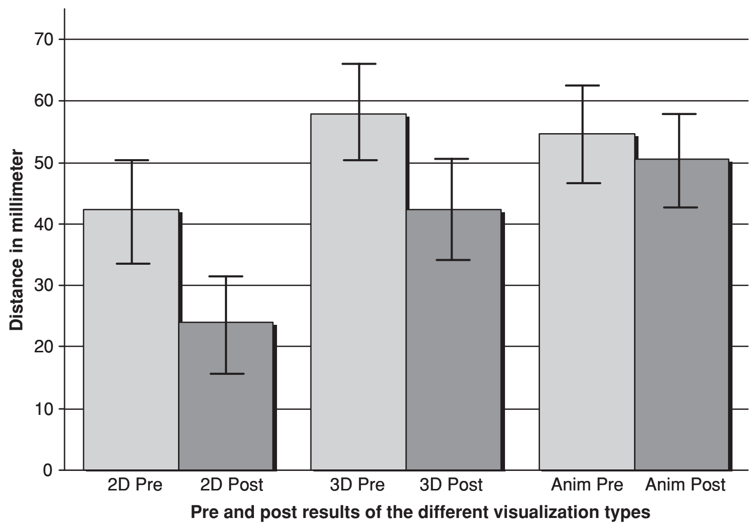

Fig. 7. Results from Experiment 1. The distances between where the participant thought the vehicles would meet and the correct position. Vertical bars denote the 0.95 confidence interval.

3.2.5 *Results.* The performance of the participants was measured by the absolute distance from the point where they had pointed out as the meeting point and the correct point. This was calculated in millimeters on the display surface. See Table II and Figure 7 for the descriptive data. The maximum possible error in each trial was always larger than 100mm and the maximum error recorded for any subject was 81.4mm. Thus, no ceiling effects were observed.

The distance data was analyzed by means of an ANOVA using a decision criterion of $p < 0.05$. The results from the ANOVA shows that there was a significant effect of visualization type, ($F(2, 27) = 8.7303$, $p < 0.01$). There was also a significant main effect of learning ($F(1, 27) = 30.7505$, $p < 0.001$). An interaction effect between learning and visualization type was also present, ($F(2, 27) = 3.6063$, $p < 0.05$) The effect of visualization type (i.e., 2D, 3D, and animation) was further analyzed by a post hoc test, the Tukey HSD test (see Table III). The result from the post hoc test revealed that the 2D condition led to a significantly better performance compared to the 3D and animation visualizations. There was no significant difference between the animation and 3D visualization.

Table III. The Results from the Tukey HSD Post Hoc Test on the Effect Type of Visualization

Visualization	2D	3D
3D	*0.01*	
anim	*0*	0.88

Italics indicates significance at 95%.

Table IV. The Results from the Tukey HSD post Hoc Test on the Interaction Effect Between Type of Visualization and Pre/Post Trials

Visualization	2D pre	2D post	3D pre	3D post	Anim pre
2D post	*0.00*				
3D pre	0.29	*0.00*			
3D post	1.00	0.14	*0.01*		
Anim pre	0.54	*0.00*	1.00	0.30	
Anim post	0.73	*0.01*	0.79	0.87	0.89

Italics indicates significance at 0.05.

The interaction effect between learning and visualization type was also analyzed by a post hoc test using the Tukey HSD test. The results revealed that there was no difference between the three visualization types in the pretest phase. In the posttest phase, subjects performed significantly better in the 2D and 3D visualizations compared with the pretest results. Subjects in the animation visualization, however, did not show any significant effects of learning. See Table IV for the full analysis.

3.2.6 *Discussion.* In this experiment, hypotheses H1 and H3 were tested. Hypothesis H1 states that if the task requires perceiving aspects of structure invariant only under Euclidean and similarity transformations, a traditional 2D visualization should lead to a performance that is better compared to space–time cube visualizations. Hypothesis H3 states that regardless of the task, a traditional 2D visualization should lead to a comparable performance when compared to a visualization using animation. The main result from this experiment is that, for our task and our subjects, the 2D visualization led to a better performance than the 3D visualization. Therefore, hypothesis H1 cannot be rejected and thus receives additional support. Hypothesis H3 must be, for our subjects and on the basis of our data, rejected. The performance of our subjects for the animation was significantly worse than for both the 2D and 3D visualizations.

The main effect of learning indicates that the participants did learn the relevant aspects of the task and were able to utilize the feedback they were given after each trial. The posthoc test of the interaction effect between display type and learning revealed that with the animation, no learning took place between pre- and posttrials, but there was a significant difference between pre- and posttrials in both 2D and 3D visualizations. This means that in the animation test, the participants did not learn the task at all. However, in both the 2D and 3D visualizations, the participants did show improvement. Although there was no main difference in performance between the 3D visualization and the animation, the fact that there was a learning effect in the 3D visualization warrants further investigation.

The poor performance in the animation test may be due to the fact that the screen used in this experiment resulted in an approximate visual angle of 30 degrees. (This value is based on the general data from the head tracker in a similar experiment using the same experimental set-up and apparatus, which revealed the minimum distance to the screen by any observer was 0.8m and the maximum distance was 1.2m). The large visual angle could be a problem because high accuracy vision is restricted to the central area of the retina, the fovea, (e.g., Ware [2000]). The consequence of this is that

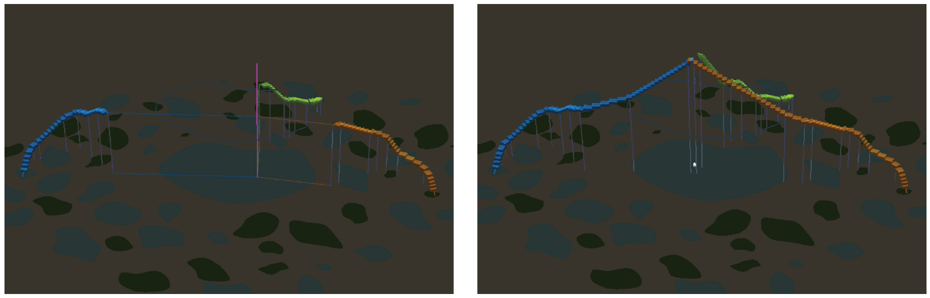

Fig. 8. The space–time cube visualization from Experiment 2. The improved 3D visualization used in Experiment 2. In the space–time cube task, a redesigned visualization of the line was used. Instead of a line, boxes were used forming a staircase. The left picture shows the stimuli as the participant saw it before he/she gave an answer regarding the meeting point. The vertical line was controlled by the participant. The right picture shows the feedback he/she received, the white point is the correct meeting point and the lines are drawn in the direction of the point the participant thought was the correct meeting point.

the amount of information that can be encoded in high quality in one glance is limited. Of course, the peripheral retina is extremely sensitive to motion (e.g., Brown [1972] and Paillard and Amblard [1985]), but estimating speed is not enough for this task. The speed must be related to the type of area the object moves over at a specific instance in time. To determine this, our subjects needed to fixate on the object. In order to solve the task in this experiment, the participants were forced to make a substantial amount of saccades when trying to compare one object with another and at the same time the visualization changed between saccades. This may be a disturbing factor because of the difficulties for observers to detect even large changes in a scene that occur during a visual disruption, like when a saccade is made. This phenomenon is know as change blindness (see e.g., Simons [2000], Simons and Chabris [1999], and Simons and Rensink [2005].

3.3 Experiment 2

The results from the first experiment indicated that the traditional 2D visualization was superior to our version of the space–time cube in this kind of task. Because the space–time cube offers the potential of directly depicting additional information compared to a 2D visualization, we asked ourselves whether there were any optimizations we could implement that could make it more useful for our class of spatiotemporal tasks. At least two things became apparent. One is that a single line in 3D is not a very salient 3D shape. For instance, according to Pizlo [1994] and Pizlo and Stevenson [1999], a 3D form is harder to perceive if the shape is formed of lines or vertices than if the shape is formed of polyhedrons. It is, therefore, possible that a redesign of the space–time cube visualization used could improve performance. Another important aspect is that by inducing a rotation of the visualization around a vertical axis, the metric properties of shape can be more accurately perceived [Bingham and Lind 2008]. Based on these findings, we redesigned our version of the space–time cube and evaluated it, like before, against the traditional 2D visualization. This time, no animations were included, since the results from Experiment 1 were reasonably clear in this respect.

3.3.1 *Task and Stimuli.* A new way of rendering the 3D visualization was developed. Instead of using a line showing the position of a vehicle in the combined space–time volume, a staircase of boxes was used. By using boxes, the criteria for easy 3D perception of shape pointed out by Pizlo [1994] and Pizlo and Stevenson [1999] were fulfilled. A depiction of these renderings is shown in Figure 8. Furthermore, when a trial began and the 3D visualization appeared, the visualization was slowly

Table V. Descriptive Statistics from Experiment 2

Visualization	Mean	Std.Err.	−95.00%	+95.00%	N
2D pre	33.98	3.95	25.69	42.27	10
2D post	25.09	2.88	19.05	31.13	10
3D pre	43.96	3.95	35.67	52.26	10
3D post	32.57	2.88	26.53	38.61	10

The numbers shown represent distances in millimeters on the table surface between the marked meeting point and the correct meeting point.

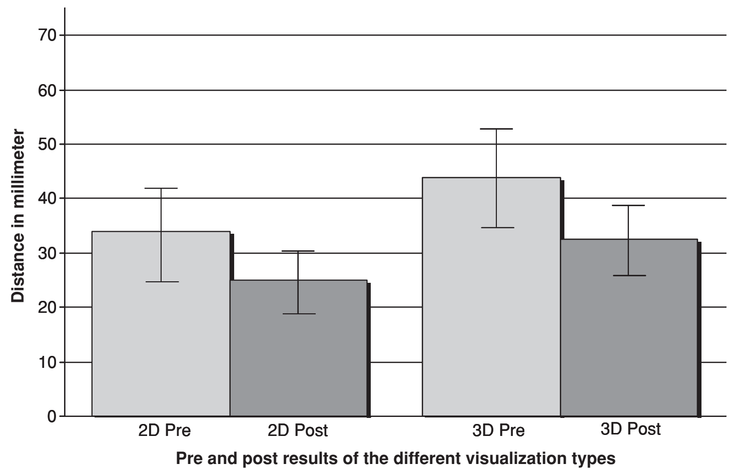

Fig. 9. Results from Experiment 2. The distances between where the participant thought the vehicles would meet and the correct position. Vertical bars denote 0.95 confidence intervals.

rotated back and forth. The amount of rotation was 60 degrees and it was performed around a centrally placed vertical axis. The 2D visualization was the same as in Experiment 1 (see Figure 4).

The task was identical to the one used in Experiment 1.

3.3.2 *Experimental Design.* The experiment used a mixed design with one between-subject factor "visualization type" with two levels (2D and 3D) and one within subject factor "learning" with two levels (pretest and posttest). In all other aspects the experiment was identical to Experiment 1.

3.3.3 *Subjects.* Twenty participants took part in the experiment. They were all male and aged between 20 and 59 years. Their median age was 25 years.

3.3.4 *Results.* As in Experiment 1, the performance of the participants was measured by the absolute distance from the point where they had pointed out as the meeting point and the correct point. This was calculated in millimeters on the display surface (Table V, Figure 9).

An ANOVA was calculated using the error measurements, that is, the distances from the expected point of the participants' marking of the meeting point to the point where they clicked. In the ANOVA, a decision criterion of $p < 0.05$ was used. A significant effect of visualization was found, ($F(1, 18) = 4.54$, $p < 0.05$). There was also a significant effect between the pre- and posttests, ($F(1, 18) = 14.63$, $p < 0.01$).

3.3.5 *Discussion.* Hypothesis H1 states that a 2D visualization should have a better performance than a 3D visualization, when the task requires users to perceive metric structures. Also in this experiment, there is a significant difference between traditional 2D visualization and the space–time cube.

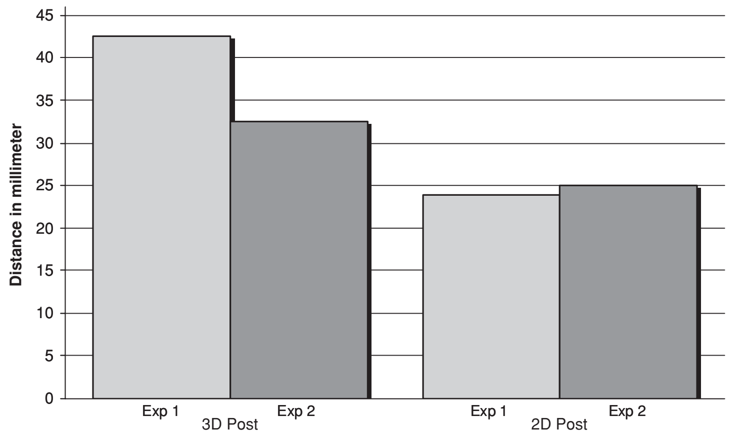

Fig. 10. Comparison between Experiments 1 and 2 of the 2D and 3D post results. For the 2D visualization, there is no major change in performance between Experiments 1 and 2. However, the result for the 3D visualization is 23% lower in Experiment 2 than in Experiment 1. The changes made to the space–time cube in Experiment 2 thus seem to have had an impact on performance.

This was found despite our effort to maximize the visual quality of the 3D visualization. In conclusion, hypothesis H1 cannot be rejected based on these data and receives even more support.

We were also interested if we could find any evidence suggesting that the changes implemented in the 3D visualization improved performance compared to in Experiment 1. As shown in Figure 10, the results in the 2D visualization are very similar in both experiments, indicating that the samples of subjects are roughly comparable with respect to performance in both these experiments. The results from the 3D visualization differ, however. The mean error in the 3D test of the second experiment is 23% lower than in the 3D test in the first. This indicates indeed that the manipulations of the 3D visualization did improve performance in this task, although it still is worse than in the 2D visualization.

3.4 Experiment 3

A clear body of results from the perceptual literature shows that metric judgments of 3D spatial layout are greatly inaccurate. Based on this, it can be predicted that (i) space–time cubes are not a suitable form of visualizing data if the task entails judgments based on properties in the visualization that are only invariant under Euclidean transformations, and (ii) a 2D visualization mapping relevant properties of the task should lead to a better performance compared to when performing the task using a 3D representation such as the space–time cube. The results from the two first experiments also strongly support these predictions. However, if the task is changed to one that can be solved based on spatial properties of the visualization that are invariant under affine transformations, the predictions are changed. In general, the space–time cube should support this task well. Second, whether it does this as well as a 2D visualization is an open question.

To test the first prediction and to measure the comparative performance in the space–time cube and our selected 2D visualization, a new task was constructed. The same basic idea of the two first experiments was kept in that the motions of individually moving vehicles were visualized. The task was changed from predicting an optimal meeting point to determining the order of arrival at a meeting point. In the space–time cube, this task can be solved based on spatial properties that are invariant under affine transformation because the user can look at the ordering of the motion tracks of the different vehicles along a vertical axis passing through the meeting point. Thus, we were predicting that the space–time cube, for this task, should lead to a good performance, that is, few errors even for short exposure times.

The comparison between 3D visualizations and 2D visualizations in general, when no specific predictions from theory can be made as described earlier, is a more difficult business. A wealth of opportunities exist in choosing different ways of visually coding data in both 2D and 3D and a not-too-daring guess is that the choice of visual codes are at least as important as whether to use a 2D or a 3D visualization. Given this, we decided to stick with the 2D visualization we had used in the previous two experiments. It had proven to be superior for the task in the first two experiments and to keep it would lead to at least some comparability over all three experiments, especially since the 3D visualization was kept unchanged as well.

The resolution of our dependent measure in the previous experiments was high and on a ratio scale (millimeters of deviation from optimum). Our task in this experiment leads to a dependent measure of much lower resolution, basically a yes/no answer. However, by increasing the number of vehicles to four and looking at the rank-order correlation between the ordering given by a participant as his or her answer and the correct ordering, a dependent measure having 11 different values on an ordinal scale was created. Another change in the procedure, compared to Experiments 1 and 2, was also introduced by the results from our own exploration of the task and a pilot study. Performance under the two visualizations (2D and 3D) seemed quite different and the combination of a fixed exposure time and a dependent measure with only 11 levels on an ordinal scale was thus far from optimal. A short exposure time would risk a floor effect in the worst case and a longer exposure time would cause a ceiling effect in the best case. Therefore, we decided this time to instruct our subjects to solve the task as quickly and correctly as possible and to measure both task time and errors (i.e., rank-order correlations). Furthermore, in an effort to reduce the task times in the 2D test, the mean number of cross-lines per motion path was reduced from approximately 40 to approximately 15.

3.4.1 *Task and Stimuli.* On the map, four different tracks were shown. Each track was the result of the movement of a vehicle. The task was to tell in which order the four vehicles had arrived at the central meeting point. The movements were visualized in two different ways: 2D-scarce (the same visualization that was used in Experiments 1 and 2 but with 15 cross-lines instead of 40), and in a 3D space–time cube utilizing the optimized track rendering from Experiment 2. As a control condition, the 2D visualization in its original form (i.e., using approximately 40 cross-lines) was also included. However, this condition was always run last for each participant, that is, after the experiment (Figures 11 and 12).

To solve the task in the 2D visualizations, the observer had to estimate the number of cross-lines per motion path. The difference between the different vehicles' times were 20 minutes (i.e., the difference between any two of the vehicles was at least one cross-line). In the space–time cube condition, the task could be solved by looking at the meeting point and give the answer from the lowest to the highest of the vehicles, since the time was represented orthogonally to the surface of the table. In the following, the space–time cube visualization will be referred to as the "3D" condition.

3.4.2 *Experimental Design.* A within-subject design was used in this experiment. The within factor was visualization type with two levels, 2D-scarce and 3D. The factors 2D-scarce and 3D were counterbalanced between subjects. As a control, each subject ran the 2D-dense visualization last in each session.

3.4.3 *Procedure.* The experiment was self-paced. When the participants pressed the space bar, the experiment started. The map appeared with the four tracks, and when a participant thought he/she had the correct answer, he/she pressed the space bar again. By using the keyboard, he/she noted the

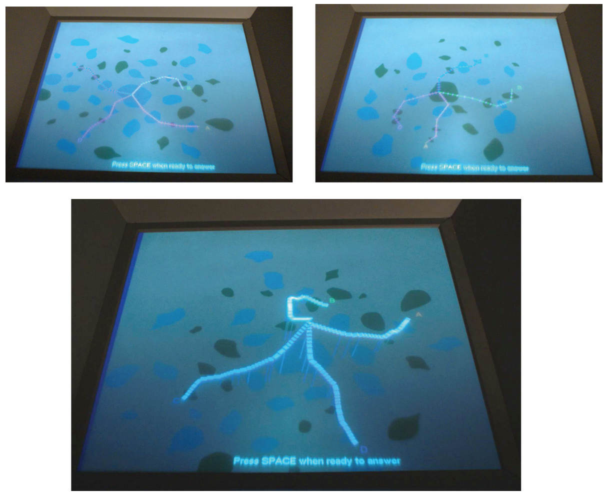

Fig. 11. The layout of Experiment 3. Top left: 2D-dense, the same 2D visualization as used in Experiments 1 and 2. Top right: 2D-scarce, this is a variation of the 2D-dense, but with fewer cross lines. Bottom: Space–time cube, same as in Experiment 2.

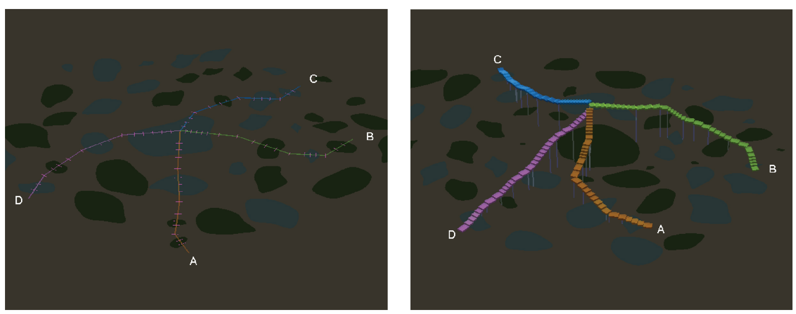

Fig. 12. **Left:** The 2D-scarce visualization from Experiment 3. The task is to give the order of the arrival of the vehicles at the meeting point. The same 2D visualization used in Experiments 1 and 2 but with fewer cross-lines. **Right:** The space–time cube visualization from Experiment 3. The task is to give the order of the arrival vehicles at the meeting point. The same 3D visualization is used here as in Experiment 2.

order that the four vehicles arrived at the meeting point (each vehicle was labeled with a letter). The participants performed 15 trials in each of the three factors. Before each session, three practice tasks were used to familiarize the participants with the task.

3.4.4 *Subjects.* Fourteen participants, five female and nine male, participated in the experiment. Their age ranged between 21 and 54 years, and their median age was 23.

3.4.5 *Results.* The participant's task was to state the order in which the four vehicles had arrived at the meeting point. In order to disregard possible practice effects as much as possible, the results were calculated using the five last of the 15 trials in each test. In each trial, we used Spearman's rho to calculate the correlation between the rank the participant had given to each of the four vehicles, and the correct rank. The median of the five rho values (from the last five trials in each test) for each subject was used in the subsequent comparisons (Table VI and Figure 13).

By analyzing the data by means of a Wilcoxon Signed Rank Test and by using a decision criterion of $p < 0.05$, we found that 3D is significantly different from 2D-scarce, ($Z = 2.31$, $N = 14$, $p < 0.05$).

Table VI. Descriptive Statistics from Experiment 3

Condition	N	Median	Minimum	Maximum	Lower Quartile	Upper Quartile
3D	14	1.00	0.80	1.00	1.00	1.00
2D scarce	14	0.80	0.40	1.00	0.60	1.00

The data refers to medians of Spearman's rho values for the last five trials.

Table VII. Descriptive Statistics of the Response Time Results in Experiment 3

Visualization	Mean	Std.Err.	−95.00%	+95.00%	N
3D	6.35	2.74	5.97	6.73	14
2D Scarce	31.73	13.33	29.87	33.6	14

The data is calculated on the last five trials in each visualization.

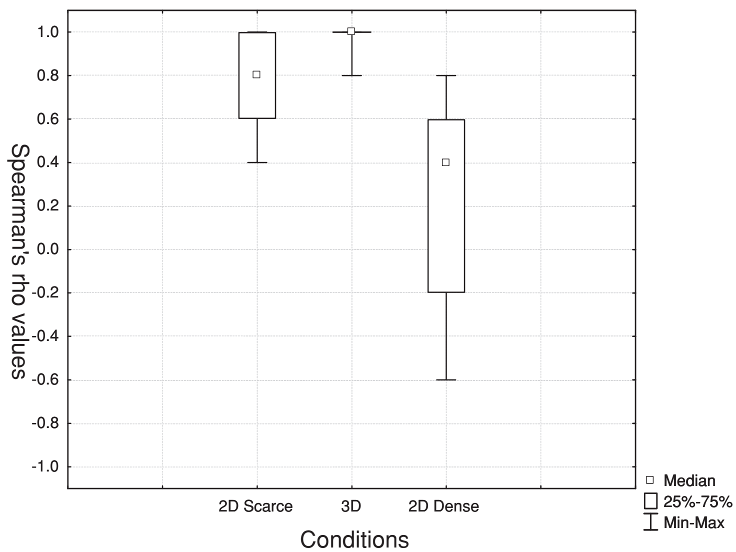

Fig. 13. Data from Experiment 3. The graph shows Spearman's rho data for the three different kinds of visualizations.

The task time data showed the same pattern but with greater accentuation. Since reaction times typically are nonnormally distributed (i.e., Heathcote et al. [1991], Hockley [1984], and McCormack and Wright [1964]), we employed a logarithmic transformation of the data from the last five trials in each visualization type before statistical testing. The mean task time for 2D-scarce was 31.73 seconds and for 3D it was 6.35 seconds, (Figure 14). That is, the task time in the 2D-scarce visualization was more than four times as long as in the 3D condition. The logarithmic time data were then analyzed by means of an ANOVA, using a decision criterion of $p < 0.05$. The result from the ANOVA shows that there was a significant difference in task time between 3D and 2D-scarce, $F(1, 26) = 98.58$, $p < 0.001$. Thus, the 3D visualization led to a better performance both in terms of fewer errors and much shorter task times. Also, there was no significant effect of trials ($F(4, 104) = 1.01$, $p = 0.405$) nor any interaction ($F(4, 104) = 1.71$, $p = 0.154$) indicating that no further learning took place during these last five trials.

The attempted optimization of the 2D visualization also seemed to have been successful in that the errors were much larger, with similar task times, in the original 2D visualization (2D-dense) run as a control condition after the experiment (see Figure 14).

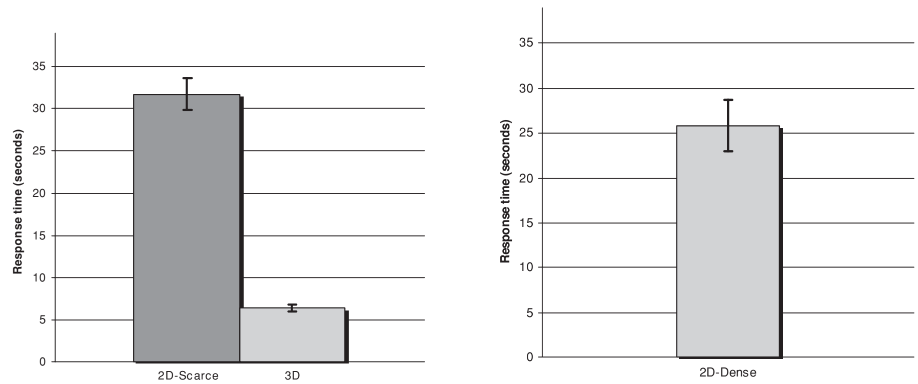

Fig. 14. Response time in seconds for the three different visualization conditions. The graph on the left is for 2D-scarce and for 3D. For comparison, 2D-dense results appear on the right. Vertical bars denote 0.95 confidence intervals.

3.4.6 *Discussion.* In this experiment, we tested hypothesis H2, which states that if a spatiotemporal task only requires perceiving a structure invariant over affine transformations, a space–time cube would be efficient. The task used was different from the one used in the two earlier experiments in precisely this respect. There was a significant difference between the 2D and 3D visualizations in both how well the participant could decide in which order the vehicles had arrived as well as in task time. Therefore, hypothesis H2 cannot be rejected and receives support.

4. GENERAL DISCUSSION

The general finding in these three experiments is, perhaps not surprisingly, that no general visualization solution exists that fits all the tasks that might be involved in time-varying geovisualizations. Our results also accentuate the fact that performance differences between different visualization types—for the same task and the same data—can be quite large. Especially interesting in this context is the failure of the space–time cube to efficiently support users in at least one type of task. The space–time cube in general is a clever idea and has some intuitive appeal in that it simultaneously shows both spatial and temporal properties. Also, in a task where only ordinal information in space–time was needed to solve the task (Experiment 3) it proved to be beneficial. However, human general inability to perceive precisely 3D metric properties limits its general usefulness. In a task where metric properties were crucial, even a fairly unsophisticated 2D visualization supported users better (Experiments 1 and 2). These results point to the need for a new taxonomy of tasks and data. This taxonomy needs to be based on empirical findings of human performance in different types of tasks as well as more general, theoretical considerations. This would facilitate a close coupling between the classes of the taxonomy and predictions about user performance when carrying out a task and facing different types of visualizations. A major component of such a taxonomy would be, based on the results presented here and previous results (e.g., Lind et al. [2003] and Todd et al. [2001]), whether a task generally requires the perceiving of metric properties of a visualization or not. If that is not the case, efficient 3D visualizations are possible, whereas 2D visualizations are required in other cases.

At least for large visual angles and for our type of task in Experiments 1 and 2, animation does not seem to be an efficient method for conveying the type of time-varying geospatial data used in our studies. Further studies are of course needed to look more deeply into the use of animation as a visualization type for these and other tasks.

Further efforts need also to be directed toward studying the efficiency of the space–time cube compared to an improved 2D visualization for the (nonmetric) task used in Experiment 3. One simple possibility is to use time labels instead of cross-lines. It could, for instance, be more efficient for a user to read a label instead of estimating the number of cross-lines. But whether this is efficient enough compared to the 3D visualization in the form of a space–time cube is another question. And, as pointed out in Section 2.1.3, whatever graphical expression we invent to show time in a 2D visualization of spatiotemporal data, this graphical expression can be used in a space–time cube as well and, within that, be used to visualize an additional property of the events depicted. In this manner, a space–time cube, given that the tasks it is used for can be solved by using properties of the visualization that are invariant under affine transformations, has the potential of effectively communicating more dimensions of data than any 2D visualization. In all other cases, 2D visualizations should be the first choice.

ACKNOWLEDGMENTS

We specially want to thank Ulrik Spak at SNDC for discussing and helping with the experiment.

REFERENCES

ALVAREZ, G. A., HOROWITZ, T. S., ARSENIO, H. C., DIMASE, J. S., AND WOLFE, J. M. 2005. Do multielement visual tracking and visual search draw continuously on the same visual attention resources? *J. Exp. Psych. Human Percept. Perform. 31*, 4, 643–667.

ANDRIENKO, G., ANDRIENKO, N., AND GITIS, V. 2003. Interactive maps for visual exploration of grid and vector geodata. *ISPRS J. Photogrammetry Remote Sens. 57*, 5-6, 380–389.

ANDRIENKO, N. AND ANDRIENKO, G. 2003. Usability tests with interactive maps. In *Proceedings of the 10th International Conference on Human-Computer Interaction*. ACM, New York, 1153–1157.

ANDRIENKO, N., ANDRIENKO, G., AND GATALSKY, P. 2000. Supporting visual exploration of object movement. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. ACM, New York, 217–220.

ANDRIENKO, N., ANDRIENKO, G., AND GATALSKY, P. 2003a. Exploratory spatio-temporal visualization: An analytical review. *J. Visual Lang. Comput. 14*, 503–541.

ANDRIENKO, N., ANDRIENKO, G., AND GATALSKY, P. 2003b. Tools for visual comparison of spatial development scenarios. In *Proceedings of the 7th International Conference on Information Visualization*. IEEE, Los Alamitos, CA, 237–244.

BERTIN, J. 1983. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, Madison, WI.

BINGHAM, G. P. AND LIND, M. 2008. Large continuous perspective transformations are necessary and sufficient for accurate perception of metric shape. *Percept. Psychophys*. 70, 3, 524–540.

BROWN, B. 1972. Resolution thresholds for moving targets at the fovea and in the peripheral retina. *Vision Res. 12*, 2, 293–304.

CHUNG, W., CHEN, H., CHABOYA, L. G., O'TOOLE, C. D., AND ATABAKHSH, H. 2005. Evaluating event visualization: A usability study of cop link spatio-temporal visualizer. *Int. J. Hum. Comput. Stud. 62*, 1, 127–157.

FORSELL, C. 2007. Perceptually motivated constraints on 3D visualizations. Ph.D. thesis, Department of Information Science. Acta Universitatis Upsaliensis.

HÄGERSTRAND, T. 1970. What about people in regional science? *Pap. Reg. Sci. 24*, 1, 6–21.

HARROWER, M. 2002. Visualizing change: Using cartographic animation to explore remotelysensed data. *Cartograph. Perspect. 39*, 30–42.

HEATHCOTE, A., POPIEL, S. J., AND MEWHORT, D. J. K. 1991. Analysis of response time distributions: An example using the strop task. *Psychol. Bull. 109*, 2, 340–347.

HOCKLEY, W. E. 1984. Analysis of response time distributions in the study of cognitive processes. *J. Exp. Psych. Learn. Memory Cogn. 10*, 4, 598–615.

HOUSE, D. H., BAIR, A. S., AND WARE, C. 2006. An approach to the perceptual optimization of complex visualizations. *IEEE Trans. Visual. Comput. Graphics 12*, 4, 509–521.

ISO 9241-11. 1998. Ergonomic requirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability. ISO 9241-11.

ITC-MINARD. 2007. What has ITC done with Minard's map? http://www.itc.nl/personal/kraak/1812/minard-itc.htm.

KAPLER, T. AND WRIGHT, W. 2004. Geotime information visualization. In *Proceedings of the Symposium on Information Visualization*. IEEE, Los Alamitos, CA, 25–32.

KLEIN, F. 1939. *Elementary Mathematics from an Advanced Standpoint: Geometry*. Dover, New York.

KOSARA, R., HEALEY, C. G., INTERRANTE, V., LAIDLAW, D. H., AND WARE, C. 2003. User studies: Why, how, and when? *IEEE Comput. Graphics Appl. 23*, 4, 20–25.

KOUA, E., MACEACHREN, A., AND KRAAK, M. 2006. Evaluating the usability of an exploratory geovisualization environment. *Int. J. Geog. Inf. Syst. 20*, 20, 425–428.

KULYK, O., KOSARA, R., URQUIZA, J., AND WASSINK, I. 2007. Human-centered aspects. In *Human-Centered Visualization Environments*, A. Kerren, A. Ebert, and J. Meyer, Eds. Springer, Berlin, 13–75.

LIGHTNER, N. J. 2001. Model testing of users' comprehension in graphical animation: The effect of speed and focus areas. *Int. J. Hum. Comput. Interact. 13*, 1, 53–73.

LIND, M., BINGHAM, G. P., AND FORSELL, C. 2003. Metric 3D structure in visualizations. *Inf. Visual. 2*, 51–57.

LIND, M. AND KJELLIN, A. 2005. Faster is better: Optimal speed of animated visualizations for decision makers. In *Proceedings of the 9th International Conference on Information Visualisation*. IEEE, Los Alamitos, CA, 896–900.

MACKINLAY, J. D. 1999. Automating the design of graphical presentations of relational information. In *Readings in Information Visualization*, S. K. Card, J. D. Mackinlay, and B. Schneiderman, Eds. Morgan Kaufmann, San Francisco, CA, 66–81.

MCCORMACK, P. AND WRIGHT, N. 1964. The positive skew observed in reaction time distributions. *Can. J. Psych. 18*, 1, 43–51.

NORMAN, J. F., TODD, J. T., PEROTTI, V. J., AND TITTLE, J. S. 1996. The visual perception of 3D length. *J. Exp. Psych. Hum. Percept. Perform. 22*, 1, 173–186.

OGAO, P. J. AND KRAAK, M. J. 2002. Defining visualization operations for temporal cartographic animation design. *Int. J. Appl. Earth Observ. Geoinf. 4*, 1, 23–31.

PAILLARD, J. AND AMBLARD, B. 1985. Static versus kinetic visual cues for the processing of spatial relationships. In *Brain Mechanisms and Spatial Vision*, D. J. Ingle, M. Jeannerod, and D. N. Lee, Eds. Martinus Nijhoff, Amsterdam, 299–330.

PETTERSSON, L. W., SPAK, U., AND SEIPEL, S. 2004. Collaborative 3D visualizations of geo-spatial information for command and control. In *Proceedings of the Annual SIGRAD Conference.*

PEUQUET, D. 1994. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Ann. Assoc. Am. Geographers 84*, 3, 441–461.

PIZLO, Z. 1994. A theory of shape constancy based on perspective invariants. *Vision Res. 34*, 12, 1637–1658.

PIZLO, Z. AND STEVENSON, A. K. 1999. Shape constancy from novel views. *Percept. Psychophys. 61*, 7, 1299–1307.

PYLYSHYN, Z. W. AND STORM, R. W. 1988. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision 3*, 3, 179–197.

SCHIPPER, L. AND VERSACE, J. 1956. Predictions of arrival sequences of simulated radar targets as a function of display size, target size, and target sharpness. WADC Tech. rep. 56-72, Laboratory of Aviation Psychology, Ohio State University.

SCHOLL, B. J. 2001. Objects and attention: The state of the art. *Cognition 80*, 1–46.

SIMONS, D. 2000. Current approaches to change blindness. *Visual Cogn. 7*, 1-3, 1–15.

SIMONS, D. AND CHABRIS, C. 1999. Gorillas in our midst: Sustained in attentional blindness for dynamic events. *Perception 28*, 9, 1059–1074.

SIMONS, D. J. AND RENSINK, R. A. 2005. Change blindness: Past, present, and future. *Trends Cogn. Sci. 9*, 1, 16–20.

SUPPES, P. 1977. Is visual space Euclidean? *Synthese 35*, 4, 397–421.

TITTLE, J. S., TODD, J. T., PEROTTI, V. J., AND NORMAN, J. F. 1995. Systematic distortion of a perceived 3D structure from motion and binocular stereopsis. *J. Exp. Psych. Hum. Percept. Perform. 21*, 3, 663–678.

TODD, J. AND BRESSAN, P. 1990. The perception of 3D affine structure from minimal apparent motion sequences. *Percept. Psychophys. 48*, 5, 419–30.

TODD, J. T., CHEN, L., AND NORMAN, J. F. 1998. On the relative salience of euclidean, affine and topological structure for 3D form and discrimination. *Perception 27*, 3, 273–282.

TODD, J. T. AND NORMAN, J. F. 2003. The visual perception of 3D shape from multiple cues: Are observers capable of perceiving metric structure? *Percept. Psychophys. 65*, 1, 31–47.

TODD, J. T., OOMES, A. H., KOENDERINK, J. J., AND KAPPERS, A. M. L. 2001. On the affine structure of perceptual space. *Psychol. Sci. 12*, 3, 191–196.

TRESILIAN, J. 1995. Perceptual and cognitive processes in time-to-contact estimation: Analysis of prediction-motion and relative judgment tasks. *Percept. Psychophys. 57*, 2, 231–245.

TUFTE, E. R. 1983. *The Visual Display of Quantitative Information*. Graphic Press, Cheshire, CT.

TUFTE, E. R. 2006. *Beautiful Evidence*. Graphics Press, Cheshire, CT.

WARE, C. 2000. *Information Visualization: Perception for Design*. Morgan Kaufmann, San Francisco, CA.

WEHREND, C. AND LEWIS, S. 1990. A problem-oriented classification of visualization techniques. In *Proceedings of the 1st Conference on Visualization*. IEEE, Los Alamitos, CA, 139–143.

WIEGMANN, D. A., ESSENBERG, G. R., OVERBYE, T. J., AND SUN, Y. 2005. Human factor aspects of power system flow animation. *IEEE Trans. Power Syst. 20*, 3, 1233–1240.

YANTIS, S. 1992. Multielement visual tracking: Attention and perceptual organization. *Cogn. Psych*. 24, 3, 295–340.

YATTAW, N. 1999. Conceptualizing space and time: A classification of geographic movement. *Cartography Geogr. Inf. Sci. 26*, 2, 85–98.

Received June 2007; revised September 2009; accepted 2009