Available online at www.sciencedirect.com

SciVerse ScienceDirect

Int. J. Human-Computer Studies 70 (2012) 812–827

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

Unpacking the temporal advantage of distributing complex visual displays

Jooyoung Jang[a,*], Susan Bell Trickett[b], Christian D. Schunn[a], J. Gregory Trafton[b]

[a]Department of Psychology, Learning Research and Development Center, University of Pittsburgh, 823 LRDC, 3939 O'Hara Street, Pittsburgh, PA 15260, USA

[b]Naval Research Laboratory, Information Technology Division, Code 5500 4555 Overlook Ave., S.W. Washington, DC 20375-5337, USA

Received 31 January 2012; received in revised form 3 July 2012; accepted 18 July 2012
Communicated by E. Mottak
Available online 31 July 2012

Abstract

Spatial arrangement of information can have large effects on problem solving. Although such effects have been observed in various domains (e.g., instruction and interface designs), little is known about the cognitive processing mechanisms underlying these effects, nor its applicability to complex visual problem solving. In three experiments, we showed that the impact of spatial arrangement of information on problem solving time can be surprisingly large for complex real world tasks. It was also found that the effect can be caused by large increases in slow, external information searches (Experiment 1), that the spatial arrangement itself is the critical factor and the effect is domain-general (Experiment 2a), and that the underlying mechanism can involve micro-strategy selection for information encoding in a response to differing information access cost (Experiment 2b). Overall, these studies show a large slowdown effect (i.e., approximately 30%) that stacking information produces over spatially distributed information, and multiple paths by which this effect can be produced.

Keywords: Information display; Information access cost; Cognitive load theory; Proximity compatibility theory

1. Introduction

Ten years ago, Grudin (2001) pointed out that "the amount of accessible information is rapidly overloading the displays [single 17-inch monitors], which have not grown at the same pace" and software supporting the partition of digital information has grown even less quickly. These days, large displays and multiple monitors are even more affordable and now prevalent in homes and offices and thus, questions are emerging around whether the newly developed visual displays promote work efficiency, whether we know how to effectively use them, and whether designers have working principles and knowledge to design hardware and software that support people using large screens.

Many researchers have examined the ways in which users interact with large displays and the performance benefits users gain in using large rather than small displays for the completion of various tasks, such as spatial orientation, reading comprehension, sense-making (i.e., integrating data or experiences to anticipate future events), programming, and window management (Andrews et al., 2010; Bailey et al., 2008; Grudin, 2001; Mynatt et al., 1999; Robertson et al., 2005; Tan et al., 2003, 2004). These studies often found that large displays are better (for a review: Czerwinski et al., 2006). However, while many people have adopted dual or large screens, others are now spending less time using large monitors or have removed desktops entirely from their work and home, relying instead on the smaller screens found in high-powered ultra-light laptops, multi-touch display tablet computers and Internet-enabled smartphones. Is this move to smaller screens detrimental to work efficiency or have new software conventions been found to mitigate the need for larger displays?

Some research suggests that larger displays are not always beneficial to problem solving. In Kroft and Wickens (2002),

*Corresponding author. Tel.: +1 412 648 1109; fax: +1 412 624 7439.
E-mail address: joj15@pitt.edu (J. Jang).

1071-5819/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijhcs.2012.07.003

student pilots who were required to use a large integrative display superimposing two different maps were significantly slower in answering questions exclusively relevant to just one map, but faster and more accurate when given integrative questions (i.e., questions that must be answered by *combining* information from two maps). Critically, the pattern of results suggests that the issue is likely not the size per se but the spatial organization and its fit to the task at hand—the previously observed effects of raw display size are likely caused by the different organizations of given information that different display sizes tend to require (Jang et al., 2011). It is likely that data should be shown in a way that assists timely perception, information integration, and comprehension. For these reasons, the spatial organization of information displays is an essential topic that needs to be explored further.

1.1. Spatial organization of visually presented information

Past work on the spatial organization of information displays has focused on whether information is presented in an integrative or a separated manner. Integrative displays combine multiple sources of information into a single source. The practical benefit of integrative displays has been shown in many domains including mathematics instruction, way finding, and flight cockpit systems (Kroft and Wickens, 2002; Ni et al., 2006; Sweller et al., 1990; Wickens, 2000; Wickens et al., 2000; Yeh and Wickens, 2001). Across these lines of work, there is general consensus that having all relevant information in a single display (i.e., integrative displays) is better than having information in a separated format, especially for problems that require integration across information sources (i.e., integrative tasks). For example, when students study a worked-out example to learn geometry, text instruction can be placed inside the accompanying picture, thus removing the cognitive load for mental integration of separated text and diagrams. The disadvantage of separated versus integrative display formats can be strong enough to obscure the benefit of well-prepared content (Tarmizi and Sweller, 1988; Wickens and Carswell, 1995). Thus, purely physical elements such as spatial organization can have a substantial effect on performance.

Note that there is also a different but closely related aspect of display organization relevant to integrative tasks when fully integrated displays are not possible: spatially distributed displays (i.e., when information sources are presented side-by-side) versus stacked displays (i.e., when information sources are sitting on top of one another in a fully overlapped manner thus only the top source fully visible). For example, when a meteorologist attempts to make a forecast by integrating information from a large number of maps (e.g., air pressure, wind speed, and cloud distribution maps by the unit time- and height-interval), a single integrative display is not a practical option because superimposing even three such information-rich maps would be enough to make the display too cluttered to search and perceive critical information. In this case, a separated but spatially distributed display can be a promising alternative. For such cases, preliminary research has found that spatially distributed displays are better than spatially stacked displays (Jang et al., 2011; Wickens and McCarley, 2008). The observed benefits included greater understanding of information, lower error rates, higher accuracy, and shorter problem solving time.

1.2. Teasing apart display size, organization, and density

To position the current study within the visual display literature and to clearly discuss underlying mechanisms, we define three important inter-related dimensions of visual displays: size, organization, and density. Display size refers to the physical size of a display, ranging from large projection displays to small smartphone displays. Display organization refers to the spatial arrangements of information windows/pages when multiple information sources are available. We posit three types of spatial organization as discussed in the previous section: integrative, distributed, and stacked (Fig. 1). The stacked displays tested in the current study specifically refer to true hiding (i.e., information sources perfectly overlapped on top of one another) instead of partial stacking (e.g., overlapping windows on a computer screen). A form of partial stacking with interactive user manipulation deploys other factors to consider such as the degree of stacking, the degree of user controllability, and user's display manipulation strategy, which can be explored in future studies once the 100% overlapping case is better understood. Display density is defined as "the number of characters divided by the total number of text spaces available on a computer screen" (Staggers, 1993) and thus it often depends on display size and organization.

The three display dimensions can be manipulated separately or together. Tan et al. (2004), for example, manipulated only display size. They compared the effectiveness of an 18″ desktop monitor against a 76″ wide by 57″ tall projected image on a wall in performing 3D navigation tasks, holding the visual angle and display density constant. They found that large displays induced more effective egocentric strategies to perform navigation tasks presumably due to a better sense of presence. Bly and Rosenberg (1986) manipulated display organization alone by comparing tiled vs. overlapping windows and found that tiled windows are better for tasks that require little window manipulation, while overlapping windows are better for tasks that require more window manipulation. By comparison, Staggers (1993) manipulated display density holding display size constant, but also changing organization to achieve this contrast. She compared low-, moderate-, and high-density screens by controlling the amount of information shown per screen. High density took the form of an integrative display whereas moderate and low densities were displayed in a stacked manner. She found that clinical nurses performed target-searching tasks

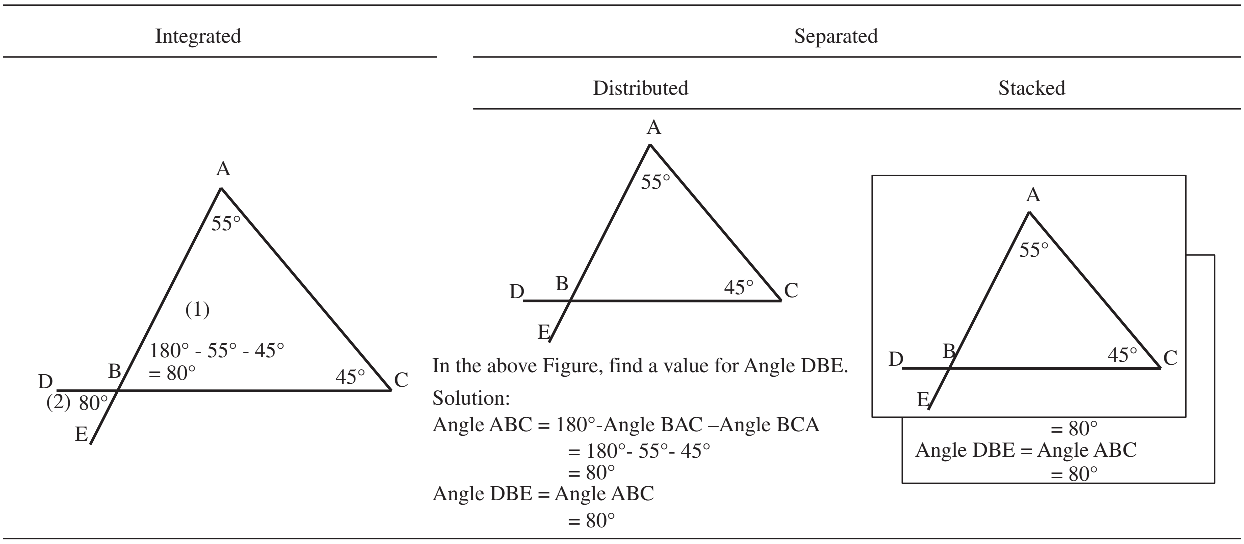

Fig. 1. Three types of display organization. The image of a stacked display here presents partially overlapped contents to reveal what was stacked but the two contents should be imagined as fully overlapped.

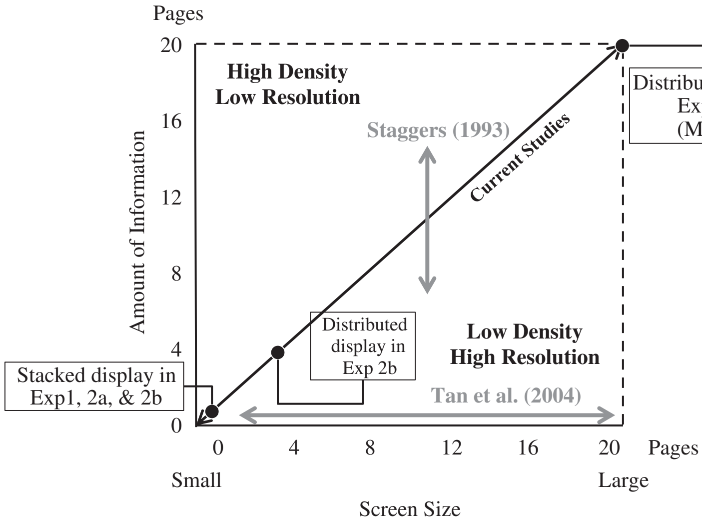

Fig. 2. Descriptive positions of previous and current studies on the dimensions of display size, organization and density.

faster on high-density screens than moderate or low, with no difference in mean accuracy and subjective satisfaction.

Unlike theses prior studies, the current paper evaluates two types of spatial organization (i.e., distributed vs. stacked), holding display density constant (Fig. 2). The density is equivalent across distributed and stacked displays because the ratio of the amount of visible information to the display size is constant (e.g., twice as much information shown across twice as large a display). Considered another way, distributed vs. stacked naturally varies both organization and display size (when density is held constant), whereas integrated vs. distributed (or stacked) naturally varies organization and density. Previous and current studies' descriptive positions across the three dimensions are summarized graphically in Fig. 2.

Although the display density is constant across distributed and stacked displays, there are two other factors that emerge when manipulating spatial organizations: display resolution and distribution ratio. Display resolution is highly relevant to the physical display size as large displays can show the same amount of information at a greater resolution relative to small displays. The importance of spatial resolution on problem solving performance has been found in prior studies (Kroft and Wickens, 2002; Wickens et al., 2000). Specifically, Kroft and Wickens (2002) compared large integrated and small integrated displays and found a sizable effect of resolution. With the larger display, responses were faster and 12% more accurate regardless of question types (i.e., whether the question requires information integration or not). In the current study, both display density and resolution are kept constant within each experiment. Across experiments, however, display resolution varied. For example, the stacked display in Experiment 2a (i.e., a full screen) was four times larger than the stacked display in Experiment 2b (i.e., a quadrant of the screen). Although it is a minor focus in the current study, the effect of display resolution is of great interest in the field, and therefore it is analyzed and discussed at the end of Experiment 2b from the perspective of the current data.

More central to the study of stacked vs. distributed displays is the distribution ratio, which is the ratio of a stacked page to the number of pages that are shown in the distributed view. For example, the ratio is 1:2 when reading a pdf document in a single page display vs. in a two-page display. Then the question becomes, how large do distribution ratios need be to show an effect on performance? Presumably, larger distribution ratios produce larger effects, up to some ceiling. But it is pragmatically important to find the minimum ratio that produces a benefit because displays that can support 1:40 distribution ratios are not always handy or possible.

Further, the role of distribution ratio is theoretically important, because the distribution ratio may affect a user's cognitive strategy, which may shift the locus of the effect. For example, when using stacked displays, turning 60 pages (Experiment 1) instead of 20 pages (Experiment 2a and 2b) is a sizable difference and the distribution ratio (1:60 or 1:20) may contribute to the overall disproportionate slowdown by affecting the time point at which the slowdown begins. Using stacked displays, participants could learn that page turning and regressions to previous pages can be a burden after they have processed a few pages (i.e., the effect may depend upon within-task feedback). Other participants may immediately predict the cumbersome nature of page turning and regressions and thus decide to slow down from the very beginning (i.e., the effect may be prediction based). Whether prediction-based or feedback-based, the overall size (rather than timing) of the slowdown is expected to be dependent upon the distribution ratio.

1.3. Exploring underlying mechanisms

The benefits of large displays and useful practical guidelines for design can be found in many studies (for a review: Czerwinski et al., 2006), but most of those studies did not focus on exploring possible underlying mechanisms. Both the studies of Staggers (1993) and Tan et al. (2004) were more focused on testing the efficiency of performance using different displays and on generating practical suggestions. Yet, Tan and colleagues speculated that large displays provide a better sense of presence and thus help users induce efficient navigation strategies. They also suggested the importance of cognitive cues and tested the idea with three monitors and a larger projection display (Tan et al., 2001). They had each monitor provide cognitive cues of information location and the projection display produced a sense of presence using ambient visuals and sounds (i.e., contextual cues). Users who used the augmented display had better memory retention for word pairs than those who used a desktop with a single monitor. Although they tested an underlying mechanism (i.e., cognitive cues), the studies were focused on multiple visual and auditory cues found in augmented displays, and therefore not applicable to unpacking distributed vs. stacked displays comparisons.

More directly relevant to distributed vs. stacked displays, Grudin (2001) has argued for the importance of using peripheral vision and provided supporting qualitative data gained from structured interviews with multiple monitor users. Users preferred to have each monitor show a unique information source consistently (e.g., extended menus and tools for drawing always on the left monitor) and have multiple information sources partitioned and placed side-by-side rather than having a primary work window dominate the entire space across the two monitors. The use of peripheral vision may sometimes explain the differences between distributed and stacked displays in terms of having a larger space vs. a smaller space. However, in the current study, we used complex problem-solving tasks that require close examination and processing of each information source. For peripheral vision to be useful, information in the peripheral area should be simple enough to be recognized peripherally or somehow alerting to draw attention or to direct the focus of a gaze; but information in the current study was only meaningful at a high spatial frequency (text values, fine-grained line patterns) that is not detectable in the periphery, similar to many complex problem-solving scenarios (e.g., weather forecasting, intelligence analysis, and scientific data analysis). Thus, the use of peripheral vision cannot account for display organization effects in the kinds of situations studied in the current research.

Two theories have explained the underlying mechanisms for the benefit of integrative displays in terms of either differing level of extraneous cognitive load that a display format imposes (Sweller et al., 1990) or combinations of search, interface manipulation, and information access cost (Wickens and McCarley, 2008). Because of the complexity found inherently in integrative tasks, even a small increase of extraneous cognitive load that comes from information spatial arrangement can matter greatly. But this load explanation begs questions about the causes and consequences of a higher load: why does one format of display produce higher cognitive load than another, and what kinds of strategic choices do people make to compensate for higher cognitive loads?

In theory, different display formats result in differing external information access costs (i.e., the time to locate information, move the head and eyes, and encode information). According to Gray and Fu (2004), when the cost of accessing external information increases, people tend to memorize information to make it readily accessible *in the head* (i.e., use a memorization strategy to keep older information accessible for integration). By contrast, when external information access cost is low, people do not bother to memorize information and instead rely on external *in the world* information (i.e., use a perceptual-motor strategy to recover older information). Consistent with this idea, Morgan et al., 2009 recently demonstrated that imposing a relatively higher access cost in a simple copying task induced improved recall, which, in turn, facilitated resumption after interruptions. People were significantly more likely to memorize target information in higher access cost situations, such as when a mouse movement is needed to uncover target information or when a mouse movement and an additional few seconds of waiting is needed, than in a low cost situation when all information was always visible. The memorization strategy exercised in higher access cost situations reduced forgetting and helped people resume the task more efficiently after interruptions. As expected, however, it took more time to complete the task, which reflects the time spent in memorization or careful encoding of target information.

Applying this construct to display formats, the stacked display presumably produces a higher external information access cost situation because information is a screen change or a page-turn away, compared to the low cost of an eye/head turn in the distributed display. Under this explanation, we expected that people in the stacked display format would experience a higher cognitive load. More importantly, we expected the adoption of different micro-strategies to compensate for the cost (e.g., becoming a memorizer or a verifier, as described by Gray and Fu). People in the stacked display condition may fixate significantly longer on information pieces on each page throughout an integrative problem-solving task (i.e., become memorizers) than those who solve the same problem using the distributed display, presumably as a micro-strategy to bypass the relatively higher information access cost in the stacked display and to reduce the need for revisits. By contrast, people in the distributed display condition may not fixate as long as the people in the stacked display would because they may choose to revisit the information as necessary rather than memorize it (i.e., become verifiers). Alternatively, if the information to be encoded greatly exceeds what can be memorized, the higher load condition may simply increase the amount of verifications or re-encodings.

Although our focus is on testing micro-strategy choice depending on the display organization, it is important to note that other studies examined the possibility of spatial memory benefits in large or distributed displays (Cockburn and McKenzie, 2001; Robertson et al., 1998; Robertson et al., 2000). In using 2D and 3D data management interfaces such as Task Gallery and Data Mountain, people demonstrated faster mean response times, a result which supports the benefit of spatial cues. Also, users showed a high percentage of correct recall on task window content and their relative locations in space when they were asked to draw the interface from memory. Applying these concepts to the distributed vs. stacked display comparison, distributed displays may provide spatial cues by showing each information page in a unique and static location, and therefore, participants may go to the desired pages faster with less error.

Explaining the general advantages of large distributed displays in terms of spatial memory benefits, however, seems to be of limited usefulness because of two factors. First, the task in the current study involves complex problem solving that requires integration of information. The consistent spatial location of information is expected to assist users partially in information search and management, but the integrative tasks studied here involve many more aspects than information search alone. Second, distributed displays do not guarantee consistent spatial mappings. For example, in a.pdf document, information viewed simultaneously in four pages is more distributed than information viewed simultaneously in just one or two pages, but the content of those four pages will change over time. Explanations of the advantage of four page over one page views of large information spaces (as in Experiment 2b) cannot depend upon consistent spatial mapping explanations. In such circumstances information access cost may provide the better account. In sum, the focus of the current studies was to investigate the effect of information access cost in explaining the distributed display benefit; the possible effects of other factors were systematically controlled (e.g., distribution ratio) or minimized (e.g., spatial memory).

1.4. Overview of the studies

To examine the generality of effects of distributed vs. stacked display formats and to further our understanding of the effects—whether information access cost is indeed a determining factor and what strategies people adopt to compensate for the access cost, we strategically used mixed populations (i.e., meteorologists-in-training and university undergraduates) and experimental tasks ranging from real complex problems and artificially developed lab tasks (Dunbar, 1995; Trickett and Trafton, 2007).

Experiment 1 was conducted to establish the phenomenon in a naturalistic context and explore possible underlying causes in that context. In this experiment, student meteorologists who had already trained for 3 years made weather forecasts using computers (i.e., a stacked display) and mapwalls (i.e., a distributed display). Experiment 2a replicated the effect in a controlled setting that eliminated a number of confounds found in the natural contrast (e.g., the role of animation) and also tested the generality of the effect with a different information integration task. Experiment 2b examined underlying causes in the second task.

In Experiment 2a, undergraduate students solved the integrative task either using computers (i.e., stacked display) or mapwalls (i.e., distributed). In Experiment 2b, undergraduate students solved the integrative task using only computers in one of two different layouts: viewing one page at a time (i.e., stacked display) or viewing four pages at a time (i.e., distributed display). An eye-tracker was used to capture differences in information processing. Experiments 2a and 2b also assess the importance of distribution ratio by implementing the ratio of 1:20 in Experiment 2a and 1:4 in Experiment 2b.

These studies differ from most studies of visual displays in two ways. First, we examine two understudied display types of spatial organization (distributed vs. stacked), holding the display density factor constant. This contrast is a natural contrast in the world, inherently combining the amount of information displayed with (effective) display size and visual angle. The goal is not to separately understand the contributions of display size, information content, and visual angle to processing, but rather to understand what changes when this natural combination occurs.

The second contribution lies in an effort to explore the underlying cognitive mechanisms, bridging the gap between cognitive theories and human-computer studies. There has traditionally been a divide between cognitive science and human computer interaction studies when it comes to testing novel visual displays. In this study, we bring cognitive theory to explain why and how one type of display can be more beneficial than another by testing the memorization strategy hypothesis, which can, in turn, benefit the development of future generations of visual displays.

2. Experiment 1

2.1. Method

2.1.1. Participants

Participants were ten meteorologists-in-training (all males), junior and senior undergraduate meteorology majors, with an average of 3 years experience in forecasting. Their ages ranged from 20 to 22 years, with a mean of 21. All forecasters had extensive experience with both the computerized display and a mapwall display used in the study as part of their regular curriculum.

2.1.2. Design

The study used a mixed design with two conditions (between-subjects: mapwall vs. computer) defined by the display layout on which users made weather forecasts. Two equivalent meteorological scenarios were created, and all participants solved both scenarios (within-subjects). Participants were randomly assigned to display condition and scenario pairings. Further, the scenario order was counterbalanced across participants. Problem-solving activities were videotaped and the key dependent measures consisted of overall time spent and prediction accuracy on each scenario, with additional process information provided by the total number of visualizations examined per scenario, the mean amount of time spent looking at each visualization, and the number and percentage of repeated visualization visits.

2.1.3. Materials and procedure

The two display layouts were constructed in the following manner. In the mapwall condition, maps were printed and laid out on a wall (i.e., in a distributed layout). The approximate size of the wall display was 176″ wide and the viewing angle subtended by the wall was approximately 140°, assuming an arm-length (32″) viewing distance. The estimated walking required to view all the pages (gross information access cost) was five footsteps; 176″ divided by 36″ (average stride length). In the computer condition, the same information was presented one screen at a time (i.e., in a stacked layout) on a 17″ monitor (estimated viewing angle subtended by the monitor was 36°–60°, depending on one's viewing distance of 11″–19″); see Fig. 3. Importantly, the computer interface mirrored modern weather prediction interfaces and it was optimized to facilitate rapid comparison across models, data types, and times (e.g., maps become visible by just placing the mouse pointer over an index, instead of requiring clicking, and animations across time could be initiated as well). Thus, the relative external information access cost was lowered by the interface design: hovering instead of clicking. Each participant solved both scenarios, one in each condition.

The two scenarios had all the information needed to make a complete forecast: a satellite image, surface analysis, upper air maps (at four heights), climatological information for the target city, and outputs across time from two different weather prediction models, for time periods 00Z (time of product creation) to 72Z (prediction of what will happen 72 h into the future). The forecasters' task was to create forecasts for each of the two locations at multiple time points into the future, writing down the following predictions for each time point and location: temperature, amount of precipitation, cloud cover, wind speed, and wind direction. The task required two types of integration: temporal and dimensional integration (Trafton et al., 2000). Since the goal of the task was to extrapolate the future weather condition, forecasters had to integrate information across time. For example, they needed to track the movements of clouds from time period 00Z to 72Z to predict chances of rain in a given area in the future. Integration across weather dimensions was also necessary because a weather change cannot be predicted by a single variable alone (e.g., a combination of pressure, cloud cover, and humidity determines the chances of rain).

After each forecast was completed, videotapes were coded for overall time spent on task and the amount of time that each student spent on each meteorological display to a temporal accuracy of 1 second. Because the two scenarios were based on actual situations from the past, the experimenters knew actual weather outcomes and could calculate prediction accuracy (absolute differences between predicted and actual for each variable averaged across all time periods).

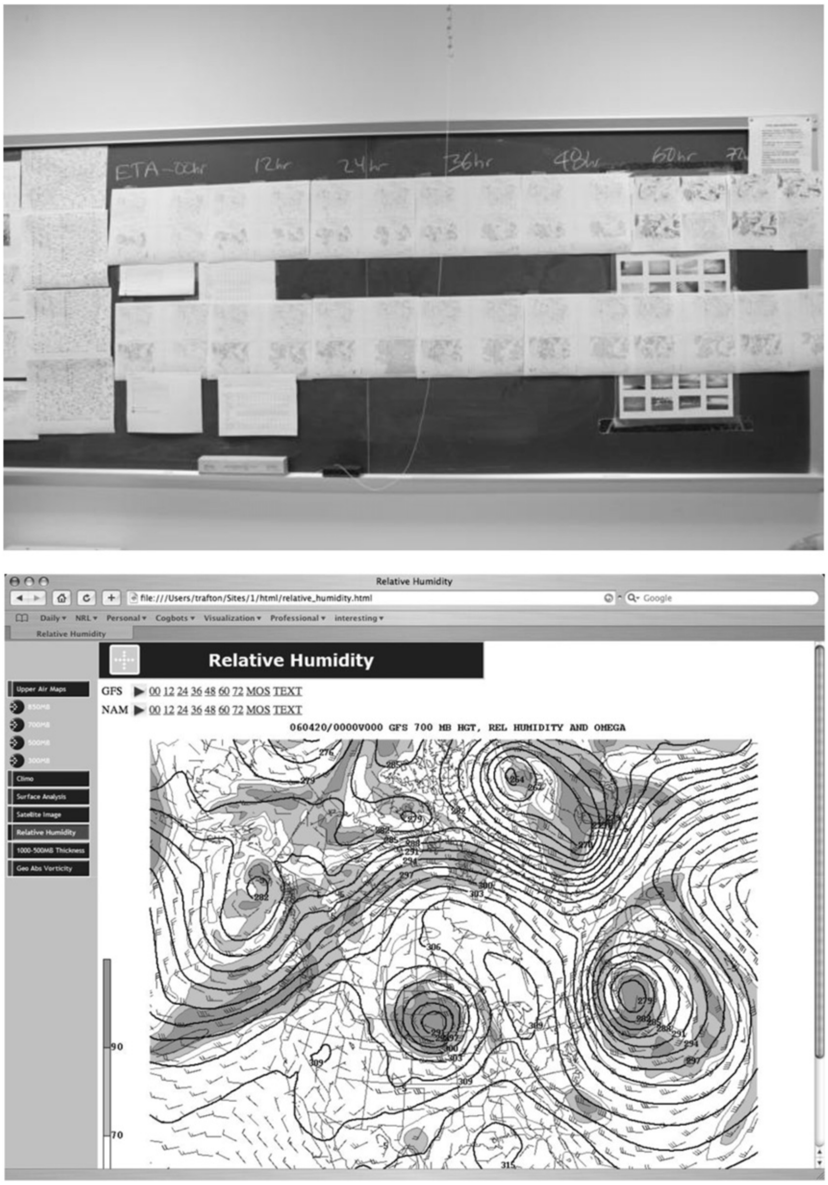

Fig. 3. Experiment 1: Spatial layout of all pages in the mapwall condition (upper panel) and one visible page in the computer condition (lower panel).

2.2. Results and discussion

A single factor (display condition: mapwall vs. computer) MANOVA was used to test the effect of display type on multiple dependent measures (i.e., time and accuracies for five weather dimensions) while controlling for Type I errors. All results were collapsed across the two forecast scenarios. Participants took 35% less time on task in the mapwall condition ($M=26.1$ min, $SE=3.9$) than in the computer condition ($M=40.7$ min, $SE=3.4$), $F(1, 9)=8.0$, $p=.02$. There were no significant differences for any of the performance variables (see Table 1), and in aggregate, participants were equally accurate across display conditions, $F(1, 9)=1.73$, $MSE=56.18$, $p=.23$ (i.e., no sign of a speed/accuracy tradeoff overall or by specific performance variable).

To explore the time difference between conditions in terms of memorization vs. perceptual-motor strategies, we examined the number of visualizations and amount of time spent looking at each visualization. Participants in the computer condition looked at three times more visualizations ($M=252.9$, $SE=50.2$) than in the mapwall condition ($M=85.8$, $SE=14.9$), t (4.7)[1] $=-3.19$, $p=.026$. Participants in the computer condition were twice as fast at examining each visualization ($M=3.2$ s, $SE=0.6$ s) as

[1] T-tests, not F-tests, were reported from here due to the violation of equal variance assumption. Corrected degrees of freedom and p values are reported. Levene's tests for equality of error variances for each of dependent variables were: the number of visualizations, $F(1, 8)=12.75$, $p=.01$; the amount of time spent looking at each visualization, $F(1, 8)=1.50$, $p=.26$; the number of repeated looks, $F(1, 8)=16.31$, $p=.00$; the percentage of repeated looks, $F(1, 8)=3.80$, $p=.09$.

Table 1
The absolute difference between forecast and actual for participants in the mapwall and computer conditions. For all measures, the closer to 0, the more accurate the prediction was.

Measure	Mapwall		Computer		
	M	SE	M	SE	*p* value
Temperature	6.12	0.91	5.12	0.54	.36
Amount of Precipitation	0.06	0.01	0.05	0.01	.47
Cloud Cover	2.26	0.25	2.1	0.14	.59
Wind Speed	4.25	0.20	5.02	0.99	.40
Wind Direction	70.69	4.56	81.69	6.85	.19

participants in the mapwall condition ($M=6.9$ s, $SE=1.2$ s), t (8)=2.71, $p=.027$. The overall 35% time effect can then be understood as three times as many visualizations ($252.9/85.8 \cong 3$) examined at twice the rate ($6.9/3.2 \cong 2$), thus 37% ($(252.9/85.8)/(6.9/3.2)=1.37$) more information access effort.

Next we explored the number and percentage of repeated looks that each condition performed, a tighter measure of relying on in-the-world access for previously encoded information. A repeat was defined as any time that a participant examined a visualization that had already been examined. Looking at a visualization repeatedly can occur because the forecasters need to refresh their memory, their understanding of the weather has changed or needs refinement, or for specific comparisons between different models. Participants in the computer condition ($M=210.8$, $SE=47.2$) had 6 times as many repeats as in the mapwall condition ($M=34.4$, $SE=9.5$), t (4.3)=4.32, $p=.019$. This difference extended to the percentage of visualizations that were repeats as well: in the computer condition most visualizations were repeats ($M=81\%$, $SE=4\%$), whereas in the mapwall condition only 38% ($SE=6\%$) were repeats, t (8)= -5.77, $p<.001$. In sum, participants in the computer condition appear to compensate for a possible higher cognitive load in the computer display by frequently going back to prior screens. Each page return was relatively quick in the computer interface, but the total volume of returns significantly slowed them down overall. With this speculation about cognitive load in mind, subjective cognitive load was explicitly measured in Experiment 2b.

Possible confounds of the naturalistic goal manipulation in Experiment 1 of distributed vs. stacked displays were static vs. animated displays and paper vs. computer medium. Experiment 2a removes the confound of static vs. animated displays and Experiment 2b removes the confound of paper vs. computer medium.

3. Experiment 2a

For Experiment 2a, we developed a lab task to test the generality of the phenomenon observed in Experiment 1 and to build a lab variation that could be manipulated with greater precision. Practically, weather forecasting requires highly specialized training, which makes experimental participants difficult to obtain. Although it has high value as a real-world task, it is hard to implement in the lab. Further, if the observed effect is not a one-time special case and has a certain consistent and meaningful underlying mechanism, the effect should be replicable in another task.

Although the benefit of distributed over stacked displays is generalizable to other tasks, it is unlikely to be universal to all tasks—very simple tasks likely can be handled with very simple displays. Thus, it is likely that a lab task must have some essential features of the weather forecasting task to replicate the layout effect, and we conservatively implemented a lab task with many shared features. Overall, the problem space in the weather forecasting task consists of many pages of information, presenting meteorological elements such as temperature and air pressure. Each information type is also presented for many time points, since forecasting involves making inferences about meteorological elements unfolding over time. This problem space produces the most important feature of the task: it requires information integration across dimensions and over time. Because meteorologists give relatively simple predictions like temperature and chance of rain, they must have interpreted and integrated information in some meaningful way based on the prediction rules they have learned. On the basis of this problem space analysis, a simplified task that requires information integration over time and space was developed and tested.

3.1. Method

3.1.1. Participants

Forty-seven undergraduates from the University of Pittsburgh participated in this study for course credit and were randomly assigned to one of the two display conditions (24 in mapwall with 12 females, and 23 in computer with 10 females). All of the participants were 18 years old except for one 26 years old.

3.1.2. Design

The study used a between-subjects design with two conditions (mapwall vs. computer) defined by the display layout on which users solved a given task. The experiment was conducted in two phases: practice and main task. The main dependent variables were practice time, task time, and task accuracy.

3.1.3. Materials and procedure

Based on the problem space analysis, a ''Deer task'' was created to involve information integration of complex visual information but not to rely on relatively rare expertise and also to remove the confounding feature of animation/quick access found only in the weather computer interface in Experiment 1. The participants' task was to compute how many adult deer and fawns survive over time in each of five regions by applying specific prediction rules. Participants were told that deer die when they lose too much weight and that weight loss is caused by extreme temperatures (i.e., either too hot or too cold) and lack of food (i.e., food access limited by snow depth).

After reading the rule pages, participants received specific data on temperature and snow depth over ten days in five different geographical regions; 20 total pages of information (see Fig. 4). Participants made predictions for the final number of surviving adults and fawns in each region.

In the mapwall condition, information pages were distributed along a wall (Fig. 4). The size of the wall display was 77″ wide and the viewing angle subtended by the wall was approximately 100°, assuming an arm's-length viewing distance (32″). The estimated walking required to view all the pages (gross information access cost) was two steps; 77″ divided by 36″ (average stride). In the computer condition, individual information pages were presented one at a time on a 17″ monitor (estimated viewing angle subtended by the monitor was 36°–60°—the same as in Experiment 1). Participants could browse through different pages of information by clicking within a table of contents on the left; note that information access cost is higher here relative to that of the weather interface without scroll-over and animation features. In addition, the distribution ratio is 1:20 (one window per screen vs. 20 printed out pages on a wall) in Experiment 2a in contrast to 1:60 (one window per screen vs. 60 printed out maps on a wall) in Experiment 1. We therefore expected a decreased relative time advantage of distributed displays in Experiment 2a on the basis of relatively fewer page transitions and regressions required to examine 20 pages instead of 60.

Participants in both conditions were first led through a practice session done on the computer, which allowed

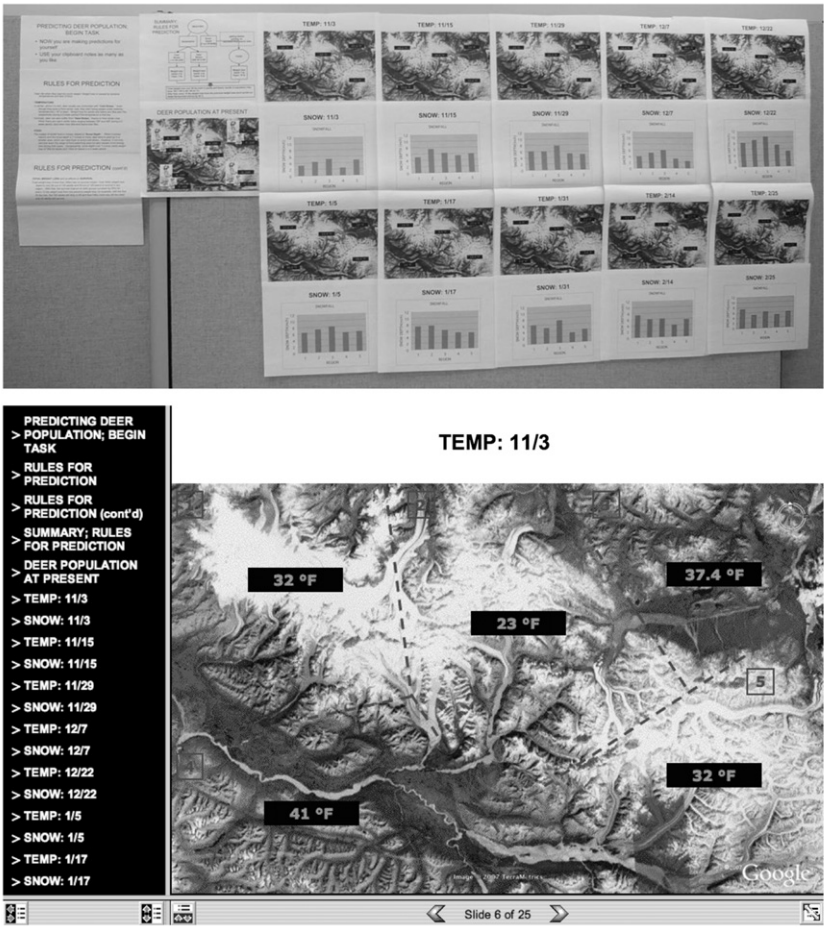

Fig. 4. Experiment 2a: Spatial layout of all pages visible in the mapwall condition (upper panel) and one visible page in the computer condition (lower panel).

them to learn how to apply the rules (practice time). Participants could refer to the rules at any point in time during practice and task sessions. Once a participant could correctly apply the rules, the experiment was allowed to begin.

The task session was done either on the computer or the wall. Participants were asked to make population predictions as they did in the practice session and let the experimenter know when they finished (task time). To assist calculation and reporting of final predictions (task accuracy), a clipboard for note taking was provided (Trafton and Trickett, 2001). Clipboard paper had lines distinguishing five regions as drawn in a temperature slide. No restrictions were given on the use of the note paper.

3.2. Results and discussion

Outliers were defined as accuracy lower than 50% correct, a sign of difficulties with the basic task calculations and/or very low task motivation. Five participants were excluded from analyses as outliers (2 males from computer, 3 females from mapwall), which left 21 participants in each condition (10 females in computer, 9 females in mapwall). Analyses were done with MANOVA.

Participants in the two conditions did not differ in their practice time, Computer: $M=12.4$ min, $SE=0.8$, Mapwall: $M=11.6$ min, $SE=0.7$, $F(1, 40)=.58$, $MSE=6.88$, $p=.45$. Remarkably consistent with Experiment 1, participants in the mapwall condition ($M=13.8$ min, $SE=1.2$) took 30% less time to complete the task than participants in the computer condition ($M=19.7$ min, $SE=1.6$), $F(1, 40)=8.28$, $MSE=366.09$, $p=.006$, Cohen's $d=0.90$. Again, this effect was not due to a speed/accuracy tradeoff: participants were equally accurate, Computer: $M=77\%$, $SE=4\%$, Mapwall: $M=79\%$, $SE=3\%$, $F(1, 40)=.151$, $MSE=38.09$, $p=.70$; see Fig. 5.

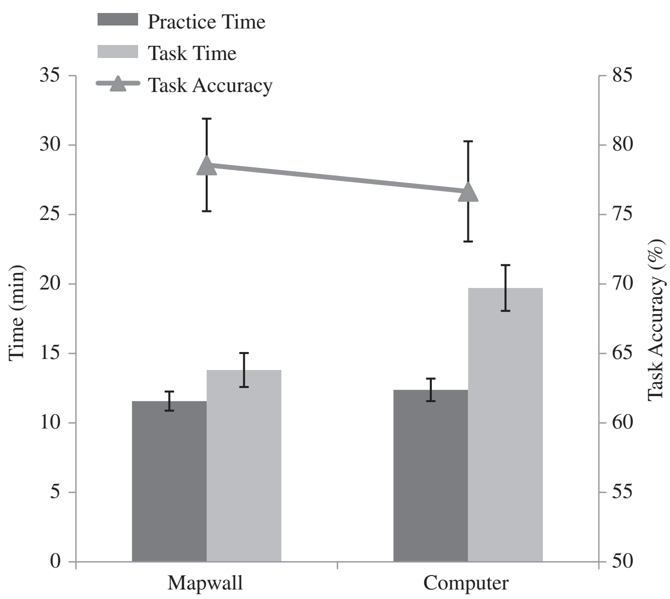

Fig. 5. Experiment 2a: Mean times and standard error bars for Mapwall and Computer displays in practice and task sessions. The four bar graphs for the means of practice and task times are contingent on the left vertical axis and the line graph for the means of task accuracy is contingent on the right vertical axis.

These results demonstrate that the effect of presentation modality is not due to the idiosyncrasies in Experiment 1 (e.g., expertise, domain-specificity, or unique visualizations). Further, it shows that the effect of simple information layout per se (i.e., distributed/mapwall vs. stacked/computer) is the issue rather than negative effects of animation, that the effect of information layout is quite robust, and that the effect is specific to time rather than accuracy differences. Experiment 2a still contains the confound of medium: paper vs. computer. Experiment 2b explores the causes of the effect using a cognitive load survey and eye-tracking data, and establishes whether layout (distributed vs. stacked) or modality (paper vs. computer) is the source of the effect by varying layout within a computer interface.

4. Experiment 2b

In Experiment 2b, we more directly tested the memorization strategy hypothesis by collecting eye-movement and cognitive load survey data. From this hypothesis, we predict that participants who use stacked displays would examine each piece of information for longer (i.e., longer average fixation durations) than those who use distributed displays as a way to overcome the relatively higher information access costs associated with stacked displays. These increased fixation durations should also coincide with a higher subjective cognitive load. We evaluated cognitive load levels using a subjective rating scale.

The cognitive load of stacked displays may be due to a combination of three possible factors: page turning, multi-tasking (e.g., note taking and information tracking), and memorization. However, unlike the first two possible factors, the memorization factor (i.e., stacked display users investing time to memorize and keep access to older information for integration) provides insights into the underlying cognitive mechanism (i.e., strategy selection depending on the degree of information access cost that a display imposes) and systematically influences the other two explanations (i.e., memorizers should turn pages and get lost less often than verifiers).

To delve into the contributions of each of these three factors, eye-movement and computer data were(Trafton and Trickett, 2001) analyzed, focusing on three key variables: the number of pages re-visited and the time spent on page revisits which corresponds to the page-turning factor, off-screen gaze durations corresponding to the multi-tasking factor, and mean fixation duration to test the memorization strategy. First, when there is a page transition, the eye-tracker automatically records the time the index was clicked and the time the page becomes available. By computing the time difference between the two and summing them up, it was possible to obtain an estimate of total time spent on page-turning activities. Also, page-turning frequency was counted. According to the memorization strategy hypothesis, it is expected that the stacked display users make fewer page transitions and thus spend less time on page-turning activities than distributed display users, which would rule out page-turning as an explanation for the stacked display performance deficit.

Second, off-screen gaze durations longer than the normal blink duration of 300–400 ms can be used as an index of time users are engaged in other activities—most commonly note taking in this study. Conservatively, we collected off-screen gaze durations longer than 2000 ms and summed them up to measure the time spent on multi-tasking (i.e., sum of the time spent on various activities beyond online visual information processing). According to the memorization hypothesis, stacked display users might spend more time taking notes, essentially an external form of memorization. If they did experience relatively higher information access cost when using stacked displays, they may take notes to reduce the need to revisit previous pages.

Finally, the mean fixation duration for on-screen eye-gazes can provide a more direct measure of the memorization hypothesis, because each fixation duration serves as an on-line measure of information processing, similar to the eye-mind assumption and immediacy assumption used in eye-tracking studies of reading processes (Carpenter and Just, 1983). The memorization strategy hypothesis predicts that stacked display users would produce longer mean fixation durations than distributed display users. That is, if they experience higher information access cost when using stacked displays, they would try to overcome the cost by spending extra encoding time to facilitate later retrievals from memory (Morgan et al., 2009). This prediction is consistent with the memory test results of Experiment 2 of Jang and Schunn (2012). In that study, stacked display users remembered more items about the last four pages they just viewed on a surprise memory test given as soon as participants reached the last information page. In addition, in contrast to the distributed display users, the stacked display users did not show a recency effect (items from most recent pages remembered at higher rates), suggesting successful memorization of information.

Note that the distribution ratio was further reduced from 1:20 (one window per screen vs. 20 printed out pages on a wall) in Experiment 2a to 1:4 (one quadrant per screen vs. four quadrants per screen) in Experiment 2b. In terms of external validity, large walls are not always available and it may be that four equal-sized spaces in a 17″ screen is an effective balance of screen space and readability. Also, if the distributed display time advantage disappears when the distribution ratio is reduced to 1:4, the applicability of this effect will be limited since many users still have only a single 17″ screen and there is an issue of how to deploy screen space effectively, or perhaps whether to move to a mobile platform with a smaller screen. We expected a decrease in the size of the time benefit in Experiment 2b. However, we predicted that a statistically significant time advantage would still be observed with the distribution ratio of 1 to 4, given that the information access cost is noticeably higher in the stacked display than that of the four distributed pages display. In the literature, even smaller differences in information access cost were sufficient to elicit strategy change (Gray and Boehm-Davis, 2000; Gray and Fu, 2004).

4.1. Method

4.1.1. Participants

Forty-six undergraduates from the University of Pittsburgh participated for course credit and were randomly assigned to condition (23 in distributed with 13 females, and 23 in stacked with 15 females). Ages ranged from 18 to 34 years, with a mean of 19.

4.1.2. Design

The study had a between-subject design with two conditions (distributed vs. stacked) defined by the display layout. As in Experiment 2a, there were two sessions: practice and task. Eye-movements were recorded only during the task session. Eye fixation data was extracted using fixation thresholds of 100 ms and 30 pixels. Also, only the gazes that fall into screen areas with task content (e.g., index, tables, and graphs) were included as on-task fixations. The dependent measures were the duration and frequency of page-turning actions, the sum of off-screen gaze durations longer than 2 s (i.e., presumably, note-taking time as a secondary task), the sum and average fixation duration on the screen content information, and overall behavioral measures such as practice time, task time, task accuracy, and cognitive load.

4.1.3. Materials and procedure

The task was the same as in Experiment 2a except that the number of information pages was lengthened slightly based on pilot testing to maximize power of the study within an hour-long participant session: 12 day with temperature and snow information for each day. Recording eye-movements required fitting both conditions within a 17″ monitor; this physical constraint changed the condition contrast from one full page vs. 20 full pages to one full page (i.e., stacked) vs. 4 full pages (i.e., distributed). In the stacked condition, one piece of information (snow or weather) was presented per window. In the distributed condition, four information pages were presented per window (snow and weather for two days). Thus, there were five windows to view in the distributed condition and 20 windows in the stacked condition. The display size of an information "page" (similar to the pages in Experiment 2a) across the conditions was equal to the one-fourth of the screen; see Fig. 6. In the stacked condition an information page was displayed in the top-left quadrant, and in the distributed condition an information page was displayed in each of the four quadrants. Participants could browse pages by clicking on a table of contents on the top; the information access cost in the stacked condition was the same as in Experiment 2a (i.e., information is a page-turn away by clicking) and higher than in Experiment 1 (i.e., clicking takes more time and effort than hovering).

The overall procedure was the same as with Experiment 2a, except that a 9-point scale survey measuring cognitive load (perceived difficulty and mental effort) was implemented after each of the two sessions (i.e., practice and task) and the computer was the display medium for both conditions. Specifically, students were asked to rate how easy or difficult the task was and how much mental effort the task took on 9-point scales, ranging from very, very easy (1) to very, very difficult (9) and from very, very low mental effort (1) to very, very high mental effort (9). Among the three commonly used cognitive load measures (i.e., subjective rating scale, dual task paradigm, and physiological test), this subjective rating scale was used based upon its well-documented validity and reliability (Ayres, 2006; Marcus et al., 1996; Paas, 1992; Paas et al., 2003; Paas and van Merriё]nboer, 1994) and ease of integration into an already complex paradigm.

Eye-movements were recorded with a Tobii 1750 remote eye-tracker. The 17″ monitor's screen resolution was 1280 × 1024. The system runs at a constant frame-rate of 50 Hz. The approximate distance between the screen and participant was 25″.

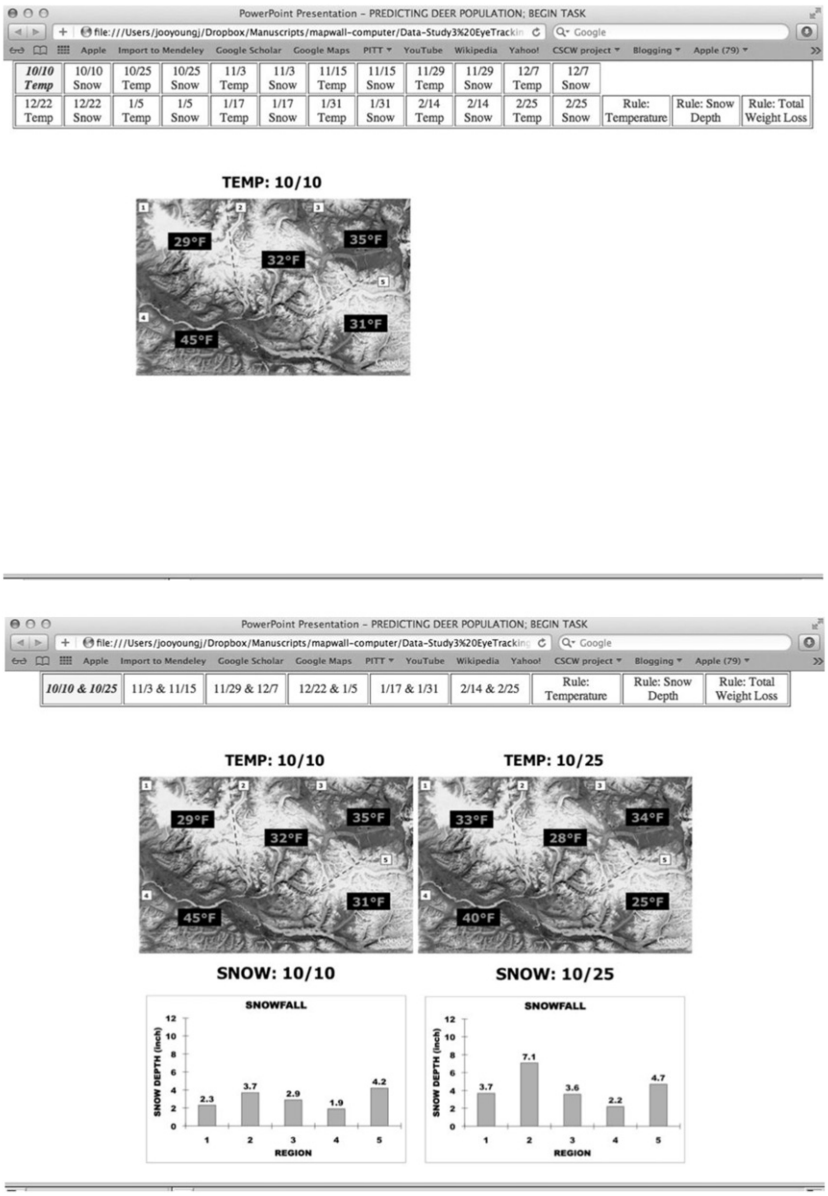

Fig. 6. Experiment 2b: Spatial layout in the stacked condition (upper panel) and four visible pages in the distributed condition (lower panel).

4.2. Results and discussion

Six participants were excluded from analyses as outliers (accuracy lower than 50% correct), leaving 20 participants per condition.

Participants in the two conditions did not differ in practice time, F (1, 38) = 1.48, MSE = 5.69, p = .23. Consistent with the Experiment 1 and 2a, participants in the distributed condition took 20% less time (M = 14.9 min, SE = 0.9) to complete the task than participants in the stacked condition (M = 18.1 min, SE = 1.2), F (1, 38) = 4.31, MSE = 98.31, p = .045, Cohen's d = 0.67. Again, this time difference was not due to a speed/accuracy tradeoff: participants were equally accurate, F (1, 38) < 1, 81% (SE = 3%) in the distributed and 82% (SE = 3%) in the stacked condition.

One participant in the stacked condition who answered 1 on all the 9-point scale questions was excluded from the cognitive load analysis. For the remaining participants, participants in the distributed condition (M = 2.5, SE = 0.2) reported lower cognitive load for the task, compared to those in the stacked condition (M = 3.1, SE = 0.2), F (1, 37) = 4.12, MSE = 98.31, p = .05, Cohen's d = 0.66. No significant difference was observed in the practice session, F (1, 37) = 1.56, MSE = 0.02, p = .22.

Two participants in the distributed and one in the stacked condition with outlier fixation durations were removed (more than two standard deviations above or below the mean), leaving 18 distributed condition and 19 stacked condition participants for analyses of the eye-tracking data.

4.2.1. Page-turning actions

As predicted, stacked display users made far fewer page transitions/revisits (M = 11, SE = 1) than the distributed display users (M = 54, SE = 5),[2] F (1, 36) = 80.21, MSE = 17065.20, p = .000, Cohen's d = 3.39. Given a similar number of total fixations in the stacked condition (M = 1128, SE = 63) and in the distributed condition (M = 1046, SE = 65), the larger number of page transitions made in the distributed condition is consistent with the memorization hypothesis; the pattern of behavior is consistent with the idea that distributed display users would rather choose to be a verifier due to the relatively lower information access cost. Further, the total time spent on page-transitions was three times longer for the distributed display users (M = 1.5 min, SE = 0.2) than the stacked display users (M = 0.5 min, SE = 0.1), F (1, 36) = 14.58, MSE = 9.11, p = .001, Cohen's d = 1.36. The results clearly rule out the possibility that page transition actions are the main source of the increased time in the stacked display condition.

[2]The mean and SEs reported here excludes first visits, which is 20 (i.e., 20 pages of information) for both conditions.

4.2.2. Multi-tasking

To measure time spent on note-taking activities, presumably the most prominent secondary task after the main task of processing screen information, off-screen gaze durations longer than 2 s were summed. The stacked display users spent significantly more time on note-taking (M = 10.8 min, SE = 1.1) than the distributed display users did (M = 7.9 min, SE = 0.7), F (1, 36) = 4.59, MSE = 72.96, p = .039, Cohen's d = 0.73. This time difference of ~3 minutes appears to account for 75% of the total time difference of ~4 minutes, but only a little more than half of the remaining time effect to be explained when additional time for the stacked condition on page turning is included. Further, the note-taking time difference is more descriptive than explanatory: Why do stacked display users spend more time on note-taking given the fact that they were working on the exact same problem as were the distributed display users? As discussed above, we suggest that the less frequent page transitions and the longer note-taking time in the stacked display condition are both caused by the tendency to better encode information pieces (e.g., memorization) as a reaction to the relatively higher information access cost of stacked displays. If this reactive memorization account is correct, the mean fixation durations on the screen content should be longer in the stacked display condition than in the distributed display.

4.2.3. Memorization strategy

The fixation duration data was analyzed as blocks of four information pages (minimal page size for the distributed condition). Both the sum (Fig. 7A) and average (Fig. 7B) fixation durations of first visits to the pages (i.e., excluding any returns) demonstrate longer fixation durations in the stacked condition. Mixed 2 (between: distributed vs. stacked) × 6 (within: blocks of four content pages) ANOVAs found significant differences between conditions for sum of fixation durations per four-page block: F (1, 35) = 12.13, MSE = 2.053×10^9, p = .001, Cohen's d = 0.87, and mean fixation durations, F (1, 35) = 6.33, MSE = 48675, p = .017, Cohen's d = 0.69. The stacked condition appears to have led participants to spend more time encoding information (e.g., memorizing) at the levels of whole pages and individual information pieces.

4.2.4. Display resolution

The effect of display resolution can be assessed by comparing the stacked displays in Experiment 2a (n = 21) and 2b (n = 20). There was no significant difference between having one information page occupy a full screen and having that same page within just a quadrant of the screen in terms of task completion time and accuracy. Participants with the full screen view (M = 19.7 min, SE = 1.6) took a similar amount of time to complete the task to participants with the quadrant view (M = 18.1 min, SE = 1.2), F (1, 39) = 0.64, p = .43. Against the idea that higher display resolution would yield better performance, participants with the full screen view (M = 77%, SE = 4%)

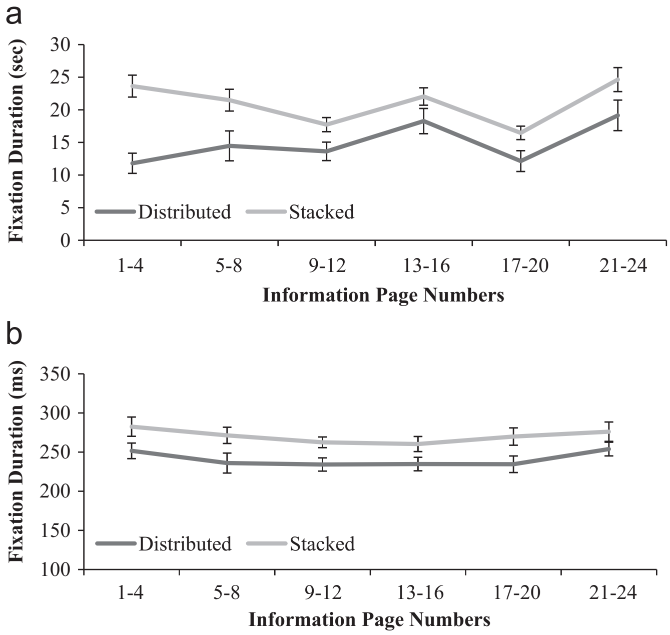

Fig. 7. Experiment 2b: (a) sum of fixation durations and (b) average fixation duration as a function of four-page blocks.

were slightly less accurate than participants with the quadrant view ($M=82\%$, $SE=3\%$), but the difference was not significant, F (1, 39)$=1.32$, $p=.26$. Overall, this data suggests that the smaller resolutions used in Experiment 2b to create 1:4 distributed displays were not a challenge for participants. However, further reductions in resolution to put further increase distribution ratios within a 17″ monitor may become a challenge.

5. General discussion

The three studies presented here found a robust and regular slowing effect of stacked displays over distributed displays. The regularity of the effect across varying materials and interfaces suggests that the effect was caused by the spatial arrangement of information per se rather than by negative effects of animation, advantages of paper over computer, or peculiarities of particular interface organizations across stacked or distributed displays. Further, these studies suggest that differing information access cost associated with information layout affects the probabilities of adopting different micro-strategies. The stacked view display makes it harder to access information that needs to be integrated and a problem-solver can compensate either by slowing down and memorizing information as observed in the deer task (Experiment 2a and 2b) or by frequently going back and forth between pages as observed in the weather interface (Experiment 1).

The choice of memorization or display search strategy likely can be explained by relative access costs; hovering vs. clicking. People in the weather interface may have been more likely to become verifiers because the interface made flipping maps very easy through hovering. By contrast, people in the deer task may have been more likely to become memorizers because they had to click and wait hundreds of milliseconds to access a content page. However, regardless of which strategy was used, the stacked display cost users a significant amount of time that could have been saved through use of a distributed display.

Can such small interface differences create such large effects on users? Seemingly small differences at each low level step of behavior accumulate across a task, particularly when the task is complex and long. Gray and Boehm-Davis (2000) demonstrated that even fairly simple interactive behaviors such as mouse movements and button clicks are affected by small features of an interface. Moreover, users were found to selectively deploy micro-strategies (i.e., combinations of mouse movements and button clicks) that add or subtract milliseconds from routine actions but collectively have large effects on total task time. Proportionally, the time required by the four micro-strategies reviewed in their study ranged from 22% to 560% having the slowest micro-strategy as the reference point.

Although we argue that the selection of different micro-strategies is the primary factor that drives the overall effect on performance and intervening cognitive processes, it is likely not the only factor underlying the effect of distributed versus stacked displays on task time. The working memory load resulting from memorization, the cost of manual interactivity (e.g., page-turning action), and the existence of concurrent non-visual tasks (e.g., keeping track of current state and goal state and taking notes) may also play a role (Wickens and McCarley, 2008), and there was some support for each of these factors in our data.

However, it important to note that the pattern of results from page transitions, note taking, and eye fixation durations in Experiment 2b was exactly matched to what the overall memorization strategy hypothesis predicted. As a way to relieve the burden of relatively higher information access cost in the stacked displays (i.e., information is a page-turn away rather than a head-turn), users spent more time on inspecting each information piece and writing the information down, and consequently they saved a considerable number of page transitions and revisits.

Further, we argue that the fixation duration effect is likely driven mainly by the time to encode information rather than other changes in processing strategies of the task itself (e.g., conducting different task computations) for three reasons. First, the task fixations were very short in both conditions, not consistent with complex mental calculations. Second, the cognitive decisions required by the deer task were relatively simple (one-step) rules. Third, in recall test results from a previous study using the deer task (Jang and Schunn, 2012), the stacked display users showed a tendency of better memory for information pieces that they have just processed. In the deer task, exact temperature and snow depth values are not useful when the problem-solvers moves to the next page. Therefore, it is unlikely that the stacked display users tried to memorize any information on purpose as part of a different complex task strategy; rather, their better memory seems to be a consequence of an implicit effort for better online encoding.

The current research has several practical implications. First, relative to how easy it is to shift from using a stacked display to using a distributed display (e.g., printing and distributing pages on a table or using two windows side-by-side on a screen), the distributed display time advantage is fairly large. The more complex and integrative a task is, the larger the effect will presumably be. Thus, we believe problem-solving time on many tasks could be improved by the transition from stacked to distributed displays, either with larger monitors (or larger paper) or multiple monitors, or by simply showing more content in smaller form within a given interface. Yet, making transitions from distributed displays to integrative displays is often up to designers rather than users, and evokes new design problems such as how to deal with increasing display density. Further, finding the boundary of the effect is necessary (i.e., how many pieces of distributed information are beneficial) to make precise practical suggestions in any given context. Relatedly, levels of expertise have not been examined in the current findings, which would be an important factor in real world applications.

Second, research about spatial arrangement of information has implications for the basic user interfaces found within an overall operating system. For example, operating systems are evolving to incorporate multiple virtual spaces to help users make the most of the fixed asset of a given screen. The current work suggests that operating systems should focus on minimizing visual information access costs, for example by dividing a given screen, rather than placing content across slower-to-access virtual screens. However, one should note that, when dividing spaces, the issue of resolution change should be addressed because distributing displays in a fixed space may cause loss of resolution. Although the loss of resolution caused by dividing 17″ inch screen into four quadrants was not found to be a critical issue in the current experiments, it may become problematic for higher distribution ratios or higher task complexity.

Third, the benefit of distributed displays was found not only in the context of complex problem solving but also in the context of learning (Jang et al., 2011). How to organize and distribute information for effective multimedia learning appears to be as vital a question as which content to include.

Acknowledgments

The research was supported by grants from the Office of Naval Research: N000140610053 to the third author and N0001406WX20091, N0001412WX21502, and N0001412WX3008 to the fourth author.

References

Andrews, C., Endert, A., North, C., 2010. Space to think: large high-resolution displays for sensemaking. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems. ACM, Atlanta, Georgia, USA, pp. 55–64.

Ayres, P., 2006. Using subjective measures to detect variations of intrinsic cognitive load within problems. Learning and Instruction 16, 389–400.

Bailey, B., Biehl, J., Cook, D., Metcalf, H., 2008. Adapting paper prototyping for designing user interfaces for multiple display environments. Personal and Ubiquitous Computing 12, 269–277.

Bly, S.A., Rosenberg, J.K., 1986. A comparison of tiled and overlapping windows, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Boston, Massachusetts, United States.

Carpenter, P.A., Just, M.A., 1983. What your eyes do while your mind is reading. In: Rayner, K. (Ed.), Eye Movements in Reading: Perceptual and Language Processes. Academic Press, New York, pp. 275–307.

Cockburn, A., McKenzie, B., 2001. 3D or not 3D?: evaluating the effect of the third dimension in a document management system. In: Proceedings of the SIGCHI Conference on Human factors in Computing Systems. ACM, Seattle, Washington, United States, pp. 434–441.

Czerwinski, M., Robertson, G., Meyers, B., Smith, G., Robbins, D., Tan, D., 2006. Large Display Research Overview, CHI 2006. ACM, Montreal, Quebec, Canada, pp. 69–74.

Dunbar, K., 1995. How scientists really reason: scientific reasoning in real-world laboratories. In: Sternberg, R.J., Davidson, J.E. (Eds.), The Nature of Insight. MIT Press, Cambridge, MA, pp. 365–395.

Gray, W.D., Boehm-Davis, D.A., 2000. Milliseconds matter: an introduction to microstrategies and to their use in describing and predicting interactive behavior. Journal of Experimental Psychology: Applied 6, 322.

Gray, W.D., Fu, W.T., 2004. Soft constraints in interactive behavior: the case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. Cognitive Science: A Multidisciplinary Journal 28, 359–382.

Grudin, J., 2001. Partitioning digital worlds: focal and peripheral awareness in multiple monitor use., CHI 2001: The ACM Conference on Human Factors in Computing Systems, pp. 458–465.

Jang, J., Schunn, C.D., 2012. Performance benefits of spatially distributed versus stacked information on integration tasks. Applied Cognitive Psychology 26, 207–214.

Jang, J., Schunn, C.D., Nokes, T.J., 2011. Spatially distributed instructions improve learning outcomes and efficiency. Journal of Educational Psychology 103, 60–72.

Kroft, P.D., Wickens, C.D., 2002. Displaying multi-domain graphical database information. Information Design Journal 11, 44.

Marcus, N., Cooper, M., Sweller, J., 1996. Understanding instructions. Journal of Educational Psychology 88, 49–63.

Morgan, P.L., Patrick, J., Waldron, S.M., King, S.L., Patrick, T., 2009. Improving memory after interruption: exploiting soft constraints and manipulating information access cost. Journal of Experimental Psychology: Applied 15, 291.

Mynatt, E.D., Igarashi, T., Edwards, W.K., LaMarca, A., 1999. Flatland: New Dimensions in Office Whiteboards, CHI '99. ACM, Pittsburgh, PA, pp. 346-353.

Ni, T., Bowman, D.A., Chen, J., 2006. Increased Display Size and Resolution Improve Task Performance in Information-Rich Virtual Environments. Canadian Information Processing Society, pp. 139–146.

Paas, F., 1992. Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. Journal of Educational Psychology 84, 429–434.

Paas, F., Tuovinen, J., Tabbers, H., Van Gerven, P., 2003. Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist 38, 63–71.

Paas, F., van Merriё]nboer, J.J., 1994. Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. Journal of Educational Psychology 86, 122–133.

Robertson, G., Czerwinski, M., Baudisch, P., Meyers, B., Robbins, D., Smith, G., Tan, D., 2005. The large-display user experience. IEEE Computer Graphics and Applications, 44–51.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., Dantzich, M.v., 1998. Data mountain: using spatial memory for document management, Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology. ACM, San Francisco, California, United States, pp. 153–162.

Robertson, G., Dantzich, M.v., Robbins, D., Czerwinski, M., Hinckley, K., Risden, K., Thiel, D., Gorokhovsky, V., 2000. The task gallery: a 3D window manager, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, The Hague, The Netherlands, pp. 494–501.

Staggers, N., 1993. Impact of screen density on clinical nurses' computer task performance and subjective screen satisfaction. International Journal of Man-Machine Studies.

Sweller, J., Chandler, P., Tierney, P., Cooper, M., 1990. Cognitive load as a factor in the structuring of technical material. Journal of Experimental Psychology: General 119, 176–192.

Tan, D.S., Gergle, D., Scupelli, P.G., Pausch, R., 2003. With Similar Visual Angles, Larger Displays Improve Spatial Performance, CHI 2003. ACM, Ft. Lauderdale, Florida, pp. 217–224.

Tan, D.S., Gergle, D., Scupelli, P.G., Pausch, R., 2004. Physically large displays improve path integration in 3D virtual navigation tasks, CHI 2004: the ACM Conference on Human Factors in Computing Systems. ACM, pp. 439-446.

Tan, D.S., Stefanucci, J.K., Proffitt, D.R., Pausch, R., 2001. The Infocockpit: Providing Location and Place to Aid Human Memory. ACM, pp. 1–4.

Tarmizi, R., Sweller, J., 1988. Guidance during mathematical problem solving. Journal of Educational Psychology 80, 424–436.

Trafton, G., Kirschenbaum, S.S., Tsui, T.L., Miyamoto, R.T., Ballas, J.A., Raymond, P.D., 2000. Turning pictures into numbers: extracting and generating information from complex visualizations. International Journal of Human-Computer Studies 53, 827–850.

Trafton, J.G., Trickett, S.B., 2001. Note-taking for self-explanation and problem solving. Human-computer interaction 16, 1–38.

Trickett, S., Trafton, J., 2007. 'What if': the use of conceptual simulations in scientific reasoning. Cognitive Science: A Multidisciplinary Journal 31, 843–875.

Wickens, C.D., 2000. Human factors in vector map design: the importance of task-display dependence. Journal of Navigation 53, 54–67.

Wickens, C.D., Carswell, C.M., 1995. The proximity compatibility principle: its psychological foundation and relevance to display design. Human Factors 37 (422), 473.

Wickens, C.D., Kroft, P., Yeh, M., 2000. Data base overlay in electronic map design: testing a computational model, IEA 2000/HFES 2000 Congress. Human Factors and Ergonomics Society, Santa Monica, CA, pp. 451–454.

Wickens, C.D., McCarley, J.S., 2008. Applied Attention Theory. Taylor & Francis, Boca Raton, FL.

Yeh, M., Wickens, C.D., 2001. Attentional filtering in the design of electronic map displays: a comparison of color coding, intensity coding, and decluttering techniques. Human Factors: The Journal of the Human Factors and Ergonomics Society 43, 543–562.