Journal of Experimental Psychology: Applied

A Framework for Unpacking Cognitive Benefits of Distributed Complex Visual Displays

Jooyoung Jang and Christian D. Schunn

Online First Publication, June 16, 2014. http://dx.doi.org/10.1037/xap0000022

Journal of Experimental Psychology: Applied

1076-898X/14/$12.00 http://dx.doi.org/10.1037/xap0000022

A Framework for Unpacking Cognitive Benefits of Distributed Complex Visual Displays

Jooyoung Jang
University of California, Los Angeles

Christian D. Schunn
University of Pittsburgh

What are the advantages and disadvantages of spatially stacked (i.e., when information sources are presented side-by-side) versus distributed (i.e., when information sources are sitting on top of one another with only the top source fully visible) organizations of information? We introduce a new theoretical decomposition of these advantages and disadvantages (information internalization, information access, and information externalization costs), along with a new analytic technique for measuring each theoretical aspects using eye tracking. Thirty-eight scientists-in-training solved a complex data interpretation problem using either a distributed or a stacked display. Display format influenced all 3 factors, but in opposing ways: stacked displays increase internalization and externalization costs but decrease information access costs. The framework reveals trade-offs among the 3 factors that can be precisely characterized to guide interface user design and optimization.

Keywords: information displays, integrative reasoning, information access costs

Given the limited cognitive capacity that humans have (for a review, see Baddeley, 2003), it is rather surprising that people can manage complex tasks and systems such as power plants and air traffic control. For instance, the space shuttle cockpit control panels are overwhelming to most people; the cockpit control panels show how complex visual tasks can be and yet a trained mind can handle these displays.

What are the basic interactions between problem solver and complex visual task? We argue that there are three common and key aspects related to the use of the large amounts of complex visual information: storing visually extracted information in our memory (information internalization) for rapid reuse later, seeking out externally available information when needed (information access), and external note taking and note manipulation (information externalization). In short, we deal with complex visual tasks by internalizing, accessing, and externalizing information. The three factors have typically been studied in isolation (Trafton & Trickett, 2001) and occasionally in pairs (Fu & Gray, 2000, 2006) to reveal cognitive bases of human information processing, but no study has examined the effects of the three factors together to consider the tradeoffs among the factors.

In addition to developing and testing this new integrative framework, the current study extends the empirical literature in two additional ways. First, we establish that strategic tradeoffs between internalized and externalized information happen for complex tasks involving many pages of information and solved in tens of minutes, in sharp contrast to the conceptually similar work by Gray and colleagues (Fu & Gray, 2006; Gray & Boehm-Davis, 2000; Gray & Fu, 2004) that involved only more much basic tasks taking place in the order of seconds or milliseconds. That is, the current study demonstrates that microstrategic behaviors happening at the millisecond level can collectively have a large impact on complex tasks as well. Second, this study examines intermediate scientists attempting to understand and interpret experimental results, a highly complex, and yet everyday task in science, which is unfortunately rarely studied in the literature. A large proportion of the experimental literature in this area involves studies of relative novices doing simple tasks. For example, many "science" tasks used in psychology labs and science education research are far from what scientists actually do (Chinn & Malhotra, 2002), simply because the pool of undergraduates and young students available for experimentation cannot do more typical science tasks. Further, undergraduates performing science tasks can be very different from scientists doing actual science tasks, just as beginning driver behaviors can be very different from racecar driver behaviors. The current research directly examines the effect of spatial organization of scientific information on a relatively large set of scientists, but still in a carefully controlled study.

We will first describe the issue of spatial organization in information display and its impact on performance. We then discuss how the three-factor framework can be systematically examined and applied to explain differences in information processing observed between the two display formats.

Spatial Organization of Visual Information for Complex Tasks

Different visual displays of the same information (i.e., informationally equivalent displays) can often yield drastically different

Jooyoung Jang, Graduate School of Education and Information Studies, University of California, Los Angeles; Christian D. Schunn, Department of Psychology, Learning Research, and Development Center, University of Pittsburgh.

Correspondence concerning this article should be addressed to Jooyoung Jang, Graduate School of Education and Information Science, University of California, Los Angeles, Moore Hall, 2005E, 405 Hilgard Avenue, Los Angeles, CA 90095-1521. E-mail: jooyoungjang15@gmail.com

task performance because different visual displays are not computationally equivalent (Carpenter & Shah, 1998; Gattis & Holyoak, 1996; Hegarty, Canham, & Fabrikant, 2010; Kroft & Wickens, 2002; Larkin & Simon, 1987; Ratwani, Trafton, & Boehm-Davis, 2008; Sanfey & Hastie, 1998). For example, two sets of informationally equivalent graphs were found to be computationally different, and the computational advantages of a new representation could even outweigh the disadvantages of unfamiliar representations (Peebles & Cheng, 2003). If visual display design has a significant impact for simple displays of small-scale data (e.g., a graph), then the impact could be even greater for complex displays of data. With regular technology-based change in the visual presentation and complexity of tasks, research on the role of visual display design in complex tasks is essential (Hegarty, 2011).

To make our larger framework concrete, we consider two particular types of spatial organization: spatially distributed displays (i.e., when information sources are presented side-by-side) and stacked displays (i.e., when information sources are sitting on top of one another with only the top source fully visible). They are common display formats that cleanly differ only in terms of organization. Note that both formats are kinds of separated displays, in contrast to integrated displays (Jang & Schunn, 2012). Integrated displays not only accommodate multiple sources of information in a single display but also present them superimposed so as to reduce the load of mental integration. In contrast, separated displays (i.e., both distributed and stacked displays) present multiple sources of information as separate pages of information, thus presumably the load of mental integration stays equivalent. The difference between the two formats lies in whether the pages are distributed in spatially close proximity or stacked on top of the each other.

Although advantages of integrated displays have been shown in many studies (Chandler & Sweller, 1991; Kroft & Wickens, 2002), the effects of distributed versus stacked displays has been less well studied. Several prior studies involving undergraduates have found large performance benefits of spatially distributed displays over stacked displays across studies in instruction designs and problem-solving domains that involve information integration (Jang & Schunn, 2012; Jang, Schunn, & Nokes, 2011; Jang, Trickett, Schunn, & Trafton, 2012). For example, undergraduate students solved integrative problems almost two times faster without any loss of accuracy when information or learning instructions were provided in a distributed format (e.g., 20 information pages printed and pinned on a wall or four pages of instructions printed on 11” $\times$ 17” paper); we have coined this phenomena the *distributed display time advantage*. It has been consistently found across three different tasks, but a coherent explanation for the effect is lacking, as the underlying mechanisms appeared to vary across studies, as we will discuss in the next section.

A Three-Factor Framework for Understanding Visual Information Processing

To unpack the effects of such displays on performance, we present and test a new tripartite framework for understanding the nature of visual information processing that happens during integrative reasoning. We argue that the three factors account for the effects of display organization on task performance—internalizing, accessing, and externalizing information—and that each of the three factors can be indexed by eye tracking data (see Figure 1). We further argue that to understand the overall effects of display organization on performance, the tradeoffs among the factors must be more precisely examined.

To illustrate the framework concretely, we apply it to the stacked versus distributed display contrast by considering three explanations for the distributed display time advantage: Do stacked displays slow down problem solvers by making problem solvers into slow memorizers (i.e., incur a higher information internalization cost), frequent page flippers (i.e., incur a higher information access cost), or note-takers (i.e., incur a higher information externalization cost)? The three possibilities are unpacked below.

First, people may choose to slow down and memorize information to keep it available in their heads rather than visually search again and reencode. Gray and his colleagues (Fu & Gray, 2006; Gray & Boehm-Davis, 2000; Gray & Fu, 2001, 2004) have shown with very simple tasks that people can shift toward memorizing when the necessary information is even just a click away. That is, when the cost of accessing external information increases (as in the case of stacked displays), people will tend to memorize information to make it readily accessible in the head (i.e., use a memorization strategy). In terms of performance accuracy, the memory strategy selection can be construed as an adaptive choice balancing accuracy and effort, because information in the world is accurate but that in the head may not be. For example, participants made more errors in a given task when they adopted the memorization strategy, but with a reduction in task time (Gray & Fu, 2004). While in the much longer scales of tasks in science there may be a greater emphasis on accuracy than on speed, there will also be the same overall tradeoff at the cumulative level because the set of possible analyses is infinite and time is finite.

In support of this first underlying factor, a recent eye-tracking study of undergraduates working on a simple problem solving task indeed suggested that problem-solvers adopt an information memorization strategy in stacked display conditions, and this memorization time could account for a large part of the previously seen stacked display time disadvantage (Jang et al., 2012). Eye tracking was used to measure the length of time eye gazes remained fixed on a location (called a *fixation duration*). Fixation durations can be a direct measure of the memorization hypothesis, because each fixation duration serves as an online measure of information processing, similar to the eye-mind assumption and immediacy as-

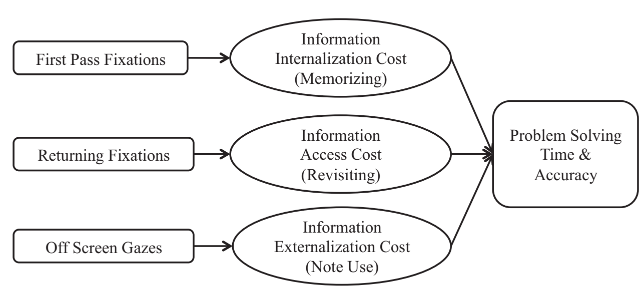

Figure 1. A tripartite framework of unpacking cognitive mechanisms underlying information visualization effects on integrative reasoning using eye-tracking data.

sumption used in eye-tracking studies of reading processes (Carpenter & Just, 1983). If stacked display users experience relatively higher information access cost, they should try to overcome the cost by spending extra encoding time to facilitate later retrievals from memory (Morgan, Patrick, Waldron, King, & Patrick, 2009). As predicted, participants in the stacked display condition fixated significantly longer on information pieces on each page during their first visit to each page (called *first-pass fixations*) than those who solved the same problem using the distributed display.

Second, a stacked display may lead to more frequent revisiting of information, relying on external/in the world information (i.e., use a perceptual-motor strategy). As demonstrated in Kroft and Wickens (2002), student pilots with spatially stacked display produced significantly more toggles between the two information sources, compared to those that had integrated displays. Similarly, student weather forecasters (Trafton, Trickett, Schunn, & Kirschenbaum, 2007) constrained to a 17-inch desktop revisited maps six times more often than those who had the map wall display (i.e., maps of meteorological information printed out and stuck on a large wall). This information-revisiting factor can be measured using returning fixations, the fixations made during return visits to previously visited information pages.

Third, in stacked displays, note taking can be used to keep track of critical information within and across pages, another type of perceptual-motor strategy. Going back to these notes may be more efficient than hunting through original information source pages, especially when the information is effortful to locate (e.g., with stacked displays). With eye-tracking, note-taking time can be approximated by summing long off-screen gaze durations (i.e., those that are clearly not eyeblinks).

Note that the effect of information display organization is likely due to a combination of these three explanations and that the factors tradeoff against one another. Specifically, the slow memorizer effect will systematically influence the other two factors (i.e., memorizers should revisit pages less often and may take/consult notes less often). Given these trade-off factors, subtle access costs differences in the display conditions can change the overall performance effect observed, and indeed the overall framework suggests that stacked displays could be beneficial under some circumstances. For example, if stacked displays may involve a small cost to memorize but large reduction in return visits or note use, then stacked displays will show an advantage.

To see how the framework is useful for better understanding display effects, consider the following seemingly contradictory prior findings. In a lab study, Jang et al. (2012) found that stacked displays lead to increased first-fixation durations (i.e., a sign of increased dependence on internal information), but their study of student weather forecasters found that the stacked display condition led to far more page revisits (i.e., a sign of increased dependence on external information). The difference in effects across studies might be explained by relative access costs from hovering versus clicking. People using the computerized weather interface in a stacked display format likely became verifiers because the interface made flipping maps very easy through hovering (simply holding the mouse over different areas changes the animation content immediately). By contrast, people in the lab study became memorizers, as they had to click and wait hundreds of milliseconds to access a content page. In both cases, the stacked display slowed users down, but the mechanism depended upon different underlying factors (increased revisits or increased memorization).

In sum, we propose three types of cognitive costs in relation to information access efforts that account for the effect of display organization, and we claim that each cost can be indexed by eye-tracking data: information internalization cost (i.e., memorizing) by measuring first time fixations, information access cost (i.e., revisiting) by measuring returning fixations, and information externalization costs (i.e., note use) by measuring off-screen gazes. We test these fine-grained predictions in the context of large group of relative experts working on a complex science task.

Method

Intermediate scientists (i.e., psychology graduate students and postdoctoral students) were given a data interpretation task in two display formats (i.e., distributed vs. stacked) and their performance was compared to examine the effects of display format. Eye-tracking data was used to examine information encoding strategies (i.e., memorization vs. perceptual-motor strategies) and information access costs.

Participants

Thirty-eight psychology graduate students and two postdoctoral students (32 female; age range 23–48) from the University of Pittsburgh and Carnegie Mellon University participated for a compensation of $25. Participants were randomly assigned to either display format, and there were no differences in the level of expertise, determined by survey responses, across the display conditions (see Table 1).

Table 1
Mean, SD, and 95% Confidence Interval (CI) of Expertise Measures Within Each Display Format Condition

	Stacked		Distributed		
Measure	*M* (*SD*)	95% CI	*M* (*SD*)	95% CI	*p* value (η^2)
Year in PhD program	4.2 (2.4)	(3.10, 5.22)	3.9 (2.2)	(2.84, 4.95)	.72 (.004)
Years of experience with behavioral data	4.6 (3.5)	(3.17, 5.99)	4.8 (2.5)	(3.43, 6.25)	.79 (.002)
Years of experience with analysis of variance	4.3 (3.3)	(2.94, 5.70)	4.6 (2.6)	(3.25, 6.01)	.75 (.003)
Number of journal publications	2.6 (2.5)	(0.76, 4.40)	2.8 (5.0)	(0.97, 4.61)	.87 (.001)
Number of nonjournal publications	2.3 (2.5)	(0.87, 3.66)	3.4 (3.4)	(1.98, 4.76)	.26 (.035)
Number of first author publications	0.7 (0.9)	(−0.47, 1.84)	1.6 (3.4)	(0.43, 2.73)	.28 (.033)
Number of publications in last 2 years	1.2 (1.3)	(0.65, 1.78)	1.1 (1.1)	(0.54, 1.67)	.79 (.002)

Design

The independent variable was display format (distributed vs. stacked). In individual sessions, participants solved a data interpretation problem using either a distributed or a stacked display. The task was self-paced. While participants work on the task, task time, page transition logs, and eye movements were recorded. A demographic survey was collected at the end of the experiment. The core dependent variables include task times (i.e., time on information window examination and time on question answering), task solution quality, and patterns of eye-movements.

Materials

Eye-tracker. 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”.

Main task materials. The main task involved data interpretation, examining research hypotheses with quantitative data presented in tables and graphs, and drawing a plausible conclusion. Data interpretation is likely influenced by information display format because it is an integrative task: It involves processing multiple pieces of information (e.g., dependent and independent variables; descriptive and inferential statistics) presented in various formats (e.g., text, tables, graphs, and diagrams) and combining the information into a single coherent story.

To avoid the complicating effects of domain-specific knowledge across differing focal areas of expertise such as cognitive, developmental, clinical psychology, or neuroscience (Schunn & Anderson, 1999), the task materials consisted of phenomena that are understandable to a broad range of research psychologists. The topic and content of the data was selected from *Psychological Science*, a psychology journal that delivers brief research reports of broad interest. Research data from real journal reports were used to maintain the plausibility of the task. The particular topic involved destination memory (e.g., remembering the person to whom one has given information) and source memory (e.g., remembering the person from whom one has received information), examining which memory is more fallible and why. Theories, study designs, and resulting data were adapted from Gopie and MacLeod (2009); Koriat, Ben-Zur, and Druch (1991), and Marsh and Hicks (2002).

The materials included a general description of the research topic, data from two studies that each provides evidence consistent with one of the two hypotheses, and thus contradict each other. A one-page paper handout motivated the general research questions and provided participants with definitions of key concepts such as destination and source memory. On the computer, 13 content pages were available: questions to be answered, Study 1 intro, Study 1 hypothesis, Study 1 methods (1) and (2), Study 1 results, Study 2 intro, Study 2 hypothesis, Study 2 methods (1) and (2), Study 2 results (1), (2), and (3).

The goal of the task was to compose short paragraphs for four questions: whether the hypothesis of the first study was confirmed and why, whether the hypothesis of the second study was confirmed and why, whether the two studies are congruent, and how to reconcile the discrepancies if they are not congruent.

Practice task materials. A simple practice task was developed to familiarize participants with the general procedure of the task. The practice topic was learning with diagrams (i.e., why and when having a diagram improves learning) and seven pages of information were provided. The content was adapted from Willows (1978).

Display formats. The two formats of information display were defined in the following manner. The distributed display divided the 17-inch screen into four equal-sized spaces (see Figure 2). Each space had a drop-down menu with which participants could choose an information page of interest. By contrast, the stacked display presents only one information-page at time, and it presents that information in the top-left one-quarter of the 17-inch screen space, leaving the other three spaces blank.

Procedure

The experiment was done in individual sessions and the session consisted of three components: practice, task, and survey. Each participant was first seated at an eye-tracker (Tobii 1750) with a chin rest. After a brief eye-calibration, participants performed a practice task in the stacked display format. Before the practice started, participants first read aloud a passage describing the topic of the practice problem and confirmed that they understood the topic. Then they were given instructions on how to navigate information pages using a drop-down menu, what the questions to solve are, and that the information window will be available at all times. Also, they were instructed to let the experimenter know when they are done with information window and ready to compose their answers. A blank letter-size paper was provided to each participant to take notes during the task. After the instructions, participants were asked to try the practice problem alone. During the practice, they were allowed to ask questions about the problem content and the procedure.

The main task was done in the same manner as the practice task. Those who participated in the distributed display condition were given additional instructions on how to use the distributed display; a sample distributed information window was shown to participants using the practice material. A new blank paper was provided for note taking. Unlike during practice, participants were not allowed to ask questions about the task content. When they indicated that they had finished with examination of the information windows, eye movement recording was stopped and saved. Then, on another workstation equipped with a keyboard, participants were provided with a new window with two tabs, one to type answers to each question and the other to display the information windows. Also in a self-paced manner, participants composed their answers for each and every question. They were able to refer back to the information window and their notes as they wrote answers.

Results

There were two outliers whose total task time was longer than two standard deviations from the mean, leaving 19 participants (15 females) in each condition.

Task Time

Two primary time measures were available: time spent processing information presented in a given window display (window time) and time spent composing answers while having the infor-

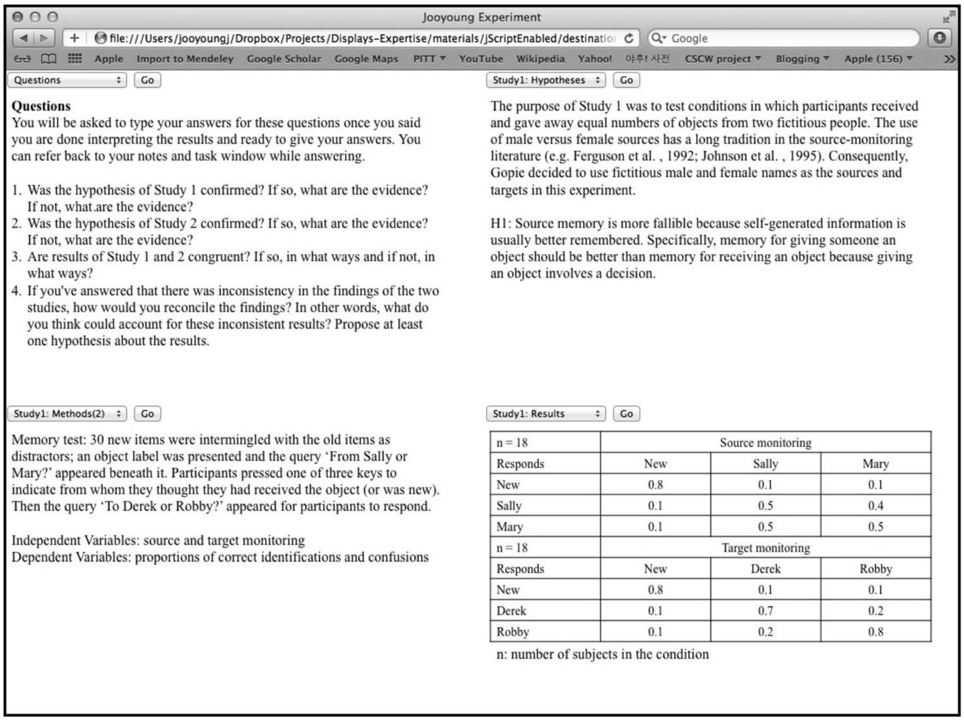

Figure 2. Layout of spatially distributed display in the current experiment.

mation window available on the second tab of a window (answering time). The practice time was not different across formats, $t(36) = 2.38$, $p = .24$, Cohen's $d = 0.38$, thus not included in the analyses. A multivariate analysis of variance was used to test the effect of display format on window time and answering time. Consistent with prior studies of the overall effect, even with more expert participants working on a complex authentic task, distributed display users tended to finish examining information more quickly ($M = 15.7$ min, $SD = 4.5$, 95% confidence interval [CI]: 12.29, 19.07) than stacked display users ($M = 20.1$ min, $SD = 9.3$, 95% CI: 16.70, 23.49), $F(1, 36) = 3.48$, $MSE = 185.24$, $p = .07$, $\eta^2 = .09$, but with high variability in the stacked condition. There was no effect of display format for answering time phase, $F(1, 36) = 0.02$, $MSE = 1.36$, $p = .88$, $\eta^2 = .001$.

Task Solution Quality

Solution quality was evaluated according to the rubric presented in Appendix A. We use the term *quality* rather than *accuracy* because open-ended complex tasks such as data interpretation do not have fixed and simple solutions. Instead, there can be a range of relevant key information that can be extracted from data to support one's analyses and explanations. The coding rubric (see the appendix) focused on salient aggregate results from each study, including some support for the given interpretation, and then providing an integrative account across the two studies. The rubric consists of eight items (1a, 1b, 2a, 2b, 3a, 3b, 4a, and 4b), two for each of the four posed questions. The first six items were assigned 5 points each while the last two items (4a and 4b) were assigned 10 points each because the last question embodies the main overall inference task.

The overall solution quality (max 50) was analyzed using a 2 × 2 analysis of covariance with answering time as a covariate to adjust for speed–accuracy trade-off effects. Distributed display users seem to generate slightly higher quality solutions ($M = 35.5$, $SD = 9.8$, 95% CI: 30.06, 41.08) than stacked display users did ($M = 31.3$, $SD = 13.5$, 95% CI: 25.76, 36.78), but the difference was not significant, $F(1, 35) = 1.25$, $MSE = 175.20$, $p = .27$, $\eta^2 = .04$. Most of the condition difference was shown on the integration questions (4a + 4b), consistent with prior work showing that the distributed display effect was specific to information integration tasks (Jang & Schunn, 2012). This pattern clearly establishes that there was no speed–accuracy trade-off; stacked display users tended to have lower quality solutions and slower task times.

The following sections examine process factors that were predicted to be influenced by the display conditions: internalization of information, external information access cost, and externalization of information.

First-Pass Fixation Durations: Internalization of Information

Two different eye movement measures (i.e., number of fixations and mean of fixation durations) are relevant to examining memorization strategy effects: the number of fixations and the mean fixation duration. Greater memorization could be shown by more fixations and longer fixation. Mean fixation durations was used in a previous study of the distributed display effect (Jang et al., 2012). In the current study, however, the sum of fixation durations (i.e., a composite measure that reflects both the number of fixations and the average fixation durations; sum = number × average) was used because it more appropriately includes various forms of memorization strategies involving either longer or repeated fixations. We only included fixations during the first visit to a page (hence the name, first-pass fixations).

In fact, stacked display users almost tripled the time spent on initial information encoding (M = 6.6 min, SD = 2.7, 95% CI: 5.65, 7.57) than did distributed display users (M = 2.3 min, SD = 1.0, 95% CI: 1.36, 3.28), $t(36) = 6.40$, $p < .001$, Cohen's d = 2.27, revealing a strong effect of display format on memorization/internalization of information.

Return Fixation Durations: External Information Access Cost

The sum of return fixation durations was computed by subtracting the sum of first-pass fixation durations from the sum of total fixation durations, which includes all regressions and returns. Whereas measures of first-pass fixations show how much time and effort people invested during the first time encoding in reaction to the relatively higher information access cost in the stacked display condition, return fixations provide an index of external information access costs. Note that total external information access cost involves planning returns, mouse movements, and eye-saccades, in addition to the fixation duration of the return but those elements are hard to capture using eye data. More importantly, because those other elements are likely a multiple of the return fixations, we treat the return fixations as the tip-of-the-iceberg estimate of external information access costs.

If the distributed display produces relatively lower information access cost (i.e., an eye-turn) than does the stacked display, distributed display users should depend more on regressions and revisits. Thus, it was predicted that stacked display users would show shorter return fixation durations than distributed display users.

As predicted, stacked display users reaccessed information for much less time (M = 2.5 min, SD = 1.6, 95% CI: 1.45, 3.57) than distributed display users did (M = 6.9 min, SD = 2.8, 95% CI: 5.87, 7.99), $t(36) = 6.0$, $p < .001$, Cohen's $d = 2.01$, which clearly demonstrates the impact of relatively higher information access cost in the stacked display condition. The larger time spent revisiting pages in the distributed display might seem counterintuitive in that finding prior information requires more work in the stacked condition. However, this pattern is highly consistent with the memorization hypothesis in that stacked display users appear to rely on their memories rather than external information. Further there seems to be an almost equal tradeoff between time spent on first-pass and return visits between the two conditions (i.e., there was a 4-min difference between conditions for both measures, but in opposite directions).

Off-Screen Gaze Durations: Externalization of Information (Note Use)

Another large possible processing time factor that could be influenced by display layout involves note use (i.e., note taking and processing of notes). Participants were able to take notes freely during the window examination phase and most participants took notes and consulted their notes throughout the task. We use the term *note use* to refer to both of the writing of new notes and the mental processing of existing notes.

Notes reflect another form of (personally constructed) external storage for this complex task. The memorization strategy hypothesis predicts that stacked display users would spend more time taking notes, essentially as an alternative form of (external) memorization. If stacked display users experience the relatively higher information access cost, they should take notes to reduce the need of revisits and thus to take advantage of cognitive offloading.

To examine the note use time, off-screen gaze durations longer than 2,000 ms were collected and summed up as a proxy measure (i.e., sum of the time spent on various activities beyond online visual information processing). In general, off-screen gaze durations longer than the normal blink duration of 300–400 ms could have been used as an index of time users are engaged in off-screen activities. A more conservative threshold of 2,000 ms was used, however, to reduce the effect of activities other than note use (e.g., looking at a computer clock to check the time, looks to the mouse after a hand was taken off of it, or head scratches).

As predicted by the memorization hypothesis, stacked display users spent twice as much time on notes (M = 5.2 min, SD = 5.6, 95% CI: 3.27, 7.17) as those who used distributed displays (M = 2.3 min, SD = 2.1, 95% CI: 0.38, 4.28), $t(36) = 2.12$, $p = .04$, Cohen's d = 0.76. It is interesting that the between-subjects variability was particularly high on this measure in the stacked display condition, which was likely responsible for the high total time variability in the stacked condition.

One may wonder whether it is plausible to consider all off-screen gaze durations longer than 2,000 ms as a valid estimate for the amount of notes taken, as opposed to simple closed eye reflection, for example. To validate the measure, the amount of notes for each participant was coded on a 4-point scale: little (*only a couple of lines*), light (*5–10 lines*), medium (*10–20 lines*), and heavy notes. The Pearson correlation at the participant level between the total off-gaze durations and the amount of notes (as a 4-point scale) was $r = .70$, $n = 38$, $p < .001$. Further, a t test on the amount of notes showed a consistent pattern found in the off-screen gaze duration. Stacked display users tended to take larger amounts of notes (M = 2.6, SD = 1.2, 95% CI: 2.02, 3.14), than did distributed display users (M = 1.9, SD = 1.2, 95% CI: 1.40, 2.50), $t(36) = 1.63$, $p = .11$, Cohen's $d = 0.53$. Interestingly, the within-group variability in amount of notes was similar across conditions, suggesting the variability was more in how much notes were used, rather than in how much content was externalized.

Integrating the Effects of the Three Main Time Elements

The overall difference observed in the window-examining time can be explained by the combination of effects across the three factors examined thus far: time spent on first-pass fixations, time spent on return fixations, and time spent on note use. As the memorization hypothesis predicted, stacked display users spent more time on first time encoding, less time on return fixations, and more time on note use (see Figure 3).

The stacked display produced more information internalization due to its relatively higher information access cost. Note that a similar pattern could be observed when analyzing the number of first-pass and return fixations, rather than fixation durations. Stacked display users made more first-pass fixations and fewer return fixations whereas distributed display users showed exactly

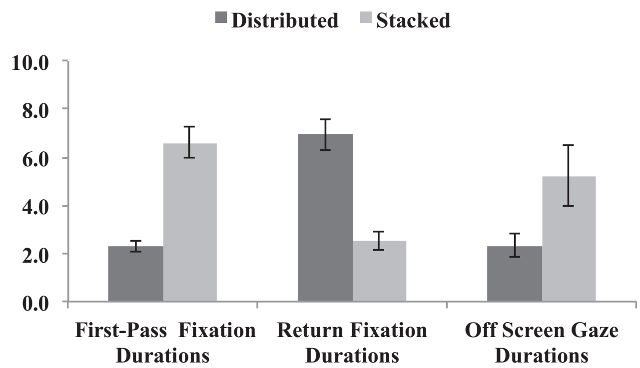

Figure 3. Means of first-pass fixation durations, return fixation durations, and off-screen gaze durations by display format with standard error bars.

the reverse pattern: number of first-pass fixations, $F(1, 36) = 67.65$, $MSE = 9622391.68$, $p < .001$, $\eta^2 = .65$, and number of return fixations, $F(1, 36) = 49.72$, $MSE = 10948358.1$, $p < .001$, $\eta^2 = .58$. Similarly, analyses of page transition frequency (i.e., a fixation-based measure including page-turning action: how many times participants moved their eyes to different pages) showed that stacked display users made far fewer page transitions ($M = 24.9$, $SD = 6.0$, 95% CI: 11.75, 38.15) than distributed display users ($M = 147.4$ $SD = 39.7$, 95% CI: 134.17, 160.57), $F(1, 36) = 176.86$, $MSE = 142375.68$, $p = .00$, $\eta^2 = .61$.

Taken altogether, the overall difference observed in the window time can be explained by the combined contribution of the three factors (see Table 2), and these factors show tradeoffs, explaining how large process effects may combine to produce smaller total time effects. The time difference between stacked and distributed displays observed in each factor was computed and summed up, then compared to the time differences observed in the window time. The mean window time difference was 20.1 – 15.7 = 4.4 min, and $(2.8/4.4) \times 100 = 64\%$ of this time difference was explained by the three factors.

About a minute and half was left unexplained. Given that the note use time computed only by off-screen gaze durations longer than 2,000 ms, the unexplained time may include time spent on multitasking other than note use such as information tracking and even some eye blinking when people make fixation jumps within/between pages. In addition, information-loading time (i.e., time required to load and view an information page whenever a selection is made) was not accounted for in the equation.

The physical page transition time (i.e., page-turning action only: how many times participants selected and changed information page to view using the drop down menu) can also explain a portion of the unexplained time. Presumably, distributed display users did not have to turn pages frequently. Stacked display users, however, had to turn pages if they wanted to look at information found on another page and they indeed turned more pages ($M = 25.7$, $SD = 6.6$, 95% CI: 22.88, 28.60) than did distributed display users ($M = 19.4$, $SD = 5.7$, 95% CI: 16.56, 22.28), $t(36) = 3.16$, $p = .003$, Cohen's $d = 1.03$. But they were obviously reluctant to turn the pages, as they did not do so four times more often than did distributed display users, relying instead on internally stored information or notes.

General Discussion

Replicating prior studies involving undergraduates working on artificial tasks (Jang & Schunn, 2012; Jang et al., 2011) and authentic tasks (Jang et al., 2012; Kroft & Wickens, 2002), the current work involving relative experts on complex tasks again finds that the organization dimension of distributed versus stacked displays can influence integrative reasoning performance. The current study carefully controlled display content to manipulate display format per se, now establishing that the dimension matters even with rich and diverse tasks and with participants who are trained in the task.

More importantly, we showed that the overall task time effects of display organization needs to be understood as the summation of three different underlying factors: information internalization, access, and externalization. The stacked display led to longer information internalization time and this effect was indexed by mean fixation times during first pass through content. Other researchers examining very simple tasks (Gray & Fu, 2004; Gray, Sims, Fu, & Schoelles, 2006) have conceptualized this effect as a memorization microstrategy that is automatically applied when there are relatively higher information access costs, like in the case of stacked displays. At the same time, in a complementary fashion, the stacked display led to shorter information access time, as indexed by shorter total return fixations. As a result of not having memorized the information, distributed display users sought external information more often, whereas stacked display users relied on internally stored information, which will typically (although not always) be accessed more quickly than external information.

It is important to note however, that the information internalization effect is larger than the information access effect, and thus, on the basis of just these two effects, there should generally be an

Table 2
Mean, SD, and 95% Confidence Interval (CI) of Times (in Minutes) in First-Pass Fixation Durations, Return Fixation Durations, Note Use Time, and Summed Time for Each Condition and the Difference Between Conditions

	Stacked		Distributed		Stacked–distributed
	M (*SD*)	95% CI	*M* (*SD*)	95% CI	
First pass fixations	6.6 (2.7)	(5.65, 7.57)	2.3 (1.0)	(1.36, 3.28)	4.3
Return fixations	2.5 (1.6)	(1.45, 3.57)	6.9 (2.8)	(5.87, 7.99)	−4.4
Off-screen gazes	5.2 (5.6)	(3.27, 7.17)	2.3 (2.1)	(0.38, 4.28)	2.9
Sum	14.3		11.5		2.8

overall time benefit of distributed displays over stacked displays. If the users were not allowed to take/use any notes, the difference between information internalization and access time presumably have been larger than it is observed here. As for the third effect, informational externalization time, stacked display users took more notes presumably to manually transform stacked information into distributed information, thereby reducing the need to reaccess information. In some ways, this effect could also be seen as an adaptive tradeoff between externalization time and information access time.

Implications

Taking an important theoretical step beyond the previous studies, the current study proposed a new framework for unpacking the underlying mechanisms of the distributed versus stacked display effect. Also, this study examined the kind of rich ill-defined tasks that are particularly important in the work place. The basic skill sets required for accurate quantitative data interpretation are common to a range of social sciences and human science fields. As a result, the trends toward using microdisplays for work applications (e.g., powerful ultralight laptops, tablet computers, and smartphones) should be reexamined carefully with respect to their impact on performance. The novel and intuitive interaction modes such as swiping, pinching, and spreading (instead of clicking) that are available with such microdisplays can be understood as minimizing external information access costs. However, it is important to note that there are benefits that only large displays can provide. For instance, the sense of location associated with certain information often helps users doing quick search and saving memory (Smith, 1979; Tan, Stefanucci, Proffitt, & Pausch, 2001).

More importantly, changes in interaction methods could lead to changes in information processing strategies (i.e., tradeoffs between information memorization and access) rather than changes in overall performance (Jang et al., 2012). For example, when considering patterns of results across prior studies that involved stacked display users with differing interaction methods (hovering vs. clicking), users accessed information much more easily and frequently with hovering than with clicking. As a result, users with clicking were more likely to memorize information but those with hovering revisited information more often. Overall, both types of stacked display users suffered relative to distributed display users, but the way in which the deficit was displayed varied by interaction method (i.e., memorization time or reaccess time).

In terms of theoretical implications, the proposed tripartite framework was found to be useful for explaining the time differences and it is expected to be useful to unpack the effects of other display format comparisons (e.g., separated vs. integrated displays or collaborative information sharing systems). The underlying framework includes factors both at the lower/perceptual level, and the higher-level cognitive processes. In addition, the overall model considers the bidirectional interaction between the information presentation environment (e.g., innate high information access cost in stacked displays) and the human's activities (e.g., strategic information encoding and note use).

Future Directions

This research suggests that spatially distributed information lead users to more efficient and effective problem solving. For future studies, boundary conditions of the effect should be explored to deepen our theoretical understanding and to provide precise recommendations for designers. For example, is there a benefit of distributing information across two large monitors rather than just one large monitor? Does the distributed benefit hold for 20-year experts working on data in their own focal area of expertise? Relevant subquestions include how individual differences in cognitive abilities and strategy adaptivity would affect the information encoding strategy. As Schunn and Reder (2001) found, some people are much less sensitive and slow to change their strategy choices.

References

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4,* 829–839. doi:10.1038/nrn1201

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). New York, NY: Academic Press.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4,* 75–100. doi:10.1037/1076-898X.4.2.75

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8,* 293–332. doi:10.1207/s1532690xci0804_2

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86,* 175–218. doi:10.1002/sce.10001

Fu, W. T., & Gray, W. D. (2000). *Memory versus perceptual motor tradeoffs in a blocks world task.* Paper presented at the Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society, Hillsdale, NJ.

Fu, W. T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology, 52,* 195–242. doi:10.1016/j.cogpsych.2005.08.002

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 231. doi:10.1037/0278-7393.22.1.231

Gopie, N., & MacLeod, C. M. (2009). Destination memory: Stop me if i've told you this before. *Psychological Science, 20,* 1492–1499. doi:10.1111/j.1467-9280.2009.02472.x

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. doi:10.1037/1076-898X.6.4.322

Gray, W. D., & Fu, W. T. (2001). *Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design*. Paper presented at the CHI '01 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

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.

Gray, W. D., Sims, C. R., Fu, W. T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113,* 461. doi:10.1037/0033-295X.113.3.461

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3,* 446–474. doi:10.1111/j.1756-8765.2011.01150.x

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 37. doi:10.1037/a0017683

Jang, J., & Schunn, C. D. (2012). Performance Benefits of spatially distributed versus stacked information on integration tasks. *Applied Cognitive Psychology, 26,* 207–214. doi:10.1002/acp.1811

Jang, J., Schunn, C. D., & Nokes, T. J. (2011). Spatially distributed instructions improve learning outcomes and efficiency. *Journal of Educational Psychology, 103,* 60–72. doi:10.1037/a0021994

Jang, J., Trickett, S. B., Schunn, C. D., & Trafton, J. G. (2012). Unpacking the temporal advantage of distributing complex visual displays. *International Journal of Human-Computer Studies, 70,* 812–827. doi:10.1016/j.ijhcs.2012.07.003

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of input and output events in memory. *Psychological Research, 53,* 260–270. doi:10.1007/BF00941396

Kroft, P. D., & Wickens, C. D. (2002). Displaying multi-domain graphical database information. *Information Design Journal, 11,* 44–52. doi:10.1075/idj.11.1.06kro

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (somtimes) worth ten thousand words. *Cognitive Science, 11,* 65–99. doi:10.1111/j.1551-6708.1987.tb00863.x

Marsh, R. L., & Hicks, J. L. (2002). Comparisons of target output monitoring and source input monitoring. *Applied Cognitive Psychology, 16,* 845–862. doi:10.1002/acp.859

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. doi:10.1037/a0018008

Peebles, D., & Cheng, P. C. H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors, 45,* 28–46. doi:10.1518/hfes.45.1.28.27225

Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology: Applied, 14,* 36. doi:10.1037/1076-898X.14.1.36

Sanfey, A., & Hastie, R. (1998). Does evidence presentation format affect judgment? An experimental evaluation of displays of data for judgments. *Psychological Science, 9,* 99–103. doi:10.1111/1467-9280.00018

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science: A Multidisciplinary Journal, 23,* 337–370.

Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General, 130,* 59. doi:10.1037/0096-3445.130.1.59

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460. doi:10.1037/0278-7393.5.5.460

Tan, D. S., Stefanucci, J. K., Proffitt, D. R., & Pausch, R. (2001). The Infocockpit: Providing location and place to aid human memory. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces (pp. 1-4)*. ACM.

Trafton, J. G., & Trickett, S. B. (2001). Note-taking for self-explanation and problem solving. *Human-computer interaction, 16*(1), 1–38. doi:10.1207/S15327051HCI1601_1

Trafton, J. G., Trickett, S. B., Schunn, C. D., & Kirschenbaum, S. (2007). Complex visualizations, spatial transformations, and uncertainty. Washington, DC: Presented at the meeting of the Naval Research Laboratory.

Willows, D. M. (1978). A picture is not always worth a thousand words: Pictures as distractors in reading. *Journal of Educational Psychology, 70*(2), 255. doi:10.1037/0022-0663.70.2.255

(*Appendix follows*)

Appendix

Grading Rubric for the Data Interpretation Task

ID	Points	Question types	Plausible answers
			Question 1: Was the hypothesis of Study 1 confirmed? If so, what are the evidences? If not, what are the evidences?
1a	5	Yes or no	• Yes, the hypothesis was confirmed
1b	5	Explanation	• Destination memory was better than source memory • Source memory was worse than target memory • Derek–Derek, Robby–Robby > Sally–Sally, Mary–Mary; more correct identification for destination memory • Confusion: Sally–Mary and Mary–Sally is higher than Derek–Robby and Robby–Derek; more confusion for source memory
			Question 2: Was the hypothesis of Study 2 confirmed? If so, what are the evidences? If not, what are the evidences?
2a	5	Yes or no	• Yes, the hypothesis was confirmed
2b	5	Explanation	• Destination memory was worse than source memory • Source memory was better than target memory • High hits and low false alarm for source memory • Low hits and high false alarm for destination memory • High correct recognition (hits–false alarm) for source memory • Low correct recognition (hits–false alarm) for destination memory • Shorter face-fact pair bar for destination memory • Condition $\times$ Item Type interaction is significant
			Question 3: Are the results of Study 1 and 2 congruent? If so, in what ways and if not, in what ways?
3a	5	Yes or no	• No, the results are not congruent
3b	5	Explanation	• Study 1: source memory worse than destination memory; Study 2: destination memory worse than source memory • Hypotheses are contradicting and yet each was confirmed • Each study shows evidence for different theories
			Question 4: If you've answered that there was an inconsistency in the findings of the two studies, how would you reconcile the findings? In other words, what do you think could account for these inconsistent results?
4a	10	Number and breadth of ideas	• 0 pt: no plausible difference presented • 2 pts: one difference found either for material or task • 4 pts: differences for material and task; one difference for theory • 6 pts: differences for material and theory, or task and theory • 8 pts: differences for material, task, and theory • 10 pts: differences for material, task, and theory; an additional dimension(s) or an idea(s) regardless of dimension overlap
4b	10	Idea elaboration	• 0 pt: no elaboration made; only thesis statements • 2 pts: elaboration attempted with 1 or 2 supporting phrases • 6 pts: elaboration made with 3 or 4 supporting phrases • 10 pts: elaboration made with 5 or more supporting phrases • Thesis statements are not counted because they were already counted towards previous 10 points; but number of supporting phrases summed up across multiple thesis statements • Phrases were defined as idea chunks that contain new/different idea even though a sentence runs on

Received March 3, 2014
Revision received May 7, 2014
Accepted May 8, 2014 ■