The effect of Interface Consistency and Cognitive Load on user performance in an information search task

Jeremy Mendel and Richard Pak
Department of Psychology
Clemson University

Although interface consistency is theorized to increase performance and user satisfaction, previous research has found mixed and often non-significant results. By controlling for the level of task difficulty, specifically the cognitive load, it is hoped that the study will explain the contradictory findings of past research. Eighty participants will be used in a 2x2x2 between-subjects factorial design. The study will manipulate the level of interface consistency, along with intrinsic and extraneous cognitive load imposed by the task. Objective performance measures of task completion time and error-rate along with the subjective measure of user satisfaction will be analyzed. It is expected that the high interface consistency condition will perform better than low interface consistency; in contrast, it is predicted that the low cognitive load condition will outperform the high cognitive load condition. Furthermore, it is predicted that the effect of consistency and cognitive load will interact causing the lack of interface consistency to be significantly more detrimental when paired with a high level of either form of cognitive load.

INTRODUCTION

Interface consistency can be described as the look and feel of an interface; it is considered a core aspect of usability (Nielsen, 1989). The earliest human factors studies examined controls of World War II aircraft (Chapanis, 1953; Fitts & Jones, 1961). Aircraft controls of the era were haphazardly designedwith one researcher calling them "fiendishly inconsistent" (Chapanis, 1953). Fitts & Jones (1961) attributed over half of pilot errors to the inconsistent control designs.

However, the problem of interface inconsistency is not limited to aviation. Many construction vehicles use a similar control layout but require inconsistent control manipulations to operate. For example, excavator and skid steer vehicles both have two control sticks and pedals; however, the excavator pedals maneuver the vehicle while the skid steer's control sticks drive the machine. These two interfaces may appear consistent but clearly are not. Although it is unclear how many accidents or deaths are due to control layout inconsistency between machinery, according to The Bureau of Labor Statistics Census of Fatal Occupational Injuries, of 481 deaths in the excavation industry from 1992-2002, vehicles caused 253, or over half of these deaths (McCann, 2006). Although a consistent interface is at the core of "good interface design" (Grudin, 1989; Shneiderman, 1998) how to best define consistency remains debated.

Dimensions of Consistency

In an effort to further clarify the concept, researchers have operationalized specific aspects or dimensions of interaction that lead to consistency: physical, communicational and conceptual (Adamson & Wallace, 1997; AlTaboli & Abou-Zeid, 2007; Ozok & Salvendy, 2000; 2004; Rhee, Moon & Choe, 2006).

Physical consistency. Physical consistency considers the visual or graphical appearance of an interface or object (AlTaboli & Abou-Zeid, 2007; Ozok & Salvendy, 2000; Rhee et al., 2006). One example of physical consistency is the location of an automobile speedometer. Nearly all automobiles have an analog speedometer with a similar design right above the steering column. This standard is so prevalent that drivers are able to gauge speed in a different car without first studying the speedometer's design.

Communicational consistency. Communicational consistency is the level of consistency between the way the user interacts with the system and the way in which the system presents information to the user (Ozok & Salvendy, 2000; Rhee et al., 2006). For example, pulling back on an airplane's throttle will always decrease the engine's output.

Conceptual consistency. Conceptual consistency refers to how a user thinks about an interface and its match to how the system presents the interface. Conceptual consistency has been described as the consistency of the metaphor applied to the system and how it is represents components an interface (AlTaboli & Abou-Zeid, 2007; Kellogg, 1987; Ozok & Salvendy, 2000; Rhee et al., 2006). An example of this is the menu bar found in most windows-based applications. This menu bar uses similar if not identical headings across programs (e.g. File, Edit, etc) and each menu contains similar commands such as File>Save or Edit>Copy.

Incomplete consistency (only addressing some of the dimensions of consistency) can be detrimental to user performance (Finstad, 2003; Rhee et al, 2006; Satzinger & Olfman, 1998). Returning to the previous example of the excavator controls compared to the skid steer controls, the two machines express an incomplete consistency. While the interfaces of the two machines are physically consistent, the interfaces are not communicationally consistent. These discrepancies can be detrimental given that consistency is generally considered good for performance (Adamson & Wallace, 1997; Eberts, 1997; Schneider, Dumais & Shiffrin, 1984; Schneider & Shiffrin, 1997; Shneiderman, 1998).

Review of Interface Consistency and Human Performance

Consistency between two interfaces may encourage learned skills to be transferred to new systems; consistency helps the user to predict system responses (AlTaboli & Abou-Zeid, 2007; Nielsen, 1989; Rhee et al., 2006; Schneider & Shiffrin, 1997). The results of interface consistency can be seen in a shortened learning process, reduced working memory demand and increased efficiency (e.g., AlTaboli & Abou-Zeid, 2007). These results are measured through shorter task completion time, reduced error-rate, and higher user satisfaction (e.g., Rhee et al., 2006).

Although interface consistency appears to be a worthy goal to achieve, with seemingly positive outcomes, the literature is actually unclear as to the benefits of user interface consistency. In practice, several studies found weak and even detrimental effects from consistent interface designs (Finstad, 2003; Ozok & Salvendy, 2000; Rhee et al., 2006; Satzinger & Olfman, 1998).

In addition to questionable performance benefits of interface consistency, the concept of interface consistency has been criticized as too abstract and vague as to what makes an interface consistent (Grudin, 1989; 1992). Grudin (1992) argued that interface consistency is so fuzzy a concept that it ultimately detracts from usability by distracting designers from other more achievable usability goals.

Supporting Grudin (1989; 1992), empirical evidence has shown that consistent interfaces can cause users to over-generalize functions within the interface (Finstad, 2003). According to Finstad (2003), this over-generalization confused users, forcing them to redevelop mental models. In this case, interface consistency was detrimental to performance (longer completion time and more errors; Finstad, 2003).

Satzinger and Olfman (1998) found mixed results of consistency. Their study manipulated consistency of the "action language syntax" and the "visual consistency" of the interface. "Action language syntax" refers to the consistency of the commands labels (e.g. F1 = help). This manipulation had a positive effect on performance. "Visual consistency", which is analogous to the physical consistency dimension, however, negatively impacted performance. It was concluded that task variety induced through the visual inconsistencies improved performance by helping users distinguish between the two interfaces (Satzinger & Olfman, 1998).

Other studies have found non-significant effects of consistency. Ozok & Salvendy (2000) found no difference for overall error-rate, completion time, or satisfaction. Rhee et al. (2006) had similar non-significant results. If interface consistency is a crucial aspect of design, then why are all these empirical studies unable to show the benefits? An explanation may reside in the tasks used in these experiments. For both of these studies, the tasks used were simple web-navigation tasks. Tasks for these two studies included combinations of clicking, data entry, reading comprehension, and word searches (Ozok & Salvendy, 2000; Rhee et al., 2006).

The literature reviewed thus far suggests a conflicted view of consistency; that is, some studies show positive effects of consistency (AlTaboli & Abou-Zeid, 2007; Ozok & Salvendy, 2003), while other studies show none (Ozok & Salvendy, 2000; Rhee et al., 2006) or even negative effects (Finstad, 2003; Satzinger & Olfman, 1998) of consistency.

This discrepancy may be due to the nature of the tasks used in these studies, more specifically, the manipulation of task difficulty. If one study used a significantly harder task than another, it could help explain the contrary findings. Previous research has not operationalized or measured the task difficulty. One method to measure task difficulty is through the cognitive load theory; cognitive load is the demand on working memory from a task (Paas & van Merriënboer, 1994). Before proceeding, it may be helpful to describe the concept of cognitive load.

Cognitive Load

Cognitive load can be defined as the burden placed on the working memory during problem solving and learning (Ayres, 2006). Cognitive load theory is divided into disparate parts to identify which aspects of the task are contributing to the difficulty. The total cognitive load include intrinsic and extraneous cognitive load. The most immediate portion of the task difficulty is the difficulty of the task itself, also known as the intrinsic cognitive load. Another piece of cognitive load, the extraneous cognitive load, is the added and unnecessary difficulty induced by the method of presentation for the material (Ayres, 2006; Bannert, 2002; Paas et al., 2003).

The effects of intrinsic cognitive load and extraneous cognitive load loads are interactive in nature; an increase in one makes an individual more sensitive to increases in the other (Schnotz & Kürschner, 2007). Sweller & Chandler (1994) demonstrated this interaction between the two loads by concluding that increased extraneous cognitive load was significantly more detrimental in tasks with high intrinsic cognitive load.

Cognitive load as an explanation for conflicting interface consistency results. Methodological issues with previous interface consistency studies may explain why consistency has shown mixed effects. In studies with nearly identical tasks, contradictory results were found. In Ozok & Salvendy (2000), no significant effects of interface consistency were found. In contrast, AlTaboli & Abou-Zeid (2007), which based the participant tasks on Ozok & Salvendy (2000), obtained significant results. While these two experiments used seemingly identical tasks, it is unclear of the cognitive load imposed in these studies. It has been theorized that the amount of load imposed by an interface is affected by screen design (Chalmers, 2003; Saadé & Otrakji, 2007; Szabo & Kanuka, 1998). If this is the case, the varied amount of cognitive load in these tasks may have caused the conflicting results of past studies.

Saadé and Otrakji (2007) found a correlation between interface design and cognitive load (as measured by a questionnaire); however, the study stopped short of directly manipulating cognitive load. If screen design, specifically interface consistency, is affected by the cognitive load of a task, then it must be manipulated understand the nature of the relationship.

Cognitive Load Theory would predict that poor interface design (more specifically interface inconsistency) would increase the extraneous cognitive load of the user. If interface consistency is one aspect that makes up the extraneous load, then the total cognitive load imposed by the task would moderate the effect of interface consistency (Chandler & Sweller, 1991; Sweller & Chandler, 1994). In the 1994 study, it was shown through multiple experiments that for tasks with a low intrinsic cognitive load, participants were not affected by increased levels of extraneous cognitive load. The explanation for this finding was that “easier” tasks required less working memory thus leaving more cognitive resources to deal with extraneous load before the participant was overloaded.

Past interface consistency research, such as Ozok & Salvendy (2000) and Rhee et al (2006), have ignored the cognitive load of the participant’s tasks; furthermore, they were relatively simple tasks. The tasks may have generated such a low level of intrinsic cognitive load that participants were easily able to deal with the additional load imposed by the inconsistent interfaces. Without controlling for varying levels of cognitive load imposed by the tasks in these studies, it is unclear if the tasks were difficult enough to produce an effect.

Current Study

Past research on interface consistency has left the cognitive load of the tasks uncontrolled (e.g., AlTaboli & Abou-Zeid, 2007). While some research has suggested that cognitive load and screen design are related, a paucity of research has combined the two concepts. Saadé and Otrakji (2007) linked cognitive load to screen design by measuring cognitive load via questionnaire. The study did not manipulate the levels of cognitive load thus the discovered relationship was merely correlational. The current study will manipulate cognitive load along with interface consistency to determine the nature of the relationship: if interface consistency interacts with the cognitive load of the tasks similar to Chandler and Sweller (1994), it could explain the lack of significant results from other studies.

For the present study, participants will be tasked with finding information on a pair of websites. The pair of websites will be designed to be either consistent or inconsistent with each other. The extraneous cognitive load and intrinsic cognitive load of the tasks will also be manipulated by creating variations of the website for each of the conditions.

It is hypothesized that the consistent interface groups will perform better and with higher satisfaction scores (e.g., Rhee et al. 2006). Also, similar to extant cognitive load literature, the two groups given less demanding tasks will perform better and show higher user satisfaction (Saadé & Otrakji, 2007).

An interaction between the level of interface consistency and the amount of cognitive load imposed by the task is also expected. It is predicted that high levels of cognitive load will result in a sharper drop-off of performance and satisfaction from the inconsistent interface designs. In a situation with a high level of interface consistency and a low level of cognitive load, performance and user satisfaction will be higher when compared to the other conditions. In contrast, the condition with an inconsistent interface along with a high level of cognitive load will result in the worst performance and user satisfaction. This interaction would occur when the increased cognitive load leaves participants less able to cope with the inconsistencies.

METHOD

Participants

Eighty undergraduate students from Clemson University will participate in the study. Participants will receive extra credit in their psychology course in exchange for participation. Individuals will be unaware of the intention of the study prior to participation.

Design

The study is a 2 (consistency, high/low) x 2 (intrinsic load, high/low) x 2 (extraneous load, high/low) factorial design manipulated between participants. Subjects will be randomly be assigned to one of eight possible conditions. Participants will first be asked to complete a demographics form. Next, they will be given series of website tasks on the first website then another set of tasks on a second website. Website presentation order will be counter-balanced between participants to control for order effects.

Independent variables. Interface consistency, extraneous cognitive load, and intrinsic cognitive load of the task will serve as independent variables; they will be directly manipulated between subjects. To control interface consistency, a series of inconsistencies will be used to alter the original design. These manipulations and justification for using them are seen in the following table (Table 1).

Extraneous cognitive load will be manipulated by including extraneous hyperlinks within the website for the high cognitive load condition. In this condition, participants will be faced with additional decisions of whether or not to click on the hyperlinks; simply having the extra links requires more decisions to be made (DeStefano & LeFevre, 2007; see Figure 1). To manipulate the amount of intrinsic cognitive load of the task, the content of the websites will also vary between groups. The high intrinsic cognitive load group will work with a website containing detailed financial information while the low intrinsic cognitive load group will work with a website containing basic travel information.

Table 1
Manipulations of Consistency Between Websites

Dimen.	Difference	e.g. A	e.g. B	Citation
Phy.	Location of navigation Bar	Left	Top	AlTaboli & Abou-Zeid, 2007; Ozok & Salvendy, 2004
Phy.	Text spacing	Single Spaced	Double spaced	Bednall (1992); Benbasat and Todd (1993); Ozok & Salvendy, 2004
Comm.	Menu systems	Text hyperlinks for navigation	Drop down box for navigation	Adamson & Wallace, 1997; Ozok & Salvendy, 2004
Comm.	Text anchors for navigation	No anchors	Yes anchors	Rhee et al., 2006; Ozok & Salvendy, 2004
Conc.	Replacing words with icons	"Four star" hotel written in text	Four stars in an icon form instead of text	Satzinger & Olfman, 1998
Conc.	Alphabetized list sorting	Alphabetically sort	Randomized order	Ozok & Salvendy, 2004

Dependent variables. The dependent variables that will be used are task completion time, error-rate, and subjective user satisfaction. These measures are representative of user performance in web browsing and have been used previously (e.g., Ozok & Salvendy, 2000). Completion time will be measured from the beginning of the first question until the end of the first website's questions then again for the second website's questions. Error-rate will be tracked for each participant and is operationalized as follows: each use of the back or home button resulted in one error; also, navigation to incorrect website using extraneous or other hyperlinks counted as a single error. User satisfaction scores will be measured using a questionnaire with a seven-point Likert scale.

Figure 1
Extraneous hyperlinks

With Extraneous Hyperlinks	Without Extraneous Hyperlinks
Clemson University is located in upstate South Carolina in Pickens County just north of Interstate 85 and Anderson, South Carolina, along the shores of Lake Hartwell.	Clemson University is located in upstate South Carolina in Pickens County just north of Interstate 85 and Anderson, South Carolina, along the shores of Lake Hartwell.

Materials/Apparatus

Six workstations running Windows XP will be used in the study. The individual workstations will retrieve the website from a local server. To record the participant data, TechSmith's Morae recorder software will be used. Morae recorded the time participants spent on each page, mouse clicks, and website navigation.

A demographics form will be used to gather age, gender, ethnicity and school major/education. A short survey adapted from the IBM Computer Usability Satisfaction Questionnaires will be used to gather user satisfaction scores from participants. (Lewis, 1995).

Procedure

Participants will be randomly assigned to one of eight conditions. Participants will be first given a demographics form. Next, they will be given a list of questions to be answered in order by using the first website. Before beginning the task, participants will be instructed to press the record button on Morae. Once pressed, Morae will record each participant's performance. Next, participants will be given a second set of questions to answer on the second website. At the end of the questions, participants will be reminded to stop the Morae recorder. When finished with the questions, participants will be given a satisfaction survey regarding the websites. When the questionnaire is complete, participants will be given a debriefing form and allowed to depart.

PLANNED ANALYSIS

Objective measures: completion time and error-rate. Task completion time and error-rate will be measured across all questions for both websites for each participant. Both objective measures will be analyzed using an analysis of variance (ANOVA) at the 0.05 level.

Subjective measure: user satisfaction. To analyze the satisfaction scores, a repeated measures ANOVA will be used.

EXPECTED RESULTS

As in past research, it is expected that high interface consistency will result in faster task completion, fewer errors and higher user satisfaction. Similarly, it is expected that the low intrinsic cognitive load and extraneous cognitive load group will also perform better (e.g. Sweller & Chandler, 1994). The unique prediction for this study is that interface consistency and the two forms of cognitive load will interact. In the high cognitive load conditions, it is expected that performance will suffer significantly more from the lack of consistency. In contrast, the low load condition will see significantly less of a decline in the inconsistent interface condition.

Acknowledgement. This work was supported by a grant from Deere & Company.

REFERENCES

Adamson, P.J. and Wallace, F.L. (1997). A comparison between consistent and inconsistent graphical user interfaces. Pre-publication Report, University of Northern Florida, Department of Computer and Information Sciences, Jacksonville, Florida.

AlTaboli, A. & Abou-Zeid, M. R. (2007). Effect of Physical Consistency of Web Interface Design on Users' Performance and Satisfaction. In J. A. Jacko (Ed.), *Human-Computer Interaction: HCI Applications and Services* (pp. 849-858). Berlin: Springer.

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

Bannert, M. (2002). Managing cognitive load - recent trends in cognitive load theory. *Learning and Instruction*, 12, 139-146.

Bednall, E.S. (1992). The Effect of screen format on visual list search. *Ergonomics, 35*, 369-383.

Benbasat, I. & Todd, P. (1993). An experimental investigation of interface design alternatives: icon vs text and direct manipulation vs menus. *International Journal of Man-Machine Studies, 38*, 369-402.

Brünken, R., Plass, J., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53-61.

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

Chandler, P., & Sweller, J. (1994). Why some material is difficult to learn. *Cognitive and Instruction, 12*, 185-233.

Chalmers, P. A. (2003). The role of cognitive theory in human-computer interface. *Computers in Human Behavior*, *19*, 593-607.

Chapanis, A. (1953). Psychology and the Instrument Panel. *Scientific American*, 301-306.

DeStefano, D. & LeFevre, J. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*, 1616-1641.

Eberts, R. (1997). Cognitive modeling. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics, 2nd Edition*. New York, NY: John Wiley & Sons, 1328-1374.

Finstad, K. (2003, January). Analogical problem solving in casual and experienced users: When interface consistency leads to inappropriate transfer. Dissertation abstracts international (UMI Dissertation Information Service, No. 3058935).

Fitts, P.M., & Jones, R.E (1961). Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls. In W.H. Sinaiko (Ed.), *Selected papers on human factors in the design and use of control systems* (pp. 332-358). New York: Dover.

Grudin, J. (1989). The Case against user interface consistency. *Communications of the ACM*, *32*(10), 1164-1173.

Grudin, J. (1992). Consistency, Standards, and Formal Approaches to Interface Development and Evaluation: A Note on Wiecha, Bennett, Boies, Gould, and Greene. *ACM Transactions on Information Systems*, *10*(1), 103-111.

Kellogg, W. A. (1987). Conceptual consistency in the user interface: Effects on user performance. In H. J. Bulinger & B. Shackel (Eds.), *Human-computer interaction; INTERACT '87IFIP Conference on Human-Computer Interaction* (pp. 389–394). Amsterdam, The Netherlands: Elsevier.

Lewis, J.R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction, 7*(1), 57-78.

Mahajan, R. & Shneiderman, B. (1997). Visual and Textual Consistency Checking Tools for Graphical User Interfaces. *IEEE Transactions on Software Engineering*, *23*(11), 722-735.

McCann, M. (2006). Heavy equipment and truck-related deaths on excavation work sites. *Journal of Safety Research, 37*(5), 511-517.

Nielsen, J. (1989). Executive Summary: Coordinating User Interfaces for Consistency. In J. Nielsen (Ed.), *Coordinating User Interfaces for Consistency* (pp. 1-7). San Francisco, CA: Morgan Kaufmann Publishers.

Ozok, A.A. & Salvendy, G. (2000). Measuring consistency of web page design and its effects on performance and satisfaction. *Ergonomics*, *43*(4), 443-460.

Ozok, A. A., & Salvendy, G. (2003). The effect of language inconsistency on performance and satisfaction on the World Wide Web: results from four experiments. *Behaviour and Information Technology*, *22*, 155–163.

Ozok, A.A. & Salvendy, G. (2004). Measuring consistency of web page design and its effects on performance and satisfaction. *Computers in Human Behavior*, *20*, 149-161.

Paas, F., & van Merriënboer, J.J.G. (1994). Measurement of cognitive load in instructional research. *Perceptual & Motor Skills, 79*, 419-430.

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive load measures as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Polson, P. (1988). The Consequences of Consistent and Inconsistent User Interfaces. *Cognitive Science and Its Applications for Human Computer Interaction*, 59-108.

Proctor, R. W. & Vu, K. L. (2006). Location and Arrangement of Displays and Control Actuators. In W. Karwowski (Ed.), *Handbook of Standards and Guidelines in Ergonomics and Human Factors* (pp. 399-409). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Reisner, P. (1990). What is Consistency? *Proceeding IFIP Third International Conference*, 175-181.

Rhee, C., Moon, J., & Choe, Y. (2006). Web interface consistency in e-learning. *Online Information Review*, *30*(1), 53-69.

Saadé, R. G. & Otrakji, C. A. (2007). First impressions last a lifetime: Effect of interface type on disorientation and cognitive load. *Computers in Human Behavior*, *23*, 525-535.

Satzinger, J., & Olfman, L. (1998). User Interface Consistency Across End-User Applications: The Effects on Mental Models. *Journal of Management Information Systems*, *14*, 167-193.

Schneider, W., Dumais S.T. and Shiffrin, R.M. (1984). Automatic and control processing and attention. In *Varieties of Attention* (pp. 1-27), New York: Academic Press.

Schneider, W. and Shiffrin, R.M. (1997). Controlled and automatic human information processing: 1. Detection, search, attention. *Psychological Review*, 84, 1-66.

Schnotz, W. & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, *19*, 469-508.

Shneiderman, B. (1987). *Designing the user interface-Strategies for effective human-computer interaction.* Reading, MA: Addison-Wesley.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human–computer interaction.* Tokyo: Addison Wesley Longman.

Smith, S. L., & Mosier, J. N. (1986). *Guidelines for designing user interface software* (Tech. Rep. No. ESD-TR-86-278). United States Air Force Electronic Systems Center, Hanscom AFB, MA.

Szabo, M. & Kanuka, H. (1998). Effects of Violating Screen Design Principles of Balance, Unity, and Focus on Recall Learning, Study Time, and Completion Rate. *Journal of Educational Multimedia and Hypermedia*, *8*(1), 23-42.

Tanaka, T., Eberts, R. E., & Salvendy, G. (1991). Consistency of human–computer interface design: Quantification and validation. *Human Factors*, *33*(6), 653–676.

Van Gerven, P.W.M., Paas, F.G.W.C., Van Merriënboer, J. J.G., & Schmidt, H.G. (2000). Cognitive load theory and the acquisition of complex cognitive skills in the elderly: Towards an integrative framework. *Educational Gerontology*, *26*, 503-521.