

December 2007

An Alternative Mechanism for the Cognitive Fit Theory in Spreadsheet Analysis

Hock Chan

National University of Singapore

Hee-Woong Kim

National University of Singapore

Suparna Goswami

National University of Singapore

Follow this and additional works at: <http://aisel.aisnet.org/icis2007>

Recommended Citation

Chan, Hock; Kim, Hee-Woong; and Goswami, Suparna, "An Alternative Mechanism for the Cognitive Fit Theory in Spreadsheet Analysis" (2007). *ICIS 2007 Proceedings*. 15.

<http://aisel.aisnet.org/icis2007/15>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

AN ALTERNATIVE MECHANISM FOR THE COGNITIVE FIT THEORY IN SPREADSHEET ANALYSIS

Hock Chuan Chan

National University of Singapore
chanhc@comp.nus.edu.sg

Hee Woong Kim

National University of Singapore
kimhw@comp.nus.edu.sg

Suparna Goswami

National University of Singapore
suparnag@comp.nus.edu.sg

Abstract

A basic task in spreadsheet analysis in order to understand the structure of a given spreadsheet is that of finding precedent cells (cells that are referenced in the formula of a given cell). The cognitive fit theory is used to analyze this task. Current applications of the cognitive fit theory assert a fit when the information emphasized by the task matches the information emphasized by the problem representation. This study applies the concept of cognitive fit from a different perspective, with the intention of assessing mental representation in problem solving. By keeping the task invariant, the study does a detailed analysis of the different problem representations (the "A1" referencing style and the "RIC1" referencing style), and the corresponding mental representations. It is experimentally shown that the problem representation affects the mental representation, the cognitive fit between the mental representation and the problem representation, and the subsequent performance.

Keywords: Spreadsheet, cognitive fit theory, mental representation, Hick's law

Introduction

Spreadsheets are used by many types of application that are critical to a company, from invoices that deal with simple calculations to analytical models for sales forecasting. Although spreadsheets have been in use for a few decades, the cognitive difficulties in analyzing and understanding spreadsheet models have persisted. Cognitive difficulties lead to errors in the use of spreadsheet. Along with documented errors in commercial spreadsheets that lead to huge financial losses, field studies, laboratory research and news reports indicate significant amounts of error (Panko and Halverson, 2001; <http://panko.cba.hawaii.edu/ssr/>). Therefore, it is important to have research directed towards understanding the cognitive difficulties that spreadsheet users face. For example, Goswami et al. (2006) studied the effects of spreadsheet visualization tools, Panko and Halverson (2001) studied the effect of user collaboration for spreadsheet development and Clermont (2003) and Burnett et al. (2002) proposed new methods for analyzing large spreadsheet models.

Cognitive fit theory is a widely used theory to analyze problem-solving performance (Vessey 1991, Vessey and Galletta 1991). "The basic model of cognitive fit views problem solving as the outcome of the relationship between the external problem representation and the problem solving task, which are both characterized by the types of

information they emphasize. ... Cognitive fit exists when the cognitive processes used to act on the problem representation and those used to complete the task match, resulting in superior problem solving performance.” (Agarwal et al. 1999, p. 542). The theory has been used to study decision-making problems such as data presentation formats of tables and graphs, multi-attribute data presentation, accounting models and software maintenance (Dennis and Carte 1998; Dunn and Grabski 2001; Smelcer and Carmel 1997; Shaft and Vessey 2006; Umanath and Vessey 1994; Vessey 1994; Vessey and Galletta 1991). The cognitive fit theory has also been used to study spreadsheet error correction by users (Goswami et al. 2006).

Previous research on cognitive fit theory has primarily analyzed fit in terms of the match between task and problem representation. Users act upon the information emphasized by the problem representation and the information emphasized by the task to form a mental representation of the problem before actually solving the problem. Thus, mental representation is a factor that is determined by both task and problem representation. However, little research has focused on how to assess mental representation. Since mental representation forms the very basis for the fit and affects performance in problem solving, it is important to get a better understanding of mental representation. Prior research has highlighted the need to examine the intermediate stages before performance outcome (e.g., Shaft and Vessey 2006). Assessing mental representation is a very important step in understanding the problem solving process underlined by cognitive fit theory.

This study demonstrates a mechanism for assessing and validating mental representation in the context of cognitive fit in spreadsheet analysis. Referencing is an important aspect of spreadsheet programs because it helps to understand the linkages between cells in the spreadsheet that are connected through formulas. According to the most popular spreadsheet software (i.e., Microsoft Excel), there are two referencing styles. That is, a spreadsheet formula (in a cell) can be shown in two different styles, the “A1” and “R1C1”referencing styles, which are illustrated in the following section. Cognitive fit theory is used to analyze how these two styles of referencing affect users’ mental representation, and ultimately performance, when they are required to perform the task of finding a spreadsheet cell that is referenced in a formula contained in a given cell. In order to assess the mental representations that are formed based on the two referencing style, Hick’s law, which gives a formula for deriving user’s response time in making a selection from a given number of choices, is used. It is argued that based on the referencing styles, users will have different number of cells to make a choice from, and their measured response time follows Hick’s law, thus indicating different mental representations driving the choice mechanisms.

The next section analyzes the referencing styles with cognitive fit theory and derives the hypotheses. This is followed by the experiment description and data analysis. Finally, the results and their contribution to cognitive fit theory are discussed, along with some practical implications for spreadsheet users and developers. The conclusion highlights the contributions of this study.

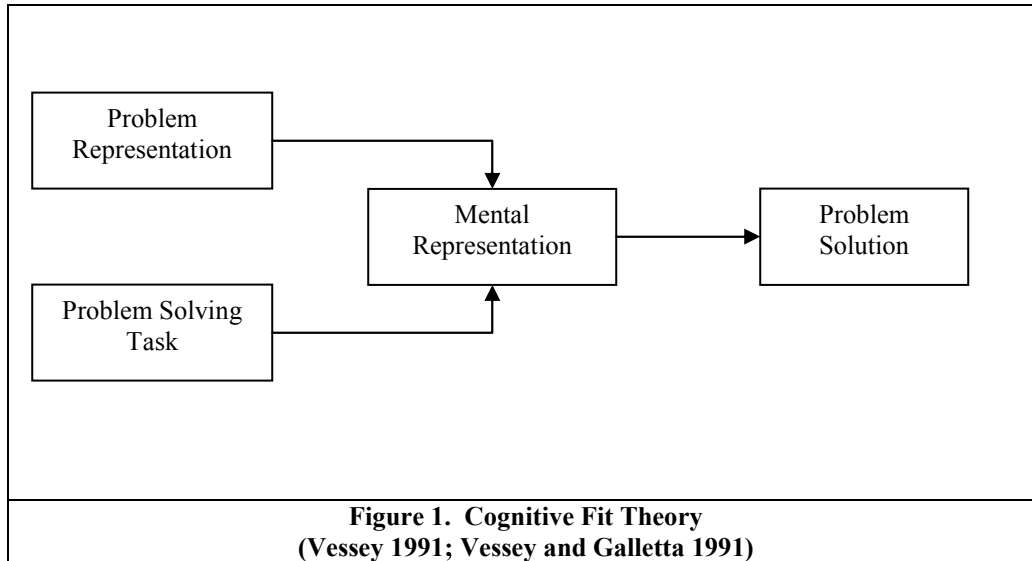
Cognitive Fit Theory and the Hypotheses

The cognitive fit model (Vessey 1991; Vessey and Galletta 1991), as shown in Figure 1, asserts that a problem solving task and a problem representation may emphasize different information. A person develops a mental representation to solve the task based on the information from both the problem-solving task and the problem representation. If the task information matches the information emphasized by the problem representation, a fit occurs and this improves problem-solving performance. Prior research has also examined the role of individual cognitive skills in studies applying the cognitive fit theory (Galletta et al. 2003, Hubona et al. 1998; Vessey and Galletta 1991). The current study, however, does not consider the role of individual differences.

When applying cognitive fit theory, studies typically examine the relationship between task, problem representation and problem solving performance. Therefore, fit is explained as a match between the information in the task and the information in the problem representation (Vessey 1991; Shaft and Vessey 2006). For example, Hubona et al. (1998, p. 708) stated, “... if the mental model formulated by the problem representation is inconsistent with the task representation then decision making performance may be impeded.” In a similar vein, Dunn and Grabski (2000) controlled fit by manipulating the experiment tasks to match different problem representations.

This study applies the cognitive fit theory from a different perspective. The primary emphasis is on understanding and assessing the mental representation created by the task and the problem representation, rather than purely measuring the performance outcome of cognitive fit.

In this study, the problem-solving task does not change and therefore, its role in determining the mental representation and hence the fit remains invariant. However, different problem representations are used. Hence the problem representation influences the mental representation, as well as provides the information that can result in either a match or mismatch with the mental representation it helps in creating.



Prior studies applying cognitive fit theory usually provide reasons for the expected mental representation, without actually attempting to empirically assess the mental representation (Chandra and Krovi 1999; Galletta et al. 2003; Hubona et al. 1998; Shaft and Vessey 2006). For example, Shaft and Vessey (2006, p. 48) noted that their study “examined the outcome of the problem-solving process rather than the process directly” and suggested that the process itself would be worth studying in future. This study analyzes the expected mental representations, and further validates them through data analysis. The following subsections provide detailed descriptions of the context and develop the hypotheses.

Task

The context of the study is spreadsheet analysis. The task is to find the cell (specifically, to click on the cell) that is referenced in the formula of a particular cell. For example, if cell C4 contains a formula “=3*A1”, then A1 is the cell referenced by C4, and the task would be to locate A1 and click on it. In spreadsheet terminology, the referenced cell (A1 in this example) is called the precedent cell and the referencing cell (C4 in the example) is called the dependent cell (Davis 1996; Goswami et al. 2006). Thus, the problem-solving task is to find the precedent cell given a dependent cell.

This task is a fundamental activity in spreadsheet analysis (Galletta et al. 1996; Goswami et al. 2006; Hendry and Green 1994; Panko 1999; Teo and Lee-Partridge 1999). In a spreadsheet model, many cells are interconnected through formulas. This interconnection is also called data dependency (Davis 1996). To understand a spreadsheet model, a user will have to find precedent cells (Goswami et al. 2006). In fact, to “understand how a formula works often requires the user to recursively track down the meaning of cell references” (Hendry and Green 1994, p. 1053). For example, in a profit calculation model, a user may select the cell showing final profit, and recursively trace precedents to understand the elements that contribute to profit.

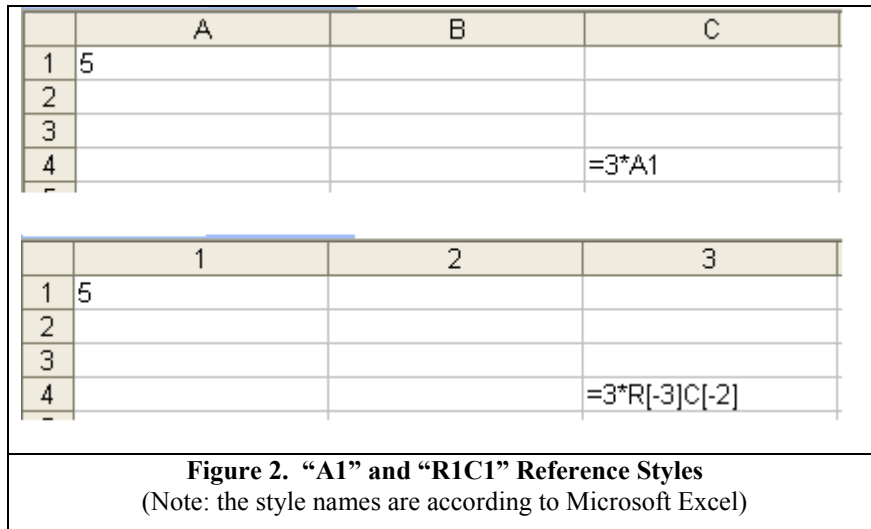
This process of tracing precedents may be considered as a backward process to determine which cells affect a particular cell. In spreadsheet analysis, it is also common to consider a forward process to determine cells that are affected by a particular cell. This is a process to trace dependent cells, and has been studied by Davis (1996). As the backward and forward processes involve different information, this study focuses only on the backward process.

Higher task level analyses of spreadsheet have also been studied, e.g. Galletta et al. (1996) studied error detection with or without formula display, and with paper or screen presentation, Teo and Tan (1999) studied qualitative and quantitative errors, and Goswami et al. (2006) relate error types to visualization tools.

Problem Representation

In Microsoft Excel, formulas can be shown in two alternative methods as illustrated in Figure 2. These are the problem representations for this study. The “A1” referencing style names a cell with a labeling system that names columns alphabetically from the left, and numbers rows sequentially from the top. The “R1C1” referencing style shows the position of the precedent cell relative to the dependent cell. For example, in Figure 2, cell C4 (column C / 3 and row 4) shows the precedent cell as “A1” using the “A1” referencing style and “R[-3]C[-2]”, which means 3 rows up and 2 columns to the left, using the “R1C1” referencing style.

In addition to the different referencing styles in the cell formula, the problem representations also have different labels for columns. The “A1” referencing style uses alphabets for the column while the “R1C1” referencing style uses numbers, as illustrated in figure 2. The column and row labels are important in the context of this study, because they are a part of the problem representation.



Mental Representation

According to cognitive fit theory, mental representation is developed through a consideration of the information required by the task and the information given in the problem representation: “The mental representation is formulated using the characteristics of both the problem representation and the task.” (Vessey 1991, p. 221).

In the context of the current study, the task is to click on the precedent cell for a given cell. The information emphasized by the task is the location of the cell. Location can be indicated in many ways, such as by using the “A1” referencing method, or the “R1C1” referencing method, or even visually. Thus, the task itself does not emphasize on any particular way of representing the location. Hence, the primary (if not the only) source of information for performing the given task is the problem representation. In other words, the problem-solving task is constant while the problem representation varies between the “A1” and “R1C1” referencing styles. In such a case, it is the problem representation that will determine the mental representation and consequently the cognitive fit.

Since the “A1” referencing style names the cell starting from the “A” (first) column and “1” (first) row, a user will have a natural tendency to decide on the location of the precedent cell by starting from column “A” and row “1”. Therefore, the corresponding mental representation will be to start looking from the cell “A1” in order to find the

precedent cell. In contrast to the “A1” referencing style, the “RIC1” referencing style indicates the precedent cell relative to the dependent cell. Thus, a user will have a natural tendency to decide on the location of the precedent cell by starting from the dependent cell. Thus the corresponding mental representation will be to start looking from the dependent cell in order to find the precedent cell.

Thus, in both representations, the decision process involved in performing the given task is that of selecting a specific column/row among a number of possible columns/rows. Users have to select the target cell from a number of possible targets. Experimental psychology and human-computer interaction research provide an equation for estimating the time taken to choose from a number of targets. This is Hick’s law (Hick 1952), which states that the response time of users making a choice varies with the logarithm of the number of possible choices. It can be expressed in the following form:

$$\text{Time taken} = \text{constant} + b * \log_2(N+1), \text{ where } N \text{ is the number of targets to choose from.}$$

Hick’s law is commonly used to study reaction time for choice among alternatives (e.g. Beggs et al. 1972; Gignac and Vernon 2004; Mahurin and Pirozzolo 1993) and has been declared as, “one of the most robust regularities that have been reported in the choice response time literature” (Usher et al. 2002, p. 704) The law was found to apply to even people with Alzheimer or Parkinson disease (Mahurin and Pirozzolo 1993) and even to pigeons (Vickrey and Neuringer 2000). A later study suggested that the logarithm term can be $b * \log_2(N)$ (Howarth et al. 1971). We use this shorter version for the subsequent calculations.

In the current study, for the two different referencing styles, users will start looking for the precedent cell by starting their search process from the A1-cell or from the dependent cell respectively and, therefore will have different number of options to make a choice from. The time as calculated using the above formula will be used to assess the mental representations that are formed for the two referencing styles.

In order to develop the time equations for the two mental representations, we first need to identify the number of possible choices for each representation. We define certain terms which are used to derive the equations. Let r_p be the row number of the precedent cell, when numbering from the topmost row, c_p be the column number of the precedent cell, numbering from the leftmost column. Similarly, let r_d and c_d be the row and column numbers for the dependent cell.

Let m_r be the number of rows between the precedent and dependent cells, inclusive, and cn_r be the number of columns between the precedent and dependent cells, inclusive. Thus,

$$m_r = | r_p - r_d | + 1$$

$$cn_r = | c_p - c_d | + 1$$

Let m_a be the number of rows between the precedent cell and cell “A1”, inclusive, and cn_a be the number of columns between the precedent cell and cell “A1”, inclusive. Thus,

$$m_a = r_p$$

$$cn_a = c_p$$

Two equations for time taken can be developed to assess the two different mental representations and the corresponding decision processes. If users try to locate the precedent cell by starting from the leftmost column and top row (i.e. cell “A1”), then the number of possible target choices that are available to them is a product of m_a and cn_a . Therefore, the time taken to find the precedent cell is expected to follow equation 1.

$$\text{Time} = \text{constant}_1 + b_1 * \log_2(m_a * cn_a) \text{ ----- (Equation 1)}$$

On the other hand, if users try to locate the precedent cell by starting from the dependent cell, then the number of possible target choices that are available to them is a product of m_r and cn_r , and the time taken is expected to follow equation 2.

$$\text{Time} = \text{constant}_2 + b_2 * \log_2(m_r * cn_r) \text{ ----- (Equation 2)}$$

For the “A1” referencing style, users’ mental representation is likely to be anchored to the cell “A1”. We call this the A1-anchored mental representation. Current studies consider the problem-solving process to occur within the mental representation (Shaft and Vessey 2006). Thus time taken is more likely to follow equation 1, and equation 2. Hypotheses 1a and 1b reflect this reasoning. The use of double-hypotheses in this context is analogous to

correlation hypotheses where high values of an independent factor are highly associated with high values of the dependent factor, but are less associated with low values of the dependent factor.

H1a: When the “A1” referencing style is used, equation 1 will be statistically significant.

H1b: When the “A1” referencing style is used, equation 2 will not be statistically significant.

On the other hand, for the “R1C1” referencing style, users’ mental representation is likely to be anchored to the dependent cell, since the representation of the precedent cells in the formula is relative to the dependent cell. We call this the Dependent-anchored mental representation. Thus, the time taken to find precedent cells is more likely to follow equation 2 and not equation 1. Therefore, we hypothesize.

H2a: When the “R1C1” referencing style is used, equation 1 will not be statistically significant.

H2b: When the “R1C1” referencing style is used, equation 2 will be statistically significant.

Cognitive Fit

In the current context, the task is that of finding a precedent cell, given a dependent cell. In order to perform the task, users need information on the position of the precedent cell. This information is contained in the problem representation. Since the task is kept invariant across the two problem representations, it is likely to play a relatively neutral role in deriving cognitive fit. Thus, for this particular context, fit will result from information consistency between the problem representation and the mental representation (which, in turn is a product of the problem representation).

The concept of fit can be explained using Figure 2 where the shaded areas show the column and rows labels for the two referencing styles. In “A1” referencing style, if the user is trying to find the cell “C7”, then this cell is going to have the column label “C” and the row label “7”, and thus there will be a direct match between the way the cell is referenced in the dependent cell’s formula, and the corresponding mental representation formed based on the row and column labeling style of the spreadsheet. The user does not have to perform any transformation or calculation based on the cell name and the column and row labeling style. Therefore, there is a cognitive fit for the “A1” referencing style and the corresponding mental representation.

In contrast, for the “R1C1” referencing style while the columns and rows are numbered starting from the leftmost column and top row, the mental representation starts from the dependent cell because the position of the precedent cell is indicated relative to the dependent cell. Thus, it will require an additional information transformation for finding the position of a precedent cell. For instance, the user will have to first calculate the numbers of labels for the “-3” row and “-2” column, and then locate them. Therefore, the level of cognitive fit is likely to be lower for the “R1C1” referencing style than for the “A1” referencing style.

To examine cognitive fit, and compare the levels of fit achieved for the two referencing styles, this study examines the time taken and error rate for each referencing style. Time taken and error rate are the most commonly measured dimensions of problem solving performance as indicated by prior research. For example, Vessey (1991) and Hubona et al. (1998) measured speed and accuracy, Agarwal et al. (1999) and Galletta et al. (2003) measured accuracy and Goswami et al. (2006) measured speed. Since the level of cognitive fit is likely to be higher for the “A1” referencing style than for “R1C1” referencing style, and a higher level of cognitive fit results in better problem solving performance, users will take less time and have lower error rates when using the “A1” referencing style than when using the “R1C1” referencing style. Hence we hypothesize:

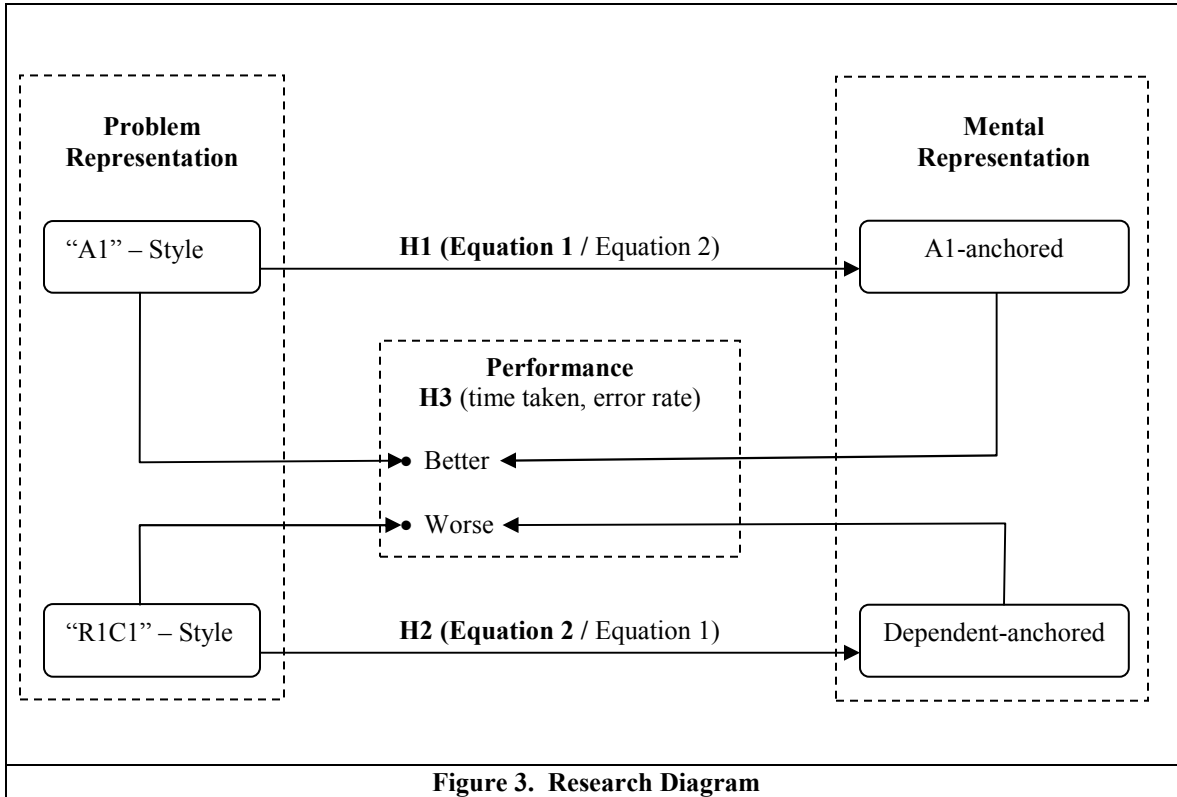
H3a: Time taken for the “A1” referencing style will be less than the time taken for the “R1C1” referencing style.

H3b: Error rate for the “A1” referencing style will be less than the error rate for the “R1C1” referencing style.

Summary

The research design for the study is summarized in figure 3. In this research design, problem representation and mental representation are two-valued constructs. The two values of problem representation are the “A1” style and the “R1C1” style. Similarly, mental representation takes on two forms – “A1-anchored” and “dependent-anchored”,

which are assessed by Equation 1 and Equation 2 respectively. “A1” style representation will result in the “A1-anchored” mental representation, whereas the “RIC1” style representation will result in the “dependent-anchored” mental representation. A good fit between the problem representation and the mental representation, results in better performance. Performance is measured in terms of time taken to perform the task and the error rate. As discussed above, the “A1” referencing style and the corresponding A1-anchored mental representation will result in a fit and hence better performance. On the other hand, “RIC1” referencing style and the dependent-anchored mental representation will not result in a fit, and hence performance will be worse compared to the “A1” style.



Research Methodology

A laboratory experiment was conducted for this study. There were two problem representations following the “A1” and the “RIC1” referencing styles. Problem representation is the independent variable with two values (i.e., the two referencing styles). The dependent variables are the time taken to click on the precedent cell and the error rate.

The experiment involved 73 subjects (56 males, 17 females) aged between 17 and 25. 80 percent had low to medium level of Excel expertise. All were volunteer students and were paid S\$30 (about US\$20) at the end of the experiment. The subjects were given sufficient time to complete the experiment.

All instructions were printed and given to the subjects at the beginning of the experiment. Subjects were given 15 minutes to read the instructions carefully. The instructions were then repeated to all subjects before the start of the experiment. The subjects could ask questions to clarify any doubt before the experiment. Specific instructions were again displayed on the screen before the start of each spreadsheet session.

A short warm-up exercise consisting of two trials for each referencing style was given prior to data collection. These trials were necessary to enable the subjects to familiarize themselves with the format of the experiment and to understand how each method was used. Subjects were instructed to click on the target cell as fast as possible, and at

the same time to minimize the number of errors. The software generated a beep when the subject clicked outside the target cell.

The spreadsheets were shown in two sizes: with the Excel default cell size, and with the cell size doubled in length and height. The default cell size had a height of 17 pixels and a width of 64 pixels. To avoid complications from multiple screen access, each spreadsheet was limited to one screen. The screen size was 36.5 cm by 27.0 cm with a screen resolution of 1024 by 768 pixels, using 17-inch monitors. Each referencing style was used for both spreadsheet sizes. Every subject had to work with both “A1” and “R1C1” referencing styles. Every subject worked through two rounds, each with four spreadsheets in the following size sequence: small (1 style), small (another style), big (1 style), big (another style). The size sequence was kept fixed, but the style sequence was randomized across subjects. The dependent cell was placed at the lower right hand corner of the screen. Clicking on the dependent cell revealed the formula (in “A1” or “R1C1” style according to the style of the session). The formula was restricted to only one precedent cell. The subject then had to click on the precedent cell. Time was measured between the click on the dependent cell and the click on the precedent cell, in milliseconds. Clicking outside the target counted as an error. Error rate was the number of error clicks divided by the number of precedent cells in the experiment. This process of clicking on the dependent cell and clicking on the precedent cell was repeated 20 times for each spreadsheet. The twenty precedent cells in each spreadsheet were fixed, but the sequence was randomized for every subject. There were compulsory rest periods before each spreadsheet. At the end of each spreadsheet, the subject was shown his/her mean time for that spreadsheet. This use of feedback was to enhance the motivation of subjects to perform optimally (Whisenand and Emurian, 1996).

Data Analysis and Results

The performance values were different for spreadsheets of different cell sizes. A within subjects analysis of the data is carried out. Table 1 summarizes the performance values for the different spreadsheet-cell sizes. Time is the average for all clicks. Also, round 2 values were better than round 1 values. The improvement in performance with experience is expected (Vickrey and Neuringer 2000). We performed hypothesis testing for each spreadsheet size, as well as for each round. The results were identical. To simplify the presentation, we combined data for different cell sizes and rounds, leaving only the independent variable of referencing style.

Problem Representation	Time (milliseconds)		Error rate (%)	
	Default-cell spreadsheet	Big-cell spreadsheet	Default-cell spreadsheet	Big-cell spreadsheet
“A1” referencing style	5196 (1056)	2914 (400)	14.8 (22.4)	6.0 (9.6)
“R1C1” referencing style	11199 (3450)	5194 (1151)	49.6 (60.2)	25.7 (29.6)

Table 2 shows the hypothesis testing for the mental representations. As described above, the default-cell and big-cell spreadsheets had 20 precedent cells each, giving a total of 40 precedent cells. Timing for each precedent cell was averaged over all subjects. Thus, each precedent cell provided a set of values for m_r , c_r , m_a , c_a and time. These were used in the regression analysis for equation 1 and 2. Natural logarithm is used instead of logarithm to base two. This affects the coefficients equally for all equations. All the hypotheses (H1a, H1b, H2a, and H2b) were supported, showing strong support for our proposition that the two referencing style lead to different mental representations.

The means and standard deviations for the performance measures are shown in Table 3. A multiple analysis of variance test between the two problem representations shows significant differences for time taken ($F=34.9$, $p=0.001$, $N=80$) and error rate ($F=127.0$, $p=.001$, $N=80$). The results support the reasoning for good fit and lack of fit for the two referencing styles.

Problem Representation	Regression	R²	F	p	Hypothesis
“A1” referencing style	Equation 1	.605	60.7	.001	H1a supported
	Equation 2	.001	1.01	.320	H1b supported
“RIC1” referencing style	Equation 1	.021	1.83	.184	H2a supported
	Equation 2	.543	47.4	.001	H2b supported

Problem Representation	Problem Solving Performance	
	Time (milliseconds)	Error rate (%)
“A1” referencing style	4055 (1503)	10.4 (8.4)
“RIC1” referencing style	8196.5 (4179)	37.7 (18.3)
Hypothesis	H3a supported	H3b supported

Implications and Conclusion

Implications

This study has important theoretical and practical implications. Till date most prior research on cognitive fit theory has focused on the match between the information emphasized by the problem solving task and the information emphasized by the problem representation and the corresponding performance outcomes (e.g., Vessey and Galletta 1991; Hubona et al. 1998; Goswami et al. 2006), and most of the findings by and large validate the theory (Vessey 2006). However, there is limited research which assess the abstract construct of mental representation or provide a detailed understanding of how the mental representation influences cognitive fit. From a theoretical perspective, this study attempts a different take on the cognitive fit theory as it is traditionally understood and applied, and delineates a mechanism for assessing the mental representation component of cognitive fit.

By keeping the task invariant (and relatively neutral in terms of the information is emphasizes), and simply varying the problem representation, this study demonstrates differences in mental representation and therefore differences in performance. A recent extension of the cognitive fit theory (Vessey 2006; Khatri et al. 2006) distinguishes between external problem representation that the users are provided with, and the internal representation of the problem domain that exists in the user’s mind, and states that the interaction between the internal and external representations contribute towards the mental representation developed to solve the problem. Although this study does not aim to assess the internal representation as well, or differentiate between the internal representation and mental representation, our research provides some form of support to this extended cognitive fit model by showing that for the same task, different problem representations give rise to different mental representations in users’ cognition, although the task remains unchanged.

This study also provides a mechanism for re-examining prior studies in a new light. For example, the question of whether table or graph is better has been studied in the context of different tasks, by matching information from task and information from table or graph. Our study indicates that it might be worthwhile to consider how the table or graph itself can influence the mental representation and thus the fit and performance.

Our study demonstrates a new approach towards assessing and validating mental representation. Current studies rely on performance differences to validate the mental representation, or more precisely, to validate the fit between mental representation and problem representation. This study is designed in a manner such that it provides an opportunity to assess the mental representation separately from the performance differences. Future studies applying cognitive fit theory can consider this approach wherever the context allows for separate validation of the mental representation. In addition, this approach provides some insight into the problem-solving process while most previous research has focused on examining the outcome of the problem-solving process (Shaft and Vessey 2006).

From a practical perspective, this study highlights that the “R1C1” referencing style is not as effective and efficient as the “A1” referencing style, especially when it comes to tracking down cell references. Developers can consider redesigning the style, for instance, the column and row could be numbered to fit the decision process that starts from the dependent cell. Developers can also examine other possibly more fitting problem representations. Future studies on spreadsheet analysis can test other problem representations for the same task of finding precedents, or other spreadsheet analysis tasks.

Limitations

The results of this study should be interpreted in the context of its limitations. The use of a controlled laboratory experiment gives rise to limitations inherent to experiments. The use of student subjects is a practical limitation; even so, this might not affect the generalizability of the results to other populations with good general training or education. Next, there is existence of a learning effect when using A1 and R1C1 referencing methods as measured in round 1 and 2. The performance may increase further over time. Although results are consistent over two rounds, there is no certainty that the results will be the same if the subjects had continued for many more rounds. The experiment was designed to be a simple one involving only one precedent cell for a dependent cell. The applicability of the result to multiple precedents needs to be further tested.

The performance difference between problem presentations may also be partly caused by factors other than fit. It is an inherent difficulty in most studies with the cognitive fit theory. It is almost impossible to constrain the problem representations to differ only in terms of fit. It is also usually not desirable to do so, e.g. when the study compares two different data models (Chandra and Krovi 1999). In this experiment, some may argue that the “A1” referencing style has higher readability compared to the “R1C1” referencing style. The experiment tried to minimize the readability effect by keeping the formula very simple: it has only one precedent cell with no other mathematical operations.

Although this study is a significant advance over previous studies in cognitive fit as it provides a method for validating the mental representation, the assertions regarding the difference in mental representation could be farther strengthened by supplementing them with findings from other ways of assessing the mental representation. For instance, in the current setting, the use of protocol analysis would have provided a better understanding of how the subjects worked through the spreadsheet in order to identify the precedent cell when given a dependent cell.

Conclusion

Research on spreadsheet analysis, comprehension and error detection is important, given the widespread existence and serious consequences of spreadsheet errors. This study examines a very basic, and yet very essential, step of spreadsheet analysis. This is the task of finding a precedent cell from the formula in a dependent cell. The study focuses on the two formula referencing styles available in Microsoft Excel, which is probably the most widely used spreadsheet software worldwide. On a practical level, the study shows how the “A1” referencing is more efficient and effective than the “R1C1” referencing style.

Equally important, the analysis of this simple task has provided a new dimension to cognitive fit theory. Going beyond previous research on cognitive fit theory, it shows how problem representation can influence mental representation, as well as the fit and the corresponding performance. By pioneering a more direct assessment of the

mental representation, this study offers researchers a fresh approach for analyzing old problems, as well as designing studies to investigate new phenomena.

On the methodology front, the experiment design allows for a validation of the reasoned mental representations from experiment data using an established law on response time, Hick's law. Thus, the experiment also provides further empirical support to Hick's law. This approach thus increases the internal validity of the cognitive fit study.

Acknowledgements

This research was supported by the National University of Singapore, research grant R-253-000-027-112.

References

- Agarwal, R., De, P. and Sinha, A.P., "Comprehending Object and Process Models: An Empirical Study", *IEEE Transactions on Software Engineering* (25:4), 1999, pp. 541-556.
- Beggs, W.D.A., Graham, J.C., Monk, T.H., Shaw, M.R.W. and Howarth, C.I., "Can Hick's Law and Fitts' Law be Combined?" *Acta Psychologica* (35:3), 1972, pp. 348 – 357.
- Burnett, M., Sheretov, A., Ren, B. and Rothermel, G., "Testing Homogeneous Spreadsheet Grids with the 'What you See Is What You Test Methodology'," *IEEE Transactions on Software Engineering* (28:6), 2002, pp. 576-594.
- Chandra, A. and Krovi, R., "Representational congruence and information retrieval: Towards an extended model of cognitive fit", *Decision Support Systems*,25(4), 1999, pp. 271-288.
- Clermont, M., "Analyzing Large Spreadsheet Programs," *10th Working Conference on Reverse Engineering*, 2003, p. 306.
- Davis, J.S., "Tools for Spreadsheet Auditing", *International Journal of Human Computer Studies* (45:4), 1996, pp. 429-442.
- Dennis, A. R. and Carte, T. A., "Using Geographical Information Systems for Decision Making: Extending Cognitive Fit Theory to Map-based Presentations," *Information Systems Research* (9:2), 1998, pp. 194-203.
- Dunn, C.L. and Grabski, S.V., "Perceived Semantic Expressiveness of Accounting Systems and Task Accuracy Effects," *International Journal of Accounting Information Systems* (1:2), 2000, pp. 79-87.
- Dunn, C. and Grabski, S., "An Investigation of Localization as an Element of Cognitive Fit in Accounting Model Representations," *Decision Sciences* (32:1), 2001, pp. 55-94.
- Galletta, D.F., Hartzel, K.S., Johnson, S., Joseph, J.L., and Rustagi, S., "Spreadsheet Presentation and Error Detection: An Experimental Study," *Journal of Management Information Systems* (13:3), 1996, pp. 45-63.
- Galletta, D.F., Durcikova, A. Everard, A., and Jones, B., "Cognitive Fit and an Intelligent Agent for a Word Processor: Should Users take all that Advice?" *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*, 2003.
- Gignac, G.E. and Vernon, P.A., "Reaction time and the dominant and non-dominant hands: an extension of Hick's Law," *Personality and Individual Differences* (36), 2004, pp. 733-739.
- Goswami, S., Chan, H.C. and Kim, H.W., "Spreadsheet Error Correction using an Activity Framework and a Cognitive Fit Perspective," *Proceedings of the 27th International Conference on Information Systems*, 2006, pp.695-710.
- Hick, W.E., "On the rate of gain of information," *Quarterly Journal of Experimental Psychology* (45), 1952, pp.188-196.

- Hendry, D.G. and Green, T.R.G., "Creating, Comprehending and Explaining Spreadsheets: A Cognitive Interpretation of What Discretionary Users Think of the Spreadsheet Model," *International Journal of Human-Computer Studies* (40:6), 1994, pp. 1033-1065.
- Howarth, C.I., Beggs, W. D. A., and Bowden, J.M., "The relationship between speed and accuracy of movement aimed at a target," *Acta Psychologica* (35:3), 1971, pp. 207 – 218.
- Hubona, G.S., Everett, S., Marsh, E., and Wauchope, K., "Mental Representations of Spatial Language," *International Journal of Human-Computer Studies* (48), 1998, pp. 705-728.
- Khatri, V., Vessey, I., Ram, S., and Ramesh, V. "Cognitive Fit between Conceptual Schemas and Internal Problem Representations: The Case of Geospatio-Temporal Conceptual Schema Comprehension," *IEEE Transactions on Professional Communication* (49:2), 2006, pp. 109-127.
- Mahurin, R.K. and Pirozzolo, F.J., "Application of Hick's Law of Response Speed in Alzheimer and Parkinson Diseases," *Perceptual and Motor Skills*, (1), 1993, pp. 107-113.
- Panko, R. R. "Applying Code Inspection to Spreadsheet Testing," *Journal of Management of Information Systems* (16:2), 1999, pp. 159-176.
- Panko, R.R. and Halverson, R.P., Jr., "An Experiment in Collaborative Development to Reduce Spreadsheet Errors," *Journal of the Association for Information Systems* (2:4), 2001, pp. 1-29.
- Shaft, T.M. and Vessey, I., "The Role of Cognitive Fit in the Relationship Between Software Comprehension and Modification," *MIS Quarterly* 30(1), 2006, pp.29-55.
- Smelcer, J. B. and Carmel, E. "The Effectiveness of Different Representations for Managerial Problem Solving: Comparing Tables and Maps," *Decision Sciences* (28:2), 1997, pp. 391-420.
- Teo, T. S. H., and Lee-Partridge J.E. "Effects of Error Factors and Prior Incremental Practice on Spreadsheet Error Detection: An Experimental Study," *Omega, International Journal of Management Science* (29:5), 1999, pp. 445-456.
- Teo, T.S.H. and Tan, M., "Spreadsheet Development and 'What-if' Analysis: Quantitative versus Qualitative Errors," *Accounting, Management and Information Technologies* (9:3), 1999, 141-160.
- Umanath, N. and Vessey, I., "Multi-Attribute Data Presentation and Human Judgement: A Cognitive Fit Perspective," *Decision Sciences* (25:5/6), 1994, pp. 795-824.
- Usher, M., Olami, Z. and McClelland, J.L. "Hick's Law in a Stochastic Race Model with Speed-Accuracy Tradeoff," *Journal of Mathematical Psychology*, (46), 2002, pp. 704-715.
- Vessey, I., "The Effect of Information Presentation on Decision making: A Cost-Benefit Analysis," *Information and Management* (27), 1994, pp. 103-119.
- Vessey, I., "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature," *Decision Sciences* (22:2), 1991, pp. 219-240.
- Vessey, I., "The Theory of Cognitive Fit: One Aspect of a General Theory of Problem Solving?" in *Human-Computer Interaction and Management Information Systems: Foundations, Advances in Management Information Systems Series*, P. Zhang and D. Galletta (eds.) M.E. Sharpe, Armonk, NY, 2006.
- Vessey, I. and Galletta, D., "Cognitive Fit: An Empirical Study of Information Acquisition," *Information Systems Research* (2:1), 1991, pp. 63-84.
- Vickrey, C. and Neuringer, A., "Pigeon Reaction Time, Hick's Law and Intelligence," *Psychology Bulletin and Review*, (7:2), 2000, pp. 284-291.
- Whisenand T.G. and Emurian H.H., "Analysis of cursor movements with a mouse," *Computers in Human Behaviour* (15), 1999, pp.85-103.