

ERIC STOWE HAARBAUER

Proximity Compatibility in Medical Diagnosis Displays

(Under the Direction of ROBERT PETER MAHAN)

The Proximity Compatibility Principle (PCP) states that a display format is well-suited to a given task if the information sources in the display are related to the same degree as information sources in the task. While experiments have shown that PCP can provide useful display design guidelines for many types of tasks, diagnosis tasks have not seemed to conform to PCP's predictions. The current experiment compared performance and user preference with integral, configural, and separable displays in three diagnosis tasks based on a medical diagnosis system. Tasks differed in variable count and differential utility. As predicted, the integral display was superior in tasks without differential utility. Differential utility diminished but did not eliminate the integral display's superiority. Users preferred the integral display and disfavored the separable display in all task conditions, especially in tasks without differential utility. The results indicate that PCP is a useful theory for diagnosis tasks, but different diagnosis tasks can differ widely in their task demands.

INDEX WORDS: Proximity Compatibility Principle, Information displays, Integral displays, Object displays, Physiologic State Severity Classification System, Diagnosis, Sepsis, Human factors, Attention

PROXIMITY COMPATIBILITY IN MEDICAL DIAGNOSIS DISPLAYS

by

ERIC STOWE HAARBAUER

B.S., The University of Alabama at Birmingham, 1993

M.S., The University of Georgia, 1998

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2001

PROXIMITY COMPATIBILITY IN MEDICAL DIAGNOSIS DISPLAYS

by

ERIC STOWE HAARBAUER

Approved:

Major Professor: Robert Mahan

Committee: Rich Marsh
Garnett Stokes
Chuck Lance
Adam Goodie

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
May 2001

© 2001

Eric Stowe Haarbauer

All Rights Reserved

DEDICATION

I dedicate this dissertation to Griffin Walker Haarbauer. Sometimes the purpose behind 21 years of formal education becomes difficult to remember. The point is simple: Baby needs a sane Daddy who loves his work.

ACKNOWLEDGEMENTS

Thanks to the entire Vrooman family for helping out with baby, wife, and life so I could complete this on time; to Chris Marino and Phil Dunwoody for putting up with my arguments and for providing some of their own; to Rich Marsh, Chuck Lance, and Gary Lautenschlager for distributing their statistical advice; to Stu Katz for input during his time on my committee; to Morbid Angel, King Diamond, Nile, In Flames, Nevermore, and AC/DC for helping me vent; and especially to Rob Mahan for education and generosity both in and out of the classroom.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
CHAPTER	
I INTRODUCTION.....	1
II THE PROXIMITY COMPATIBILITY PRINCIPLE	4
Stimulus Dimensions and Attentional Demands.....	4
Extending Dimensional Integrality	5
Experimental Evidence and PCP Interactions.....	9
Parallel Processing vs. Emergent Features.....	12
III DIAGNOSIS AND OBJECT DISPLAYS	15
The Physiologic State Severity Classification System.....	18
PSSC Diagnosis vs. Previous Experiments.....	20
The Present Experiment	22
IV METHOD.....	25
Participants	25
Diagnosis Tasks.....	25
Prototype Profiles.....	25
Profile Sets	26
Display Formats	26
Experimental Design	28
Procedure.....	28
V RESULTS.....	30
Normal Diagnosis Performance	31
Variable Count	32

	Differential Utility	33
	User Preference	35
VI	DISCUSSION.....	37
	Normal Diagnosis Performance	37
	Variable Count	39
	Differential Utility	39
	User Preference	41
	Conclusion.....	42
	REFERENCES	46

CHAPTER I

INTRODUCTION

Technology has improved many aspects of daily life, but it has also created a dramatic increase in the amount of information that people have to process. For example, today's aircraft provide safer and faster transportation than the planes of decades ago, but now pilots must cope with sensor displays that cover almost every square inch of the cockpit walls. Thus, the human factors community must answer the need for carefully designed information displays for a wide range of complex tasks. This need, coupled with advances in computer graphics over the past thirty years, has resulted in unprecedented levels of research on new display formats.

Where older displays generally featured "single-sensor, single-display" designs, many modern designs combine the output of several sensors (Bennett & Flach, 1992). For example, an artificial horizon display combines information on an aircraft's pitch and roll. The *object display* carries this idea one step further. In an object display, several or even dozens of related variables are mapped onto features of a single perceptual object with the idea that automatic, efficient perceptual organization can be brought to augment effortful processing of the display. One such design is Chernoff's (e.g., 1975) face display, in which variables are mapped to features of a cartoon-like face. "Chernoff faces" have been used in statistics (Wainer & Thissen, 1981) and marketing (Huff, Mahajan, & Black, 1981), and other object displays have been used in nuclear power control rooms (Woods, Wise, & Hanes, 1981), aircraft cockpits (O'Hare & Roscoe, 1990), digital signal processing software (DSP-FX, 1993), and medicine (Hesselvik, Carlsson, Brodin, Jorfeldt, & Schildt, 1985; Motoki & Honda, 1977; Siegel et al., 1980).

No one display format can serve all purposes, so designers must identify those cases in which object displays support task performance better than, say, bar graphs or

digital readouts. To help guide these decisions, Wickens and others (Carswell & Wickens, 1987; Wickens & Andre, 1990; Wickens & Carswell, 1995) developed the *Proximity Compatibility Principle* (PCP). In the PCP framework, both tasks and displays have a characteristic called *proximity* that describes the relatedness of the components that make up the task or display. Task proximity refers to the degree to which different variables must be integrated to perform the task. Display proximity refers to the perceptual relatedness of the displayed variables; therefore, an object display is a high proximity display because all components are strongly related through their participation in the formation of a single object. The central hypothesis of PCP is that high proximity tasks are best served by high proximity displays, and low proximity tasks are best served by low proximity displays.

Although experiments have provided persuasive support for PCP, investigators studying *diagnosis* (also called *profile recognition* or *classification*) have repeatedly reported results that contradict PCP's predictions. In a diagnosis task, the participant observes several attributes of a system and then classifies the system as exhibiting one of several prototype states. Medical diagnosis is one example: The physician observes the physiological attributes of a patient and then classifies the patient as suffering from a certain illness. Within the PCP framework, diagnosis qualifies as a high proximity task because it involves the integration of several related variables. Therefore, one might expect a high proximity display, such as an object display, to provide optimal support for diagnosis. Yet, several studies have shown that *lower* proximity displays—such as bar graphs and numeric tables—allow for better diagnosis performance than object displays (Boulette, Coury, & Bezar, 1987; Coury & Boulette, 1992; Coury, Boulette, & Smith, 1989). However, the specific task and display variations featured in past experiments may not represent the entire spectrum of object displays or diagnosis tasks. This paper will cover the specifics of these validity issues later; for now, just understand that the

human factors community still awaits a more complete evaluation of display formats for diagnosis tasks.

To answer this need, this paper describes an experiment that tested the relative merits of an object display, a bar graph, and a digital display in several tasks based on a type of real medical diagnosis, the Physiologic State Severity Classification (PSSC) system developed by Siegel and his associates (e.g., Rixen, Siegel, Espina, & Bertolini, 1997; Siegel et al., 1979; Siegel & Coleman, 1986). The PSSC task has many characteristics that differ from those used in previous studies of display formats for diagnosis tasks, and this experiment incorporated those characteristics to maximize external validity. Furthermore, this design avoided a confound that may have affected previous experiments. In any case, the present results contrast with those of earlier experiments. To understand why object displays might serve PSSC better than previously studied diagnostic tasks, one must fully understand PCP, beginning with a review of the relevant PCP literature.

CHAPTER II

THE PROXIMITY COMPATIBILITY PRINCIPLE

Stimulus Dimensions and Attentional Demands

The origins of PCP grew out of Garner's work (later extended by Pomerantz) on the interaction of attention and the dimensional relationships of visual stimuli (Carswell & Wickens, 1987). Garner (1970) sorted tasks into two categories according to the use of attention. His *selective attention* tasks are those in which one focuses on one particular feature of a stimulus and ignores the other features. In contrast, *divided attention* tasks require attention to many features of a stimulus.

Pomerantz and Garner (1973) classified stimulus dimensions according to the interactions between them. *Separable* dimensions are distinguished by a lack of interaction. Color and form are an example of a pair of separable dimensions because if a polygon's color changes, it causes no change to its form, and vice versa. *Integral* dimensions show strong interactions; in fact, if two dimensions are integral, one cannot exist without the other (Garner, 1970). Hue and brightness are an example of a pair of integral dimensions. *Configural* dimensions show a degree of interaction between separable and integral dimensions. Configural dimensions do not require the presence of the one another to exist, but they do interact to form *emergent features*—features that exist only through the interaction of multiple individual features. These features can be *global*, involving the entire stimulus, such as symmetry and parallelism. Emergent features may also be *local* to a certain area of the stimulus, such as the angle formed between two line segments (Pomerantz & Schwaitsberg, 1975).

In a series of experiments, Garner and Pomerantz found that stimuli with separable dimensions best support selective attention tasks, but stimuli with integral or configural dimensions best support divided attention tasks (for review, see Pomerantz &

Pristach, 1989). Pomerantz and Garner (1973) argued that to focus on one of a set of integral or configural dimensions (a selective attention task), one must strategically reorganize the perception of the stimulus to filter out the unnecessary elements.

However, if the stimulus consists of separable features, perceptual reorganization is not needed, and so the selective attention task requires less cognitive effort. These findings formed the nucleus around which PCP developed.

Extending Dimensional Integrality

Wickens and others adapted and extended Garner's and Pomerantz's work to offer the human factors community a set of design guidelines for information displays. Wickens' work differs from that of Garner and Pomerantz in two important ways. First, Garner and Pomerantz did not choose their stimuli to represent useful information displays; typically, they used sets of parentheses configured in different patterns. In contrast, Wickens' studies use real information displays, such as bar graphs and object displays, for stimuli. Therefore, Wickens' experiments are more likely to produce results that can be directly applied to display design. Second and more importantly, Wickens expanded Garner and Pomerantz's task and stimulus classification systems into the concept of *proximity*.

In the PCP framework, both tasks and displays have the property of proximity, but the term has a slightly different meaning for each. Task proximity is very similar to Pomerantz and Garner's classification of tasks by attentional requirements. A low proximity, or *focused attention*, task is essentially the same as Garner and Pomerantz's selective attention task: It requires processing of only one variable while ignoring other available sources of information (Wickens & Andre, 1990). Likewise, a high proximity, or *integrative*, task is similar to a divided attention task in that the operator must integrate multiple variables. Wickens and Carswell (1995) point out two variations of high proximity tasks: computational tasks and Boolean integration tasks. Computational tasks require some sort of mental arithmetic to be carried out on a range of variables, whereas

Boolean integration tasks require checking whether task variables meet a set of complex conditions.

Although task proximity in PCP is very similar to the Garner and Pomerantz system of classifying tasks, display proximity is a more inclusive concept than Garner and Pomerantz's stimulus dimension categories. A display shows higher proximity when its component variables show *any kind* of similarity or relatedness (Carswell & Wickens, 1987). Therefore, high-proximity displays are composed of highly related components, and low-proximity, or *separable*, displays are made up of less related elements. Wickens and Andre (1990) note that the types of relatedness involved in display proximity fall into two categories: physical metrics and objectness (for review, see Wickens & Carswell, 1995).

Physical metrics covers a wide range of characteristics that include most conventionally recognized variables in display design (Bailey, 1982). Two examples are *color* and *closeness in space*. Another type of physical metric is the *specifier homogeneity*. A specifier is the feature of a display that changes in response to changes in the underlying variable, such as length, color, or brightness (Carswell, 1992). Less obvious types of physical metrics also contribute to display proximity. Displays can vary in *temporal* proximity by revealing elements sequentially or simultaneously. Also, display elements may show *connectedness*; that is, line segments may connect related display elements. For example, four meters might be connected by four lines forming a square. Explicit line segments are not necessary to connect display components. Instead, the connections may be implied by orienting the elements along a single axis. It is important to realize that the connections themselves, whether explicit or implicit, do not change in response to any underlying variable. The connections simply highlight a fixed grouping among the display elements.

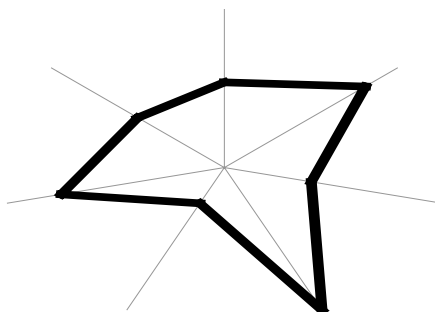


Figure 1. Example of a radial polygon

Objectness is somewhat similar to connectedness in that the display elements involved physically touch one another. But object displays are designed so that changes to the underlying variables result in overall changes to the displayed object: Unlike static connecting lines, the connecting features of an object display *are* mapped to dynamic system variables. An example that has received a great deal of attention in the literature is the *radial polygon*.¹ In a radial polygon, each variable is graphed on one of several axes that radiate out from a single point. A point on each axis represents the current value of a system variable, and those points are connected to form a polygon (see Figure 1). Thus, in an object display, elements are grouped not because of their similarity, but because they form a perceptual object whose shape represents the system state.

According to Wickens and Carswell (1995), objectness is the strongest form of display proximity because of the integral relationship of the features that make up the object. Just as Pomerantz and Garner's (1973) integral dimensions cannot exist independently of one another, an object display cannot exist independently of its

¹ Radial polygons have also been called spoke diagrams (Hesselvik et al., 1985), polar graphics (Munson & Horst, 1986), radial diagrams (Green, Logie, Gilhooly, Ross, & Ronald, 1996), sequential circle diagrams (Siegel et al., 1979), radar charts (Excel, 1999), and polygon displays (Boulette et al., 1987).

constitutive elements. To acknowledge this dependency, object displays are often called *integral displays*.

Just as Pomerantz and Garner (1973) described configural dimensions as showing slightly weaker interactions than integral dimensions, *configural displays* feature slightly less objectness than integral displays. Where the elements of an object display explicitly form a single perceptual object, the components of a configural display only *suggest* the existence of an object. A bar graph is the archetypical example of a configural display. The individual bars of a bar graph are not part of a single object, but their relative heights imply the existence of a contour line that links the bars (see Figure 2).

Recognizing the different forms of task and display proximity is important because of the interactions between them. The central hypothesis of PCP is that task proximity and display proximity interact in a certain way: Task performance is maximized when the proximities of the task and the display are compatible. Thus, high-proximity displays best serve high-proximity tasks and low-proximity displays best serve low-proximity tasks (Wickens & Andre, 1990). Wickens and Andre (1995) write that the mechanism behind the PCP hypothesis is *information access cost*, the sum of movements of attention, the eyes, and the head required to extract the information for a given task. But before discussing how PCP works any further, it is important to know whether PCP makes accurate predictions.

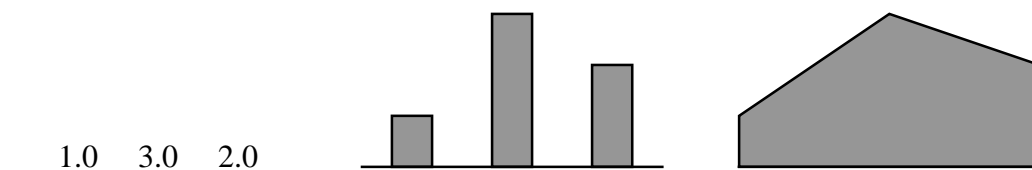


Figure 2. (L to R) Separable, configural, and integral displays. All three displays convey the same information in different formats. Note that the bar graph suggests the contour that is explicitly shown in the integral display,

Experimental Evidence and PCP Interactions

Experimental tests of PCP have provided convincing—although not unanimous—support for the central PCP hypothesis. Most PCP-related experiments compare performance in different types of tasks using several of displays. Low-proximity tasks usually fall into three categories. *Detection* tasks require the participant to search for an abnormal value in any of the displayed variables. *Differentiation* tasks are similar to detection tasks, but the participant must not only determine *whether* a value is abnormal, but also *which* variable is abnormal. Finally, *recall* tasks call for the participant to recall a specific variable's value from memory.

One may classify most high-proximity tasks into two categories. An *operative integration* task requires that the participant perform some combination of the computational and Boolean integration tasks described by Wickens and Andre (1995). In a *diagnosis* task, the participant must classify a variable array as one of a number of predefined system states.

Experiments testing PCP also use displays covering a range of proximity values. Researchers often use an alphanumeric display as the lowest-proximity condition and an object display as the highest-proximity condition. In addition, many investigators use a bar graph to represent a medium degree of proximity.

It is important to realize that PCP predicts an interaction between display type and task type. In its “strong” form, PCP predicts that separable displays support focused tasks better than integrative tasks, and that integral displays support integrative tasks better than focused tasks (see Figure 3). However, some integral displays may facilitate integration more than they hinder focused attention. In such cases, one may observe a “weak” PCP interaction in which the integral display is superior for all types of tasks, but less so for separable tasks (see Figure 4; Wickens & Carswell, 1995). In the following discussion, *PCP interaction* refers to the entire set of interactions that PCP predicts, both weak and strong.

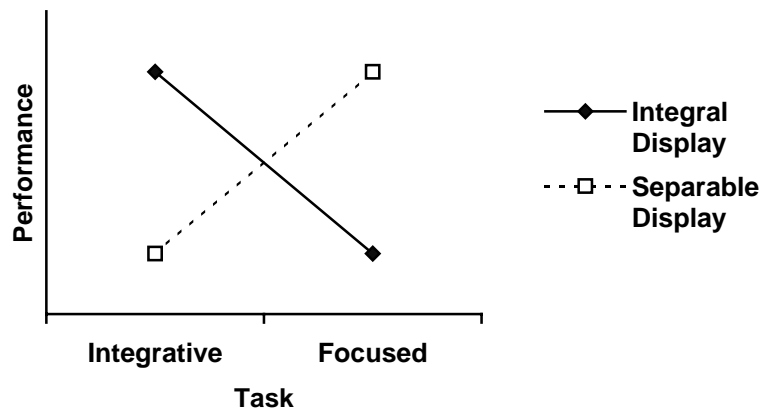


Figure 3. Idealized "strong" PCP interaction: performance as a function of task proximity and display proximity.

The different forms of display proximity described in the literature have varied widely in their adherence to PCP's predictions. The evidence regarding closeness in space has been especially weak. Vincow and Wickens (1993) reported that the degree to which increasing spatial closeness of tabular displays improved performance increased with the complexity of a computational integration task. This pattern of results does seem consistent with PCP. But Andre and Wickens (1988) found that increased spatial closeness supported both focused and integrative tasks. Furthermore, Barnett and Wickens (1988) found that manipulations of closeness in space in a bar graph display had no significant effect on either detection or operative integration tasks. Wickens and

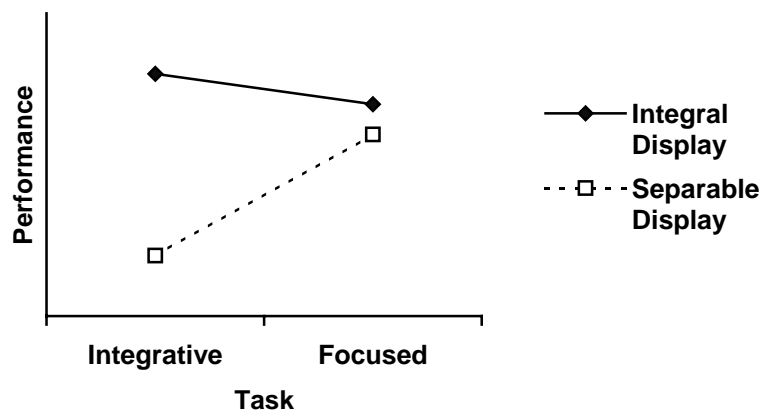


Figure 4. Idealized "weak" PCP interaction: performance as a function of task proximity and display proximity.

Andre (1990) found no effect of closeness in space in a similar experiment using meter displays. One could view these experiments as showing that either PCP is wrong or closeness in space is a weak form of display proximity.

The few experiments involving color similarity have produced slightly more support for PCP. Andre and Wickens (1988) found that color similarity produced a trend towards a PCP interaction, although the trend was not statistically significant. Wickens and Andre (1990) found that color similarity produced a marginally significant PCP interaction in measures of accuracy, but no significant interaction in response time. This evidence indicates that color is not an effective form of display proximity.

The strongest support for PCP comes from experiments involving object displays, in which researchers have often observed the PCP interaction. Carswell and Wickens (1987) found that a radial polygon supported an operative integration task better than a bar graph, but the bar graph was superior for a detection task. Wickens and Andre (1990) reported similar results when comparing a complex object display to a bar graph. Bennett, Toms, and Woods (1993) also found the PCP interaction when they compared an object display with a bar graph in operative integration and recall tasks. Goettl, Kramer, and Wickens (1991), too, found the PCP interaction in a comparison between an object display and a bar graph in detection and computational integration tasks.

Bennett and Flach (1992) found evidence of the PCP interaction in a meta-analysis of 39 experiments concerning graphical displays and attentional resources. The authors concluded that indeed, separable displays served focused attention tasks better than integrative tasks. But object displays seemed to yield better performance than separable displays no matter which type of task was involved. However, the special advantage of object displays in integrative tasks was suggested in a trend of marginally significant and nonsignificant results across several studies. Wickens and Carswell (1995) argue that this overall trend is evidence of a weak PCP interaction, but Bennett

and Flach did not perform the rigorous statistical analysis (such as MTMM) required to confirm the trend.

Not all experimental designs involving object displays have been capable of discovering the PCP interaction, but several studies have demonstrated the advantage of object displays in integrative tasks. For example, Goldsmith and Schvaneveldt (1984) found that several object displays, including a radial polygon, supported accuracy in an operative integration task better than a bar graph. Also, MacGregor and Slovic (1986) reported that a Chernoff face display allowed for higher accuracy in an operative integration task than did a bar graph. Furthermore, Jones, Wickens, and Deutsch (1990) compared a radial polygon display and a bar graph in a computational integration task and found that the radial polygon supported more consistent and accurate performance.

So although PCP's worth is doubtful with regard to physical metrics such as color and spatial proximity, PCP appears to be a useful design principle to guide the choice between separable and integral display formats. But the *nature* of the object display's advantage has been the subject of some debate.

Parallel Processing vs. Emergent Features

Wickens (Wickens & Andre, 1990; Wickens & Carswell, 1995) explains the superiority of object displays in terms of information access cost, which he breaks down into two factors: parallel processing and useful emergent features. By *parallel processing*, Wickens refers to the fact that attention seems to favor whole perceptual objects rather than individual features of objects. In object recognition, the separable features of an object are processed more or less in parallel (Kahneman, Treisman, & Gibbs, 1992; Kahneman & Treisman, 1984). Therefore, in an integrative task, one can match the object display *as a single unit* in memory. The alternative would be to serially process the set of variables in the display. Supporting the notion of parallel processing in object displays, several researchers (Greaney & MacRae, 1996; Greaney & MacRae, 1997; Munson & Horst, 1986) have found that accuracy and speed in detection and

discrimination tasks are independent of the number of variables shown in an object display. Note that based solely on the idea of parallel processing, the nature of a given integrative task is irrelevant to choosing between different object display formats. So long as object displays are salient and discriminable, all integral formats are equally well-suited to all tasks.

In contrast, the details of an integrative task are crucial to evaluating an object display's use of emergent features. A emergent feature can reduce effortful integrative tasks to automatic visual perception if and only if the feature in question is mapped to an important variable; in other words, a well-mapped emergent feature enables perception to replace computation. For example, consider the task of determining the distance traveled by a car given its speed and travel time. If the speed and time are displayed numerically, one multiplies the two variables to find the distance. But if the speed and time are displayed as the width and height of a rectangle, then the task reduces to estimating the area of the rectangle. No computation is involved because the important variable is directly available through the emergent feature. Emergent features do not appear only in object displays; configural displays can have useful emergent features as well. But according to Wickens and Carswell (1995), object displays tend to have especially salient emergent features because of the high proximity of the constituent components.

Although early versions of PCP (for example, as described by Carswell & Wickens, 1987) emphasized parallel processing over useful emergent features, studies soon revealed that emergent features may be the most important if not the *only* important benefit of object displays. Sanderson, Flach, Buttigieg, and Casey (see also Buttigieg, Sanderson, & Flach, 1988; 1989) showed that bar graphs with well-mapped configural emergent features supported an operative integration task better than object displays without such an emergent feature. Sanderson et al. suggested that previous work that supposedly showed the superiority of objectness actually showed the superiority of useful emergent features. In fact, several studies (e.g., Carswell & Wickens, 1987; Jones et al.,

1990) confounded objectness with the presence of emergent features by comparing object displays to “staggered” bar graphs designed specifically to eliminate salient emergent features. Buttigieg and Sanderson (1991) strengthened this notion when they demonstrated that the presence or absence of a useful emergent feature in an object display determined its superiority over a bar graph in an operative integration task. In response to this line of research, Wickens (Wickens & Andre, 1990; Wickens & Carswell, 1995) modified PCP to highlight the role of useful emergent features in both configural and object displays.

It seems that high proximity displays benefit performance by providing emergent features that are well-mapped to the demands of the task. Higher proximity displays have more salient emergent features and thus lead to better performance. Unfortunately, the benefits of emergent features in object displays have not been clear for all types of integrative tasks. In particular, studies of diagnosis tasks have produced puzzling results that deserve close examination.

CHAPTER III

DIAGNOSIS AND OBJECT DISPLAYS

In a diagnosis task, one classifies a system state as belonging to one of a set of prototypical states. Because diagnosis involves many system variables, it is an integrative task within the framework of PCP and so should benefit from the use of an object display. In this case, the useful emergent feature is the overall shape of the object. Different system states give rise to different characteristic object shapes, potentially reducing diagnosis to a perceptual pattern matching task. One could expect configural displays to provide similar benefits, albeit to a lesser degree due to the relatively low salience of the emergent feature.

Object displays do indeed lead to better diagnosis performance, at least in comparison to numeric displays. Coury, Boulette, Zubritzky, and Fisher (1986) had participants classify four-variable profiles into one of four system states using either a tabular numeric or radial polygon display. The authors also measured the profiles' distances from the closest prototype profile in four-variable space, referring to profiles with small distances as *certain* classifications and those with larger distances as *uncertain* classifications. Coury et al. found that the radial polygon group outperformed the digital display group during both learning and extended practice.² The radial polygon's advantage diminished (but did not disappear) only for the most uncertain classifications. Boulette, Coury, and Bezar (1987) used a method similar to that of Coury et al., but added a factor of time stress. The authors confirmed the advantage of radial polygons over digital displays, and observed that this advantage increased with increasing time

² In all of Coury's studies examined in this paper (Boulette et al., 1987; Coury & Boulette, 1992; Coury et al., 1989; Coury & Purcell, 1988), the authors defined performance only in terms of response time. Accuracy was near 100% for all groups.

stress. So both Coury et al.'s and Boulette et al.'s results conform to the predictions of PCP: A high proximity display maximized performance on a high proximity task.

In contrast, comparisons of integral and configural displays in diagnosis tasks have not yielded results consistent with PCP. Recall that the radial polygon, as an integral display, has a higher display proximity than a bar graph, which is generally considered a configural display. On this view, radial polygons should allow for better performance on a diagnosis task than bar graphs. However, Coury and Purcell (1988), duplicating the technique used by Coury et al. (1986), compared a digital display, a bar graph, and a radial polygon in a four-variable diagnosis task. Looking only at the data for the digital display and the radial polygon, Coury and Purcell's results mirrored those of Coury et al., but participants using a bar graph outperformed both of the other groups. Coury and Purcell also performed another experiment identical to the first, except the profile sets featured a non-uniform distribution of distances from the correct prototypes biased towards larger distances and thus more uncertain classifications. With the biased profile sets, the advantage of the bar graph over the digital display disappeared, but both formats supported better performance than the radial polygon. Coury, Boulette, and Smith (1989) replicated Coury and Purcell's experiments with a larger group of participants and recorded similar results.

Coury et al. (1989) attributed the superiority of bar graphs in their and Coury and Purcell's (1988) experiments to two factors. First, they argued that the configural pattern formed by the bars in the bar graph was at least as salient as the shape of the radial polygon. Their opinion is consistent with that of Greaney and MacRae (1997), who showed that participants used parallel processing with bar graphs in a detection task just as readily as they did with object displays. Second, Coury et al. asserted that the values of the individual variables were easier to extract from the bar graph than from the radial polygon. This was especially important in the more uncertain classifications, which required participants to carefully examine individual values to diagnose the profiles.

Although Coury and Purcell (1988) and Coury et al. (1989) seem at first to present a convincing case for the bar graph's superiority over radial polygons in diagnosis tasks, their experiments do not offer a definitive answer for several reasons. Most importantly, their display designs gave the bar graph an unfair advantage. Their bar graphs featured four horizontal grid lines labeled with numeric values that corresponded exactly to the cutoff values that defined the four profiles. For example, one system state was defined in part by a certain variable falling between 25 and 50, and the bar graph featured grid lines labeled 25 and 50. Neither the digital display nor the radial polygon contained comparable demarcation information despite the fact that many radial polygon designs (e.g., those of Beringer, Howard, & Jenkins, 1986; Hesselvik et al., 1985; Hughes & MacRae, 1994) feature *reference polygons*, faint outlines of polygons showing reference values (see Figure 5). Reference polygons contain the same information as horizontal gridlines on a bar graph, making it easier to extract individual variable values. So Coury and Purcell's and Coury et al.'s experimental designs may have confounded display format with the amount of directly available information.

Another possible issue with Coury and Purcell's (1988) and Coury et al.'s (1989) studies lies in the number of variables they used. Researchers have observed that in some circumstances, performance with radial polygons and similar displays increases with the

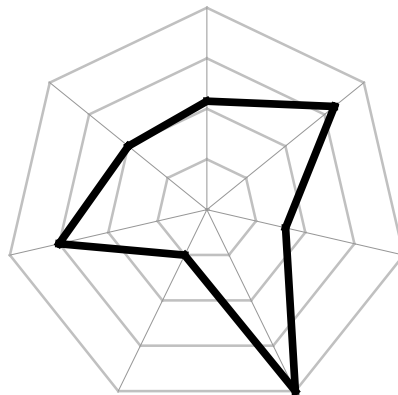


Figure 5. Example of a radial polygon with reference polygons

number of variables mapped to the display (Green et al., 1996; Hughes & MacRae, 1994). Hughes and MacRae suggest that with larger numbers of variables, the “spikes” of a radial polygon become sharper and thus more salient. In addition, Elvers and Dolan (1995) note that the polygons formed in displays with few variables are all simple perspective transformations of one another, and thus difficult to discriminate. Not all studies have shown a positive correlation between variable count and performance in object displays, but *no* study has shown performance to increase with the number of bars in a bar graph (see Greaney & MacRae, 1997). It is possible, then, that Coury and Purcell’s and Coury et al.’s results are a special case of systems with few variables, and do not generalize to systems with larger numbers of variables.

Coury and Purcell’s (1988) and Coury et al.’s (1989) designs do not invalidate their results, but neither do they settle the matter of object displays and diagnosis tasks. Other issues with these experiments spring not from the experimental designs per se, but from a possible limitation on external validity. In particular, Coury and Purcell’s and Coury et al.’s findings might not apply to PSSC diagnosis.

The Physiologic State Severity Classification System

Siegel and his collaborators have developed PSSC for over 25 years to serve as a diagnostic tool in an intensive care environment. The PSSC focuses on *sepsis*, a condition in which a localized infection enters the blood stream and spreads throughout the entire body. Almost any serious infection can lead to sepsis, which in turn can lead to shock and even death. Intensive care patients have a particularly high risk of developing sepsis, making it especially important for physicians to remain vigilant for the condition (Intelihealth, 2000).

The development of PSSC has involved two parallel activities. The first is enumerating the physiologic variables important to the prediction and description of the course of sepsis. (Examples include cardiac index, heart rate, and venous pH.) The second is plotting a typical sepsis patient’s trajectory over time through the

multidimensional “physiologic hyperspace” defined by the chosen variables. Once the typical trajectory is plotted, researchers use cluster analysis to reveal prototypical pathophysiological states, or “waypoints” along the trajectory. Early versions of PSSC used 11 variables and defined 4 prototypical states (Siegel et al., 1979). The current version uses 17 variables and 6 states (Rixen et al., 1997; Rixen, Siegel, & Friedman, 1996), although work is underway to reduce the variable count to 13 (J. H. Siegel, personal communication, May 26th, 2000).

In addition to representing a typical patient’s trajectory, each prototypical state also calls for a particular form of treatment. Physicians use PSSC by monitoring patients for the pathophysiological states and administering the proper treatment if such a state develops. Based on a 5 year trial at Buffalo General Hospital, Siegel et al. (1980) estimated that using PSSC can save the lives of 4.6% more surgical patients and 14.1% more trauma patients than traditional diagnostic techniques. Not surprisingly, PSSC’s acceptance in intensive care units is on the rise (J. H. Siegel, personal communication, May 26th, 2000).

For the present purpose, the most important aspect of PSSC is its display format (see Figure 6). In a PSSC system, each variable is mapped to a spoke of a radial polygon on a computer screen. In order to display all variables on the same scale, values are transformed to z-scores based on each variable’s distribution in an intensive care patient population. A reference polygon or circle marks each integer between –4 and 4 inclusive; the reference is solid for 0 and dashed for other values. Each variable’s label and numeric value are displayed at the end of that variable’s spoke. In addition to the display representing the patient’s current state, the polygon for each prototypical state is shown in the margins.

Siegel (personal communication, May 26th, 2000) chose radial polygons for PSSC not on the advice of the human factors literature, but based on his own intuition that radial polygons facilitate pattern recognition. In contrast, physiologic variable display in

traditional intensive care consists of tabular numeric charts called “flowsheets,” sometimes accompanied by a line graph depicting body temperature over time (T. Lopez, personal communication, May 11th, 2000). However, Coury and Purcell’s (1988) and Coury et al.’s (1989) results suggest that bar graphs, not objects or numbers, are the optimal format for a diagnosis tasks. Despite these findings, it is now possible to identify several reasons why Siegel may have chosen correctly.

PSSC Diagnosis vs. Previous Experiments

Coury and Purcell’s (1988) and Coury et al.’s (1989) experiments featured extra information in their bar graphs and a low number of state variables in their tasks, but PSSC does not have either of these characteristics. As shown in Figure 6, the PSSC polygon contains labeled reference lines, which the Coury polygons lacked. Furthermore, the PSSC polygon uses 17 variables, over four times the number used in the Coury experiments. Recall that systems with higher numbers of variables may give radial polygons an advantage over bar graphs.

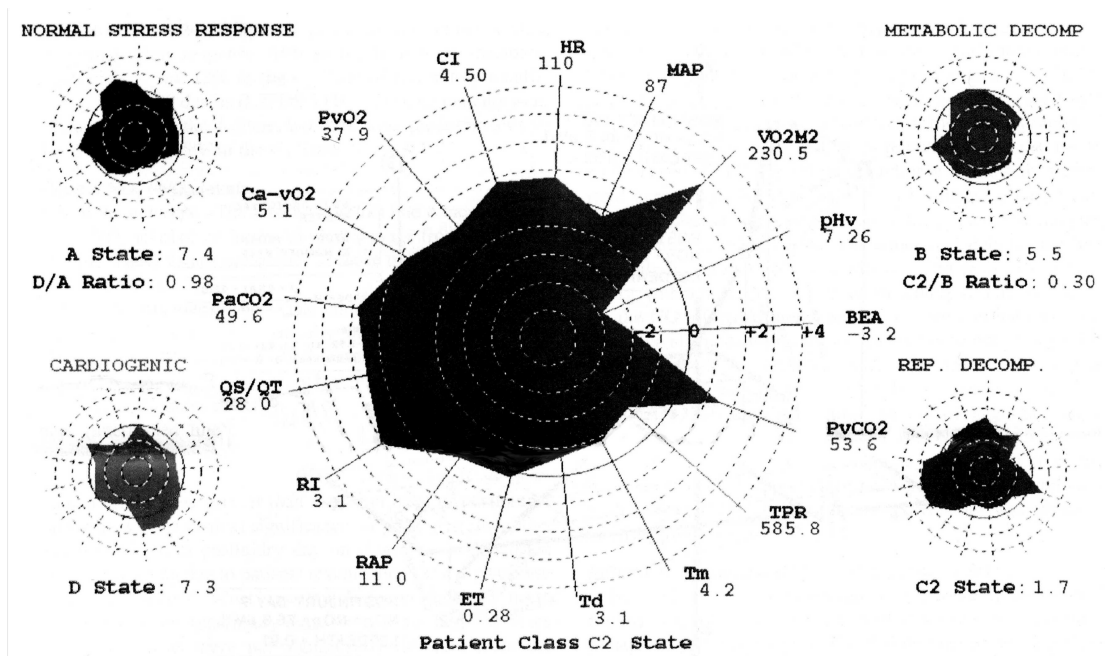


Figure 6. Example of a PSSC display showing a patient state (center) and four prototypical pathophysiological states (adapted from Rixen, 1996).

The PSSC task differs in other ways as well. Recall that Coury and Purcell (1988) and Coury et al. (1989) biased profile sets towards uncertainty in some conditions and found that performance with the object display suffered most in uncertain classifications. However, the prototypical states in PSSC are chosen specifically to *reduce* uncertainty—they are the most likely physiologic profiles for any given intensive care patient to exhibit. So the “profile sets” of real patients are biased not towards but *away from* uncertainty. In addition, uncertain classification carries a very different meaning in PSSC. Each trial in Coury and Purcell’s (1988) experiment had a single correct answer, and in the uncertain classification trials, participants had to resort to a focused attention strategy to determine that answer. But physicians using PSSC do not have to choose a single prototype if the patient is not easily classified. Instead, the physician may customize treatment to suit the patient’s particular condition. Therefore, physicians using PSSC do not have the same incentive to use a focused attention strategy that Coury and Purcell’s participants did. This state of affairs implies that an object display might indeed be the optimal display format for PSSC diagnosis.

Radial polygons may be well-suited for PSSC diagnosis for several reasons. But there is one reason why radial polygons may *not* be a good choice for other diagnostic tasks: differential variable utility. Each of the 17 physiologic variables in PSSC contributes to the definition of all 6 of the prototypical states, but the importance of each variable is not equal across all states; that is, some states show more variability in certain variables than other states do. Although Siegel (personal communication, May 26th, 2000) maintains that differential utility is small in PSSC and does not create a problem for the technique, it remains an interesting question for display research. To the author’s knowledge, in all published experiments involving diagnosis tasks and object displays, all variables in a profile contribute equally to the classification of that profile. No one has examined an object display in which certain features of the object have more diagnostic

value than others. The need for such a study, together with the questionable validity of previous experiments, leads to the present experiment.

The Present Experiment

The study described in this paper had three primary goals. The first goal concerns PCP in general and its relevance to PSSC in specific. This experiment investigated whether a radial polygon does indeed allow for better performance in PSSC diagnosis than a bar graph or digital display, where better performance is operationally defined as higher accuracy and lower reaction time (RT). The experimental task maximized external validity by capturing many important aspects of the real PSSC task. Both the current 17-variable version and the future 13-variable version of PSSC were represented in the experiment. The Coury experiments notwithstanding, PCP predicts that the radial polygon should support superior performance for both the 17- and 13-variable tasks. Thus this experiment tested PCP's predictive power while also evaluating PSSC's display design.

The study's second primary goal was to examine how participants use the display formats for diagnosis. Recall that PCP states that integral and configural displays benefit integrative tasks by encouraging parallel processing and by providing emergent features. If this is the case, then the variable count should have no effect on the RTs of participants using the graphical displays. In contrast, RTs with the digital display should be longer in the 17-variable condition as the participants focus attention on individual variables.

The third primary goal of this study was to examine the suitability of different display formats for diagnosis tasks featuring differential variable utility. To accomplish this, in one condition, the prototypical states were defined in terms of only 13 of the 17 available variables, and so participants had to focus on a subset of the total set of variables. In the framework of PCP, the role of focused attention means that this "partial diagnosis" task has a lower task proximity than "normal diagnosis." If the task has a lower proximity, PCP predicts that lower proximity displays should yield optimal

performance because emergent features (such as the shape of the radial polygon or the implied contour of the bar graph) are irrelevant to the task. If this is true, the presence of differential utility should be detrimental to performance with the graphical displays, but have relatively little impact on the digital display. Additionally, PCP characterizes configural displays as “optionally separable” and thus better suited than integral displays to low proximity tasks. If this is true, differential utility should have a larger impact on performance with the radial polygon than with the bar graph.

To summarize, this experiment tested four primary hypotheses:

1. *Normal diagnosis performance is best with the radial polygon.*
2. *Increasing the variable count in normal diagnosis increases RT with the digital display, but not with the graphical displays.*
3. *The performance decrement of partial diagnosis versus normal diagnosis is greater with the graphical displays than with the digital display.*
4. *The performance decrement of partial diagnosis versus normal diagnosis is greater with the radial polygon than with the bar graph.*

The first hypothesis addresses *whether* radial polygons are the superior display for PSSC. Evaluating the second hypothesis gives some insight into *why* radial polygons are (or are not) superior. The third and fourth hypotheses ask if this display superiority might *generalize* to other diagnosis tasks with high differential utility.

This experiment also tested exploratory hypotheses concerning the effect of proximity compatibility on user preferences. User preferences are important to study in part because preference is a strong indicator of willingness to adopt a given technology, at least in the short run (Knutson, 1998). In addition, user preference is often correlated with performance (Coll & Wingertsman, 1990; Gerhardt-Powals, 1996; Kang & Muter, 1989; Rahman & Muter, 1999) so preference data may help to confirm findings based on the performance variables. However, researchers have shown that users sometimes prefer less effective displays, tending to prefer graphical displays over digital displays

regardless of performance (Antin, 1988; Karat, McDonald, & Anderson, 1986; Knutson, 1998). For this reason, strong hypotheses regarding the relationship between user preference, performance, and proximity compatibility are not possible. For exploratory purposes, it was hypothesized that participants prefer the display with the highest proximity compatibility with their task: the radial polygon for normal diagnosis and the digital display for partial diagnosis. It was further hypothesized that participants *least* prefer (or *disfavor*) the displays with the lowest proximity compatibility with their task: the digital display for normal diagnosis, and the radial polygon for partial diagnosis. In any case, this experiment was the first to examine the relationship between proximity compatibility and user preference.

CHAPTER IV

METHOD

Participants

A total of 234 students (81 men, 153 women, mean age = 19.3 years) from the University of Georgia research pool participated in the experiment in return for credit in undergraduate psychology courses. The ratio of men to women reflects the makeup of the University of Georgia research pool. Each experimental group contained 27 men and 51 women.

Diagnosis Tasks

Participants in each group performed one of three diagnosis tasks. In each trial of all three task variations, participants were shown three prototype profiles and one target profile on a computer screen. Participants classified the target profile as belonging to one of the prototypes by pressing the 1, 2, or 3 keys on the computer's keyboard. The prototype profiles remained the same throughout all trials.

In the *17-Variable Normal Diagnosis (ND17)* and *13-Variable Normal Diagnosis (ND13)* tasks, each of the variables contributed equally to the classification of target profiles. In the *Partial Diagnosis (PD)* task, only 13 of the 17 available variables were relevant to the classification of each prototype state. The irrelevant variables were different for each prototype.

Prototype Profiles

The three prototype profiles were adapted from the *A*, *C*₂, and *D* states³ in PSSC (see Rixen et al., 1997) and relabeled states 1, 3, and 2, respectively (see Table 1). Variables in the PSSC states were relabeled *A* through *Q*. Each PSSC prototype was

³ State *A* indicates a normal stress response; state *C*₂, respiratory insufficiency; and state *D*, cardiogenic decompensation.

Table 1
Prototype Profiles

State	Variable																
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>	<i>L</i>	<i>M</i>	<i>N</i> [*]	<i>O</i> [*]	<i>P</i> [*]	<i>Q</i> [*]
1	1	-1	1	-1	-1	0	-2	-2	-2	-1	0	0	2	0	-2	1	2
2	0	-1	-2	0	-1	0	1	2	2	-2	-1	0	-1	-1	1	-2	-2
3	1	-1	0	-2	-1	2	-2	-1	-1	-2	0	2	2	2	-1	0	0

* Variables not present in the *ND13* task.

modified so that exactly 4 variables had a value of 0. In the *PD* task, only the non-zero variables were relevant to diagnosing that particular state. In the *ND13* task, variables *N* through *Q* were removed from the profiles.

Profile Sets

Each profile set for a task consisted of a block of 9 training profiles and a block of 21 testing profiles for a total of 30 trials. Within each block, an equal number of profiles were correctly classified as each of the three prototypes. The profiles' presentation order within each training and testing block was randomized.

Each target profile in the *ND17* task was created by adding a different random number between -1 and 1 to each variable in a prototype profile. Profiles in the *ND13* task were identical to those in the *ND17* task except variables *N* through *Q* were removed. Profiles in the *PD* task were identical to those in the *ND17* task except the irrelevant variables were assigned random values between -4 and 4 .

Display Formats

Each participant performed the same diagnostic task using three different display formats: *Radial Polygon (RP)*, *Bar Graph (BG)*, and *Tabular Digital (TD)*. In all formats, three small prototype "disease" profiles were lined up down the left edge of the screen and labeled 1 through 3. A large target "patient" profile appeared to the right.

The target and prototype profiles subtended visual angles of 7.4° and 3.2° , respectively.

Each display type is presented in Figures 7 through 9.

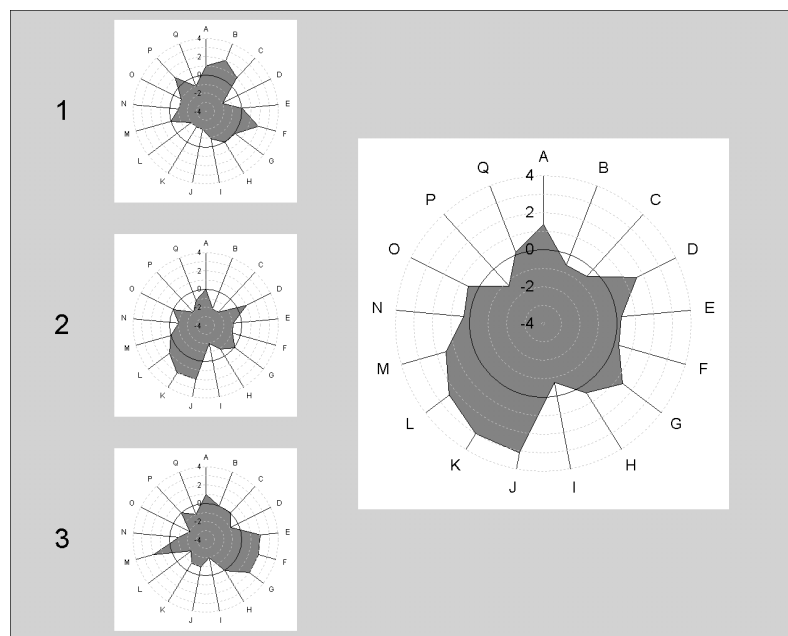


Figure 7. The radial polygon display

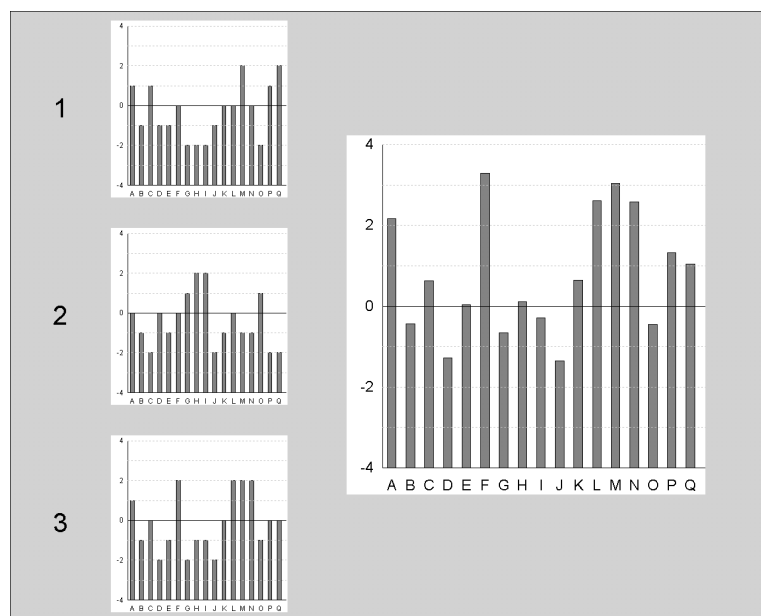


Figure 8. The bar graph display

A	B	C	D	E	F
1.0	-1.0	1.0	-1.0	-1.0	0.0

G	H	I	J	K	L
-2.0	-2.0	-2.0	-1.0	0.0	0.0

M	N	O	P	Q
2.0	0.0	-2.0	1.0	2.0

A	B	C	D	E	F
0.0	-1.0	-2.0	0.0	-1.0	0.0

G	H	I	J	K	L
1.0	2.0	2.0	-2.0	-1.0	0.0

M	N	O	P	Q
-1.0	-1.0	1.0	-2.0	-2.0

A	B	C	D	E	F
1.0	-1.0	0.0	-2.0	-1.0	2.0

G	H	I	J	K	L
-2.0	-1.0	-1.0	-2.0	0.0	2.0

M	N	O	P	Q
2.0	2.0	-1.0	0.0	0.0

A	B	C	D	E	F
1.6	-0.2	0.9	-1.4	0.2	2.7

G	H	I	J	K	L
-0.9	-0.2	-0.3	-1.3	1.1	2.7

M	N	O	P	Q
3.0	3.3	-0.2	1.4	1.4

Figure 9. The tabular numeric display

Experimental Design

The experiment featured a mixed 3×3 factorial design. It was judged that switching between task types might lead to more confusion and differential carryover than switching between display formats; therefore, task was the nonrepeated factor and display type was the repeated factor. The presentation order of display formats was counterbalanced within each task. All levels of both factors was treated as fixed.

Procedure

Participants were told they would act as a doctor diagnosing patients by looking at the patients' medical charts. Participants were instructed to choose which of three diseases most closely matched each patient's condition. The experimenter then explained how to read all three display formats using examples from handouts given to each participant. An example classification was given for each display type. Finally, the experimenter asked that the participants diagnose the patients as quickly and accurately as possible.

Each participant completed a total of 90 trials in 3 blocks of 30 for each display type. Within each display type block, the first 9 trials were training trials following the same sequence: A target profile and three prototype profiles were displayed; the participant pressed a key indicating the diagnosis; a feedback screen appeared showing the participant's diagnosis, the correct diagnosis, and the number of correct diagnoses in this set of training trials; the participant pressed a key; and the sequence repeated. The 21 testing trials were identical to the training trials except that only the participant's diagnosis was shown on the feedback screen. After all trials were completed, a preference elicitation screen appeared, and participants selected their preferred display and their least preferred display. A program written in C++ running on Dell computers with 17" color monitors displayed the stimuli and collected response data.

CHAPTER V

RESULTS

Participants were excluded from analysis if their total number of correct diagnoses on all 63 testing trials was not greater than chance ($p = .05$). The probabilistic structure of the task dictated that participants must correctly answer 26 testing trials to meet this criterion. Two participants in the *PD* condition were replaced because they did not meet this criterion.

Tests for an effect of sex on RT and accuracy did not reveal any significant main effects, nor did sex interact with task type or display type on RT or accuracy. Thus, sex was not considered a factor in subsequent hypothesis testing.

Much of the analysis involves interaction contrasts. For some of the interaction contrasts, the *NDI7* and *NDI3* conditions are grouped into a single *normal diagnosis* (*ND*) condition. Similarly, the *RP* and *BG* conditions are sometimes grouped into a single *analog graphic* (*AG*) condition.

The overall patterns of RT and accuracy results are shown in Tables 2 and 3, respectively. Heteroscedasticity is unacceptably high for both dependant variables (RT

Table 2
Reaction Time in Milliseconds for Task by Display

Display ^a	<i>NDI7</i>		<i>NDI3</i>		<i>PD</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>RP</i>	5157	2284	4550	2325	15528	7729
<i>BG</i>	9161	4220	8307	4472	18997	9148
<i>TD</i>	16914	7895	14034	5501	23998	10297

^a Display is a repeated factor; $n = 78$ for each group.

Table 3
Percentage Accuracy for Task by Display

Display ^a	<i>ND17</i>		<i>ND13</i>		<i>PD</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>RP</i>	97.2	4.4	94.8	5.3	87.3	10.9
<i>BG</i>	94.6	5.4	92.4	6.6	84.9	11.3
<i>TD</i>	92.2	8.8	86.0	13.0	84.3	14.6

^a Display is a repeated factor; $n = 78$ for each group.

$F_{max} = 20.3$, accuracy $F_{max} = 11.2$). In the RT data, variance is roughly proportional to the group mean, and Winer, Brown, and Michels (1991) suggest the transformation $x' = \log x$ for such situations. In the accuracy data, heteroscedasticity is caused by a ceiling effect, as many participants approached perfect performance. For data such as accuracy that are composed of proportions, Winer et al. recommend the transformation $x' = 2\arcsin[x - (2n)^{-1}]^{1/2}$, where n is the number of observations that contribute to the proportion ($n = 21$ for this experiment). The transformed data for RT and accuracy feature more reasonable levels of heteroscedasticity (RT $F'_{max} = 1.5$, accuracy $F'_{max} = 3.1$), especially considering that no test involves more than two levels of any factor. All ANOVA tests use the transformed data, and graphs show both untransformed and transformed data. The alpha level for each test is .05 unless otherwise noted.

Normal Diagnosis Performance

Hypothesis one states that the *RP* group will have the best performance in the *ND* conditions. The hypothesis predicts that *RP* accuracy will be higher and *RP* RT will be lower in each comparison.

The *RP* group reacted faster than the *BG* group, $F(1, 154) = 163.3, p < .001$ (see Figure 10). The *RP* group also reacted faster than the *TD* group, $F(1, 154) = 675.3, p < .001$. Furthermore, the *RP* group was more accurate than the *BG* group, $F(1, 154) =$

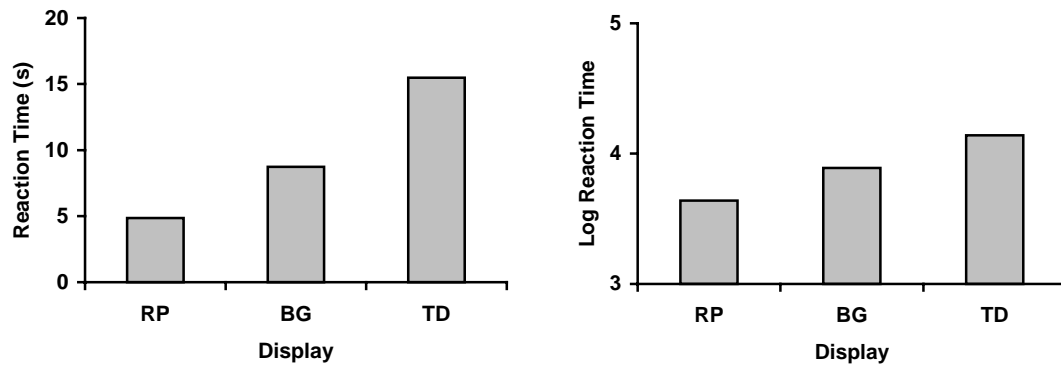


Figure 10. Untransformed and transformed mean reaction times for *RP*, *BG*, and *TD* displays within the *ND* group ($n = 156$).

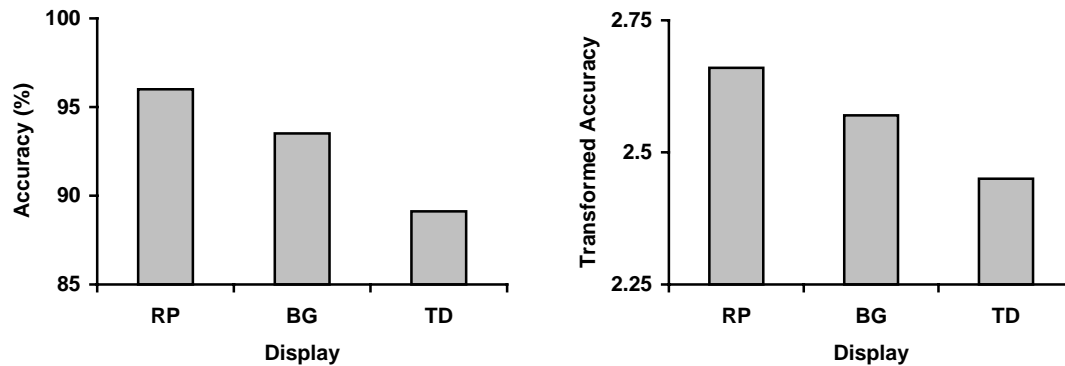


Figure 11. Untransformed and transformed mean accuracies for *RP*, *BG*, and *TD* displays within the *ND* group ($n = 156$).

18.8, $p < .001$ (see Figure 11). Finally, the *RP* group was more accurate than the *TD* group, $F(1, 154) = 56.1$, $p < .001$. Each effect is significant and in the predicted direction.

Variable Count

Hypothesis two states that increasing the variable count in normal diagnosis will increase RTs only with the digital display. However, an interaction contrast comparing RTs of the *AG* and *TD* conditions in the *ND13* and *ND17* conditions reveals no significant interaction, $F(1, 154) = .234$, $p = .63$ (see Figure 12).

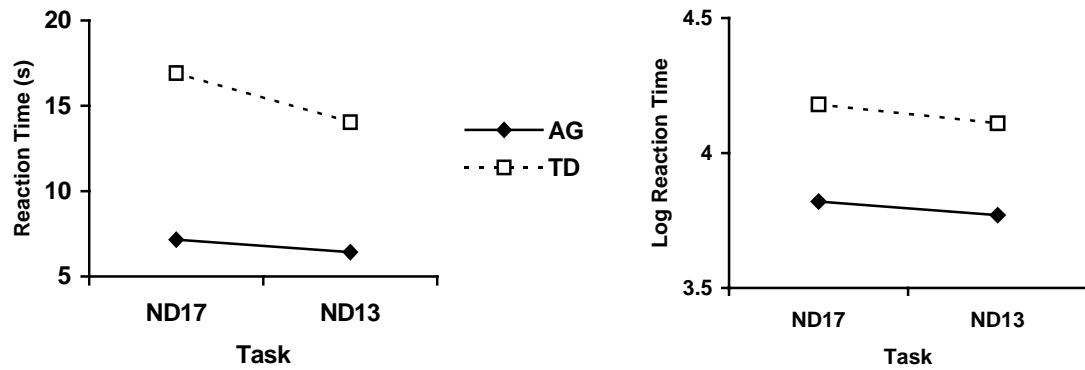


Figure 12. Untransformed and transformed mean reaction times for AG and TD displays in the ND17 ($n = 78$) and ND13 ($n = 78$) groups.

Differential Utility

Hypothesis three states that the performance decrement of partial diagnosis compared to normal diagnosis will be greater with the graphical displays than with the digital display. Interaction contrasts involving the AG, TD, ND, and PD conditions reveal significant interactions for both RT, $F(1, 232) = 37.8, p < .001$ (see Figure 13) and accuracy, $F(1, 232) = 7.6, p < .01$ (see Figure 14). As predicted, the AG-ND—AG-PD difference is greater than the TD-ND—TD-PD difference for both RT and accuracy.

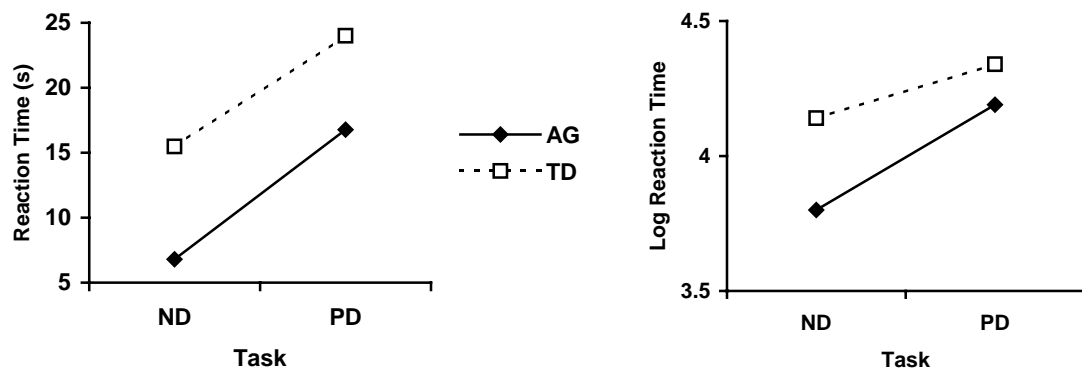


Figure 13. Untransformed and transformed mean reaction times for AG and TD displays in the ND ($n = 156$) and PD ($n = 78$) groups.

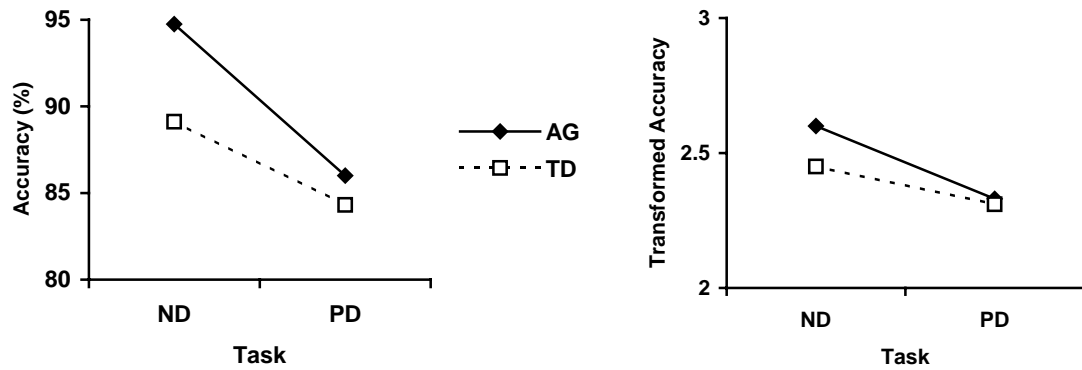


Figure 14. Untransformed and transformed mean accuracy for AG and TD displays in the ND ($n = 156$) and PD ($n = 78$) groups.

Hypothesis four states that the performance decrement of partial diagnosis compared to normal diagnosis will be greater with the radial polygon than with the bar graph. Interaction contrasts involving the *RP*, *BG*, *ND*, and *PD* conditions reveal a significant interaction for RT, $F(1, 232) = 21.5, p < .001$ (see Figure 15), but not for accuracy, $F(1, 232) = .41, p = .52$ (see Figure 16). As predicted, the *RP-ND—RP-PD* difference in RT is greater than the *BG-ND—BG-PD* difference.

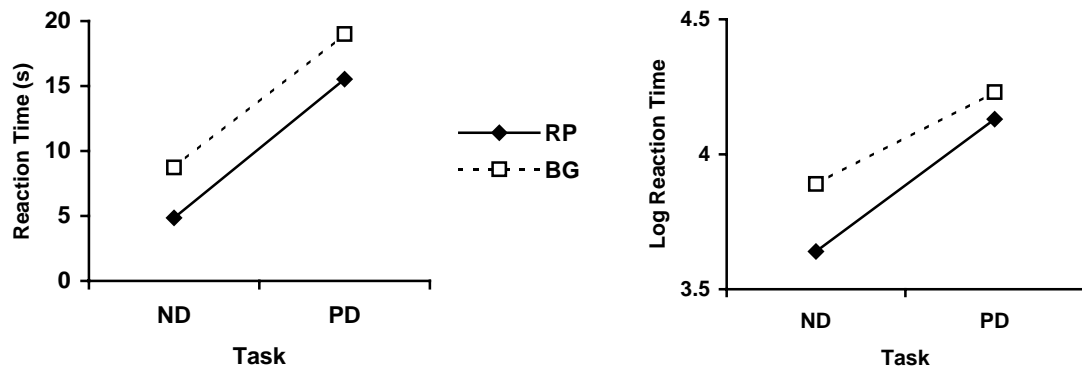


Figure 15. Untransformed and transformed mean reaction time for RP and BG displays in the ND ($n = 156$) and PD ($n = 78$) groups.

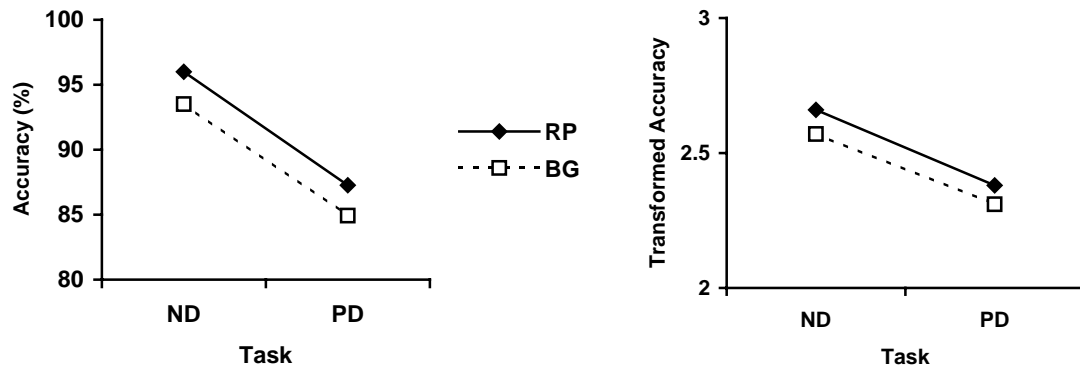


Figure 16. Mean accuracy for RP and BG displays in the ND ($n = 156$) and PD ($n = 78$) groups.

User Preference

The exploratory hypotheses state that participants will prefer the display with the highest proximity compatibility with their task and disfavor the display with the lowest proximity compatibility with their task. The analysis takes the following form for both the preference and disfavor data: First, a “gateway” χ^2 test for independence is used to test whether the distributions of the ND and PD groups differ; if the distributions are different, two “follow-up” χ^2 tests are used to test whether the ND and PD distributions

Table 4
Frequency and Percentage of Display Preference by Task

Measure and task	Display		
	RP	BG	TD
Frequency			
ND	119	30	5
PD	29	24	10
Percentage			
ND	77.3	19.5	3.2
PD	46.0	38.1	15.9

Table 5
Frequency and Percentage of Display Disfavor by Task

Measure and task	Display		
	<i>RP</i>	<i>BG</i>	<i>TD</i>
Frequency			
<i>ND</i>	5	14	135
<i>PD</i>	17	12	34
Percentage			
<i>ND</i>	3.2	9.0	87.8
<i>PD</i>	27.0	19.0	54.0

differ from a random distribution. For each follow-up χ^2 test, the alpha level is set to .025. Note that the preference questionnaire was introduced partway into data collection, and as a result, 2 *ND* participants and 15 *PD* participants did not receive the questionnaire.

For the preference data, the *ND* and *PD* distributions differ, $\chi^2(2) = 22.9, p < .001$. Furthermore, the *ND* preference distribution differs from a random distribution, $\chi^2(2) = 139.9, p < .001$. The *PD* preference distribution also differs from random, $\chi^2(2) = 9.2, p < .01$. However, the *PD* participants did not tend to prefer the *TD* display as the hypothesis predicted; they preferred the *RP* display, albeit to a lesser degree than *ND* participants.

For the disfavor data, the *ND* and *PD* distributions differ, $\chi^2(2) = 35.1, p < .001$. The *ND* disfavor distribution differs from random, $\chi^2(2) = 205.3, p < .001$. The *PD* disfavor distribution also differs from a random distribution, $\chi^2(2) = 12.7, p < .005$. But once again, the *PD* participants did not tend to disfavor the *RP* display as the hypothesis predicted; *PD* disfavored the *TD* display, but less so than the *ND* participants.

CHAPTER VI

DISCUSSION

All hypotheses evaluated in this experiment have implications for two domains. First, the results have bearing on the PSSC diagnosis technique. The physicians who developed PSSC have used radial polygons for over 20 years, but this experiment was the first empirical test of whether they have chosen correctly. Second, this experiment helps determine the usefulness of PCP in designing displays for diagnosis tasks. Although many experiments have confirmed PCP's predictions with regard to integral and separable displays in many different types of tasks, PCP has fallen short in tests with diagnosis tasks. This experiment was designed to retest PCP in a diagnostic context while building on previous research on the subject. The following discussion of each hypothesis involves both applied and theoretical issues.

Normal Diagnosis Performance

The first hypothesis states that the *RP* group should outperform the *BG* and *TD* groups in the *ND* conditions, and the results support this hypothesis. Two factors make this support especially strong. First, the *RP* display bested the other two displays in both RT and accuracy; the data show no speed-accuracy tradeoff. Second, performance with the *BG* and *TD* displays was inferior despite the participants' many years' experience with both formats. A short explanation and nine training trials with a radial polygon were enough to overcome an entire grade-school education working with numbers and bar graphs. Clearly, these results suggest that among the displays tested, the *RP* display is the best format for PSSC-style diagnosis.

Both the RT and accuracy data are important for PSSC diagnosis. The importance of accuracy in medical diagnosis is plain: The physician needs to make an accurate diagnosis to choose an effective treatment. The importance of RT is less

obvious, especially because most of the observed differences in RT are under three seconds. However, PSSC diagnosis is intended for an intensive care environment where seconds can literally make the difference between the life and death of a patient. Taken together, the RT and accuracy results make a strong case that PSSC's designers made the correct choice for their information display format.

The *RP* format's superiority also implies that hospitals should continue to work towards a "computer at every bed" standard for easy graphical display of physiologic data. Note that real displays, unlike their experimental counterparts, need not display information in only one format. The real PSSC display, for example, shows numeric data as well as graphical data, and allows the operator to switch to a tabular numeric format at the press of a button. So making graphical data available for diagnosis does not necessarily preclude the use of numeric tables for providing precise values of individual variables, as has been the concern at some hospitals (T. Lopez, personal communication, May 11th, 2000).

The *ND* performance data also have important implications for PCP. Unlike previous experiments examining object displays for diagnosis tasks (e.g., Coury & Boulette, 1992; Coury et al., 1989), the results of this study conform to the predictions of PCP: A high-proximity display yielded the best performance for a high-proximity task. This finding adds to the large body of work that has shown PCP to be a valid display design guideline for many types of tasks—tasks that may include diagnosis after all.

But why do the present results seemingly conflict with those of earlier investigations? While this experiment cannot provide a definitive answer, the difference between the past and present experimental tasks suggests that diagnosis tasks can vary widely in their task proximity. Recall that the Coury tasks had several features that encouraged or even required participants to focus on individual variables. In contrast, the *ND* tasks in the present study offered no such encouragement. And yet, the Coury tasks and the *ND* tasks are all clearly diagnosis tasks despite their different levels of task

proximity. The goals of the tasks are the same, but the processes needed to achieve those goals differ. If one takes into account these differing levels of proximity, then the apparent conflict between the results disappears. In the present study, a high-proximity diagnosis task was best served by a high-proximity display (an object display); in the Coury studies, a lower-proximity diagnosis task was best served by a lower-proximity display (a bar graph). The predictions of PCP hold equally well in both cases.

Variable Count

The primary reason for using two versions of the normal diagnosis task was to represent two versions of PSSC: the 17-variable version that is in use and the 13-variable version currently under development. But using these two tasks also provided the opportunity to test PCP's accounting of why integral display formats benefit integrative tasks. According to PCP, integral and configural displays encourage parallel processing of information through the use of emergent features. Therefore, the second hypothesis states that only the *TD* group should show a longer RT in the *ND17* condition than in the *ND13* condition. However, the results do not support this hypothesis.

In all likelihood, the range of variable counts in the two *ND* tasks was insufficient to produce positive results. But if these results provide little information about the processes that underlie PCP, they at least show the generality of PCP: Proximity compatibility applies equally well across the range of variable counts represented in this experiment. For PSSC, this means that the display need not change when the technique transitions from 17 variables to 13.

Differential Utility

Taken together, hypotheses three and four state that the performance impact of partial diagnosis versus normal diagnosis should be least with the digital display and greatest with the radial polygon. Unlike normal diagnosis, partial diagnosis requires the participant to ignore several individual variables. This lessens the benefits of emergent features because only portions of the radial polygon's shape and the bar graph's contour

are relevant to the diagnoses. So in the PCP framework, partial diagnosis has a lower proximity than normal diagnosis, indicating that performance in the *PD* condition should differ from performance in the *ND* conditions. Displays with higher proximity should yield larger changes in performance.

The results support these hypotheses. The presence of differential utility hindered performance with the *AG* displays more than the *TD* display in both RT and accuracy. Furthermore, differential utility slowed RTs with the *RP* display more than the *BG* display (although accuracy with the two displays was affected equally). But while each of these interactions occurred in the direction that PCP predicts, the interactions are ordinal and so do not change the rank order of performance with the three display formats. This pattern of results resembles the weak PCP interaction described by Wickens and Carswell (1995).

While the results support PCP, one might still ask why the rank order of the three displays did not change in the *PD* condition. At least two explanations might account for this finding. First, the *RP* and *BG* displays might allow for easier focused attention on individual variables than their display proximities imply; that is, the formats may facilitate integration more than they hinder focused attention. Such results are not unusual; for example, Bennett and Flach (1992) found more evidence for benefits than for disadvantages of object displays. Yet many studies (e.g., Carswell & Wickens, 1987; Gillie & Berry, 1994; Greaney & MacRae, 1997) have clearly shown that high-proximity displays are inferior for some low-proximity tasks. So the question remains why the higher-proximity displays are superior for *this* particular task.

Perhaps the best answer is that the proximity of *PD* task is not very low after all. Even though the task requires filtering out four variables for each prototype state, participants must still integrate the remaining 13 variables. And while the overall shape of the *RP* display was of limited utility, several regions of the display had characteristic contours for each prototype state: for state 1, *O-E* (note that the *RP* display “wraps” from

Q back to *A*), and *G-J*; for state 2, *G-K* and *M-Q*; and for state 3, *D-J* and *L-O*.

Participants may have treated these regions as emergent features in their own right. In contrast, studies showing a strong PCP interaction have often featured very low-proximity tasks (e.g., differentiation) that require far less integration than the *PD* task. The present results make perfect sense if one takes into account the limited role of integration in the *PD* task: Integration was helpful but not vital to the *PD* task, and likewise high display proximity was helpful but not as critical as it was in the *ND* tasks.

On a practical level, the results of hypothesis three and four indicate that display designers should consider the effect of differential utility on the task proximity of diagnosis tasks. But differential utility should not be the only factor nor the most important factor in such designs. Even the extreme level of differential utility in the *PD* task did not eliminate the advantage of the higher proximity displays, and smaller ranges of utility would likely have an even smaller impact. Still, designers might want to group together important variables to create larger regions of high utility within a display. These regions may maximize the benefits of high-proximity displays.

User Preference

Although the main purpose for this study was to investigate the interaction of task type and display type on performance, the user preference data provide some interesting insight into PCP and its relationship to PSSC. According to the hypothesis, participants should prefer the displays that PCP predicts are best suited to their task; that is, participants should prefer the radial polygon in the *ND* conditions and the digital display in the *PD* condition. Conversely, participants should least prefer the displays that PCP predicts are worst suited to their task: the digital display in the *ND* conditions and the radial polygon in the *PD* condition.

The user preference data support these hypotheses to some degree. Contrary to the predictions, participants tended to prefer the *RP* display and disfavor the *TD* display in both task conditions, not just in the *ND* conditions. At first glance, this pattern seems to

reflect only the ubiquitous tendency to prefer graphical displays. However, the task conditions affected both preference and disfavor distributions in the *direction* that PCP predicts, just not enough to change the most or least favored displays. In the lower-proximity *PD* task, smaller proportions of participants preferred the highest-proximity display and disfavored the lowest-proximity display.

The user preference results help to confirm other findings in this paper. The first hypothesis predicted that the *RP* display would yield the best performance in the *ND* conditions. Not only did *ND* participants perform better with the *RP* display, they tended to prefer it as well. And the third and fourth hypotheses predicted that differential utility would impact higher-proximity displays more than lower-proximity displays. Differential utility reduced not only the performance advantage but also the preference for the *RP* display. So both the performance data and the user preference data support PCP as a valid display design principle.

But the user preference data do not just buttress the preference data; they are also valuable in their own right. Preference improves the willingness to adopt new technologies (Knutson, 1998), and the results of this experiment indicate that users tend to prefer displays with the highest proximity compatibility with their task. Therefore, maximizing proximity compatibility not only optimizes the use of information, it may also improve the chance that the information is used *at all*. This is especially important for systems like PSSC that provide user-configurable display formats. Users can easily select a traditional tabular display, so they need a compelling reason to choose the object display. This study suggests that physicians have such a reason: They like the object display better.

Conclusion

This experiment answers several important questions for both practitioners and theoreticians. For the practicing medical community, this study examined the impact of using different types of display formats on diagnoses using the PSSC system. The

experimental tasks captured many crucial aspects of PSSC that previous experiments have not represented: The prototype patterns were always visible; the variable count was high (17 or 13); and profile sets were biased towards certain classifications. Furthermore, the displays in this experiment represented two formats that hospitals currently use (radial polygon and numeric table) as well as another format shown to be superior in prior experiments (bar graph; Coury et al., 1989; Coury & Purcell, 1988). The results point to four major conclusions for practitioners:

1. *The radial polygon is the best format for PSSC diagnosis.* Intensive care diagnosis requires high accuracy and quick response. In the *ND* conditions, participants excelled in both measures while using the *RP* display. This indicates that PSSC's designers made the right choice in using a radial polygon for the technique.
2. *Reducing PSSC's variable count to 13 will not change the optimal display.* Although the current form of PSSC uses 17 variables, work is underway to reduce that number to 13. In this experiment, the variable count affected RTs with the *AG* and *TD* displays equally. Although this test was predicted to show an interaction, retaining the null hypothesis suggests that the PSSC display need not change when the variable count changes.
3. *Increasing differential utility reduces but does not eliminate the advantage of the radial polygon in PSSC diagnosis.* In this experiment, increasing differential utility did not change the rank order of performance with the three displays although the advantage of high-proximity displays was reduced. In the present version of PSSC, differential utility is small and need not influence the display design. But for other diagnosis tasks, designers should be aware that differential utility has a real—but not necessarily overwhelming—impact on task proximity.
4. *The radial polygon format improves users' preference as well as performance in PSSC diagnosis.* Research shows that PSSC can save lives, but physicians must

invest time and effort to learn the technique. They may be more likely to make that investment if they prefer to use PSSC's display over standard hospital flowsheets. Participants preferred the *RP* display, particularly in the *ND* conditions, suggesting that physicians will indeed favor the PSSC display.

All told, this study gives an excellent prognosis for the PSSC diagnosis technique. Still, some aspects of PSSC diagnosis were not represented in the design either because they were not relevant to PCP or because they were not practical for undergraduate participants. For instance, the participants had not undergone years of training on the task with a different display, nor did they experience high levels of stress or distraction. Future studies should examine diagnosis displays in actual intensive care environments.

For the research community, this experiment tested the predictions of PCP in several types of diagnosis tasks. The concept of matching display proximity to task proximity has proven useful in choosing between integral and separable displays for many types of tasks. However, the superiority of bar graphs in previous studies of diagnosis displays (Coury et al., 1989; Coury & Purcell, 1988) has cast doubt on PCP's usefulness. This experiment retested PCP with refined display designs and tasks that more closely resembled a technique of medical diagnosis. Furthermore, this study tested PCP in a partial diagnosis task to investigate the impact of differential utility on proximity compatibility. The results lead to four major conclusions for researchers:

1. *PCP is a valid theory for diagnosis displays.* In almost every hypothesis, the results bear out the predictions of PCP. The only exception is the lack of interaction between variable count and display on RT, and this is probably due to the limited range of variable counts represented in the experiment.
2. *Differential utility decreases task proximity in diagnosis tasks.* Although the rank order of the displays did not change in the *PD* condition, the advantage of high display proximity was reduced in both performance and preference. As predicted, differential utility apparently decreased the proximity of the diagnosis task.

3. *Different types of diagnosis tasks have different task proximities.* The superiority of the *RP* display contrasts sharply with Coury and Purcell's (1989) findings. This makes sense, considering that Coury and Purcell's task encouraged serial processing while the *ND* tasks did not. Clearly, diagnosis tasks can vary widely in task proximity.
4. *Proximity compatibility influences display preference as well as performance.* In the *ND* tasks, participants tended to prefer the most compatible display. When the compatibility was decreased (in the *PD* condition), preference for the *RP* display decreased as well. This is the first study to show that proximity compatibility can impact display preference.

In sum, the results show that PCP is as useful a display design guideline for diagnosis tasks as for other tasks; however, just because a task involves diagnosis does not always mean that it has a high task proximity.

The relationship of PCP to diagnosis tasks deserves further study. Future experiments should broaden the number of diagnostic variables to get a better picture of how variable count influences proximity compatibility. Also, the impact of differential utility on proximity compatibility needs to be better understood. Researchers should check whether different patterns of differential utility have different effects on task proximity. Furthermore, the effect of grouping together high-utility variables in an object display needs investigation. Finally, researchers should start exploring broader ranges of tasks in each experiment. In reality, one often uses a single information display for many related tasks, some of which may be more compatible with the display than others. PSSC accommodates this fact by including numeric data in the radial polygon and by allowing physicians to switch to a tabular display. But this experiment—along with most other studies of PCP—used a single display for a single task. PCP's usefulness in display design is by now well established, so now human factors researchers should concentrate on applying PCP under more complex and conflicting task requirements.

REFERENCES

- Andre, A. D., & Wickens, C. D. (1988). The interaction of spatial and color proximity in aircraft stability information displays, *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 1371-1375). Santa Monica, CA.
- Antin, J. F. (1988). An empirical comparison of menu selection, command entry and combined modes of computer control. *Behaviour & Information Technology*, 7(2), 173-182.
- Bailey, R. W. (1982). *Human performance engineering: A guide for system designers*. Englewood Cliffs, N.J.: Prentice-Hall.
- Barnett, B. J., & Wickens, C. D. (1988). Display proximity in multicue information integration: The benefits of boxes. *Human Factors*, 30(1), 15-24.
- Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors 249: Special Issue: Visual displays*, 34(5), 513-533.
- Bennett, K. B., Toms, M. L., & Woods, D. D. (1993). Emergent features and graphical elements: Designing more effective configural displays. *Human Factors*, 35(1), 71-97.
- Beringer, D. B., Howard, F., & Jenkins, J. (1986). Putting information in the visual periphery: It beats a pointed stick in the eye, *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 613-617). Santa Monica, CA.
- Boulette, M. D., Coury, B. G., & Bezar, N. A. (1987). Classification of multidimensional data under time constraints: Evaluating digital and configural display formats, *Proceedings of the Human Factors Society's 31st Annual Meeting* (pp. 116-120). Santa Monica, CA.

- Buttigieg, M. A., & Sanderson, P. M. (1991). Emergent features in visual-display design for 2 types of failure-detection tasks. *Human Factors*, 33(6), 631-651.
- Buttigieg, M. A., Sanderson, P. M., & Flach, J. (1988). Object vs. separate displays for process failure detection: The emergent features approach, *Proceeding of the Human Factors Society's 32nd Annual Meeting* (pp. 1340-1344). Santa Monica, CA.
- Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors*, 34(5), 535-554.
- Carswell, C. M., & Wickens, C. D. (1987). Information integration and the object display: An interaction of task demands and display superiority. *Ergonomics*, 30(3), 511-527.
- Chernoff, H., & Rizvi, M. H. (1975). Effect on classification error of random permutations of features in representing multivariate data by faces. *Journal of the American Statistical Association*, 70(351), 548-554.
- Coll, R., & Wingertsman, J. C. (1990). The effect of screen complexity on user preference and performance. *International Journal of Human-Computer Interaction*, 2(3), 255-265.
- Coury, B. G., & Boulette, M. D. (1992). Time stress and the processing of visual displays. *Human Factors*, 34(6), 707-725.
- Coury, B. G., Boulette, M. D., & Smith, R. A. (1989). Effect of uncertainty and diagnosticity on classification of multidimensional data with integral and separable displays of system status. *Human Factors*, 31(5), 551-569.
- Coury, B. G., Boulette, M. D., Zubritzki, M. C., & Fisher, D. L. (1986, October). *Classification of multidimensional data: Evaluating displays to aid identification of system data*. Paper presented at the Proceedings of the IEEE Conference of Systems, Man, and Cybernetics, Atlanta, GA.

- Coury, B. G., & Purcell, J. (1988). The bar graph as a configural and separable display, *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 1361-1365). Santa Monica, CA.
- DSP-FX. (1993). (Version 5.1). Brisbane, CA: Power Technology.
- Elvers, G. C., & Dolan, N. J. (1995). A comparison of the augmented bar display and the object display. *Ergonomics*, 38(4), 777-792.
- Excel. (1999). (Version 2000). Redmond, WA: Microsoft.
- Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, 25(4), 350-358.
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1(3), 225-241.
- Gerhardt-Powals, J. (1996). Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human-Computer Interaction*, 8(2), 189-211.
- Gillie, T., & Berry, D. (1994). Object displays and control of dynamic systems. *Ergonomics 249: Special Issue: Cognitive ergonomics*, 37(11), 1885-1903.
- Goettl, B. P., Wickens, C. D., & Kramer, A. F. (1991). Integrated displays and the perception of graphical data. *Ergonomics*, 34(8), 1047-1063.
- Goldsmith, T. E., & Schvaneveldt, R. W. (1984). Facilitating multiple-cue judgments with integral information displays. In J. C. Thomas & M. L. Schneider (Eds.), *Human Factors in Computer Systems* (pp. 243-270). Norwood, NJ: Ablex.
- Greaney, J., & MacRae, A. N. (1996). Diagnosis of fault location using polygon displays. *Ergonomics 249: Special Issue: Cognitive Ergonomics V*, 39(3), 400-411.
- Greaney, J., & MacRae, A. W. (1997). Visual search and the detection of abnormal readings in graphical displays. *Acta Psychologica*, 95(2), 165-179.

- Green, C. A., Logie, R. H., Gilhooly, K. J., Ross, D. G., & Ronald, A. (1996). Aberdeen polygons: Computer displays of physiological profiles for intensive care. *Ergonomics*, 39(3), 412-428.
- Hesselvik, F., Carlsson, C., Brodin, B., Jorfeldt, L., & Schildt, B. (1985). A hemodynamic profile diagram and its application in septic patients. *Acta Anaesthesiologica Scandinavica*, 29(7), 706-711.
- Huff, D. L., Mahajan, V., & Black, W. C. (1981). Facial representation of multivariate data. *Journal of Marketing*, 45(4), 53-59.
- Hughes, T., & MacRae, A. W. (1994). Holistic peripheral processing of a polygon display. *Human Factors*, 36(4), 645-651.
- Intelihealth. (2000, Dec. 17th, 1996). *Sepsis*, [Web site]. Available: <http://www.intelihealth.com> [2000, May 30th].
- Jones, P. M., Wickens, C. D., & Deutsch, S. J. (1990). The display of multivariate information: An experimental study of an information integration task. *Human Performance*, 3(1), 1-17.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175-219.
- Kahneman, D., & Treisman, A. M. (1984). Changing views of attention and automaticity. In R. Parasuraman & R. Davies & J. Beatty (Eds.), *Varieties of Attention*. New York: Academic.
- Kang, T. J., & Muter, P. (1989). Reading dynamically displayed text. *Behaviour & Information Technology*, 8(1), 33-42.
- Karat, J., McDonald, J. E., & Anderson, M. (1986). A comparison of menu selection techniques: Touch panel, mouse and keyboard. *International Journal of Man-Machine Studies*, 25(1), 73-88.

- Knutson, J. F. (1998). The effect of the user interface design on adoption of new technology. *Dissertation Abstracts International: Section B: The Sciences & Engineering*, 59(3-B), 1399.
- MacGregor, D., & Slovic, P. (1986). Graphic representation of judgmental information. *Human-Computer Interaction*, 2, 179-200.
- Motoki, R., & Honda, K. (1977). Hemodynamics in sepsis: Its relationship to pulmonary complication. *Tohoku Journal of Experimental Medicine*, 123(2), 109-124.
- Munson, R. C., & Horst, R. L. (1986). Evidence for global processing of complex visual displays, *Proceedings of the 30th Annual Meeting of the Human Factors Society* (pp. 776-780). Santa Monica, CA.
- O'Hare, D., & Roscoe, S. N. (1990). *Flightdeck performance: The human factor* (1st ed.). Ames: Iowa State University Press.
- Pomerantz, J. R., & Garner, W. R. (1973). The role of configuration and target discriminability in a visual search task. *Memory & Cognition*, 1(1), 64-68.
- Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception & Performance*, 15(4), 635-649.
- Pomerantz, J. R., & Schwaitzberg, S. D. (1975). Grouping by proximity: Selective attention measures. *Perception & Psychophysics*, 18(5), 355-361.
- Rahman, T., & Muter, P. (1999). Designing an interface to optimize reading with small display windows. *Human Factors*, 41(1), 106-117.
- Rixen, D., Siegel, J. H., Espina, N., & Bertolini, M. (1997). Plasma nitric oxide in posttrauma critical illness: A function of "sepsis" and the physiologic state severity classification quantifying the probability of death. *Shock*, 7(1), 17-28.
- Rixen, D., Siegel, J. H., & Friedman, H. P. (1996). "Sepsis/SIRS" Physiologic Classification, severity stratification, relation to cytokine elaboration and outcome

- prediction in posttrauma critical illness. *Journal of Trauma-Injury Infection and Critical Care*, 41(4), 581-598.
- Sanderson, P. M., Flach, J. M., Buttigieg, M. A., & Casey, E. J. (1989). Object displays do not always support better integrated task- performance. *Human Factors*, 31(2), 183-198.
- Siegel, J. H., Cerra, F. B., Coleman, B., Giovannini, I., Shetye, M., Border, J. R., & McMenamy, R. H. (1979). Physiological and metabolic correlations in human sepsis - invited commentary. *Surgery*, 86(2), 163-193.
- Siegel, J. H., Cerra, F. B., Moody, E. A., Shetye, M., Coleman, B., Garr, L., Shubert, M., & Keane, J. S. (1980). The effect on survival of critically ill and injured patients of an ICU teaching service organized about a computer-based physiologic care system. *Journal of Trauma-Injury Infection and Critical Care*, 20(7), 558-579.
- Siegel, J. H., & Coleman, B. (1986). Computers in the Care of the Critically Ill Patient. *Urologic Clinics of North America*, 13(1), 101-117.
- Vincow, M. A., & Wickens, C. D. (1993). Spatial layout of displayed information: Three steps toward developing quantitative research models, *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 348-352). Santa Monica, CA.
- Wainer, H., & Thissen, D. (1981). Graphical data-analysis. *Annual Review of Psychology*, 32, 191-241.
- Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and objectness on information integration. *Human Factors*, 32(1), 61-77.
- Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37(3), 473-494.

- Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical Principles in Experimental Design* (3rd ed.). New York: McGraw-Hill.
- Woods, D., Wise, J., & Hanes, L. (1981). An evaluation of nuclear power plant safety parameter display systems. In R. G. Sugarman (Ed.), *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 110-114). Santa Monica, CA.