EFFECTS OF TASK STRUCTURE ON GROUP PROBLEM SOLVING

By

Gbemisola Adebimpe Abimbola

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of

Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2006

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

This thesis investigates the effect of problem structure on performance and behavioural variety in group problem solving. In addition, it examines the effects of problem solving strategy in group problem solving.

Previous researchers have focused their efforts on individual problem solving with minimal reference to groups. This is due to difficulties such as the presence of distributed information, the coordination of people and the large scale of work that typified group problems. Specifically, the effect of problem structure in group problems has been rarely studied due to the absence of an encompassing theory.

In this thesis, the effect of problem structure on group performance is studied using the fundamental characteristics of structure such as detour, redundancy, abstraction and degree of homogeneity. These characteristics were used in conjunction with existing problem solving theories (such as Information processing system, Gestalt approach and Lewin's lifespace approach) and Heider's balance theory to understand the effects of task structure on group performance and behavioural output.

Balance theory is introduced as a conceptual framework in which the problem solving process is viewed as a dynamic progression from cognitive imbalance towards a state of structural balance corresponding with the solution. This theoretical approach captures both incremental search processes and insight associated with cognitive restructuring, typical of existing problem solving approaches in the literature. It also allowed the development of unique measures for studying the effect of structure in group problem solving.

A Laboratory experiment was conducted using 153 undergraduate and 3 graduate students in groups of 4 subjects. The experiment examined the effect of task structure on groups' performance and behavioural variety. The stimulus used for the experiment was a categorization problem consisting of sixteen cards with two objects each shared equally among four participants. The objective was to form four groups of items with no cards left unused. The groups' performance data was collected and analyzed to verify the postulated hypotheses.

The results indicate that both increased problem structure complexity and the introduction of a restructuring dimension in the problem structure were associated with reduced performance and increased behavioural variety. With respect to problem solving strategy, early discussion in problem solving was associated with better performance and less behavioural variety. Finally, the results support the premise that group problem solving processes tend to be in the direction of attaining higher states of balance.

ACKNOWLEDGEMENTS

I came back to school to learn how to think and I am glad that on completing this thesis, I can say that I can think through problems much better than when I got into the program. The gratitude goes to my supervisor, Dr. Rob Duimering and his supervisor, Dr. Frank Safayeni. Thank you for your time, for the great questions and the timeless research. Although painful and frustrating at times, the end result is worth the stress.

Also, I will like to thank my reader and lecturer, Prof. Jewkes for the taking the time to read through and for insisting that I revise my presentation. I feel much better with the end result and I am proud of the work.

I will also like to thank my brother – Yemi for his support and willingness to do whatever it took to get me back to school; I appreciate the sacrifice.

To the love of my life – Adefemi Adejumo, I can only imagine how frustrating it is to have a fiancée that has always been a student and has put the education above all – I respect your courage and dedication. Thank you very much

To my one and only Lolade – you are the best and the greatest – I should be receiving two masters degree for this program – one for you and the other for me. Thank You from the bottom of my heart.

To my wonderful friends and family – Sitwat, Deji, Remi, Dami, Marilyn, BTO, Dupe, Lisa – you guys were my strength when I was weak – which was quite often. You believed in me and supported me all the way – I am indeed very grateful.

I will also like to thank Bell University Laboratory and Ontario Government for the financial support and exposure to some organizational problems.

And ofcourse to my parents – you have always made me feel like the best thing that exist on earth – thank you for the vote of confidence and for the endless labour and support.

DEDICATION

This thesis is dedicated to Almighty God without whom I would probably still be writing.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

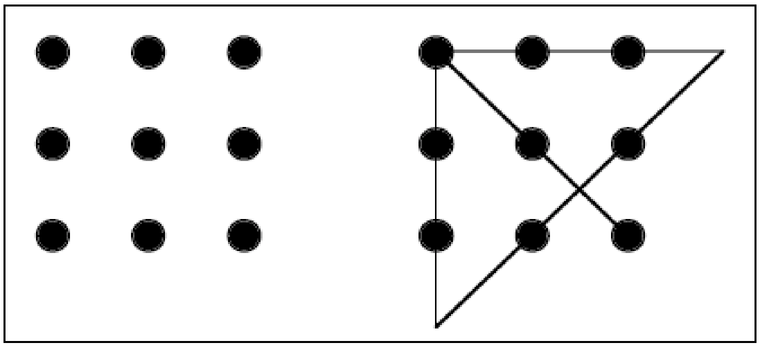
1 INTRODUCTION

Every day, in companies around the world, groups of individuals are working together to solve problems. These groups are faced with a myriad of challenges, examples of which are personality differences, information exchange and resource availability. The literature predominantly looks at the dynamics of the individuals in the group, such as personality factors (Chiu, 2000; Swinth and Tuggle, 1971) or communication effects (Bavelas, 1950; Leavitt, 1951), to explain the effectiveness of the group's performance. While such dimensions have proven to be relevant, they do not fully capture the reason behind a group's effectiveness. Problem structure, while shown to be critical in problem solving, has not received equal attention in research (Hoffman, 1965; McGrath, 1984). This research, therefore, focuses on the importance of understanding the structure of the problem, and shows that problem structure has a direct impact on a group's performance and effectiveness.

What is problem solving? Dunbar (1988) defined problem solving as "the movement through a problem space". In this definition, a problem space consists of the initial state, final state or goal state and the operators involved in the move across these two extreme states. An example of this could be two departments within an organization deciding whether or not to develop a new product. The initial state would be what to develop among the possible choices. The goal would be to develop an appropriate product within the available resources. The operators are the processes involved in making the decision of achieving the goal within the tight constraints of satisfying all parties.

Wertheimer (1982) defines problem solving as a process through which an individual understands the structure and constraints of a situation. An example of this is the nine-dot problem as illustrated in Figure 1. The nine-dot problem requires nine dots, which are arranged in a square, to be connected by four straight lines without lifting the pen from the paper or retracing the lines (Maier, 1930). In order for an individual to solve this problem, the individual must recognize that the boundaries of the line to be drawn are not restricted by the location of the dots. In coming to this realization, the individual can more easily solve the problem as he now has an understanding of the problem structure.

Figure 1: Sample Nine dot problem before and after solving adapted from Training from Insight: The Case of the nine dot problem (Kershaw and Ohlsson, 2001)



Two primary areas of research in problem solving are group and individual problem solving (Bottger and Yetton, 1988). Group problems differ from individual problems in terms of the scale of the problem and the skills required to resolve it. Solving an arithmetic problem could be an example of an individual problem. On the contrary, a problem such as developing a product on an assembly line is classified as a group problem due to the magnitude of work involved and because the knowledge required to solve the problem is distributed among several group members rather than found in one individual. Therefore group problems differ from individual problems in terms of complexity, distributed information or knowledge, and the number of people involved (Chiu, 2000). As a result of these complexities, group problem solving is less commonly researched compared to individual problem solving (Cohen and Bailey, 1997). However, due to the increasing trend in organizations to utilize the wealth and inter-dependence of their human resources, there is an increasing need to research the effectiveness of groups in problem solving (Bonner, 2004).

In addition to looking at task structure in group problem solving, this research also attempts to understand the effect that a group's choice of strategy, when approaching the problem, has on the effectiveness of the group's performance. Hackman and Oldham (1980 p.179) described strategy as "the choices group members make about how they will go about performing a given task". When a group of individuals approach a problem that they must solve together, they can make a choice as to whether they should attempt to understand the problem collectively, or each try to understand the problem individually. This research proposes that the choice of strategy will impact the effectiveness of the group in problem solving.

In summary, this thesis shows the effects of fundamental characteristics of problem structure and the choice of strategy on group performance and effectiveness. It is presented in the order listed below:

Chapter 2 reviews prior research work in problem solving. It adopts terminologies and concepts from Gestalt studies, Information processing theory, Lewin's lifespace approach and Heider's balance theory.

Chapter 3 explains the characteristics of task structure and theoretical framework on which this work is based.

Chapter 4 discusses the methodology used by providing a detailed description of the experimental sets.

Chapter 5 presents and analyses the results.

Chapter 6 concludes this research work and identifies areas for future research as well as limitations and assumptions of the study.

2 LITERATURE REVIEW

The main focus of this study is the effect of task structure in group problem solving. The literature discussion starts with general problem solving, and proceeds to discuss the different approaches to individual problem solving. It analyzes and describes the effect of task structure within these problem solving approaches and how these individual approaches can be applied to group problem solving. The literature further describes Heider's balance theory as a problem solving approach while noting the similarities to existing theories. This section concludes by introducing task structure and performance as variables in problem solving.

2.1 General Problem Solving

Problem solving is one of the most important activities performed by individuals in a group. A problem exists when there is a perceived difference between the initial state and the end final state. It can be further explained as a process through which one attempts to understand the structure of a situation in order to predict the necessary course of action to solve the problem (Swinth and Tuggle, 1971). Problem solving has been studied extensively from the individual perspective by focusing on individual skills, social process, personality types and their impact on ability to solve problems (Larson and Christensen, 1993; Laughlin et al., 2003). Some theories have been developed in order to understand the fundamental processes involved in problem solving. These approaches are listed below and are discussed in more detail in the next section:

1. The Information processing system,
2. Gestalt approach
3. Lewin's approach

2.2 Problem solving approaches

This section discusses the different problem solving approaches and their relevance to this thesis.

2.2.1 Information Processing System (IPS)

The Information Processing System (IPS) is a comprehensive problem solving theory developed by Newell and Simon (Newell and Simon, 1972b). It essentially views problem solving as a process of developing a problem space and conducting search within the constructed problem space. This search is influenced by the task environment (i.e. the physical environment, the goal and the problem as perceived by the subject), which determines the direction or scope of search. Although, this theory acknowledges the role of task structure, the main focus is on the method of problem resolution rather than the structure of the task. This theory further suggests that objective problem space representation is almost impossible as representation will always be subjective from the subjects' point of view. The only acceptable representation will be that of the stimulus or use of an exhaustive problem space including all possible moves.

The limitation of the IPS approach is that humans may not have the memory capacity to deal with all possible representations as expected using the search process. In addition, it is difficult if not impossible to represent the problem search spaces for all individuals involved in group problem solving. Finally, Information processing system (IPS) is still unable to properly explain what guides the search, how the search is constrained and what happens when the solution is unavailable within the available problem representation as often observed in insight or restructuring[1] problems (Simon and Kaplan, 1990).

2.2.2 Gestalt approach

The Gestaltists propose that *the whole is different from the sum of its parts* and thus approach problems from the whole form rather than its components parts, contrary to other scientists and psychologists. They believe that in viewing an object, the details are not relevant but rather the focus is on the larger wholes, separated by form and related to one another in a given experience (Wertheimer, 1982). The relationships observed in larger wholes are usually based on similarity, proximity, closure and continuity as

[1] Restructuring is a problem characteristic that involves a change in problem spaces representation often typical of insight problems

defined in the laws of perceptual organizational. Examples of these relationships are shown in the figure below.

Figure 2: Gestalts Law on Similarity

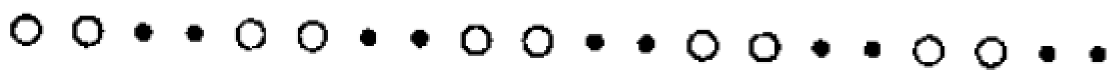
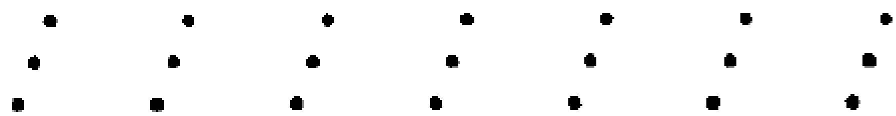


Figure 3: Gestalts Law on Proximity



According to Gestaltists, individuals are likely to group the circles in Figure 2 together rather than the circle and the dot because the mind groups similar objects together. In the proximity example (Figure 3), individuals will likely group the three slanted dots together as opposed to the row of dots due to the proximity effect on items. As a result of this perceptual conceptualization, Gestaltists put lots of emphasis on proper problem representation, as it is believed to affect problem resolution. This infers that once the problem is properly represented and the structure well understood solving becomes a mechanical process.

The idea of understanding the whole rests on the concept that interactions occur among interconnected or inter-related components. The approach of analyzing individual components or specific parts of a concept may thus lead to inappropriate or inaccurate results. Problem solving is approached mainly from understanding the underlying concept and viewing all characteristics as inter-dependent. The Gestaltist research is often based on perceptual views of observing phenomenal data and generalizing their findings to affect the way people think, feel, behave, perceive, remember and even solve problems

(Kohler, 1969). This perceptual view represents their understanding of the task structure and its effects on problem solving.

Kohler (1969) explained in the task of Gestalt psychology that solving a problem starts with establishing relations between the problem and the solution. The ability to represent the problem in terms of the goal or solution is the first step in structuring the task. Although, Kohler (1969) acknowledged that solutions to problems are not always brought about by insight, proper structuring of the problem is seen as essential to solving or establishing appropriate relations and is dependent on their perceptual views.

The first limitation of this approach is the inability to explain how restructuring or insight is achieved in problem solving. Kohler (1969 p.49) addressed this limitation by suggesting that “thorough considerations of various parts of the given situation and the intensity of our desires often lead to insight and restructuring”. Another limitation of this approach is that insight is not a characteristic of all problem types and as such the Gestaltist’s approach cannot be applied to all problem types. In comparing this approach to the IPS approach by Newell and Simon (1972), it is important to note that the problem representation as defined in both may be different. This is because the IPS approach does not necessarily assume that complete understanding of the task structure is a requirement for problem solving as inferred in Gestaltists approach to problem solving.

2.2.3 Lewin’s lifespace approach

Lewin’s problem space approach is usually expressed in terms of the individual, their psychological environment, forces, regions and life space. Lewin’s approach to problem solving and understanding behaviour can be viewed from his life space representation. Lifespace or problem space of any individual is defined as a representation of both the person and the psychological environment that the individual belongs to. Both the person and his environmental factors are believed to be interdependent and interrelated (Lewin, 1951). Each individual’s behaviour is proposed as a function of the person and the situation in which the person currently exists i.e. $B = f(P, E)$ and $B = f(LSp)$. In this formulation, Behaviour is represented as B, Person as P and Environmental factors as E.

In comparing, the second equation to the first, the function of a person and his environment is represented as the psychological lifespace (LSp). Each individual is believed to be in a psychological life space at every point in time. This psychological life space has various regions depending on what is currently happening. For instance, a student currently on a paid job and who has an exam the next day may have more and distinct regions of activities within his life space than a full time student who is currently studying for the same exam. Also, the nature of children's problem space differs from that of adults. Children have less distinct and probably fewer regions of activity than adults due to the fact that their psychological environment is still developing and a lot of their activities are yet to be differentiated.

In resolving a problem, an individual needs to establish a path between the initial state and the goal state. This allows the psychological movement from current state to the desired state. This may require restructuring the life space to overcome an existing barrier or choosing another path against which the barrier is weaker (Lewin, 1936). Lewin's approach defines force as either positive or negative valence or tension and portrays its effect on the desired goal. While this is a good representation of the dynamic nature of human problem solving and locomotion within the psychological space in general, it poses a difficult problem of representing and measuring forces and its effects. For instance, we do know and agree that there is some force pulling one in the direction of the goal when solving the problem but how do we measure or quantify the force in question? In using the analogy presented here within the group context, the group may be viewed as operating within the same lifespace and attempting to work through the problem space to reach the goal state while overcoming barriers and constraints collectively. While this is a dynamic approach to problem solving, the question of how the lifespace is defined and constructed for a group still remains unresolved.

2.3 Heider's balance theory

Heider's balance theory was proposed about half a century ago to explain an individual's cognitive state. It has experienced a few modifications since then (Cartwright and Harary, 1956; Festinger, 1950; Newcomb, 1968; Osgood and Tannenbaum, 1955; Rosenberg and Abelson, 1960) but it has rarely been used in recent times.

In 1946, Heider's balance theory suggested that "a triad consisting of either three people or two people and an attitudinal object, and the relations between these three entities, might be balanced or unbalanced". A balanced relationship occurs when Person P likes Person O and both like Object X. It is represented as PLO, PLX and OLX where L can imply like, love or own relationship. The relationship is in a state of balance if all the three relationships are positive or one is positive and the other two are negative as in PLO, O~LX and P~LX. Heider further postulated that mutual dependency among elements of thought, forming states of "order and coherence" is a characteristic of structural dynamics (Heider, 1946a). Heider also emphasized that elements that "go-together" form stable structures while dissimilar elements generate forces and reorganize in an attempt to reach a state of stability (Heider, 1960).

In recent times, additional work has been done on generalizing some of the concepts mentioned above and in creating new ones. Cartwright and Harary (1956) generalized Heider's theory to include support for symmetric relations, units containing more than three entities and negative relations in a bid to allow empirical formulations using the theory. They suggested future research in the "systematic treatment of relations of varying strength".

Although, Heider's theory appears to be a basic model, it appeals to one's intuition and can be applied to various scenarios such as triad relationship analyzes and problem-solving processes as illustrated in this thesis. There are various means of cognitive dissonance reduction or elimination that can be adopted when solving a problem. Initially, when one is faced with a problem, the individual is in a state of high dissonance. In the process of solving the problem, the person attempts to move, change or modify the current state due to the discomfort associated with imbalance. Hence, Heider's approach is a dissonance reduction process for problem solving.

The advantage of this approach is that it provides a dynamic and psychological view of problem solving. It also shows the psychological movement process from the beginning to the end of the problem solving process. Finally, the process is easy to represent and quantify and thus serves as a good approach to viewing problem solving in both individual and group settings.

2.4 Group Problem Solving

Research in group problem solving has continued due to the perception that group problem solving and decision-making results in increased productivity and performance in various organizational settings (Bavelas, 1950; Hendrick, 1979; Laughlin et al., 2003). In demonstrating this perception, various researchers have compared individuals to groups in problem solving. They have demonstrated that group performances are better than that of individuals in intellectual and judgmental problems (Laughlin et al., 2003). Laughlin, Zander, Knievel, & Tan (2003) showed that the collaborative nature of groups resulted in fewer equations, higher number of known variables and faster results in mathematical letters-to-numbers problems.

Other researchers have looked into the effect of communication structures on group problem solving (Leavitt, 1951; Shaw, 1954). The wheel pattern refers to a group structure or pattern in which there is always a central person through which other members must communicate. The wheel pattern of communication has been seen as resulting in faster and more efficient times in problem solving (Harshbarger, 1971). The circle, as the name implies provides a means through which all neighbouring subjects can communicate directly with each other. The circle is usually perceived as a better approach due to its ability to allow equal participation. This is synonymous to democracy in organizational terms.

In addition to communication structures, the task structure is also known to have an effect on group problem solving behaviour. The following section provides a more detailed analysis of task structure and the effect on problem solving.

2.5 Task Structure

Tasks have been identified as a variable that affects organizational performance (Faucheux and Mackenzie, 1966; Newell and Simon, 1972a), organizational structure and organizational behaviour in problem solving. Tasks contribute to the determination of various organizational effects like effectiveness, productivity and degree of cohesion. Tasks have been defined in various forms due to the difficulty associated with isolating the objective concept of interest. Often, tasks are defined in terms of the behaviour expected from the participant or behaviour emitted or personal abilities required in performing the task (Hackman, 1987). Each of these forms of definitions has their advantages and disadvantages depending on the purpose of the research. However, this complexity makes it difficult to develop a general theory of task structure and compare the results across multiple studies.

Some literatures have identified tasks as an intervening variable between organizational structure and performance (Shaw, 1954). This implies that the researchers showed that particular kinds of tasks were more efficient in certain organizational structures than in others. Although there have been many speculations on the effects of task structure and organizational structure or performance, very little conclusive results have been provided. This is partly due to the difficulty associated with the definition of task and the nature of the task type (Faucheux and Mackenzie, 1966).

Aside from understanding tasks, defining and measuring the dimensions of tasks along a continuum has been challenging due to the inability to develop a generally acceptable definition of task (Roby and Lanzetta, 1958; Simon and Newell, 1972; Zajonc, 1965). However, the definition of task in terms of the objective stimulus and use of fundamental task dimensions were adopted in this study. The use of the objective task definition eliminates any bias due to the observed behaviour.

The different forms of task structures adopted in this study are detour, redundancy, restructuring and heterogeneity. They motivation for these task characteristics and their abstraction will be described in detail in subsequent chapters.

2.6 Performance

Performance and Effectiveness or Behavioural outputs are used interchangeably in this thesis and are known to fulfill functions required by a task (Cartwright and Zander, 1960). This implies that they are dependent on the task structure or characteristics. Almost all studies involving organizations are interested in understanding performance or effectiveness and how they can be controlled because they are a means of ranking organizations and their processes.

In terms of productivity, several aspects of productivity increase such as quality and performance improvement have also been researched. This has resulted in the search for the factors that are necessary for such improved results. Some of these have been linked to the size of the group (Holloman and Hendrick, 1971), social factors, personality as well as people's perceptions due to participation or involvement in organizational decision making. In this thesis, the use of performance as an output or dependent variable serves as a measure for group effectiveness and problem solving efficiency.

2.7 Summary of the Literature

In summary, the literature has confirmed the relevance of problem structure as an important aspect of problem solving. Extensive work has been done on individual problem solving leading to some widely accepted theories and approaches to problem solving. While these approaches are laudable, they do not cover all aspects of problem solving especially with regards to group problem solving. Some of the identified theories are consistent and similar to one another while others appear to address different aspects of problem solving (e.g. Gestalt and IPS). This thesis does not intend to create a new theory of problem solving. Rather it adapts the identified theories and Heider's balance theory to different aspects of group problem solving and draw inference on the effect of task structure.

3 TASK STRUCTURE

The description of the task structure and the theoretical underpinnings of this study are closely interwoven and thus the description of one cannot be done without the other. The following sections describe the task stimulus in relation to the underlying theory. Heider's balance theory (Heider, 1946b) is also introduced as an approach to problem solving and related to existing problem solving methodologies and theories. Problem solving performance and behaviour are affected by task structure and strategy. Prior to discussing and analyzing performance and effectiveness, the representation and quantification of the structural characteristics will be discussed.

3.1 Stimulus Description

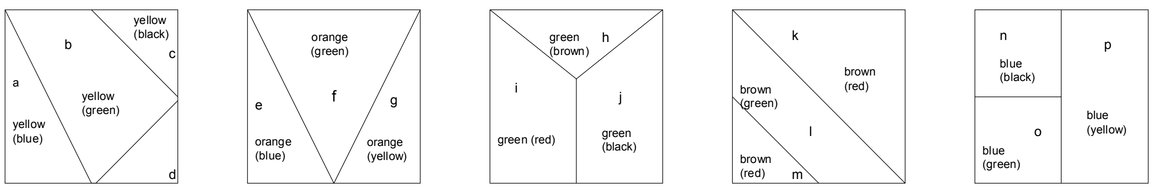
3.1.1 Background: Five squares problem

The experimental task stimulus was adapted from the "five squares" problem originally developed by Alex Bavelas (1973). The five squares problem was originally developed for the study of group cooperation under different conditions of communication network structures and rewards (Bavelas, 1950; Leavitt, 1951). The stimulus consisted of sixteen geometric shapes used to form five squares (see Figure 4 below). Researchers have adopted variations of this problem to study cooperation and creative problem solving in groups and the effects of team spirit and trust (Steinback et al, 2000). Executives, managers and administrators have also used it as a team building exercise.

The sixteen pieces are designed such that there is only one way of forming five squares of the same colour with no pieces left over. There are many possible variations of squares that lead to incomplete or inappropriate solutions thus acting as "detours" and preventing a group from finding the correct solution. For instance, some distributions of the shapes among five participants usually lead to three easy squares by two or three group members while others are unable to complete their squares. The only way to solve the problem would be for those with complete squares to break their squares to allow other members to complete their squares. This is usually difficult due to constraints such as time limit, incentive and the group's lack of awareness of a possible solution (Bavelas, 1973).

The geometric pieces are usually distributed among five subjects but some experimenters have used between three and six subjects for this task. Examples of pieces of geometric shapes distributed to each participant could be jbo, fmi, dcke, pa, ghnl etc. In some variations of the five squares problem the geometric card pieces have different colours on each side to further complicate the task.

Figure 4: The Five Squares Experimental puzzle



One benefit of using the five squares problem for studying group problem solving is the ability to illustrate the phenomenon of distributed information, whereby each member or participant has partial or incomplete information about the task thus necessitating group collaboration. Most problem solving studies have used either logic based problems (Laughlin et al., 2003) or arithmetic problems both of which are unable to incorporate collaboration in group problem solving. In addition, this task requires descriptions and negotiation among participants, hence ensuring that participants communicate with one another to achieve the final result. Communication helps participants understand each other's thought processes and understanding of the problem structure.

3.1.2 Limitations of the Five Squares Problem

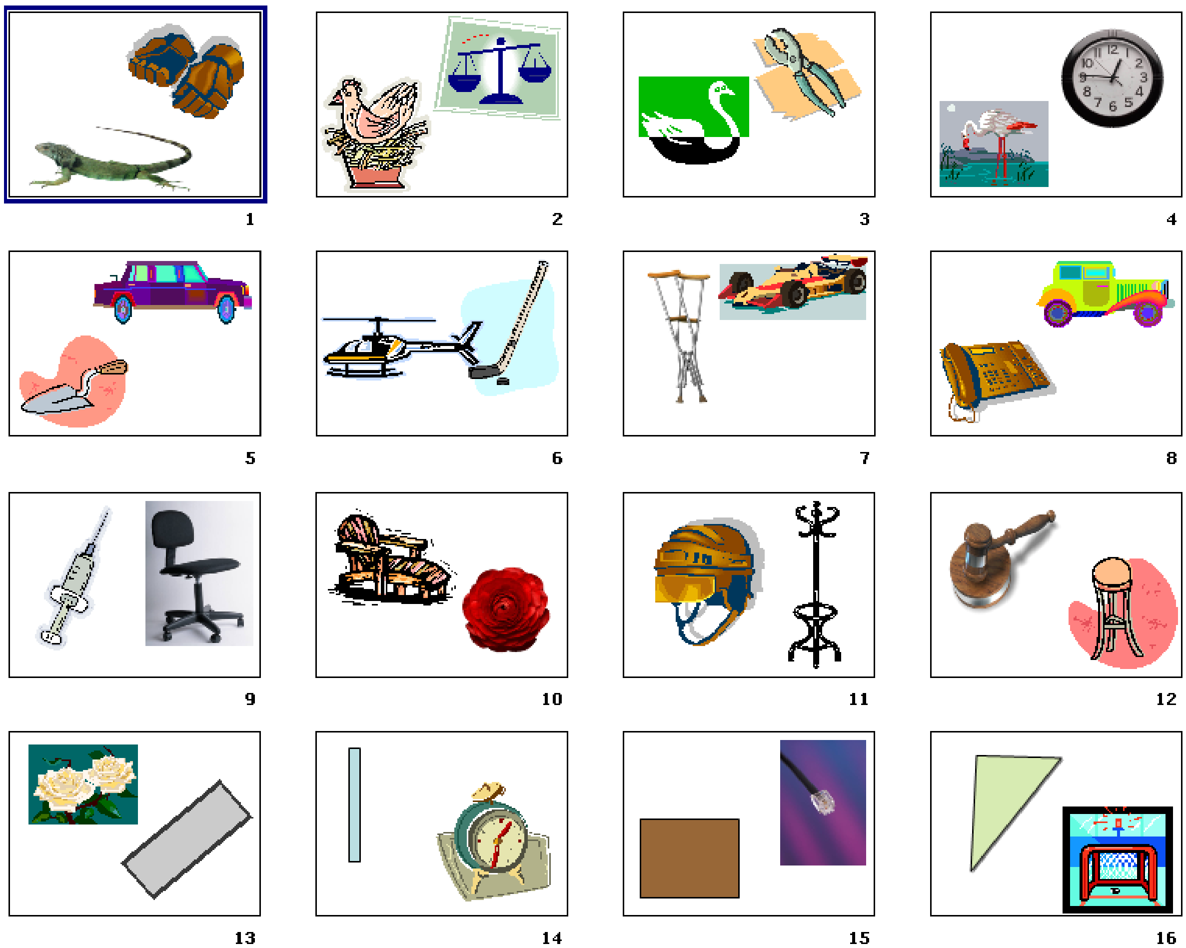
The five squares problem has several limitations from the perspective of the goals of this thesis research. One limitation is the difficulty associated with identifying the geometric pieces since some of the shapes are not regular shapes such as squares or rectangles. These irregular shapes are difficult to describe to other group members, a drawback that hinders the facilitation of card exchanges required to reach the desired goal. Another limitation is the inability to introduce new dimensions into the existing problem structure. The task itself is rather rigid in design and only external factors can be varied easily, such as the time allowed to solve the problem, or the reward structure for individuals or the group. A final limitation of the five squares problem is the quantification of structural properties. Since the shapes are not easily defined, quantifying, describing or comparing them to one another is difficult. Because of these limitations, a new experimental stimulus inspired by the original five squares problem was designed and used for this thesis.

3.2 Experimental task stimulus

The designed experimental stimulus is a categorization problem similar in structure to the five squares problem. It has sixteen cards but rather than using shapes and colours, pictorial objects that can be easily described were used as the basis for grouping cards into categories. An example of the stimulus is illustrated in Figure 5. Each card has two objects on it and subjects were required to use all sixteen cards to form four sets or categories of objects making sure that no unused cards were left over. The four categories of correct cards in Figure 5 are the horizontal sets of {1-4, 5-8, 9-12 and 13-16} representing sets of animals, vehicles, furniture and shapes respectively. Only one object on each card is used in category formation. For instance, the category of animals in cards 1 – 4 consists of the iguana, chicken, swan and ostrich. This implies that the other objects are redundant and not required in category formation. Most of these objects were included as distraction sets of paired items, which further complicate the task since they prevent participants from easily identifying the correct sets of objects. Examples of distraction pairs from Figure 5 include the medical category (hypothermic needle and crutches) in cards 7 and 9 or time pieces (wall clock and alarm timer) in cards 4 and 14. Note that the effects of these sets of random pairs were not explicitly tested in this thesis.

The cards were initially distributed such that subject 1 always received cards {1, 6, 11 and 15}, subject 2 received cards {2, 5, 9 and 13}, subject 3 received cards {3, 7, 10 and 14} and subject 4 received {4, 8, 12 and 16}. Different variations of the experimental stimuli were developed to incorporate different problem structures as explained in subsequent sections.

Figure 5: Sample Experimental Stimulus



3.2.1 Structural dimensions in the Experimental Stimuli

The different structural dimensions identified in chapter 2 were implemented in different versions of the experimental stimulus. The motivation for the structural dimensions were observations from real world problem solving in the context of new product development at a large Canadian telecommunications firm. A collaboration centre that specialized in bringing representatives of multiple divisions together to brainstorm on new initiatives had been established within the firm. Some of these initiatives include new product development, customer service improvement and process development. In a number of cases, these initiatives were not implemented according to specification and the project team members blamed one another for the failure. For instance, technical staff blamed their non-technical counterparts for incompetence or lack of knowledge while the non technical complained that their technical counterparts did not understand the customer's requirement.

However, we noticed that the initiatives had typical characteristics which were related to the nature of the problem structure. Some of these characteristics include 1) narrow individual perspectives associated with specialization, 2) communication difficulties also associated with specialization, 3) personal or departmental motives that diverged from collective organizational goals, and 4) competing objectives or presence of noise. Because it was difficult to study these typical characteristics of organizational problem solving with participants within the actual organization, we attempted to implement these characteristics in a laboratory experiment.

Narrow individual perspectives was implemented as homogeneity because specialized staff seem to always see problems in a specific way. The communication issue was implemented as restructuring as representatives of multiple departments are required to represent the information in a unified mode. Personal motives was represented as detour which usually derails the groups due to specific department's goal, and the noise factor is the redundancy characteristic where there is unnecessary information in the system. The task characteristics are explained in more detail using Figures 5 and 6, which illustrate the implementation of the experimental design.

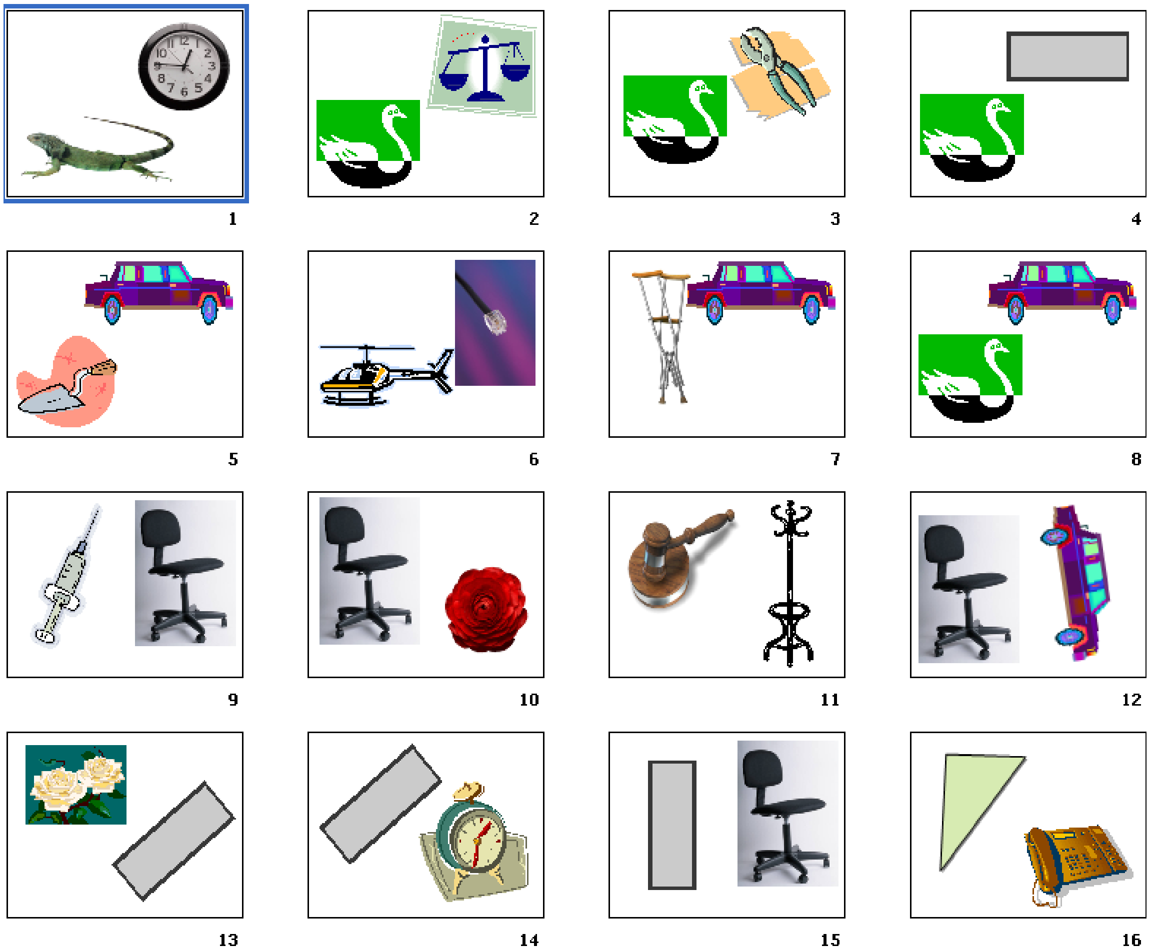
3.2.1.1 Detour

The problem solving literature discusses both detour problems and detour behaviour. A detour problem refers to a task that has an attractive and seemingly obvious solution approach, but which leads to the wrong solution. Detour behaviour refers to the ability to reach a goal by moving around an interposed object or barrier (Zucca et al., 2005). This situation is described in the hen and fence problem (Bavelas, 1973) where the food is placed directly across the fence and the hen is required to reach the goal, which is its meal. The ability to realize that walking through the fence is impossible or will not get one to the goal represents a phenomenal feature in insight problems. Both detour problems and detour behaviour have been proposed as forms of insight learning and psychologists are interested in understanding their cognitive effects on humans and other animal species (Regolin et al., 1994; Zucca et al., 2005). In this research, the presence of a detour problem is a core dimension of problem task structure that has been implemented in the experimental stimuli.

The detour property is implemented in the experimental stimulus set based on the distribution of the items on cards. For instance, the detour category in Figure 5 is the hockey items on cards {1, 6, 11 and 16}. When an individual collects this incorrect set of cards, the other three group members are unable to complete their sets of four, just as the completion of an incorrect square in the original five squares problem prevented the completion of all five squares. Based on the initial card distribution, the individual with the detour begins with three of the four cards (i.e. 1, 6 and 11) required to form the hockey category and typically searches for a fourth one to complete the set. (Recall that the cards are distributed in all cases such that subject 1 gets cards {1, 6, 11 and 15}, subject 2 gets {2, 5, 9, and 13}, subject 3 gets {3, 7, 10, and 14} and subject 4 gets {4, 8, 12 and 16}).

It is obvious to someone with the full structural knowledge that each required category is distributed across all participants resulting in each having a partial information and knowledge of the task structure. However, the participants are not aware of the full task structure so the tendency is for the individual with the majority of hockey items (subject 1) to search for the last item (on card 16) to complete the set. The three objects in the detour set are paired with objects from the correct solution sets, but the relative similarity of the detour objects makes it the category of choice. Psychologically in this experiment, detour creates a tension between individual and group goals. It has the property of apparently moving in the direction of the desired solution (i.e. four of a kind), but selecting the detour set will lead to an inability to complete the additional three desired sets.

Figure 6: Experimental task design illustrating redundancy, restructuring and homogeneity characteristics



3.2.1.2 Redundancy

Redundancy as described in Shannon's information theory (Shannon, 1948) refers to the provision of additional information which does not provide corresponding knowledge. This implies that the solution or the goal can be reached without this additional amount of information. This feature has been identified as a dimension of problem task structure and the effects are examined in the experimental study.

Redundancy is implemented in the experimental stimulus with the addition of one extra item per category as illustrated in Figure 6. The presence of these redundant items tends to make the problem easier by providing an alternate path to the correct solution. That is, the problem can be solved either by collecting cards in the four horizontal categories shown in Figure 8 {1, 2, 3, 4; 5, 6, 7, 8; 9, 10, 11, 12 and 13, 14, 15, 16} or by collecting the following sets: {1, 2, 3, 8; 5, 6, 7, 12; 9, 10, 11, 16 and 13, 14, 15, 4}.

The implementation of redundancy in our experimental design can also lead to increased confusion since it increases the search space by creating the possibility of collecting five different sets of four for each category. For instance, in Figure 6 the category of animals can be formed using any four of the cards {1, 2, 3, 4 and 8}, creating the following possible combinations: {1, 2, 3, 4; 1, 2, 3, 8; 1, 2, 4, 8; 1, 3, 4, 8 or 2, 3, 4, 8}. However, only two sets of these combinations enable the group to solve the complete problem, due to other structural constraints present in the stimulus design. Specifically, note that the set corresponding with each horizontal row in Figure 6 has an extra item that intersects with another set of objects on another row, and that the cards {4, 8, 12 and 15} each have objects belonging to two different sets. For instance, the animal category has an extra swan {card 8} that intersects with the vehicle category on the second row. The vehicle category also has an extra limousine {card 8} that intersects with the furniture category on the third row. Based on the initial distribution of cards described earlier, subject 4 with cards {4, 8, 12 & 16} receives pairs of items belonging to different categories. If subject 4 decides to collect any of these pairs to form a set, then another person in the group will be unable to form a complete set. For instance, if subject 4 collects animals using cards 4 and 8, one of the other group members will be unable to complete their set of vehicles, furniture or shapes. Redundancy therefore provides an alternative path to the solution, but also complicates the problem by increasing the size of the search space.

3.2.1.3 Restructuring

Restructuring is a problem characteristic that involves a change in problem space representation often typical of insight problems. It is a means of re-organizing the task or seeing the problem in a different way (Wertheimer, 1982). It helps to overcome detour situations and other structural features in problem solving. Restructuring is implemented in the experimental stimulus through the use of two different levels of category abstraction, and subject groups must shift from a more obvious lower level of abstraction to a less obvious higher level of abstraction in order to solve the problem. More specifically, the description or labeling of objects on cards does not always occur on the same level. Some labels apply to a narrower collection of objects in comparison to others. For instance, it is possible to categorize any card objects at what Rosch called the "basic level" of categorization, corresponding to labels like chair, table, bird, etc. (Rosch, 1975; Rosch and Mervis, 1975). However, restructuring is implemented through the introduction of objects that force subjects to group apparently dissimilar items together into more abstract, super-ordinate categories. For instance, in Figures 5 and 6 the addition of a coat rack to the chair category requires a higher level classification of these objects as furniture. The addition of an iguana to the bird category requires categorization at the level of animals. Restructuring by changing the level of abstraction affects the complexity of the categorization task and how people perceive the object collection within a set.

3.2.1.4 Homogeneity

Homogeneity refers to the degree of similarity among objects in a set as illustrated in Figures 5 and 6. In Figure 5, the four solution sets have a lower degree of homogeneity (or are heterogeneous) with different kinds of birds, cars, chairs and rectangles. In Figure 6 the solution sets contain identical (i.e. higher degree of homogenous) objects such as swans, limousines, chairs and rectangles. As discussed in chapter 2, Gestalt psychologists have demonstrated that humans tend to naturally group items based on such factors as similarity and proximity. In the present study then, subjects would be expected to perceive homogeneous items as a set more easily than heterogeneous items. Also it should be more difficult for them to break apart a set of homogenous items than a set of heterogeneous items.

In this study, homogeneity of solution sets was only varied in the task structure in relation to experimental conditions involving restructuring. Difficulty is experienced either when dissimilar items must be grouped together, or when identical items must be separated to solve the problem. The ability of participants to restructure the homogeneous swan category to animals in Figure 6 is expected to be more difficult than restructuring the heterogeneous bird category to animals in Figure 5, which already has some variation within the bird category. Homogeneity therefore implies a greater degree of restructuring than heterogeneity.

3.2.1.5 Distraction items

In addition to the structural characteristics described above, distraction items were included in the stimuli for all experimental conditions. These items come in pairs and are used as the second object on cards where no other structural property is included. In the Figures above, examples of paired items include wall clock and alarm clock, which can be grouped together as timepieces, and the letters K and N, which can be grouped as letters of the alphabet. In some cases, the pairs of random objects act as mini sets or detours of two items. This leads to a search for wrong categories during the process of solving the problem. The effect of the distraction sets in the experimental study was not studied in this thesis.

3.3 Theoretical Analysis of Complexity

To analyze the complexity of the task structure, the theoretical framework from existing problem solving approaches described earlier were adopted. This section provides an overview of how existing theories are applied to the experimental stimulus in this thesis, highlights aspects of the problem that cannot be described with existing theories and also identifies areas of overlap among existing problem solving theories. For instance, in reference to the Information processing systems approach (Simon and Newell, 1972), problem solving is based on the creation of a problem space and search within the defined space in determining the solution to a problem. In the current study, the perceived complexity of the task will be based on the size of the space that needs to be searched.

Likewise, the amount of force or degree of tension as explained using Lewin's theory is applicable in this situation, since each card has some degree of tension acting on it at each point in time, based on the objects on the card, the categories being formed and the number of possible objects within each category (Lewin, 1936). In a scenario where we have a stimulus set with only four groups and eight pairs of distraction items, there is very little tension acting on each card as the resultant force is usually in the direction of the right set due to the greater pull (i.e. higher number of similar cards) in that direction. However, in a detour situation, the tension or force is greater on the detour cards {1, 6, 11 and 16} since there seems to be equal and opposing forces acting on those cards (see Figure 5 above). The greater the force acting, the more difficult it is to eliminate an incorrect path and the more intense the search within that problem space.

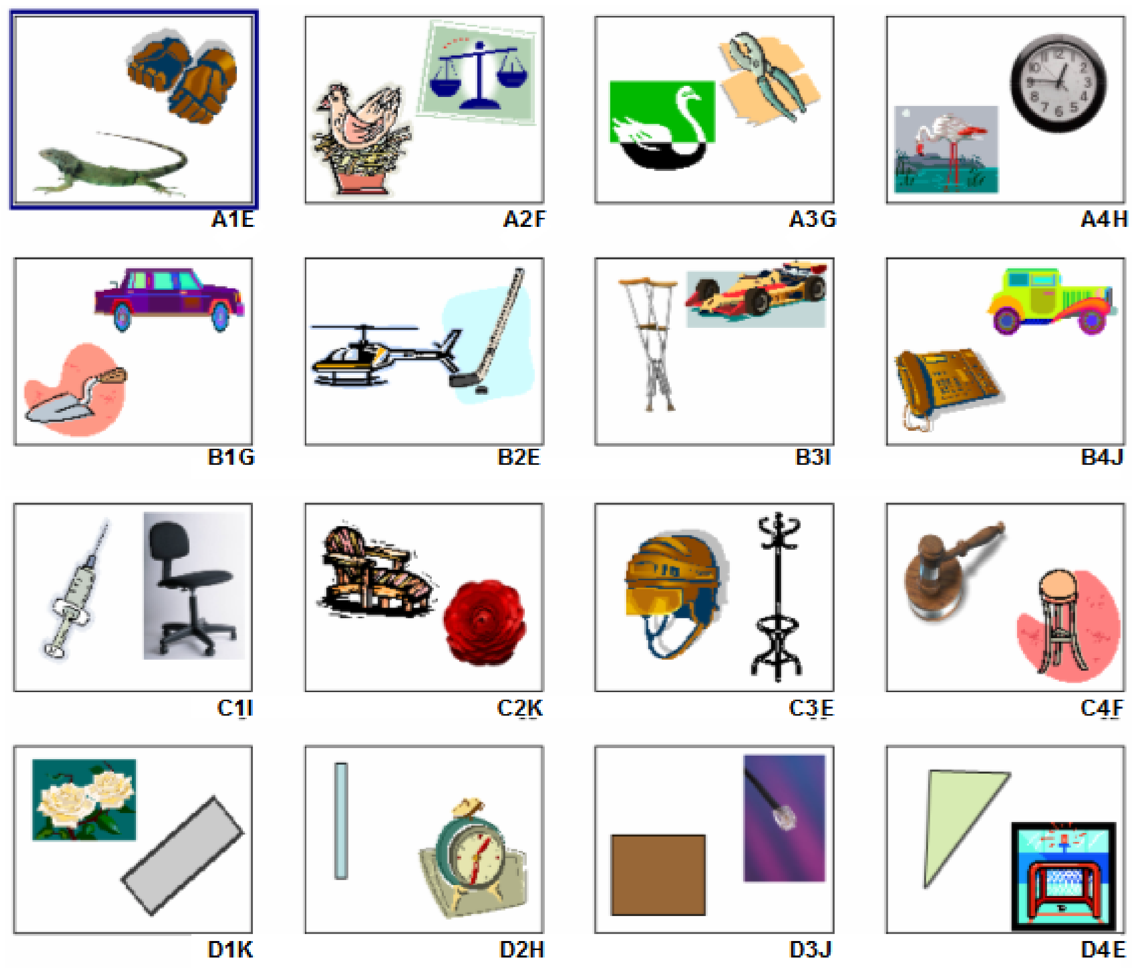
Although the theories above can account for the complexity of some task dimensions, other task dimensions within the experimental design are not captured by these theories. For instance, restructuring in terms of a higher degree of category abstraction within a stimulus set cannot be properly explained. While it may be possible that the search space approach using the Information processing system can be applied, a stimulus set with higher degree of abstraction does not necessarily have a bigger or smaller problem space to search. It merely has a different search space to consider. The Gestalt concept of restructuring captures this effect more accurately since it is based on formation of appropriate problem structures. It argues that the most important part of any problem solving is forming the right problem representation. This implies that the difficulty associated with formation of problem structures affects perceived problem complexity and vice versa. The greater the task involved in forming the right problem structure, the greater the perceived complexity of the task. This view captures both the abstraction tasks and the degree of homogeneity. In comparing a task stimulus without any of these structural dimensions to a stimulus that has such dimensions, the perceived difficulty is assumed to increase with the addition of structural dimensions, since each added dimension should make it more difficult to conceptualize the right problem structure. Similarly, Lewin's life space approach captures the restructuring characteristic, as the tension existing in the problem area would increase with an increase in perceived dissimilarity between current state and desired goal state. This is typical of problems requiring restructuring such as insight problems. However, the quantification or operationalization of these theoretical constructs is very challenging.

In addition to the theories identified above, Heider's balance theory captures the dynamic aspect of problem solving without focusing on any specific construct. Rather, problem solving as stated earlier is perceived as synonymous to dissonance reduction or continuous change in state (Heider, 1946a). Thus every attempt at problem solving is a means of approaching a state of balance through continuous dissonance reduction until the balance state is attained. With any of the structural dimensions, Heider's problem solving approach shows that each exchange or decision is a form of problem solving based on reducing dissonance. This is achieved by moving towards the goal state and working within the available constraints. For instance, using the detour problem shown in Figure 5, the decision to exchange any cards (e.g. cards 1 and 2 between persons 1 and 2) is motivated by what each person is attempting to collect based on their current cards and the goal. If person 1 decides to go with the hockey items, then he is reducing the dissonance by collecting what he has and creating the group of four of a kind. In using Heider's balance theory, one can compare the effects of the task dimensions across different stimuli sets through quantification of degree of balance at each stage of problem solving. The perceived complexity of the task structure will be based on the balance state, which affects the approach to reducing dissonance. This can be observed in the form of card exchanges and the time spent in solving the task, which are discussed in more detail in subsequent sections.

3.4 Categorization of Stimuli Sets

To further describe the experimental stimuli prior to categorizing the stimuli types, Figure 7 has been provided to demonstrate the relationship between task structure and stimulus types. Each card is represented by three alphanumeric characters as shown in the figure below. In the card A1E for example, the first letter (A) represents the group that the card belongs to, the middle digit (1) represents the card position in the group and the last letter (E) represents distraction pair or incorrect group. For instance letter A represents animals while E represents hockey items (detour category). Also letter B represents vehicles while letter I represents medical equipment (distraction pair). The relationship between the cards is based on the first and last letters (A or E). If card representations in individual sets contain similar alphabets such as E in A1E and B2E or B in B3H and B2L then they are perceived by the participant to be related.

Figure 7: Associated codes in Experimental Stimulus



Further to the description of task characteristics in the previous section, the different stimuli combinations affect the perceived degree of difficulty (see Appendix C).

Table 1: Experimental Design Chart

Cell No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Homogeneity	Y								N							
Detour	Y				N				Y				N			
Redundancy	Y		N		Y		N		Y		N		Y		N	
Restructuring	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N
Stimulus Set	**G**		**J**		**K**			**A**	**F**	**E**	**H**	**C**	**I**	**D**		**B**

Table 1 summarizes the possible combinations of stimuli sets included in the experimental study. The last row represents the stimuli sets included in the study. For instance, column 3 indicates that the stimulus set has the homogeneous, detour, and restructuring dimensions included. This stimuli set is labeled as stimulus J.

Stimulus B (column 16) has no detour, no redundancy and no restructuring characteristic; it has the desired four groups and eight pairs of random objects. Stimulus C (column 12) has the detour characteristic and six pairs of random objects while Stimulus D (column 14) has the redundancy characteristic and six pairs of random objects. Stimulus E (column 10) has detour and redundancy characteristics with four pairs of random objects. Stimuli H and I are the heterogeneous stimuli while stimuli J and K are the homogeneous version of the abstracted or restructured stimuli. In addition, stimuli H and J have the detour characteristic while Stimuli I and K have the redundancy characteristic. Stimuli H, I, J and K are represented in columns 11, 13, 3 and 5 respectively.

The different combinations of structural dimensions result in multiple forms of stimuli among which eight {B, C, D, E, H, I, J, K} were selected for the purpose of this thesis. Stimulus A is the training set and the results were not included in data analysis; stimuli F and G were also omitted due to the difficulty associated with resolving the tasks. Stimuli set B is the base structure with none of the structural dimensions, stimulus set C has the detour characteristic, stimulus set D has the redundancy characteristic and stimulus set E has both the detour and redundancy characteristics. Stimuli sets H, I, J & K are experimental sets with the restructuring dimension. The representations of these stimulus sets are provided in Appendix A.

The variations in characteristics and properties affect the perceived complexity in the structure of the stimuli and lead to differences in performance or behavioural effectiveness. The results or goals of all stimuli sets are the same with the exact same combination of cards and objects (see appendix A). In addition to the basic structural differences, there also seems to be interacting properties among the structural elements leading to more complicated results. For instance, stimulus E has both detour and redundancy properties in one stimulus set while Stimulus H has detour and abstraction. The results from these interactions may be not be the direct additive relationship among the properties and thus raises other interesting questions on the interacting relationships.

3.5 Task structural measures

Task structure can be measured in different ways using representations such as complexity, difficulty and similarity based on the characteristics of the structure. The ability to quantify and compare the structural characteristics remains a difficult and challenging task due to the varied nature of problem situations (Zajonc and Taylor, 1963).

Researchers have characterized the problem structure using different dimensions such as the amount of knowledge required to solve the problem, manipulated task 'load', problem representation, familiarity of the domain, size of problem space and length of minimum solution path (Derbentseva, 2002; Newell and Simon, 1972a). For instance, using the amount of knowledge required to solve the problem, the greater the amount of knowledge, the more complicated the task. Likewise, the fewer the solution paths that are required to solve the problem, the easier the task in question and vice versa.

While these are valid characteristics of problem structure difficulty or complexity, the factors are not generally applicable to all problems. This implies that results generated based on these measures are difficult to apply in other scenarios if the structural measures or dimensions are not universal. In this thesis, the measures used are based on more generalized concepts (defined in earlier sections) such as detour, redundancy, abstractness and homogeneity which can be adapted to other problem situations.

Two measures were used to rate the degree of complexity of various experimental stimuli:

1. Number of pairs
2. Number of structural dimensions

3.5.1 Number of pairs

This refers to the number of possible pairs (i.e., two-item categories) that can be formed in a given stimulus. The measure assumes that the minimum step in any of the categorization tasks is the grouping of items into pairs, thus the structural complexity of a stimulus can be measured in terms of the number of available pairs. The general formula to calculate the total number of pairs in any stimulus is $\left({}^{n}_{2}C\right)$ where n is the number of items in each category. For instance, Stimulus B has a combination of four sets of correct groups with four objects each and eight sets of paired distraction sets (i.e. $\left[\left({}^{4}_{2}C\right)*4+\left({}^{2}_{2}C\right)*8\right]$). The representation $\left({}^{4}_{2}C\right)$ denotes the number of pairs (i.e. 2) that can be formed from the correct group of four objects. Since there are four correct groups in every stimulus set, $\left({}^{4}_{2}C\right)$ is multiplied by 4. The representation $\left({}^{2}_{2}C\right)$ calculates the number of pairs that can be formed from each set of two (distraction pairs in this case). This number is multiplied by eight since there are eight distraction pairs in Stimulus B. In every group of a kind with four objects, there are six possible combinations of two. Since there are four groups of 4 within each stimulus, then there would be 24 possible pairs in addition to the 8 possible pairs of random objects included resulting in a total of 32 possible pairs.

For a detour case, there is a combination of $\left[\left({}^{4}_{2}C\right)*5+\left({}^{2}_{2}C\right)*6\right]$ which equals 30 and 6 with a total of 36. For a redundancy case, there is $\left[\left({}^{5}_{2}C\right)*4+\left({}^{2}_{2}C\right)*6\right]$ which equals 46 (i.e. (10*4) + 6). Lastly, for a case with redundancy and detour, there is $\left[\left({}_{2}^{5}C\right)*4+\left({}_{2}^{4}C\right)*1+\left({}_{2}^{2}C\right)*4\right]$ which equals 50.

Since this form of measurement is unable to measure the abstraction or restructuring characteristic directly, it was decided that the numbers of pairs of the restructured items would be treated in two stages. The first is based on characteristics identified above for non-restructured stimuli (i.e. detour or redundancy features), and the second part is based on the number of possible pairs without including the restructured item. For instance, detour with restructuring (Stimulus H or I) has 36 pairs $\left[\left({}_{2}^{4}C\right)*5+\left({}_{2}^{2}C\right)*6\right]$ for just the detour dimension and an additional 24 pairs $\left[\left({}_{2}^{3}C\right)*4+\left({}_{2}^{4}C\right)*1+\left({}_{2}^{2}C\right)*6\right]$ when ignoring the restructured item (i.e. assuming paired items and including 6 extra pairs which total 60 pairs) Redundancy with restructuring has 46 pairs $\left[\left({}_{2}^{5}C\right)*4+\left({}_{2}^{2}C\right)*6\right]$ with an additional 24 and 6 extra pairs $\left[\left({}_{2}^{4}C\right)*4+\left({}_{2}^{2}C\right)*6\right]$ which totals 76 pairs.

This approach shows that the degree of complexity increases in order of magnitude from Stimulus B to Stimulus K. This measure however is unable to capture the homogeneity characteristic and as such treats all restructured stimuli as the same.

3.5.2 Number of structural dimensions

This measure simply counts the number of task dimensions in each Stimulus. Stimulus B has a structural dimension of 1 based on forming four of a kind. Stimulus C has a dimension of 2, which constitutes the four of a kind and the detour characteristic. Also, Stimulus D has 2 comprising of the redundancy feature and four of a kind category while Stimulus E has a structural dimension of 3 comprising of four of a kind, redundancy and detour. Stimuli H and J have structural dimensions of 3 comprised of four of a kind, either detour or redundancy, and restructuring. Stimuli K and I however have structural dimensions of 4 comprising of four of a kind, either detour or redundancy, restructuring, and the homogeneity dimension. The limitation of this measure is that all structural dimensions or characteristics are assigned the same quantitative value, which implies that the effect of each dimension on problem solving complexity is assumed to be the same.

3.5.3 Abstraction T-Test

The abstraction t-test, which is described in more detail in the methodology section 4.2 (Experiment II), measures the degree of abstraction or restructuring and the degree of homogeneity across the different stimuli sets. This measure uses the t-test to evaluate the differences in selected participants' ratings of the degree of similarity of object sets. The degree of similarity is based on a scale of 1-10 with 1 being the lowest and 10 being the highest. For example, the rating is expected to be higher when comparing a black chair to a green chair as opposed to comparing a coat rack with a green chair.

The abstraction t-test magnitude (as shown in Table 4 – section 5.1.2) demonstrates the abstraction effect across different sets of items. This is achieved by comparing the t statistic value of sets of items with and without restructured item (e.g. iguana or helicopter in the animal or vehicle category). This abstraction test shows two forms of task effects. For the restructuring dimension, it shows the difference in the perceived similarity between objects in non-restructured categories (e.g., different types of chairs) and objects in categories requiring restructuring (e.g., chairs and coat rack). For the homogeneity dimension, it shows the relative difference between the perception of homogeneous items (e.g., identical office chair) compared to the restructured item (e.g., coat rack) and the perception of heterogeneous items (e.g., three different types of chairs) compared to the restructured item (e.g., coat rack). The different forms of structure representation and measures help in understanding and analyzing the experimental results in later sections. It also addresses the limitations of the structural measures described in sections 3.5.1 and 3.5.2. The two measures adopted here are:

1. Degree of abstraction
2. Degree of homogeneity

The measures of the task structure used in this thesis are not without limitations. Similar to the theories of problem solving discussed earlier, each measure captures some but not all of the characteristics of problem structure implemented in the experimental stimuli.

For instance, some measures capture the degree of complexity while others capture the degree of abstraction and homogeneity.

3.6 Measuring dependent variables

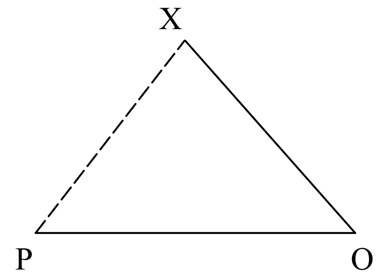
The dependent variables in this thesis are problem solving performance and behavioural variety. In some instances these variables are used interchangeably; however, performance is more quantitative in nature while behavioural variety is more qualitative. Both variables rely on Heider's balance theory as a fundamental theory and as such the heuristic will be discussed prior to describing the measures.

3.7 Heider's balance heuristic

Cognitive consistency has been the subject of a long research tradition in psychology (Festinger, 1950; Heider, 1946a). Consistency theories are based on the premise that people prefer balanced or coherent cognitions to unbalanced or incoherent cognitions. Heider's balance theory (Heider, 1946a) considers the preference for balance in the context of a triad network structure called a P-O-X network. The P-O-X network represents a person (P), another person (O), and a third unit (thing, person, etc.) and positive or negative relations among these three units. Heider discussed two kinds of relationships between units: sentiment relations (e.g. liking or disliking) or unit relations (similarity or dissimilarity). For instance, A likes (or is similar to) B is represented as a positive relationship (+) while A dislikes (or is not similar to B) is represented as a negative relationship (-). Triads are balanced if the product of the three relationship signs is positive and unbalanced if the product is negative. An unbalanced triad is illustrated in Figure 8 using solid lines to represent positive relations and a dashed line to represent a negative relation.

In 1956, Cartwright and Harary generalized Heider's ideas to the concept of structural balance using graph theory to represent more complex structures and situations. Specifically, signed graphs, or s-graphs were used to represent the structural social dynamics of balance theory (Zhong, 2005).

Figure 8: S-graph representation of an unbalanced triad configuration



While the simple computation above is sufficient for a triad structure, the balance state of more complicated structures had to be defined in terms of semi cycles. Cartwright and Harary (Cartwright and Harary, 1956) proved that a structure is balance if every semi cycle within the structure were positive, where a positive semi cycle refers to one in which the product of all link signs is positive. While this generalization is laudable, it is difficult to compute in social networks due to the number of possible semi cycles (which is dependent on the number of nodes) in any network structure.

They also showed that a balanced network can be partitioned into two disjoint subsets, where the nodes within each subset are connected by positive links and the nodes in different subsets are connected by negative links. Davis generalized this finding by proposing the concept of k-balance which corresponds to a network that can be partitioned into k disjoint subsets, such that point within the same subset are linked by positive relations and points in different subsets are linked by negative relations (Davis, 1967).

Since most empirically observed networks are unbalanced, there was the need to understand the relative degree of balance. Cartwright and Harary defined this as “the ratio of the number of positive semi-cycles to the total number of semi-cycles” (Cartwright and Harary, 1956). This measure is also difficult to apply due to the limitation described earlier where the number of semi cycles increases exponentially with the number of nodes in the structure. An alternate measure that addressed these limitations is the line index of balance (LIB), which is the minimum number of lines in a network whose sign reversal will result in a balanced graph (Harary, 1959; Harary et al., 1965). LIB is far easier to compute, and thus has been adopted in most studies of social networks. In this research LIB is used to understand balance states of individuals and groups while solving the experimental task. We propose that problem solving can be viewed as a dissonance reduction process from an initial state of imbalance to a solution state of balance. In terms of the present experiment, each card exchange corresponds with a restructuring of the network to different states of balance until the best balance state is achieved and the solution state corresponds with maximum balance where LIB = 0.

The laboratory experiment in this thesis is designed such that the overall balance is based on the sixteen cards, which are initially distributed among the participants as fours cards each. Balance in this case is based on Heider's unit relation of similarity, whereby items in the same category are positively related and items in different categories are negatively related. Thus, the correct solution corresponds to a k-balanced graph where k = 4. During the course of problem solving, the number of cards with each participant either increases to as many as five or reduces to a minimum of one. In some groups, Participant 1 may have as much as five cards, Participants 2 and 3 will have four cards and there will be only three cards with the last person. Total balance is calculated as the sum of individual line index of balance (LIB): $b_T = b_1 + b_2 + b_3 + b_4$ where b_n refers to LIB for each individual.

Due to the fact that balance state is dependent on one's perception of relationship among units, two forms of measures are devised from the line index of measure (LIB). They are 1) Subjective measure, which is based on participant's view of relationship among objects on the cards, and 2) Objective measure, which is based on the researcher's view of relationship among objects on the cards. Both forms of measure differ from each other in terms of the individual and amount of information available at any point in time.

The researcher has total information knowing objects that can be categorized together but the participants have limited knowledge and thus base their categorization on similarity as defined by them. This differentiation between the two views is important as it allows an estimation of each group's degree of understanding of the structure at each point in time.

The perceived relationship between two cards is represented as a positive link in the subjective balance. For instance, detour cards A1E and B2E will be represented as related (i.e., positive link) in subjective balance if a group collects both cards. It is assumed that they are grouped together because of the perceived relation based on group E (e.g. hockey items). However, the relationship between two cards can only be a positive link in the objective balance if it is a relationship based on the first letter of the card representation such as A1E and A3G or D1K and D2H, which will form the correct categories.

Any set of four cards can have six possible relationship links based on all the possible relationship among cards. For instance, the relationship links are between cards 1 and 2, 2 and 3, 3 and 4, 4 and 1, 1 and 3, 2 and 4. If there are only three cards, there would be three relationship links between cards 1 and 2, 1 and 3, 2 and 3 and any two cards will have only one link.

The line index of balance computed for each individual's balance state is calculated as the total number of negative links within the group of cards based on k-balance measure where $k = 1$ for each individual and $k = 4$ for the group. For instance, if Participant 1 has four cards and none of them is related in any way then the balance is six since there will be six negative links in the set as shown in Figure 9a. If there are only two related or linked cards within a set of four that will indicate a single positive link and five negative links and the total balance would be five as shown in Figure 9b. If there are two sets of similar cards with no relationship across the two distinct groups, the total balance would be four as shown in Figure 9c. If there are three related cards with one dissimilar card then the total balance is three as shown in Figure 9d. Table 2 below summarizes the balance state of similar and dissimilar cards. The most balanced network has a value of zero, which occurs in the situation where all cards belong to the same category.

Note however that in computing the balance in some situations, there could be two completing categories (since there are two objects on each card) where one has to make a choice on which group to consider. The two possible relationships in a group with cards A1E, B2E, C3E and A3F may be such that A1E is linked with either B2E & C3E or A3F. The rule here is similar to the figure and ground principle by Gestalt psychologists (Wertheimer 1982).

Figure and ground in this case implies that when categorization is based on a particular concept, the other theme on the card becomes invisible. For instance, with a detour stimulus set, once the focus is on the detour object such as hockey item, the other object (such as animal, vehicle, furniture or shape) on the card becomes the ground item and will not be noticed. This implies that if the participant chooses to link A1E as a positive relationship with B2E and C3E (due to the E) then it cannot be linked with A3F (due to A) and vice versa.

However, in either of the balance measure, the choice is based on the category with the greater amount of balance since the objective is to reach a state of complete balance. For instance, in the scenario described above, the category E will be chosen over A in the subjective balance due to the perception that it has LIB = 3 for category E and LIB = 5 for category A (see S1 and S2 in Figure 10 below). However, the objective balance has preference for the A category over E due to the fact that the LIB = 6 when category E is chosen and 5 when category A is chosen (see illustration in O1 and O2 in Figure 10 below).

Figure 9: Diagrammatic representation of possible states of balance

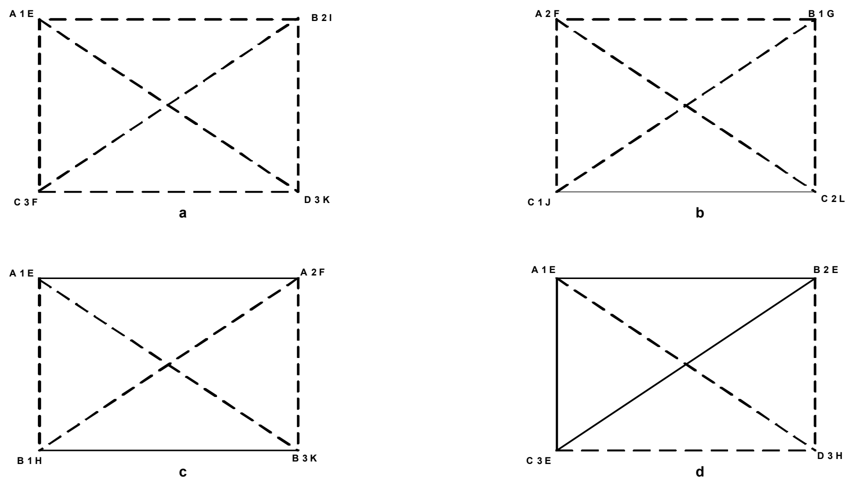
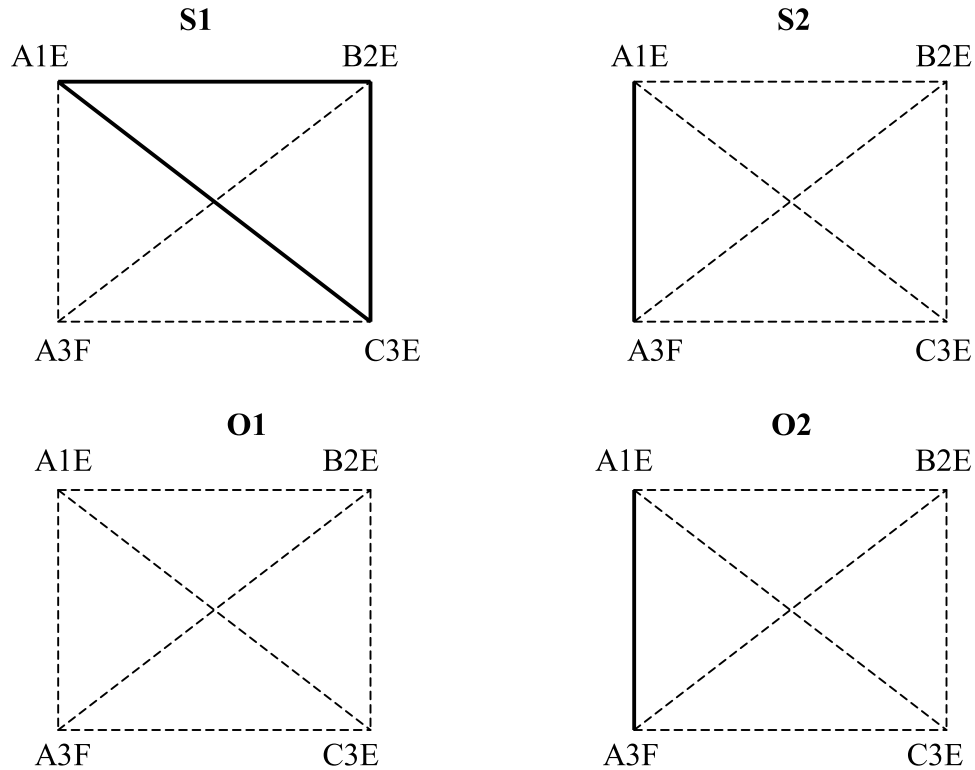


Table 2: Balance states based on Heider's Heuristics

Similar Pairs	Dissimilar Pairs	Balance
4	0	0
3	1	3
2	2	4
1	3	5
0	4	6

Figure 10: S-graphs representing subjective and objective states of balance



3.7.1 Card movement heuristics

It is assumed that participants solve the experimental problem by making decisions based on the identified goal. The goal as described to subjects was to come up with four of a kind. Four of a kind as a goal is deliberately abstract and vague. This goal was developed due to the nature of the task, which varied according to the distribution of available objects. For instance, a set of cards may be formed as snow items or the set may have to be broadened to the weather category to allow all group members to attain the desired goal. The identified goal guided participants' decisions such that four of a kind was based on similarity. The subjects tended to form a category in which they had the greatest number of similar objects based on their perceptions. For instance, participants usually started their tasks by asking if anyone had any pairs of similar objects, for example saying "I have two birds, does anyone have another bird and can you give me your bird?" While this approach was usually appropriate for simpler tasks, it could sometimes lead to difficulties for more complicated tasks. For instance, in a task with a detour category, a participant collecting the detour items may form his or her four of a kind but will prevent other group members from forming their four of a kind.

Figure 11: Sample of Heider's Balance Movement within a particular Task

	Card Movement 2E																									
	Task ID: 050103 22nd March 2005 9:00																									
Time	Person 1				PB	AB	Person 2				PB	AB	Person 3				PB	AB	Person 4				PB	AB	TPB	TAB
0:11:30	A1E	B2E	C3E	D3C	3	5	A2F	B1G	C1H	D1I	6	6	A3G	B3H	C2I	D2F	6	6	A4D	B4A	C4B	D4E	4	5	19	22
0:12:01	A1E	B2E	C3E	D3C	3	5	A2F	B1G	C1H	D1I	6	6	C4B	B3H	C2I	D2F	5	5	A4D	B4A	A3G	D4E	3	4	17	20
0:12:11	A1E	B2E	C3E	D3C	3	5	D4E	B1G	C1H	D1I	5	5	C4B	B3H	C2I	D2F	5	5	A4D	B4A	A3G	A2F	0	3	13	18
0:12:14	A1E	C2I	C3E	D3C	3	3	D4E	B1G	C1H	D1I	5	5	C4B	B3H	B2E	D2F	3	3	A4D	B4A	A3G	A2F	0	3	11	14
0:12:22	C1H	C2I	C3E	D3C	0	0	D4E	B1G	A1E	D1I	5	5	C4B	B3H	B2E	D2F	3	3	A4D	B4A	A3G	A2F	0	3	8	11
0:12:48	C1H	C2I	C3E	D3C	0	0	D4E	B1G	B4A	D1I	4	4	C4B	B3H	B2E	D2F	3	3	A4D	A1E	A3G	A2F	0	0	7	7
0:13:04	C1H	C2I	C3E	D3C	0	0	D4E	B1G	D2F	D1I	3	3	C4B	B3H	B2E	B4A	0	3	A4D	A1E	A3G	A2F	0	0	3	6
0:13:38	C1H	C2I	C3E	D3C	0	0	D4E	B4A	D2F	D1I	3	3	C4B	B3H	B2E	B1G	0	0	A4D	A1E	A3G	A2F	0	0	3	3
0:13:40	C1H	C2I	C3E	D3C	0	0	D4E	A4D	D2F	D1I	0	0	C4B	B3H	B2E	B1G	0	0	B4A	A1E	A3G	A2F	0	0	0	0

Figure 11 shows both the card movement and the Heider's balance computation. Using some of the card exchanges above, the columns show each individual's collection of four cards at the beginning of the experimental task and the set at the end of the problem solving. The columns represented as Perceived Balanced (PB) and Absolute Balance (AB) represent the individual subjective and objective computations respectively. The Total Perceived Balance (TPB) and Total Absolute Balance (TAB) are the total subjective and objective balance measures for the group at each point in time. It can be shown that the objective balance score is always equal to or less than the subjective balance score. This is usually as a result of the imperfect information from the participants' point of view and attempt to use any seemingly correct category for their individual collections of four of a kind. The first row of cards by Participant 1 shows that the initial collection and balance computation is based on the 'E' objects, which corresponds with the detour group. As a result of this misconception, the subjective balance is based on only three negative links from the last card (i.e. D3C) to the other three cards (A1E, B2E, C3E) since card D3C is perceived as incompatible with the others. Similar to the Gestalt's figure and ground rule, the counting rule amounts to picking the better balance score per person if multiple scores are available.

Figure and ground in this case implies that when categorization is based on a particular concept, the other theme on the card becomes invisible. For instance, with a detour stimulus set, once the focus is on the detour object such as hockey item in Figure 7, the other object (such as animal, vehicle, furniture or shape) on the card becomes the ground item and will not be noticed. This implies that if the participant chooses to link A1E as a positive relationship with B2E (due to the E) then it cannot be linked with A3F and vice versa.

3.8 Performance measures

Performance is one of the dependent variables identified in this experiment. Performance can be defined as the quantitative measure of how the groups react to the structural properties based on their perception of the task structure. Three measures of performance were used in this study:

1. Time taken
2. Number of card exchanges
3. Rate of card exchange

3.8.1 Time taken

This is the length of time (in minutes and seconds) that it takes a group to complete a particular task. Time is usually a good measure of performance and it is proposed that the performance varies directly with the level of difficulty of each task. The simpler the task, the less time it should take to complete and the more difficult the task, the longer time it should take to complete. The level of complexity of each task is based on the measures explained earlier.

3.8.2 Number of card exchanges

Card exchanges provide a means of movement towards the desired solution. Card exchange measurements are based on the total number of cards that had to be exchanged for the group to reach the desired solution. The greater the perceived difficulty of the stimulus, the greater the variation existing between the perceived structure and the real structure. This results in a higher number of card exchanges. Thus, the number of cards exchanged demonstrates how easy or difficult it is to get to the solution.

The minimum number of card exchanges required to solve any task is 12 with the exception of any stimuli with redundancy characteristic, which can be resolved with 10 exchanges. Although only one task was solved with ten card exchanges, Figure 12 below shows the typical card exchanges and the solution with ten card exchanges.

Figure 12: Sample Card Exchanges

	Card Movement 2C																									
	Task ID: 070404 24th March 2005 16:00																									
Time	Person 1					PB	AB	Person 2					PB	AB	Person 3					PB	AB	Person 4				
0:26:10	A1E	B2E		C3E	D3J	3	6	A2F	B1G		C1I	D1K	6	6	A3G	B3I	C2K		D2H	6	6		A4H	B4J	C4F	D4E
0:28:39	A1E	B2E		C3E	D3J	3	6	A2F	B1G		C1I	D1K	6	6	A3G	B3I	C2K		**A4H**	5	5		**D2H**	B4J	C4F	D4E
0:28:47	A1E	B2E		C3E	**B4J**	3	5	A2F	B1G		C1I	D1K	6	6	A3G	B3I	C2K		A4H	5	5		D2H	**D3J**	C4F	D4E
0:28:50	A1E	B2E		C3E	B4J	3	5	**C2K**	B1G		C1I	D1K	5	5	A3G	B3I	**A2F**		A4H	3	3		D2H	D3J	C4F	D4E
0:28:54	A1E	B2E		C3E	B4J	3	5	C2K	B1G		C1I	**C4F**	3	3	A3G	B3I	A2F		A4H	3	3		D2H	D3J	**D1K**	D4E
0:28:58	**B3I**	B2E		C3E	B4J	3	3	C2K	B1G		C1I	C4F	3	3	A3G	**A1E**	A2F		A4H	0	0		D2H	D3J	D1K	D4E
0:29:03	B3I	B2E		**B1G**	B4J	0	0	C2K	**C3E**		C1I	C4F	0	0	A3G	A1E	A2F		A4H	0	0		D2H	D3J	D1K	D4E

Time	Person 1					PB	AB	Person 2						PB	AB	Person 3						PB	AB	Person 4				
0:26:05	A1E	B2E		C3E	D3C	3	5	A2F	B1G			C1H	D1I	6	6	A3G	B3H	C2I			D2F	6	6		A4D	B4A	C4B	D4E
0:30:02	A1E	B2E		C3E	D3C	3	5	A2F	B1G			C1H	D1I	6	6	A3G	B3H	C2I			**B4A**	5	5		A4D	**D2F**	C4B	D4E
0:30:27	A1E	B2E		C3E	D3C	3	5	A2F	B1G			C1H	**C4B**	5	5	A3G	B3H	C2I			B4A	5	5		A4D	D2F	**D1I**	D4E
0:30:31	A1E	B2E		C3E	D3C	3	5	**B3H**	B1G			C1H	C4B	3	3	A3G	**A2F**	C2I			B4A	3	3		A4D	D2F	D1I	D4E
0:30:38	A1E	**C1H**		C3E	D3C	3	3	B3H	B1G			**B2E**	C4B	0	0	A3G	A2F	C2I			B4A	3	3		A4D	D2F	D1I	D4E
0:30:42	**C2I**	C1H		C3E	D3C	0	0	B3H	B1G			B2E	C4B	0	0	A3G	A2F	**A1E**			B4A	0	0		A4D	D2F	D1I	D4E

3.8.3 Rate of card exchange

This is derived from the two measurements above and it is a measure of card exchanges per minute. It predicts the ease with which people exchange cards. It is proposed that the easier the task, the higher the exchange rate.

3.8.4 Hypotheses of the effect of Task Structure on Performance

The following section lists and describes the proposed hypotheses:

Hypothesis 1:

H_1: Problem solving performance varies inversely with structural complexity.

This implies that performance will decrease with increasing structural complexity such that the greater the perceived complexity of the task, the lower the performance.

Hypothesis 2:

H_{2a}: Task performance reduces with the introduction of abstraction or restructuring.

This implies that tasks with no abstraction have better performance than tasks with different order items requiring restructuring.

H_{2b}: In restructured tasks, task performance reduces with increasing degree of homogeneity.

Homogeneity is only varied in conditions involving restructuring and its presence or absence is expected to affect group performance. Homogeneity implies that the objects in each solution set are identical (e.g. only swans) as compared with the heterogeneous version, which has different kinds of same order objects (e.g., birds). The introduction of homogeneity is expected to have slightly different effects in the restructured detour and restructured redundancy stimuli sets as described below:

In the detour restructuring condition (stimuli H and J), homogeneity has only one effect, related to the relative difficulty of grouping the different order item (e.g. iguana) with a set of homogeneous or heterogeneous same order items (e.g. swans versus birds). The hypothesis postulates that it would be more difficult for participants to group the different order object with a set of identical objects than with a set of heterogeneous objects. That is, grouping three different kind of birds with an iguana to form the animal category is expected to be easier than grouping three swans with iguana to form the animal category. The reason is that the iguana is expected to be perceived as relatively more similar to the birds than to the swans.

In the redundancy restructuring condition, the same effect described above occurs but there is also an additional effect related to the relative difficulty of breaking apart a homogeneous category of four identical items compared to a heterogeneous category of 4 items. That is, due to the redundancy characteristic, there is the possibility of forming two wrong categories of four lower order objects (e.g. birds or swans), giving the impression that a group is on the correct path leading towards the solution. These wrong categories must be broken up and restructured as more abstract categories for the problem to be solved. In this case the higher similarity of homogeneous versus heterogeneous items should make it relatively more difficult to break apart homogeneous sets than heterogeneous sets. Thus, the difficulty associated with breaking apart the homogeneous category of four identical swans is expected to be greater than breaking apart a group of four different birds to group three of them with the different order item (e.g. iguana).

3.9 Description of behaviour

Behaviour can be described as the physical reactions exhibited as a result of one's psychological state. The understanding of the behavioural component of this research is achieved using Heider's balance theory. Heider's balance theory states that individuals attempt to maintain a state of psychological balance at any point in time and thus postulated the P-O-X relationship, which was explained in Figure 9. In extending this theory to group actions, we suggest that all group units tend to move towards a state of balance in solving a problem. The behavioural effects are operationalized in the two ways shown below:

1. Card movement behaviour
2. Verbal protocol behaviour

3.9.1 Card movement behaviour

During the course of solving a problem, the current state of the group depends on who has which cards and which category of objects each member intends to collect. The card movement behaviour can be represented either as the objective (ideal) movement or the subjective (perceived) movement based on the perspective in question. The objective movement refers to the experimenter's view of an ideal card exchange based on complete knowledge of the task structure. Since the experimenter has complete knowledge of the task structure, he/she knows the end state and the optimal path to the end state. However, the participants do not necessarily have a full knowledge of the task and thus base their card exchange on their current state and level of understanding of the structure. The subjective and objective patterns would be equal if all subjects had full and complete information.

The proposition here is that the structure of the task affects the pattern of card exchanges and movement towards the goal state. It is assumed that with complete information, the subjective and objective card movement patterns should be the same. However, with incomplete information, the difference between the subjective and objective card movement patterns increases as the complexity of the task increases.

The objective and subjective states of balance are well represented in Figures 13 and 14, which show the trajectories in two different stimuli sets from the initial card distribution to the end state when the problem was solved. Stimulus B is a simpler task and as such the difference in trajectories between the subjective and objective balance is smaller than in stimulus E where the detour property tends to erroneously portray the notion of moving towards the goal.

Figure 13: Card Movement for a Simple Task – Stimulus B

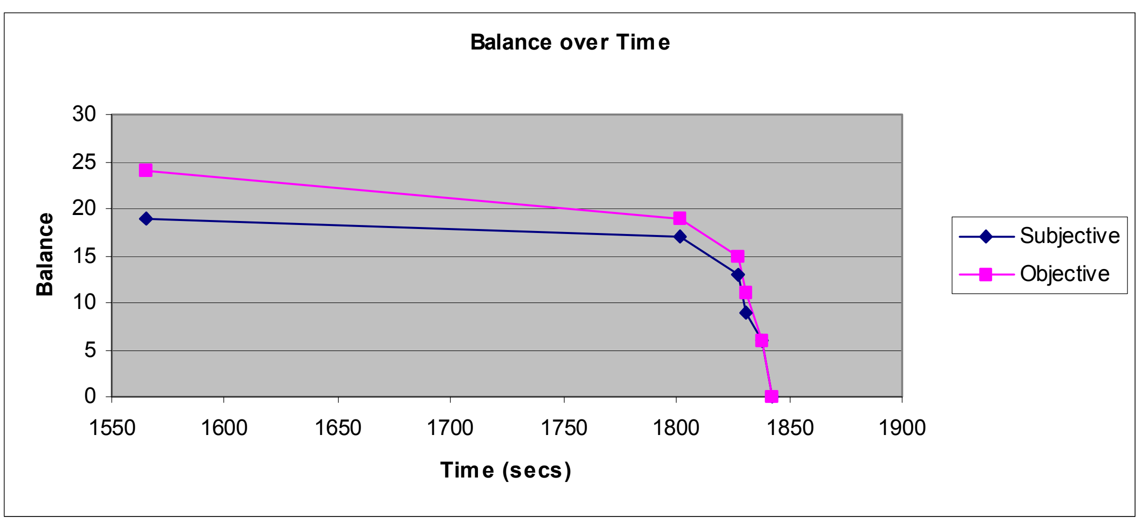
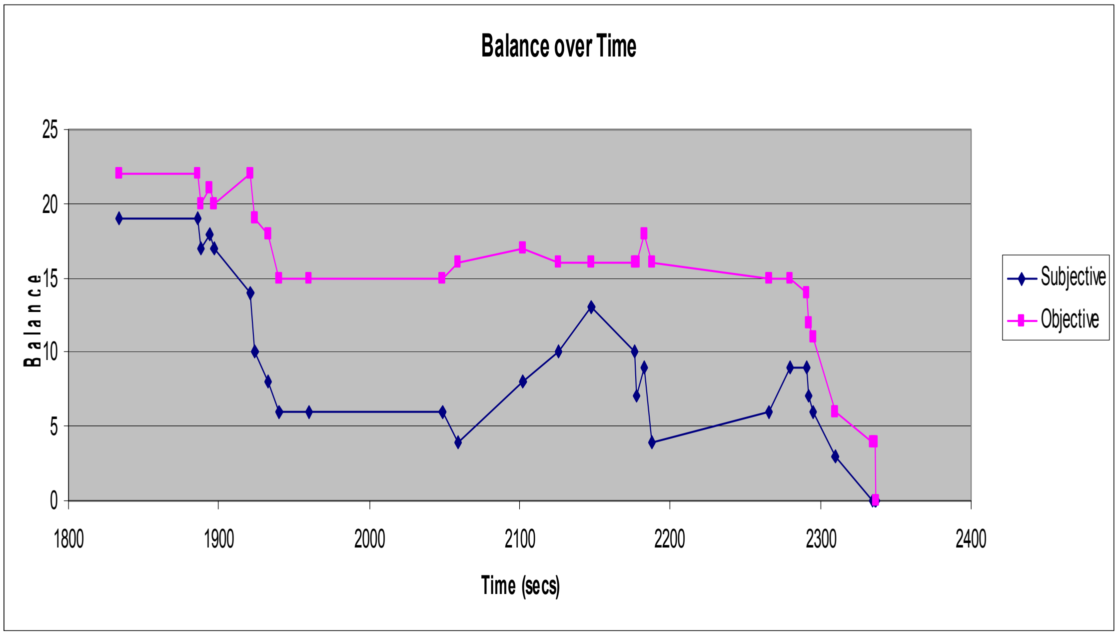


Figure 14: Card Movement for a Difficult Task - Stimulus E



3.9.1.1 Area between objective and subjective balance trajectories

The area between objective and subjective balance curves (see Figures 13 and 14 above) measures the difference between the perceived and absolute total balance (TPB & TAB) based on the computed line index of balance (LIB). Since the group has incomplete knowledge of the stimuli, the total perceived balance (TPB), which is based on the group's view, provides a rough estimate of how the group understands the structure at any point in time. The total absolute balance (TAB), meanwhile, is based on the experimenter's view. The difference between the two balance measures provides an indication of the degree to which the group's understanding of the structure varies from the objective structure. The task complexity is expected to affect this measure as the source of difficulty in the task can be attributed to the inability of the group to understand the structure. The difficulty associated with understanding the task structure is manifested in the difference in total balance between the two trajectories.

3.9.1.2 Card exchange reversals

The number of card exchange reversals is another form of behavioural measurement that is proposed. Conceptually this measure is based on the idea that more difficult problems should be associated with more wrong moves during the solution process, as groups figuratively move through the problem search space from the initial state towards the solution state. The measure counts the number of direction changes in a group's balance trajectory during problem solving, as illustrated in Figures 13 and 14 above. As can be seen in Figure 13 the balance trajectory for a simple problem structure exhibits no direction reversals, which implies that all card exchanges tend towards the balance state collectively demonstrating some degree of consistency and knowledge of the structure. In Figure 14 the balance trajectory for the more complex problem structure has eight reversals suggesting misunderstanding of the problem structure and various wrong moves during the solution process.

3.9.2 Verbal protocol behaviour

In each experimental condition, the information required to solve the problem is distributed across the four subjects with each having approximately one-quarter of the information at any given time. Subjects must communicate and exchange information to develop an understanding of the problem structure and to solve the problem. In this study three modes of communication were used: card exchanges, verbal communication, and visual communication (e.g., holding up a card so others in the group can see, nodding one's head, etc.) All three forms of communication behaviour provided means of understanding the problem structure and reducing dissonance by moving towards solution. Card exchanges are easiest to track and provide the most objective measure of problem solving behaviour in the experiment. A limitation of card exchanges as a behavioural measure is that card exchanges do not reflect the full range of search space exploration undertaken by a group during problem solving. Verbal and visual/nonverbal communication compliment card exchanges by providing a means of mentally exploring the problem space without committing to taking concrete steps in a given direction by actually exchanging cards. Visual communication was not included in this analysis since it is the most difficult of the three types of communication to track and also the least objective requiring significant interpretation on the part of the researcher. This section will explain and describe measures developed for tracking verbal communication among group members during problem solving. Verbal communication behaviour was operationalized in terms of the two measures below:

1. Number of unique verbal categories
2. Verbal reversals

3.9.2.1 Number of unique verbal categories

During the course of their discussions while problem solving, participants put labels on card objects in order to create the desired four of a kind. These labels are based on their perception of the objects and their current understanding of the categories that could form part of the solution to the problem. The variety of these labels is likely to increase proportionally with the perceived complexity of the task. The total number of different category labels mentioned by each group is counted as verbal categories for each task resolution. For instance, a group may mention chairs, seats, wooden stool, bench etc. These are four ways of describing chairs and will be counted in the number of verbal categories. The more difficult it is for the group to conceptualize the structure of the task, the harder they try to understand the problem structure and the greater the variety of labels used.

3.9.2.2 Verbal reversals

Verbal reversals are analogous to card exchange reversals and measures the number of direction changes in the trajectory of “verbal balance” during the process of problem solving. This measure is determined using a computation of verbal balance movement, which is analogous to the measures of card exchange balance described earlier.

Verbal balance movement and card exchange balance are both based on a conceptualization of problem solving in terms of Heider’s balance theory as a process of movement towards perceived balance. However, the verbal balance measure does not directly operationalize Heider’s concept of balance in the verbal communication domain.

Verbal balance is calculated based on the relative proportion of correct category labels mentioned during verbal discussion throughout the problem solving process. The verbal balance movement suggests that in addition to card exchanges, verbal discussion among participants moves the group towards a state of balance in solving the problem. This implies that the group attempts to reduce dissonance during verbal discussions by eliminating irrelevant categories while focusing on the correct categories. Hence, the greater the complexity of the task, the more locomotion there is along incorrect paths and thus the larger the proportion of verbal discussion about the incorrect categories.

Verbal reversals thus track the changes in the proportion of correct talk (correct category labels) mentioned during discussion while problem solving. For instance, with a detour problem, the group's object of discussion at some point in time may be the items in the detour category. Each time the discussion is focused on the wrong set, it implies that they may be moving in the wrong direction and have not yet identified the correct categories or correct items within these categories. This implies that the more complicated the task, the more difficult it is to conceptualize its structure and the more verbal reversals encountered in solving the problem. Reversals occur in all forms of communication but it is possible that through verbal discussions or card exchanges in attempting to solve the problem, reversals in one mode may help to avoid reversals in the other mode. For instance, reversals during verbal communication may help to understand the structure better and consequently avoid further reversals during card exchange.

This measure is calculated as follows. First, all the relevant words used by group members to label item categories during the verbal discussion are listed in the sequence they were mentioned. These category labels are then coded as either correct or incorrect, based on whether or not they correspond to one of the correct solution categories for a given experimental stimulus. A running total is then computed of the relative proportion of correct versus incorrect categories mentioned during discussion. This is done by defining a moving ten word window, beginning with each word in the list as a point of reference and including the next nine category labels mentioned during discussion. For each such window, the number of correct categories mentioned out of ten is counted and recorded. When these numbers are plotted in sequence the resulting curve provides a graphical representation of the trajectory of group discussion during problem solving. Verbal reversals are then defined as any time the slope of this curve changes in sign either from positive to negative or from negative to positive. While the number of verbal reversals could be determined from the curve at this stage, an additional step of smoothing the curve was added to eliminate noise associated with trivial reversals such as the mentioning of a single category label that deviated only slightly from the general discussion trend. Smoothing also reduced noise associated with transcription coding difficulties, such as cases when two or more group members were speaking at the same time making it difficult to determine the precise sequence. The smoothing effect was achieved by taking a moving average of the number of correct categories for each point of reference and the corresponding numbers immediately before and after this point. The resulting values show the trend of the verbal behaviour for correct and incorrect categories during problem solving. Figures 15 and 16 illustrate verbal balance movement and verbal reversals (peaks and valleys in the trajectory) over the problem-solving period.

Figure 15: Verbal Balance Movement for a simple stimulus B

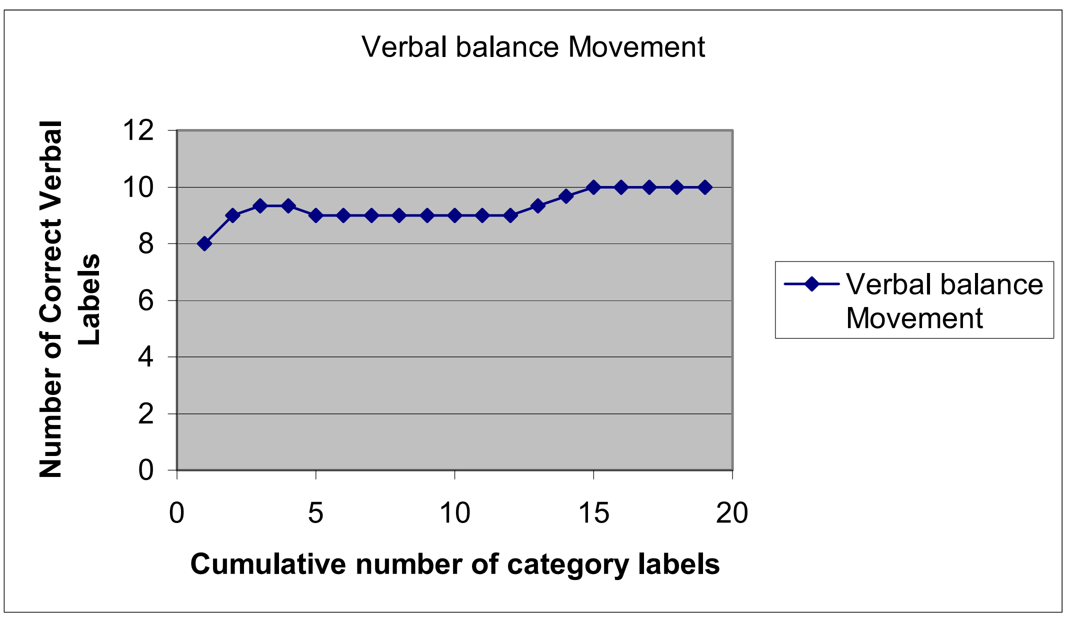
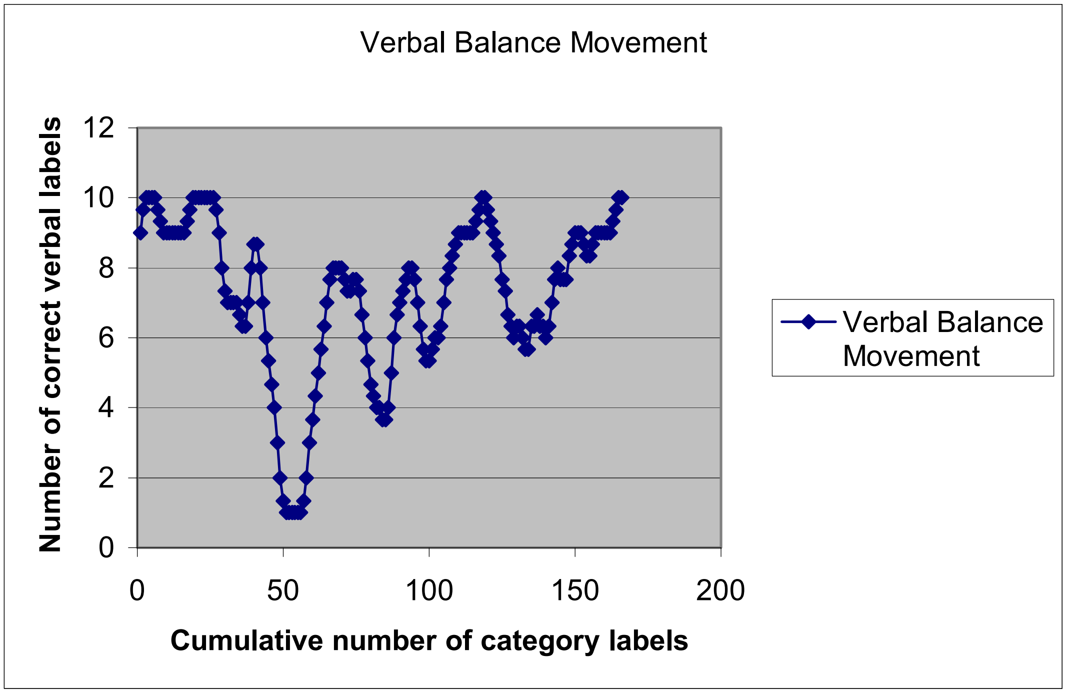


Figure 16: Verbal Balance Movement for a complex stimulus E



3.9.3 High frequency measure

This measure is based on Lewin's (1936) arguments about the effects of psychological forces on movement within life spaces and on the general view developed in this thesis that problem solving corresponds to a progression towards perceived structural balance (Heider, 1946b). The high frequency measure is designed to examine the effect of the relative frequency of card objects in a participant's possession on their subsequent card exchanges during the problem solving process. It is proposed that participants will tend to collect objects similar to the highest frequency objects currently in their possession, due to the perceived force of attractiveness for the person towards additional cards from the same set. In terms of the line index of balance (LIB) measure described earlier, this is equivalent to the idea that participants are attracted to cards that incrementally increase their relative degree of balance. Specifically, after the first card exchange (which is often a random choice) we argue that each subsequent card exchange is a function of the participant's previous state. For instance, a participant with two bird cards would likely look for more bird cards until he or she completes the task, gets stuck or the group runs out of such cards.

In order to test the effect of locomotion in the direction of higher frequency, a random choice of eight problem solving sessions (i.e., one from each stimulus set) was selected and each card exchange was analyzed to verify if it supports this proposition. In collecting the data, each problem solving session is considered from the first exchange time slot to verify if the members exchanged cards in which they had the highest number of similar cards. If the exchange is inline with the proposition, then the decision slot has a value of 1 other wise it has a value of 0. The collected data is then analyzed using chi-squared (χ^2), to verify if the number of high frequency decision fit the proposition.

3.9.4 Hypotheses of the effect of task structure on behavioural variety

Hypothesis 3:

H_3: Behavioural variety increases with structural complexity.

This implies that behavioural variety varies directly with structural complexity such that less complicated tasks will have less behavioural variety such as reduced balance and verbal reversals.

The argument is somewhat related to Ashby's law of requisite variety (Ashby, 1958) in which more complex structures are associated with a higher variety of possibilities that need to be explored by the problem solver (i.e., larger problem space to search), so more behavioural variety is required to explore and understand these structures.

Hypothesis 4:

H_{4a}: Behavioural variety increases with the introduction of restructuring.

Tasks with no abstraction result in less behavioural variety such as fewer verbal reversals or card exchange reversals

H_{4b}: In restructured tasks, increasing degree of homogeneity will result in higher behavioural variety.

Since homogeneity is only varied in restructured tasks, homogeneity in such tasks will vary directly with behavioural variety such that tasks with lower degrees of homogeneity will exhibit less behavioural variety.

Hypothesis 5:

H_5: Problem solving behaviour is in the direction of increased perceived structural balance.

This implies that cards will be exchanged in the direction of the higher frequency of similar objects or greater degree of perceived balance.

3.10 Problem solving strategy

Apart from the effect of problem structure on performance, there is also the issue of the problem strategy adopted during problem solving. It has been argued that the choice of strategy adopted affects the group performance even within the same stimuli set. The choice of strategy is the difference between members of a group understanding the problem structure collectively (verbal discussion) and individually (card exchange). The idea behind this measure is that groups might perform better if they spent time early in the problem solving process discussing and trying to understand the problem structure, rather than blindly exchanging cards in the hope of finding the solution by trial and error. Theoretically this view corresponds with a Gestalt notion of solving the problem by understanding the problem structure rather than by simply moving about within an assumed problem search space. Such early discussion is likely to be particularly important in experimental conditions with more complex problem structures, such as detour or restructuring.

It is believed that if a group spends sufficient time working through the problem verbally, there is a greater chance that the number of card exchanges will be reduced since the verbal discussion would have narrowed down the possible group options. This will also reduce the number of card exchange reversals during the course of the problem solving process. Alternatively, if the group starts to exchange cards immediately with minimal or no verbal discussion, then the card exchange frequency is likely to be high.

This measure was developed to examine the effect on performance of different problem solving strategies adopted (particularly, the impact of early discussion vs. card exchanges) by groups solving the same stimulus task. This phenomenon was tested in this thesis by comparing two forms of strategy (card exchange and verbal discussion) in understanding the structure and solving the problem.

The measure computes and ranks the duration of the time intervals between card exchanges during the problem solving process, with the longest time interval ranked first and the shortest time interval ranked last. In order to normalize these ranks across groups with different numbers of time slots and card exchanges, the average of the first quartile of the ranks was used. For instance, if the group solved the task with 16 exchanges, then they would have eight discussion time intervals. These discussion times are ranked in a decreasing order of magnitude with the longest discussion time ranked as 1. The first quartile of the rank order is computed, which is a quarter of the number of time intervals. For instance, in a group with 8 discussion time intervals, the first quartile of the rank order is 2. The average rank and duration of these two longest intervals are used as the rank index and time component of the analysis respectively (see Figure 17). This value is then correlated with behavioural and performance measures such as number of card exchanges, behavioural area difference, verbal frequency category and verbal reversals to examine the relationship between performance and the time spent in discussion. The correlation value and the degree of significance are used to verify the relationship between discussion time and problem solving performance and behaviour. The following hypotheses explain the effect of task structure and choice strategy on behaviour.

Figure 17: Excerpt illustrating rank ordering

																		Card Movem
															Task ID: 080205			30th
			Time	Person 1					PB'	PB	AB	Person 2					PB'	PB
SN	Rank Order	Discussion Time	0:29:53	A'1E	B'2H		C'3F	D3C	6	5	5	A2F	B1G		C1I	D1J	6	6
1	6	0:00:34	0:30:27	A'1E	B'2H		C'3F	**B1G**	6	5	5	A2F	**D3C**		C1I	D1J	5	5
2	7	0:00:24	0:30:51	A'1E	B'2H		C'3F	B1G	6	5	5	**C4B**	D3C		C1I	D1J	3	3
3	16	0:00:05	0:30:56	A'1E	B'2H		C'3F	B1G	6	5	5	C4B	D3C		C1I	**C2J**	0	0
4	13	0:00:10	0:31:06	A'1E	B'2H		C'3F	B1G	6	5	5	C4B	D3C		C1I	C2J	0	0
5	1	0:04:19	0:35:25	A'1E	B'2H		C'3F	B1G	6	5	5	C4B	**B3I**		C1I	C2J	3	3
6	11	0:00:16	0:35:41	A'1E	B'2H		**B4A**	B1G	5	3	3	C4B	B3I		C1I	C2J	3	3
7	10	0:00:17	0:35:58	A'1E	B'2H		B4A	B1G	5	3	3	C4B	B3I		C1I	C2J	3	3
8	3	0:01:10	0:37:08	A'1E	B'2H		**A3G**	B1G	5	4	4	C4B	B3I		C1I	C2J	3	3
9	2	0:01:21	0:38:29	A'1E			A3G	B1G	2	2	2	C4B	B3I		C1I	C2J	3	3
10	12	0:00:12	0:38:41	A'1E	**D1J**		A3G	B1G	5	5	5	C4B	B3I		C1I	C2J	3	3
11	13	0:00:10	0:38:51	A'1E	D1J		A3G	**C2J**	5	4	5	C4B	B3I		C1I	**B1G**	4	4
12	9	0:00:20	0:39:11	**A4D**	D1J		A3G	C2J	4	4	5	C4B	B3I		C1I	B1G	4	4
13	8	0:00:22	0:39:33	A4D	D1J		A3G	C2J	4	4	5	C4B	B3I		**B4A**	B1G	0	0
14	4	0:00:55	0:40:28	A4D	D1J		A3G	C2J	4	4	5	C4B	B3I		B4A	B1G	0	0
15	5	0:00:35	0:41:03	A4D	**A2F**		A3G	C2J	3	3	3	C4B	B3I		B4A	B1G	0	0
16	15	0:00:09	0:41:12	A4D	A2F		A3G	C2J	3	3	3	C4B	B3I		B4A	B1G	0	0

3.10.1 Hypotheses of the effect of choice of strategy on performance or behavioural variety

Hypothesis 6:

H_6: The greater the initial discussion time spent solving a task, the lower the behavioural variety

Controlling for problem structure, task performance varies based on the adopted problem solving strategy. The greater the time spent at the early stages of the problem solving process learning about the structure through verbal discussion, the quicker the understanding across group members and the lower the behavioural variety.

4 METHODOLOGY

The methodology consisted of two experiments as described in the subsequent sections.

4.1 Experiment I

The section provides a detailed and more comprehensive description of the experimental procedures for the first experiment, which investigates the effect of structure on performance and behavioural variety.

4.1.1 Sample

One hundred and sixty eight subjects (42 Groups of 4 subjects each) participated in this experiment; twelve were graduate students who were used for the pilot study. Two graduate students and one hundred and fifty four undergraduate students participated in the actual experiment. The graduate students were paid for their participation while the undergraduate students (excluding one) were compensated with three bonus marks in an organizational behaviour undergraduate course.

The undergraduate subjects were first year to fourth year students. Most were engineering students with a few majoring in mathematics, arts and environmental sciences. The tasks were theoretically unrelated to educational background and thus allowed for general participation in the laboratory experiment. Participation in this experiment was voluntary and students had the option of completing alternative course work for the same bonus marks.

The experiment was set up in the Uncertainty Lab within the Management Sciences department of University of Waterloo. The room is equipped with five digital cameras and three suspended microphones connected to a multiplexer and digital encoder for capturing video and voice data from participants. The participants were made aware of the use of recording equipment prior to the start of the experiment and signed consent agreements. Each experimental group comprised of four subjects who either signed up for the same time slot or were randomly assigned.

4.1.2 Stimulus Set

Twelve different stimulus conditions were used in the experiment (Training 1, Training 2, B, C, D, E, F, G, H, I, J and K). The twelve stimuli differed from one another based on structural characteristics as defined in Table 1. Each experimental group received the same two sets of training stimuli plus a combination of three of the other stimuli sets in a random order (e.g. BCD, BDI, CHK, and DHJ). These combinations resulted in 14 stimulus B experiments, 15 stimulus C experiments, 14 stimulus D experiments, 17 stimulus E experiments, 13 stimulus H experiments, 11 stimulus I experiments, 11 stimulus J experiments and 13 stimulus K experiments. All stimuli sets were created using Microsoft PowerPoint, printed on matte finish paper using a coloured printer and made into approximately 9.5 x 12.5 cm cards as shown in Appendix A.

4.1.3 Procedure

Each of the twelve stimulus conditions contained sixteen cards with two objects on each card as previously described. There were three different versions of each stimulus condition, referred to as design sets I, II, III as illustrated in Appendix A. Design set I had birds, cars, chairs and rectangles in the main categories and hockey items as the detour category. Design set II had facial parts, computer parts, snow item and medical people as the main categories and church items as the detour category. Design set III had fruits, numbers, sports and utensils as the main categories and musical instruments as the detour category. Although, the design sets had different objects on the cards, the structural property for each stimulus condition is identical. Three design sets were used such that each experimental group of participants would solve three different conditions using different design sets. Thus groups could not transfer knowledge of the solution categories from one task condition to the next. In addition, each group performed two training design sessions (training I and II) prior to starting the three experimental conditions. The training sets had four homogeneous categories with no other structural properties (stimulus version A) and data from the training sessions are not included in this analysis. Results from stimuli sets F (heterogeneous with restructuring, detour and redundancy) and G (homogeneous with restructuring, detour and redundancy) were also not included in the analysis. This is because they resulted in approximately 18% task completion (due to excessive difficulty) after administering these stimuli sets to six groups of participants. The compiled data from eight stimuli sets were analyzed and the results are presented for this research work.

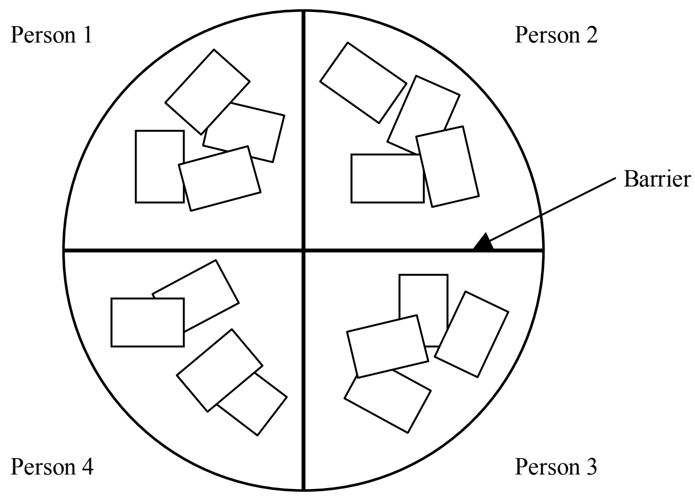
Upon arrival at the lab, the group members were introduced to each other (in cases where they were complete strangers) and nametags were provided for identification purposes.

Students were informed prior to the start of the experiment that all sessions would be video recorded. Once the preliminary introduction was completed, students were provided with a set of verbal instructions (Appendix C) and clarifications as required. The participants were asked not to disclose the information about the experiment to other potential subjects or to their classmates since the experiment was being conducted over a period of 3 weeks. After the completion of each task, subjects were asked to fill a short questionnaire to describe some of their experiences. The questionnaire questions include rating the difficulty of the task, specifying the source of difficulty and identifying the leader of the group if any. In total, each group completed two training sessions at the start of the experiment, three experimental conditions, and three short questionnaires. All students were debriefed after the completion of all experimental studies by explaining the aim of the study and their assigned marks via email.

The first training session used stimulus 1 (Appendix A). In this training, cards were initially distributed such that each individual was provided with two identical cards to provide a hint on how to form card categories (i.e., one item per card used as the basis for grouping). In addition to this, all participants were allowed to see each other's cards without any barriers on the table, so that subjects would understand the experimental goals. They were allowed to ask questions and the experimenter provided them with feedback on the correct categories during the training sections.

During the second training session, the experimenter introduced a barrier across the table (a wooden cross of about 48 inches long for each slab and 12 inches high) as shown in Figure 18. The barrier allowed subjects to see one another and exchange cards easily, but prevented them from seeing one another's cards. The aim was to encourage the participants to discuss or verbalize their thoughts during the problem solving process.

Figure 18: Overhead view of the Table showing demarcating structure



Participants in the second training were given homogeneous but non-identical cards. They had to describe their cards in order to find out what other group members had, which took slightly longer than the first training session. The procedure used in the second training was identical to that used in the actual experimental conditions except that subjects were not required to complete a questionnaire afterwards.

In the actual experiments, each individual was told they would receive up to three bonus marks in the course for their participation. They were told they would receive half a mark for each of the three experiments if they individually were able to come up with four of a kind. They would receive an additional half mark per experiment if every other member of the group was also able to come up with their 4 of a kind. Thus, an individual could receive 0 marks if (s)he did not get four of a kind in any of the three conditions, 1.5 marks if (s)he got four of a kind on all three conditions, and the full 3 marks if (s)he and the other members of the group got four of a kind on all three conditions. However, during debriefing afterwards all participants were actually given three full marks regardless of their individual or group performance.

4.1.3.1 Experimental constraints

In order to control this experiment, the following constraints were introduced:

1. The distribution of the initial set was held constant such that subject 1 always received specific cards {i.e. 1, 6, 11, 15}; subject 2 received cards {2, 5, 9, 13}; subject 3 received cards {3, 7, 10, 14} and subject 4 received cards {4, 8, 12 and 16}. This invariably led to each person having at least one card for each potentially correct set. For instance, the first correct set is {1, 2, 3 and 4} and this card set is distributed among the four participants.

2. A demarcating wooden structure (barrier) was introduced for three reasons:
   i. To prevent any participant from having full task knowledge
   ii. To allow each individual take responsibility for their cards and decision on the choice of card exchange
   iii. To force subjects to verbalize their thoughts during the problem solving process, enabling subsequent analysis of their verbal communication behaviour.

3. There were no time constraints in solving the tasks except when requested by the participants themselves to accommodate other commitments. Most experiments were completed in less than one hour, though some lasted a bit longer.

4. All participants were allowed to exchange cards with whomever and in whatever form they chose since there were no restrictions on communication patterns.

5. Individual participants were permitted to hold as many as five cards at once, but no less than one card at any time while problem solving.

4.2 Experiment II

The second experiment served three purposes:

1. To serve as a validation tool to test the variation of card items and groups used in the experimental study.
2. To measure and test the effect of the level of abstraction.
3. To understand the effect of the degree of homogeneity on abstraction.

The detailed description of the experiment and findings is presented in the following sections.

4.2.1 Sample

Nine students participated in this experiment. All participants were graduate students in the faculty of Engineering. Five of them had previously participated in the pilot studies for Experiment I and one had participated in Experiment I. Participation was voluntary and there was no compensation or benefit in any form.

4.2.2 Questionnaire

A questionnaire was designed to capture the similarity among category objects in Experiment I (see Appendix D). All card items that were part of the correct solution sets of the first experiment were printed individually. The first card item was labeled as the "anchor" and the others labeled as items I-VI (see Appendix C). The participants were required to rate the degree of similarity between the anchor card and each of the other cards in the group on a scale of 1-10. Identical items were graded as 10 and very different items (on the other extreme) were graded as 1. All cards were created using Microsoft PowerPoint, printed on matte finish paper using a colored printer and made into approximately 9.5 x 12.5 cm cards as shown in Appendix C. Each card had only one object with the exact same size as the object used in the first experiment to avoid any bias.

4.2.3 Procedure

Nine participants were selected randomly and asked if they would be willing and available to participate in a short study that would last approximately 45 minutes. Once the participants' time and availability were confirmed, they were given a questionnaire and 15 envelopes. Each set of cards was placed in a separate envelope. There was no explanation of why different card items were grouped together to prevent bias, as participants were required to judge the similarity of items based on their own chosen criteria. The documents were left with the subjects who were required to return them after completion. Due to the restriction that only a single set of cards were designed, the study was conducted individually except when two participants shared the same office space so that envelopes of cards can be swapped as required. In 90% of the cases, the researcher was not in the same room as the subjects to avoid any bias while trying to explain the meaning of similarity. The questionnaire was worded in a way that would be easy to comprehend (see Appendix D). For instance, in the body parts category, six cards comprised of lips, eye, nose, mouth, ear and hand were labeled from I – VI and the nose was also identified on another card as the anchor item. Each subject then rated the perceived similarity of the objects on the cards to the anchor object with one extreme as very different and the other as identical. The questionnaires were returned to the researcher after completion of the rating task.

5 RESULTS

5.1 Measures of task structure

This chapter presents the findings of the experiments based on the measures defined in Chapter 3. It also provides additional information to analyze the postulated hypotheses.

5.1.1 Degree of Complexity

The results obtained from the experiments show the effects of task structure and strategy on performance and behavioural variety. Due to the unavailability of a universal measure of structure, the degree of structural complexity of the task was operationalized using the following two measures: the number of pairs of card items and the number of structural dimensions in a given stimulus. The number of pairs was calculated using the formula $\left({}^{n}_{2}C\right)$ while the number of structural dimensions was obtained from description in section 3.5.2. These results are presented in the Table 3 below:

Table 3: Task Structure Measures

Stimulus	Number of Pairs	Number of Structural dimensions
B - Basic	32	1
C – Detour	36	2
D – Redundancy	46	2
E – Detour with Redundancy	50	3
H – Heterogeneous Detour	60	3
I – Heterogeneous Redundancy	76	3
J – Homogeneous Detour	60	4
K – Homogeneous Redundancy	76	4

5.1.2 Degree of Abstraction or Restructuring

The level of abstraction and degree of homogeneity were not captured explicitly by the measures listed above. The inability to properly operationalize these variables led to the design of experiment II. The results of the abstraction t-test measure (see section 3.5.3) are summarized in Table 4. The t-tests measure the degree of perceived similarity between the anchor item (i.e., the identical items used in each homogeneous category; e.g., swan) and the other items in the corresponding heterogeneous category (e.g., birds), and between the anchor item (e.g., swan) and the restructured item (e.g., iguana) for each category. The results indicate that in 83.3% of the cases (i.e., 10 out of 12 categories) the heterogeneous items were perceived as significantly different from restructured items ($p<0.005$) and in 100% of the cases (all 12 categories) the homogeneous anchor items were perceived as significantly different from the restructured items (with $p<0.005$).

This experiment is mainly a test of the validity of the experimental manipulations under the restructuring conditions. The results help to establish the following facts:

1. That shifting to a higher level of category abstraction amounts to some degree of cognitive restructuring, since the "restructured" items are perceived as significantly different from the other basic level items in the sets.
2. That homogeneous sets are in fact more homogeneous than heterogeneous sets.
3. That restructured items are perceived as relatively more dissimilar from items in homogeneous sets than they are from items in heterogeneous sets, so restructuring should be more difficult for homogeneous conditions.

These results also provide empirical validation to back up the theoretical argument on which the homogeneity hypothesis 2b and 4b are based. That is, these analyses do not constitute a test of H2b or H4b, since those hypotheses deal with the relative difficulty of problem restructuring under homogeneous and heterogeneous conditions, not perceptual similarity judgments. Rather, these results help to justify why H2b and H4b were postulated in the first place.

Table 4: T-test showing varying degree of abstraction

Questionnaire T-test

S/N	Category	Objects	Participants										
			1	2	3	4	5	6	7	8	9	Heterogeneous T-test	Homogeneous T-test
1	Animals	**swan****	10	9	10	10	9	7	10	10	7	t = -2.98 stdev=2.64 df = 52 p=0.0044 Reject Null	t = -5.20 stdev=2.36 df = 16 p = 0.0001 Reject Null
		dove	7	5	5	3	8	4	8	3	8		
		chicken	6	5	6	4	6	4	9	3	7		
		ostrich	8	6	7	7	4	6	9	4	2		
		*Iguana**	2	1	1	1	2	2	4	8	9		
		penquin	4	2	2	6	7	4	7	1	10		
2	Vehicle	truck	7	6	9	8	8	6	7	8	10	t = -5.82 stdev=1.97 df = 52 p=0.0001 Reject Null	t = -5.99 stdev=2.04 df = 16 p = 0.0001 Reject Null
		Car	9	8	9	9	8	7	8	9	7		
		Sports car	7	6	7	6	6	4	6	7	6		
		*helicopter**	3	1	1	2	7	2	4	2	8		
		pickup	8	7	9	7	4	6	9	7	1		
		limousine**	10	9	10	10	9	8	10	10	6		
3	Furniture	high chair	7	7	9	9	7	6	8	2	8	t = -7.46 stdev=2.05 df = 52 p=0.0001 Reject Null	t = -17.8 stdev=0.928 df = 16 p = 0.0001 Reject Null
		lawn chair	7	6	9	8	8	6	7	2	7		
		office chair**	10	9	10	10	9	8	10	10	10		
		wooden chair	9	8	9	9	7	6	9	4	9		
		stool	7	5	2	7	7	5	8	3	8		
		*coat rack**	1	1	1	1	4	2	3	1	2		
4	Shape	*triangle**	2	1	2	7	6	6	6	3	1	t = -4.87 stdev=1.95 df = 52 p=0.0001 Reject Null	t = -6.70 stdev=1.76 df = 16 p = 0.0001 Reject Null
		blue rectangle	8	7	3	8	7	7	9	6	6		
		multi many rec	3	5	2	7	5	6	7	5	5		
		grey rectangle**	10	9	10	9	9	9	10	9	9		
		brown rectangle	7	6	7	8	6	8	8	7	6		
		black rectangle	9	8	8	8	8	8	9	8	7		
5	Body parts	lips	8	8	1	6	7	7	7	8	7	t = -1.68 stdev=2.60 df = 52 p=0.098 Accept Null	t = -4.16 stdev=2.15 df = 16 p=0.0007 Reject Null
		eye	8	8	1	7	7	7	9	8	7		
		Nose**	10	9	10	10	9	9	10	10	10		
		mouth	2	7	1	8	7	5	8	8	3		
		ear	8	5	1	8	7	7	9	8	7		
		*hand**	7	1	1	7	3	6	9	8	7		
6	Telecommunication	printer	7	6	1	6	6	6	6	4	2	t = -0.305 stdev=2.79 df = 52 p=0.76 Accept Null	t = -3.73 stdev=2.15 df = 16 p=0.0018 Reject Null
		phone cables**	10	9	10	10	9	8	10	10	10		
		diskettes	7	1	1	4	7	6	8	4	2		
		webcam	7	5	1	8	8	6	6	8	8		
		*satellite**	2	4	1	7	8	5	9	9	7		
		keyboard	7	5	1	7	8	6	6	5	2		
7	Weather	**mittens****	9	9	10	10	9	9	10	10	10	t = -4.07 stdev=2.59 df = 52 p=0.0002 Reject Null	t = -7.76 stdev=1.82 df = 16 p=0.0001 Reject Null
		snowboard	4	5	1	8	7	6	7	8	5		
		snowman	1	6	3	9	7	7	9	7	6		
		*woman in the rain**	1	1	1	3	8	3	6	2	1		
		snow mountain	1	6	2	8	6	5	9	8	6		
		boy ice skating	1	8	5	9	7	7	8	8	7		
8	Profession	medical team	5	7	9	9	8	8	6	9	7	t = -9.47 stdev=1.61 df = 52	t = -10.9 stdev=1.38 df = 16
		optician	7	7	7	9	8	8	8	9	3		
		surgeon	7	5	9	9	8	8	9	9	7		

		pediatrics	5	8	9	9	7	7	6	9	9	p=0.0001 Reject Null	p=0.0001 Reject Null
		camera man*	4	1	1	1	6	3	4	1	1		
		doctor**	9	9	10	10	9	9	10	10	10		
9	**Food**	pear	5	7	9	9	7	7	8	9	4	t = -8.83 stdev=1.63 df = 52 p=0.0001 Reject Null	t = -11.2 stdev=1.28 df = 16 p=0.0001 Reject Null
		watermelon	5	5	9	8	7	7	8	9	4		
		orange	5	5	9	8	7	8	8	9	7		
		basket of fruit	7	8	9	9	8	7	9	9	9		
		orange + banana**	9	9	10	10	9	7	10	10	10		
		fish + lemon*	2	1	1	3	4	4	5	1	2		
10	**Alphanumeric**	letter H*	5	1	1	8	6	9	4	1	1	t = -3.99 stdev=2.20 df = 52 p=0.0002 Reject Null	t = -3.74 stdev=2.77 df = 16 p=0.0018 Reject Null
		number 1	8	8	10	9	8	8	9	6	8		
		number 1	7	7	9	8	7	7	7	6	5		
		roman number 5	6	6	4	8	7	6	6	5	2		
		number 9**	9	9	10	10	9	3	10	10	10		
		911	4	6	6	8	8	5	5	8	7		
11	**Competition**	running	5	5	9	7	7	5	5	4	3	t = -3.31 stdev=2.12 df = 52 p=0.0017 Reject Null	t = -4.53 stdev=2.13 df = 16 p=0.0003 Reject Null
		trophy*	2	1	8	7	6	9	4	7	2		
		arm wrestling	8	7	9	9	8	8	8	8	4		
		tug of war**	10	10	10	10	9	8	10	10	10		
		pulling rope	9	9	9	9	8	7	9	9	9		
		fencing	7	7	9	8	8	5	6	7	4		
12	**Utensils**	Knife**	10	9	10	10	9	9	10	10	10	t = -3.92 stdev=2.14 df = 52 p=0.0003 Reject Null	t = -6.84 stdev=1.69 df = 16 p=0.0001 Reject Null
		dishes	4	6	5	9	8	7	7	8	4		
		wine glasses	2	4	6	8	8	5	7	7	4		
		cutleries	8	8	8	9	8	3	9	8	8		
		colander*	2	2	7	8	6	3	5	2	3		
		chopsticks	6	7	8	9	7	3	8	8	7		

<u>**Note:**</u> Anchor cards are in bold**. Restructured items are in Italics *. The rating scale ranged from 1 (very different) to 10 (identical).

5.2 Results from performance measures

Problem solving performance is operationalized in terms of three measures: the time taken to solve the problem, the number of card exchanges required to solve the problem, and the number of cards exchanged per unit time.

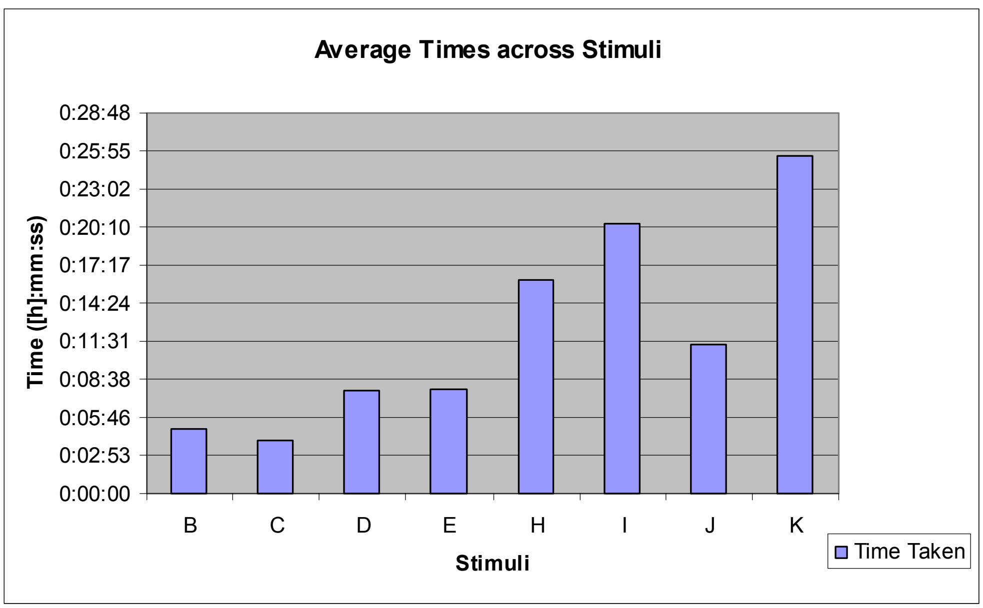
5.2.1 Time as a performance measure

The time taken to complete the task is the difference between the start time and the completion time. Table 5 shows the results of the average times across the 8 stimuli and Figure 19 provides a graphical representation of the same results. The results indicate that the average time taken for any of the tasks was less than thirty minutes.

Table 5: Average Time across Stimuli

Stimulus	Average Time (hh:mm:ss)	No. of Groups *Sample Size (n)*	Standard deviation (hh:mm:ss)
B – Basic	00:04:53	14	00:06:01
C - Detour	00:03:58	15	00:01:12
D - Redundancy	00:07:43	14	00:03:19
E – Detour and Redundancy	00:07:53	17	00:03:48
H – Heterogeneous Detour	00:16:13	13	00:02:04
I – Heterogeneous Redundancy	00:20:23	11	00:04:45
J – Homogeneous Detour	00:11:19	11	00:02:41
K – Homogeneous Redundancy	00:25:35	13	00:10:47

Figure 19: Bar Chart showing Average Time across Stimuli



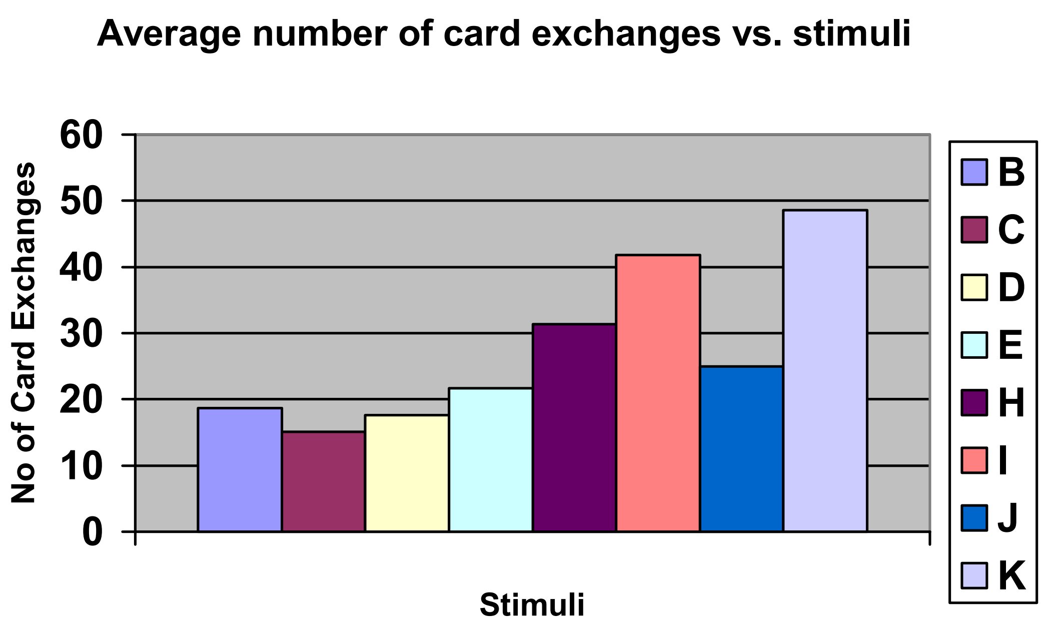
5.2.2 Using card exchanges to measure performance

The number of card exchanges required to complete each task is shown in Table 6. Figure 20 provides a graphical representation of Table 6.

Table 6: Average Card Exchanges across Stimuli

Stimulus	Average Card Exchanges	No of Groups *Sample size (n)*	Standard Deviation
B – Basic	18.64	14	15.84
C – Detour	15.07	15	6.41
D - Redundancy	17.64	14	18.80
E – Detour and Redundancy	21.71	17	11.39
H – Heterogeneous Detour	31.31	13	27.22
I – Heterogeneous Redundancy	41.82	11	35.92
J – Homogeneous Detour	25	11	18.44
K – Homogeneous Redundancy	48.54	13	39.26

Figure 20: Bar chart showing average number of card exchanges across stimuli



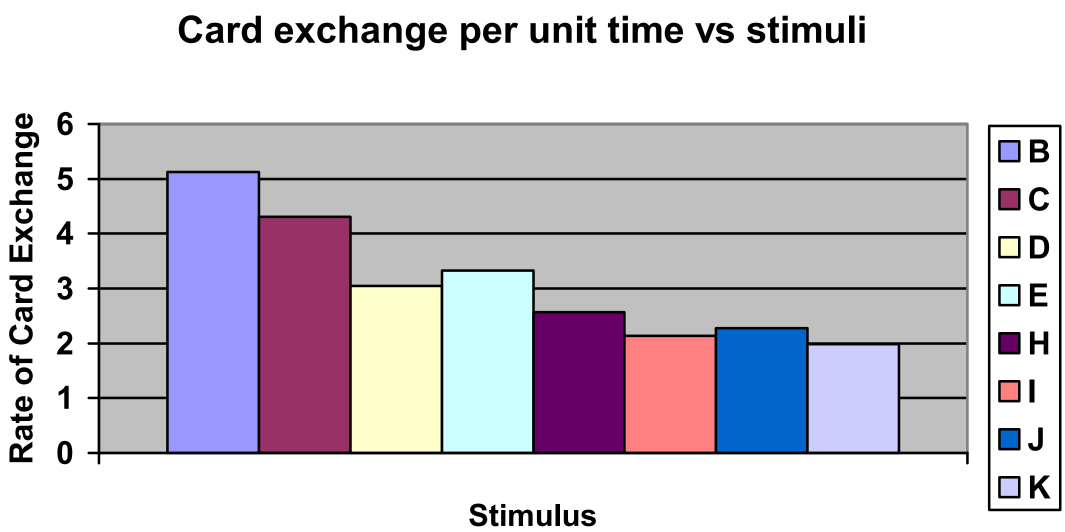
5.2.3 Rate of card exchange as a measure of performance

The rate of card exchanges (i.e., card exchanges per minute) provides an indication of perceived problem solving difficulty. If the task structure is easy to comprehend, then participants are expected to exchange cards quickly while working towards the goal. Card exchange rate is expected to drop when participants find the task structure difficult to comprehend. Hence the ease of movement may be a reflection of the degree of perceived difficulty of the task and groups' relative understanding of the problem structure. Table 7 and Figure 21 below show the distribution of this variable across all stimuli sets.

Table 7: Card exchange per unit time

Stimulus	Average Card Exchange per unit time (per min)	No. of Groups *Sample size (n)*	Standard deviation
B – Basic	5.12	14	2.68
C - Detour	4.31	15	1.46
D - Redundancy	3.05	14	1.21
E – Detour and Redundancy	3.33	17	1.56
H – Heterogeneous Detour	2.57	13	2.16
I – Heterogeneous Redundancy	2.14	11	1.47
J – Homogeneous Detour	2.28	11	0.59
K – Homogeneous Redundancy	1.98	13	1.30

Figure 21: Card exchange per unit time across stimuli



5.3 Results from behavioural variety

Behavioural variety is the second dependent variable in addition to performance. The two forms of behavioural variety are examined in this analysis: *card exchange behaviour* and *verbal behaviour*.

5.3.1 Card exchange measures

Two card exchange measures are used in this experimental study:

1. The area between objective and subjective balance in card exchanges
2. The number of card exchange reversals in subjective balance

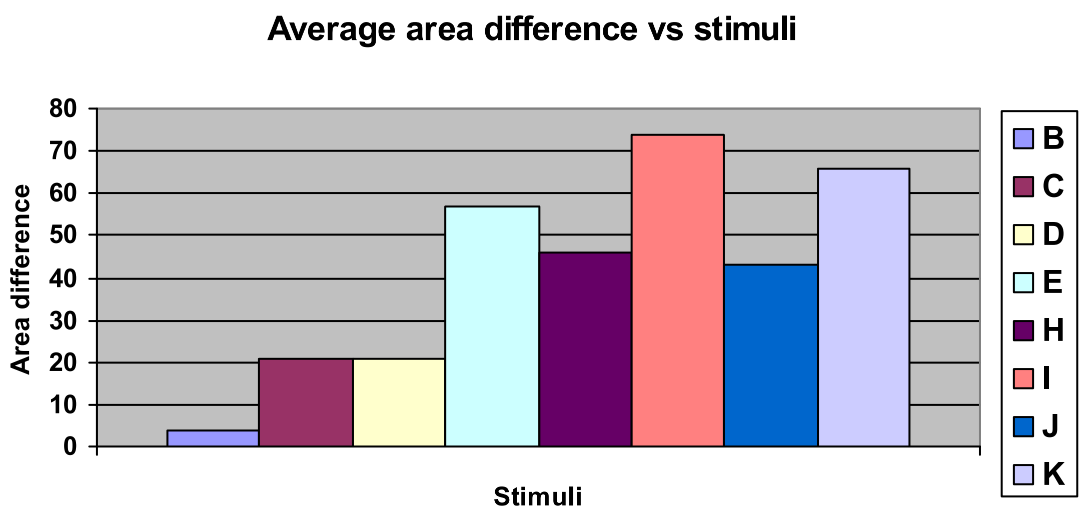
5.3.1.1 Area between objective and subjective balance

This measure shows the difference in calculated balance between the objective (TAB) and subjective (TPB) views of the stimulus. The results of the differences are presented in Table 8 and Figure 22 below.

Table 8: Area difference between objective and subjective Balance

Stimulus	Average Area difference between objective and subjective balance	No. of Groups *Sample Size (n)*	Standard deviation
B – Basic	3.79	14	4.52
C - Detour	21.27	15	22.95
D - Redundancy	20.64	14	26.97
E – Detour and Redundancy	55.18	17	52.92
H – Heterogeneous Detour	45.54	13	61.40
I – Heterogeneous Redundancy	74.83	11	82.54
J – Homogeneous Detour	43.18	11	34.66
K – Homogeneous Redundancy	66.08	13	51.11

Figure 22: Average area difference across stimuli



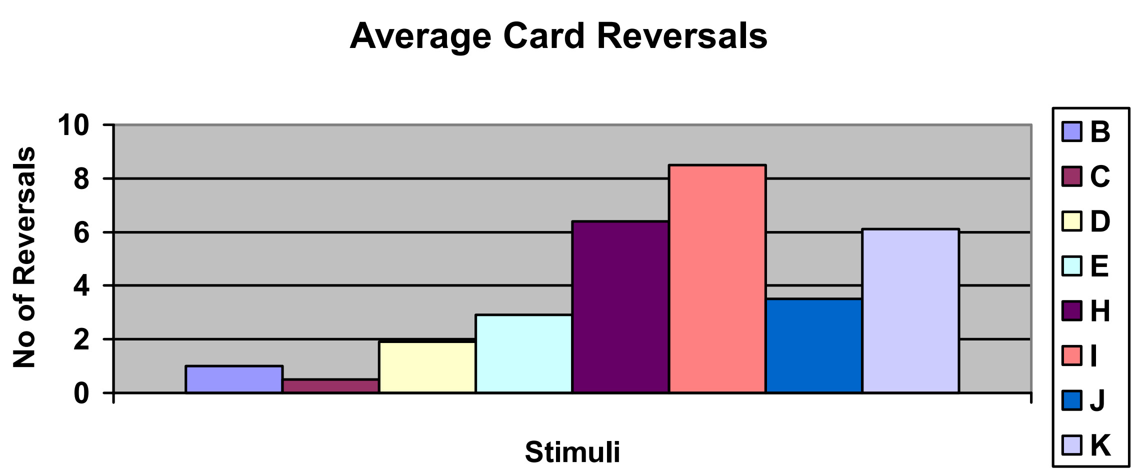
5.3.1.2 Card exchange reversals

Table 9 and Figure 23 below show the average number of card exchange reversals encountered during problem solving. The easier the task, the fewer balance reversals are expected.

Table 9: Average Card Exchange Reversals

Stimulus	Average Card Reversals	No. of groups *Sample size (n)*	Standard deviation
B – Basic	1.0	14	2.18
C - Detour	0.5	15	1.13
D - Redundancy	1.9	14	2.41
E – Detour and Redundancy	2.9	17	3.39
H – Heterogeneous Detour	6.4	13	7.58
I – Heterogeneous Redundancy	8.5	11	11.67
J – Homogeneous Detour	3.5	11	3.62
K – Homogeneous Redundancy	6.1	13	5.52

Figure 23: Average Card Reversals across Stimuli



5.3.2 Verbal Protocol Measure

A group's verbal communication provides an indication of the constraints encountered as groups search the problem space while trying to understand the problem structure. Similar to the card exchanges, the discussions are based on the participants' perceived understanding of the problem structure and path towards the solution. Verbal protocol was measured using two methods:

1. Number of unique categories mentioned
2. Number of verbal reversals

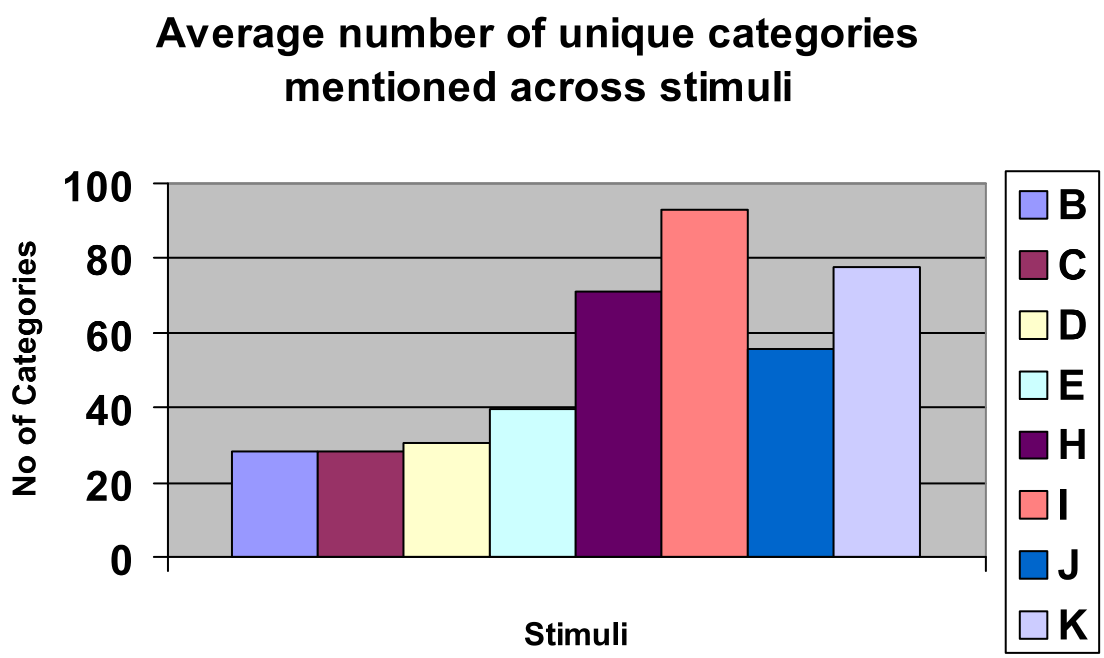
5.3.2.1 Number of unique categories mentioned

This is a measure of the total number of distinct category names mentioned by group participants during the problem solving process. The results are summarized in Table 10 and Figure 24 below.

Table 10: Average number of unique verbal categories mentioned

Stimulus	Average number of unique categories mentioned	No. of Groups *Sample Size (n)*	Standard deviation
B – Basic	28.21	14	12.29
C - Detour	28.4	15	13.90
D - Redundancy	30.53	14	8.54
E – Detour and Redundancy	39.82	17	14.43
H – Heterogeneous Detour	71	13	32.82
I – Heterogeneous Redundancy	92.91	11	29.80
J – Homogeneous Detour	55.45	11	28.56
K – Homogeneous Redundancy	77.77	13	15.19

Figure 24: Average Number of unique categories mentioned



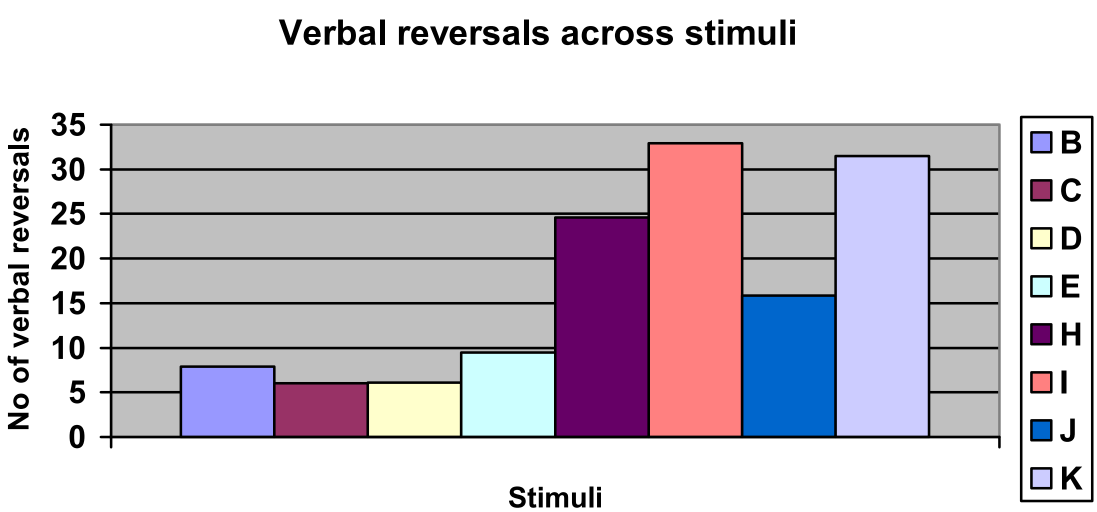
5.3.2.2 Number of verbal reversals

Verbal reversals are similar to card exchange reversals. The difference is that the card exchange measure tracks reversals in subjective balance while verbal reversals tracks changes the number of times the conversation topic switches from discussing the correct categories to discussing the incorrect categories, and vice versa, during the problem solving process. The results are presented below in Table 11 and Figure 25:

Table 11: Verbal reversals across stimuli

Stimulus	Number of Verbal Reversals	No. of Groups Sample size (n)	Standard deviation
B – Basic	7.86	14	10.18
C - Detour	6	15	4.90
D - Redundancy	6.13	14	6.42
E – Detour and Redundancy	9.47	17	6.82
H – Heterogeneous Detour	24.62	13	22.57
I – Heterogeneous Redundancy	32.92	11	14.87
J – Homogeneous Detour	15.82	11	10.96
K – Homogeneous Redundancy	31.46	13	12.00

Figure 25: Average number of Verbal Reversals across Stimuli



5.4 Tests of Hypotheses

This section tests the postulated hypotheses based on the results reported above. The main focus of the study is the influence of problem structure and problem solving strategy on problem solving behaviour and performance. The majority of hypotheses are tested using pair-wise comparisons between two different experimental conditions due to orthogonal relationships among the various structural dimensions included in the stimuli. For instance, in testing the effects of abstraction, a pair-wise comparison was made between stimulus C (detour) and stimulus E (detour and abstraction). The tables below report ANOVA results comparing the various performance and behavioural measures across the different stimuli and also spearman rank correlations among the various measures. All 108 problems solved by the various groups in the study were included in these analyses to provide a general overview of the differences between experimental conditions and the relations among variables.

Table 12: One Way ANOVA

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Average time	Between Groups	20234543	7	2890649.068	10.801	.000
	Within Groups	26763375	100	267633.753		
	Total	46997919	107			
No of card exchanges	Between Groups	13572.841	7	1938.977	3.826	.001
	Within Groups	50677.234	100	506.772		
	Total	64250.074	107			
Card exchange rate	Between Groups	116.895	7	16.699	5.922	.000
	Within Groups	281.965	100	2.820		
	Total	398.860	107			
Difference in balance area	Between Groups	55862.367	7	7980.338	3.740	.001
	Within Groups	213378.4	100	2133.784		
	Total	269240.8	107			
Card exchange reversals	Between Groups	743.000	7	106.143	3.677	.001
	Within Groups	2886.667	100	28.867		
	Total	3629.667	107			
No of verbal categories mentioned	Between Groups	56746.579	7	8106.654	19.449	.000
	Within Groups	41681.300	100	416.813		
	Total	98427.880	107			
No of verbal reversals	Between Groups	11344.581	7	1620.654	11.398	.000
	Within Groups	14219.160	100	142.192		
	Total	25563.741	107			

Table 13: Correlations between independent and dependent variables

Correlations

			No of pairs	No of structural dimensions	Average time	No of card exchanges	Card exchange rate	Difference in balance area	Card exchange reversals	No of verbal categories mentioned	No of verbal reversals
Spearman's rho	No of pairs	Correlation Coefficient	1.000	.885**	.735**	.577**	-.567**	.616**	.549**	.782**	.682**
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000	.000
		N	108	108	108	108	108	108	108	108	108
	No of structural dimensions	Correlation Coefficient	.885**	1.000	.628**	.485**	-.498**	.609**	.472**	.653**	.569**
		Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000	.000	.000
		N	108	108	108	108	108	108	108	108	108
	Time	Correlation Coefficient	.735**	.628**	1.000	.773**	-.745**	.692**	.728**	.882**	.874**
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000	.000
		N	108	108	108	108	108	108	108	108	108
	No of card exchanges	Correlation Coefficient	.577**	.485**	.773**	1.000	-.233*	.826**	.857**	.687**	.693**
		Sig. (2-tailed)	.000	.000	.000	.	.015	.000	.000	.000	.000
		N	108	108	108	108	108	108	108	108	108
	Card exchange rate	Correlation Coefficient	-.567**	-.498**	-.745**	-.233*	1.000	-.273**	-.267**	-.680**	-.634**
		Sig. (2-tailed)	.000	.000	.000	.015	.	.004	.005	.000	.000
		N	108	108	108	108	108	108	108	108	108
	Difference in balance area	Correlation Coefficient	.616**	.609**	.692**	.826**	-.273**	1.000	.762**	.597**	.575**
		Sig. (2-tailed)	.000	.000	.000	.000	.004	.	.000	.000	.000
		N	108	108	108	108	108	108	108	108	108
	Card exchange reversals	Correlation Coefficient	.549**	.472**	.728**	.857**	-.267**	.762**	1.000	.602**	.646**
		Sig. (2-tailed)	.000	.000	.000	.000	.005	.000	.	.000	.000
		N	108	108	108	108	108	108	108	108	108
	No of verbal categories mentioned	Correlation Coefficient	.782**	.653**	.882**	.687**	-.680**	.597**	.602**	1.000	.889**
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.	.000
		N	108	108	108	108	108	108	108	108	108
	No of verbal reversals	Correlation Coefficient	.682**	.569**	.874**	.693**	-.634**	.575**	.646**	.889**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.
		N	108	108	108	108	108	108	108	108	108

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The ANOVA table in Table 12 shows that there are significant differences across all stimuli sets for each of the behavioural and performance measures, at or above the 0.001 level. The results show that there are structural differences across each of the stimulus type for each dependent measure. It confirms the general hypothesis (or research question) of this thesis, which state that task structure affects performance and behavioural variety. It thus provides a basis for further test to verify each of the specific hypotheses and confirm the effect of each structural type on the dependent variables.

In addition to the ANOVA test, the correlation results in Table 13 show more specific relationship across the different measures. It shows a high correlation between the two structural measures (no of pairs and number of structural dimensions), which implies that these variables are measuring similar effects in task structure. Also, the result provides an overview of the relationship between the structural measures and the dependent measures.

For instance, it shows that there is a positive relationship between each of the structural measures and the dependent measures (apart from the rate of card exchange). It implies that each of the dependent measure is expected to increase as the structural measure increase. This implies that time and no of card exchanges are expected to increase as task complexity increases while the rate of card exchange decreases. This is intuitively correct and in line with the hypothesis, which states that, the performance and behavioural variety are expected to increase with structural complexity. Similarly, the introduction of restructuring (which is perceived as more complexity) is expected to result in lower performance measure and behavioural variety. Finally, if homogeneity is considered more complexity (as the hypothesis imply), then it should result in lower performance measure and behavioural variety as well.

The ANOVA and correlation result provide a general and high-level overview to confirm the structural effects on performance and behavioural measures and act as a premise for further analysis and testing of the hypotheses in this thesis.

5.4.1 Test of Hypothesis 1

Hypothesis 1:

H_1: Problem solving performance varies inversely with structural complexity.

Hypothesis 1 deals with the effect of task complexity on performance. As indicated in earlier sections, the three different performance measures adopted in this thesis are the time taken, the number of card exchanges and the card exchange rate. Less complex tasks are expected to result in better performance and vice versa.

To test this hypothesis, stimulus B (basic) was compared to stimulus D (redundancy) and stimulus C (detour) was compared to stimulus E (detour and redundancy). Stimuli D and E are more complex than stimuli B and C, respectively, based on the complexity measures defined in sections 3.5.1 and 3.5.2. Stimuli D and E are thus expected to result in lower performance measures than B and C respectively. The descriptive statistics in sections 5.2.1, 5.2.2 and 5.2.3 shows the expected effects for the measures in question except for average card exchange.

Student t-tests were performed on these data using SPSS 13.0 and the results are provided in Table 14a and 14b below. In Table 14a, stimulus B is compared to stimulus D using the three performance measures and the t test analysis showed only one significant result in column 3 ($t = 2.637$, $p = 0.017$). The two other measures were non significant with ($t = 0.206$, $p = 0.838$) for card exchanges and ($t = -1.113$, $p = 0.276$) for time taken respectively. These results partially support Hypothesis 1 based on the first measure but not the other two measures.

The lack of support for the Hypothesis can be attributed to two outlier groups that were among the fourteen sets that performed the stimulus B test. One of these two groups experienced English language communication problems in all of their tasks and thus continued exchanging cards thereby resulting in lower performance measures. The second group did not communicate at all at the beginning of the problem, but simply exchanged all sixteen cards among all four group members as a way of sharing information, resulting in 64 card exchanges compared to only 12 for most other groups who performed this task. The analysis in Table 14a1 (which does not include the two outlier groups for stimulus B) showed a statistically significant result with ($t = -2.152$, $p = 0.047$) for time, ($t = -2.173$, $p = 0.048$) for no of card exchanges and ($t = 2.953$, $p = 0.008$) for card exchange rate.

Table 14b compares stimulus C to stimulus E using the same performance measures. The t test analysis shows only differences in solution time were significant at the .05 level ($t = 2.764$, $p = 0.011$). The results for card exchanges ($t = -1.968$, $p = 0.060$) and card exchange rate ($t = 1.829$ $p = 0.077$) were significant at the 0.1 level. Together these results provide partial support for the hypothesis.

One reason for the inconclusive result in this case may be the fact that redundancy as implemented in the experimental stimulus provides groups with an alternate path leading to the solution. As a result, the minimal number of card exchanges required to solve tasks with redundancy is only 10 instead of 12 than for the other stimuli. Thus, even though redundancy adds structural complexity, it may also make the problem relatively easier to solve. Also, this result may be attributed to the small sample size. With a larger sample size the outliers above would have less effect and a smaller signal in the case of redundancy would be picked up as significant.

Table 14: Pair wise independent sample t-test results

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus B (Basic) and Stimulus D (Redundancy)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Time	Equal variances assumed	.658	.425	-1.113	26	.276	-170.14286	152.89301	-484.419	144.13322
	Equal variances not assumed			-1.113	24.967	.276	-170.14286	152.89301	-485.053	144.76707
No of card exchanges	Equal variances assumed	1.533	.227	.206	26	.838	1.000	4.844	-8.957	10.957
	Equal variances not assumed			.206	20.327	.838	1.000	4.844	-9.094	11.094
Card exchange rate	Equal variances assumed	2.528	.124	2.637	26	.014	2.07000	.78485	.45672	3.68328
	Equal variances not assumed			2.637	18.103	.017	2.07000	.78485	.42177	3.71823

a

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus B (Basic) and Stimulus D (redundancy) without including the two outliers						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Time	Equal variances assumed	5.794	.024	-2.014	24	.055	-268.86905	133.47888	-544.356	6.61782
	Equal variances not assumed			-2.152	15.776	.047	-268.86905	124.91471	-533.982	-3.75640
No of card exchanges	Equal variances assumed	10.179	.004	-2.009	24	.056	-5.143	2.560	-10.426	.140
	Equal variances not assumed			-2.173	13.320	.048	-5.143	2.367	-10.244	-.042
Card exchange rate	Equal variances assumed	.865	.362	3.033	24	.006	1.73976	.57364	.55582	2.92370
	Equal variances not assumed			2.953	19.471	.008	1.73976	.58906	.50887	2.97065

a1

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus C (Detour) and E (Detour with Redundancy)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Time	Equal variances assumed	7.309	.011	-2.655	30	.013	-235.16471	88.57407	-416.057	-54.27233
	Equal variances not assumed			-2.764	23.593	.011	-235.16471	85.06630	-410.893	-59.43622
No of card exchanges	Equal variances assumed	5.371	.027	-1.902	30	.067	-6.404	3.368	-13.282	.474
	Equal variances not assumed			-1.968	25.551	.060	-6.404	3.254	-13.099	.291
Card exchange rate	Equal variances assumed	.030	.864	1.822	30	.078	.97631	.53586	-.11805	2.07068
	Equal variances not assumed			1.829	29.868	.077	.97631	.53369	-.11382	2.06645

b

5.4.2 Test of Hypothesis 2a

Hypothesis 2:
H_{2a}: Task performance reduces with the introduction of abstraction or restructuring.

The second hypothesis is similar to the first except that it focuses solely on the abstraction or restructuring dimension of task structure. The restructuring or abstraction feature is implemented in the task by introducing one item into each of the four correct categories that is significantly different from the other objects in the category. The inclusion of these items is expected to make the structure more complicated and the task more difficult to solve since groups are forced to define categories abstractly at the super-ordinate level rather than at the more obvious basic level at which the items are initially perceived.

To test this hypothesis, t-tests used to performance results for stimulus C (detour) were compare to those for stimulus H (detour and restructuring), and to compare results for stimulus D (redundancy) to those for stimulus I (redundancy and restructuring) as shown in Table 15a and 15b below. Stimuli H and I are identical to stimuli C and D respectively, except for the addition of the restructuring characteristic. Thus, stimuli H and I are relatively more complex and were expected to result in lower performance measures. The descriptive statistics in sections 5.2.1, 5.2.2 and 5.2.3 are consistent and generally indicate the expected effects for the measures in question.

T test results comparing performance in stimuli C and H are presented in Table 15a. Results for all three performance measures were significant with $t = 2.101$ ($p = 0.033$) for card exchanges, $t = -2.454$ ($p = 0.023$) for card exchange rate, and $t = 2.882$ ($p = 0.013$) for time taken. These results support the hypothesis that the introduction of abstraction or restructuring as a dimension of the problem structure results in lower problem solving performance.

T test results comparing performance in stimuli D and I are presented in Table 15b. Results were significant for two measures: card exchanges ($t = 2.181$, $p = 0.052$) and time taken ($t = 3.589$, $p = 0.002$). The results were not significant ($t = -1.660$, $p = 0.113$) for rate of card exchange, although the means for the two stimulus conditions did differ in the expected direction.

In summary, these results support the hypothesis that problem solving performance reduces with the introduction of restructuring as a dimension of problem structure. These results also support the general argument of this thesis that problem structure affects problem solving performance.

Table 15: Pair wise independent sample t-test results for restructuring

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus C (Detour) and H (Detour with Restructuring)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Time	Equal variances assumed	14.721	.001	3.092	26	.005	734.93846	237.65829	246.42486	1223.452
	Equal variances not assumed			2.882	12.566	.013	734.93846	255.03155	182.03376	1287.843
No of card exchanges	Equal variances assumed	11.298	.002	2.246	26	.033	16.241	7.230	1.379	31.103
	Equal variances not assumed			2.101	13.155	.055	16.241	7.729	-.435	32.918
Card exchange rate	Equal variances assumed	.091	.765	-2.522	26	.018	-1.73636	.68840	-3.15139	-.32133
	Equal variances not assumed			-2.454	20.636	.023	-1.73636	.70766	-3.20960	-.26312

a

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus D (Redundancy) and I (Redundancy with Restructuring)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Time	Equal variances assumed	.133	.718	3.710	23	.001	760.48701	204.96027	336.49438	1184.480
	Equal variances not assumed			3.589	18.278	.002	760.48701	211.89838	315.78978	1205.184
No of card exchanges	Equal variances assumed	4.329	.049	2.440	23	.023	24.175	9.909	3.677	44.673
	Equal variances not assumed			2.181	10.947	.052	24.175	11.083	-.233	48.584
Card exchange rate	Equal variances assumed	.059	.811	-1.701	23	.102	-.91084	.53557	-2.01876	.19707
	Equal variances not assumed			-1.660	19.278	.113	-.91084	.54864	-2.05803	.23634

b

5.4.3 Test of Hypothesis 2b

H_{2b}: In restructured tasks, task performance reduces with increasing degree of homogeneity.

Hypothesis 2b proposes that the degree of homogeneity has an effect on performance such that the greater the degree of homogeneity, the lower the performance for similar tasks. This proposition is based on the assumption that identical objects within a task will be more difficult to split to form a larger group than when using varied objects.

Problem solving performance on stimulus H (heterogeneous detour and restructuring) was compared to performance on stimulus J (homogeneous detour and restructuring) and performance on stimulus I (heterogeneous redundancy and restructuring) was compared to performance on stimulus K (homogeneous redundancy and restructuring) to test the effects of homogeneity on restructuring under detour and redundancy conditions respectively. Stimuli J and K were expected to result in relatively worse performance. The descriptive statistics in sections 5.2.1, 5.2.2 and 5.2.3 showed the expected effects for the redundancy case but a contradictory result for the detour case.

T test results are provided in Table 16a and 16b below. For the detour case (Table 16a) differences between stimuli H and J were insignificant for all performance measures: $t = 0.673$ ($p = 0.508$) for card exchanges, $t = -0.463$ ($p = 0.650$) for card exchange rate and $t = -1.009$ ($p = 0.326$) for time taken. Similarly, for the redundancy case (Table 16b) differences between stimuli I and K were insignificant for all performance measures: $t = 0.438$ ($p = 0.666$) for card exchanges, $t = -0.289$ ($p = 0.775$) for rate of card exchanges, and $t = 1.246$ ($p = 0.226$) for time taken.

These results lead to rejection of the hypothesis that the presence of homogeneity results in worse performance under the detour and redundancy conditions. The findings leave open the possibility of an interaction effect requiring further research. In general, the hypothesis on the effect of the degree of homogeneity is not supported and thus leaves room for more detailed work.

Table 16: Pair wise independent sample t-test results for restructuring

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus H (Heterogeneous detour with restructuring) and J (Homogeneous Detour with restructuring)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Time	Equal variances assumed	2.343	.140	-.962	22	.346	-294.90210	306.50222	-930.549	340.74460
	Equal variances not assumed			-1.009	18.974	.326	-294.90210	292.19163	-906.523	316.71859
No of card exchanges	Equal variances assumed	.868	.362	-.651	22	.522	-6.308	9.684	-26.391	13.776
	Equal variances not assumed			-.673	21.101	.508	-6.308	9.377	-25.802	13.186
Card exchange rate	Equal variances assumed	3.579	.072	-.430	22	.671	-.28958	.67338	-1.68609	1.10693
	Equal variances not assumed			-.463	14.107	.650	-.28958	.62488	-1.62886	1.04969

a

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus I (Heterogeneous Redundancy with Restructuring) and Stimulus J (Homogeneous Redundancy with Restructuring)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Time	Equal variances assumed	.876	.359	1.236	22	.230	311.34266	251.91830	-211.104	833.78923
	Equal variances not assumed			1.246	21.863	.226	311.34266	249.83794	-206.978	829.66363
No of card exchanges	Equal variances assumed	.508	.484	.434	22	.668	6.720	15.478	-25.379	38.819
	Equal variances not assumed			.438	21.839	.666	6.720	15.359	-25.145	38.586
Card exchange rate	Equal variances assumed	.027	.870	-.292	22	.773	-.16503	.56450	-1.33574	1.00567
	Equal variances not assumed			-.289	20.201	.775	-.16503	.57069	-1.35471	1.02464

b

5.4.4 Test of Hypothesis 3

Hypothesis 3:

H_3: Behavioural variety increases with structural complexity.

The third hypothesis deals with the effects of task complexity on behavioural variety. It is similar to hypothesis 1 except that it refers to behavioural measures rather than performance measures. The four behavioural measures used in this analysis are difference in balance area, card exchange reversals, no of verbal categories and verbal reversals. Less complicated tasks are expected to result in lower behavioural variety and vice versa.

To test this hypothesis, t tests were used to compare behavioural measures between stimulus B (basic) and stimulus D (redundancy), and between stimulus C (detour) and stimulus E (detour and redundancy).

Recall that Stimuli D and E were defined to be more complex and therefore were expected to result in relatively higher behavioural variety. The results are summarized in Table 17a and 17b. The results of the analysis for stimuli B and D (Table 17a) indicate a significant difference for only one of the behavioural measures (difference in balance area; t = -3.403; p = 0.004). The other three measures were not significant with (t = 0.511, p = 0.615) for card exchange reversals, (t = 0.679, p = 0.504) for number of verbal categories and (t = -0.985, p = 0.334) for verbal reversals. These results indicated partial support for the hypothesis. As discussed in section 5.5.1 this result can be attributed to the two outlier groups for stimulus B.

The result in Table 17a1 partially confirms the claim that the outlier groups may be responsible for the insignificant results as the analysis shows significant results for two of the behavioural measures – (t = -3.772, p = 0.002) for difference in balance area and (t = -2.231, p = 0.04) for card exchange reversals.

The results of the t test analysis for stimuli C and E (Table 17b) indicated significant differences ($p < 0.05$) for three behavioural measures: difference in balance area (t = -2.515, p = 0.020), card exchange reversals (t = -2.771, p =0.012) and no of verbal categories (t = -2.279, p = 0.03). The other result for verbal reversals (t = -1.667, p = 0.106) was insignificant except when considered at 0.1 level of significance.

Overall, the preceding analyses provide partial support for the hypothesis that the degree of complexity varies directly with the behavioural variety.

Table 17: Pair wise independent sample t-test results for the effect of structure on behavioural variety

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for stimulus B (Basic) and Stimulus D (Redundancy)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference in balance area	Equal variances assumed	14.674	.001	-3.403	26	.002	-16.857	4.953	-27.039	-6.675
	Equal variances not assumed			-3.403	14.643	.004	-16.857	4.953	-27.438	-6.277
Card exchange reversals	Equal variances assumed	.549	.465	-.985	26	.334	-.857	.870	-2.645	.931
	Equal variances not assumed			-.985	25.745	.334	-.857	.870	-2.646	.932
No of verbal categories mentioned	Equal variances assumed	.884	.356	-.679	26	.503	-2.714	3.998	-10.933	5.504
	Equal variances not assumed			-.679	23.177	.504	-2.714	3.998	-10.982	5.553
No of verbal reversals	Equal variances assumed	.234	.632	.511	26	.614	1.643	3.217	-4.971	8.256
	Equal variances not assumed			.511	21.933	.615	1.643	3.217	-5.031	8.317

a

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus B (Basic) and Stimulus D (Redundancy) without the outlier groups						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference in balance area	Equal variances assumed	17.232	.000	-3.487	24	.002	-18.226	5.227	-29.013	-7.439
	Equal variances not assumed			-3.772	13.315	.002	-18.226	4.832	-28.641	-7.811
Card exchange reversals	Equal variances assumed	8.645	.007	-2.091	24	.047	-1.524	.729	-3.028	-.020
	Equal variances not assumed			-2.231	16.068	.040	-1.524	.683	-2.971	-.076
No of verbal categories mentioned	Equal variances assumed	.035	.853	-1.452	24	.159	-5.512	3.795	-13.345	2.321
	Equal variances not assumed			-1.425	20.849	.169	-5.512	3.867	-13.557	2.533
No of verbal reversals	Equal variances assumed	3.603	.070	-.666	24	.512	-1.381	2.072	-5.658	2.896
	Equal variances not assumed			-.697	20.428	.494	-1.381	1.982	-5.510	2.748

a1

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus C (Detour) and Stimulus E (Detour with Redundancy)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference in balance area	Equal variances assumed	6.326	.017	-2.407	30	.022	-35.557	14.774	-65.729	-5.385
	Equal variances not assumed			-2.515	22.383	.020	-35.557	14.136	-64.845	-6.269
Card exchange reversals	Equal variances assumed	11.943	.002	-2.631	30	.013	-2.416	.918	-4.291	-.541
	Equal variances not assumed			-2.771	19.894	.012	-2.416	.872	-4.235	-.597
No of verbal categories mentioned	Equal variances assumed	.004	.948	-2.273	30	.030	-11.424	5.025	-21.686	-1.161
	Equal variances not assumed			-2.279	29.750	.030	-11.424	5.013	-21.664	-1.183
No of verbal reversals	Equal variances assumed	2.543	.121	-1.633	30	.113	-3.471	2.126	-7.812	.871
	Equal variances not assumed			-1.667	28.892	.106	-3.471	2.082	-7.730	.789

b

5.4.5 Test of Hypothesis 4 a

Hypothesis 4a:

H_{4a}: *Behavioural variety increases with the introduction of restructuring.*

In testing hypothesis 4a, the effects of adding a restructuring dimension to the problem structure (i.e, the need to categorize items at a higher level of abstraction) are examined. The restructuring dimension is expected to increase behavioural variety since groups are likely to search the apparent problem space until obvious potential solutions have been exhausted and eventually recognize the need to restructure their understanding of the problem. Thus, relative to stimuli without the restructuring dimension, more wide-ranging search is expected, corresponding to higher ratings on measures of behavioural variety.

To test this hypothesis, t tests were used to compare behaviour variety measures between stimulus C (detour) and stimulus H (detour and restructuring), and between stimulus D (redundancy) and stimulus I (redundancy and restructuring) as shown in Tables 18a and 18b. The descriptive statistics in sections 5.3.1 and 5.3.2 are consistent and showed the expected effects for the measures.

The results of the analysis for stimuli C and H (Table 18a) indicate significant differences for three of the behavioural measures: card exchange reversals (t = 2.789, p = 0.016), no of verbal categories (t = 4.354, p = 0.001) and verbal reversals (t = 2.914, p = 0.012). Differences were insignificant for difference in balance area (t = -1.346, p = 0.198).

The results of the analysis for stimuli D and I (Table 18b) also indicate significant differences (at the 0.05 level) for three of the behavioural measures: difference in balance area (t = 2.112, p = 0.059), verbal reversals (t = 5.352, p < 0.001), and no of verbal categories (t = 6.686, p = 0.00). Differences were significant for card exchange reversals (t = 1.869, p = 0.089). These results provide partial support for the hypothesis that the presence of abstraction or restructuring results in higher behavioural variety.

In summary, most of the behavioural variety measures support the hypothesis that behavioural variety increases with the introduction of abstraction or restructuring.

Table 18: Pair wise independent sample t-test results for effect of restructuring on behavioural variety

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus C (Detour) and Stimulus H (Detour with restructuring)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference in balance area	Equal variances assumed	4.073	.054	1.424	26	.166	24.272	17.046	-10.766	59.310
	Equal variances not assumed			1.346	14.894	.198	24.272	18.031	-14.184	62.727
Card exchange reversals	Equal variances assumed	11.869	.002	2.995	26	.006	5.918	1.976	1.857	9.979
	Equal variances not assumed			2.789	12.459	.016	5.918	2.122	1.314	10.522
No of verbal categories mentioned	Equal variances assumed	12.843	.001	4.585	26	.000	42.600	9.291	23.501	61.699
	Equal variances not assumed			4.354	15.693	.001	42.600	9.785	21.824	63.376
No of verbal reversals	Equal variances assumed	9.461	.005	3.119	26	.004	18.615	5.969	6.346	30.885
	Equal variances not assumed			2.914	12.981	.012	18.615	6.387	4.814	32.417

a

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus D (Redundancy) and Stimulus I (Redundancy with restructuring)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference in balance area	Equal variances assumed	4.835	.038	2.369	23	.027	53.539	22.595	6.797	100.281
	Equal variances not assumed			2.112	10.747	.059	53.539	25.347	-2.411	109.488
Card exchange reversals	Equal variances assumed	5.337	.030	2.099	23	.047	6.688	3.186	.097	13.280
	Equal variances not assumed			1.869	10.674	.089	6.688	3.578	-1.217	14.594
No of verbal categories mentioned	Equal variances assumed	7.524	.012	7.442	23	.000	61.981	8.329	44.751	79.210
	Equal variances not assumed			6.686	11.294	.000	61.981	9.270	41.641	82.320
No of verbal reversals	Equal variances assumed	2.627	.119	5.835	23	.000	25.695	4.403	16.586	34.804
	Equal variances not assumed			5.352	12.934	.000	25.695	4.801	15.318	36.071

b

5.4.6 Test of Hypothesis 4b

Hypothesis 4b:

H_{4b}: In restructured tasks, increasing degree of homogeneity will result in higher behavioural variety.

The second part of hypothesis 4 deals with the relationship between degree of homogeneity and behavioural variety. Increasing the degree of homogeneity was expected to result in greater resistance in forming categories at the super-ordinate level of abstraction, making it more difficult to recognize that restructuring to a higher level of abstraction was needed to solve the problem. Thus, increased homogeneity should lead to higher behavioural variety as groups spend more time searching for an appropriate solution.

To test this hypothesis, measures of behavioural variety were compared between stimulus H (heterogeneous, detour and restructuring) and stimulus J (homogeneous, detour and restructuring), and between stimulus I (heterogeneous, redundancy and restructuring) and stimulus K (homogeneous, redundancy and restructuring). Although stimuli J and K were expected to result in relatively higher behavioural variety, the descriptive statistics in sections 5.3.1 and 5.3.2 indicate the opposite results, as all measures of behavioural variety were lower for homogeneous stimuli than for the corresponding heterogeneous stimuli. The results of the t test analyses are summarized in Tables 19a and 19b. In comparing stimuli H and J with the detour condition, differences in all measures of behavioural variety were insignificant (see Table 19a) with ($t = -1.243$, $p = 0.230$) for verbal reversals, ($t = -0.118$, $p = 0.907$) for difference in balance area, ($t = -1.241$, $p = 0.228$) for no of verbal categories and t = (1.238, $p = 0.232$) for card exchange reversals. Similarly, in comparing stimuli K and I with the redundancy condition, differences in all measures of behavioural variety were again insignificant (see Table 19b) with ($t = -0.080$, $p = 0.937$) for verbal reversals, ($t = -0.283$, $p = 0.781$) for difference in balance area, ($t = -1.526$, $p = 0.149$) for no of verbal categories and ($t = -0.643$, $p = 0.531$) for card exchange reversals. These results lead to the rejection of hypothesis 4b that homogeneity leads to increased behavioural variety in restructured tasks.

While the hypothesis on the degree of homogeneity is not supported by any of the data, the findings suggest that what is captured by behavioural measures may differ from what is captured by performance measures. In re-analysing the behavioural outcomes in retrospect, these findings seem to be accurate and valid. For instance, in re-assessing the verbal reversals, it would be fair to state that the number of reversals in the homogeneous category will truly be lower since the groups are likely to be in the right direction anyway due to the higher number of identical items, which are truly members of the right set, in each category. The only constraint here is in the collection of sets or groupings, not the items or labels as the case would be for heterogeneous items. This implies that the heterogeneous stimuli may be seen as having multiple layers of searches such as finding what the right items are, the right groups and the right labels. The first part of the search is reduced for the homogeneous group since the items have been somewhat highlighted.

Table 19: Pair wise independent sample t-test results for effect of homogeneity on behavioural variety

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus H (Heterogeneous, Restructuring and Detour) and Stimulus J (Homogeneous, Restructuring and Detour)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference in balance area	Equal variances assumed	.945	.342	-.113	22	.911	-2.357	20.899	-45.698	40.985
	Equal variances not assumed			-.118	19.432	.907	-2.357	19.980	-44.113	39.399
Card exchange reversals	Equal variances assumed	2.326	.141	-1.172	22	.254	-2.930	2.501	-8.116	2.256
	Equal variances not assumed			-1.238	17.781	.232	-2.930	2.368	-7.909	2.048
No of verbal categories mentioned	Equal variances assumed	2.166	.155	-1.226	22	.233	-15.545	12.683	-41.848	10.757
	Equal variances not assumed			-1.241	21.972	.228	-15.545	12.531	-41.535	10.444
No of verbal reversals	Equal variances assumed	2.262	.147	-1.177	22	.252	-8.797	7.471	-24.291	6.697
	Equal variances not assumed			-1.243	17.948	.230	-8.797	7.080	-23.674	6.080

a

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means for Stimulus I (Heterogeneous, Restructuring and Redundancy) and Stimulus K (Homogeneous, Restructuring and Redundancy)						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Difference in balance area	Equal variances assumed	.471	.500	-.294	22	.771	-8.105	27.548	-65.237	49.027
	Equal variances not assumed			-.283	16.126	.781	-8.105	28.642	-68.784	52.574
Card exchange reversals	Equal variances assumed	1.603	.219	-.680	22	.504	-2.469	3.630	-9.998	5.060
	Equal variances not assumed			-.643	13.725	.531	-2.469	3.838	-10.715	5.778
No of verbal categories mentioned	Equal variances assumed	2.729	.113	-1.606	22	.123	-15.140	9.428	-34.692	4.412
	Equal variances not assumed			-1.526	14.306	.149	-15.140	9.924	-36.383	6.103
No of verbal reversals	Equal variances assumed	.044	.836	-.082	22	.936	-.448	5.481	-11.815	10.920
	Equal variances not assumed			-.080	19.197	.937	-.448	5.583	-12.125	11.230

b

5.4.7 Test of Hypothesis 5

Hypothesis 5:

H_5: Problem solving behaviour is in the direction of increased perceived structural balance.

This hypothesis deals with the question of what forces influence the next move at any stage during the problem solving process. Since these moves amount to card exchanges in the present experiment, the hypothesis deals with the forces influencing the next card move during card exchanges. As the problem solving progresses from one state to the next, the forces acting on participants are assumed to change according. As discussed earlier, since each individual's goal is to collect four of a kind, card collection is expected to be based on the relative number of items per category currently possessed by each group member. If similar items are represented as having attractive forces and dissimilar items are represented as having repelling forces then each card is expected to move in the direction of the greatest force. For example, a subject with two birds is likely to exchange for another bird. Furthermore a subject who currently possesses three birds should be more likely to collect a fourth bird than a subject with just one bird is to collect a second bird, since three items are assumed to exert a greater attraction force than just one item. This is the basis of the argument and this force idea could be further developed to explore in more depth the effects of problem structure on problem solving behaviour.

To test this concept, a random sample of eight groups (one selected from each experimental condition) was used to test the effect of higher frequency of a category on the choice of card exchange. Chi-Square goodness of fit test (see Appendix F) was performed to test the hypothesis that the next card selection or exchange is a function of each individual's current cards. This is to verify that participants individually collect the next card based on their current largest set. The overall Chi-Square value with 1df was $\chi^2 = 6.426$. This high Chi squared value with significance of 0.011 supports the hypothesis that card exchanges are based on the force of attraction and the frequency of the objects within a category operating in a group's problem solving. Table 20 below shows the Chi-Square (χ^2) values and associated probabilities.

Table 20: Chi-Square value for high frequency tests

Decision

	Observed N	Expected N	Residual
0	47	61.0	-14.0
1	75	61.0	14.0
Total	122		

Test Statistics

			Decision
Chi-Square[a]			6.426
df			1
Asymp. Sig.			.011
Monte Carlo Sig.	Sig.		.016[b]
	99% Confidence Interval	Lower Bound	.012
		Upper Bound	.019

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 61.0.

b. Based on 10000 sampled tables with starting seed 803356894.

5.4.8 Test of Hypothesis 6

Hypothesis 6:

H_6: The greater the initial discussion time spent solving a task, the lower the behavioural variety

Problem solving strategy refers to the approach taken by a group at the beginning of the problem solving process in order to acquire an understanding of the task structure. Two general types of strategies are considered in this thesis, which differ in terms of the way groups exchange information in order to learn about the problem structure. One strategy involves exchanging information primarily through discussions (i.e., verbal protocol). The other strategy involves exchanging information primarily through card exchanges. Since all groups used both discussion and card exchanges to solve the problem, these two strategies represent extreme cases and the measure identifies how close a group's particular strategy was to either extreme.

Researchers such Wertheimer (1982) have emphasized the importance of proper understanding of problem structure and individuals are advised to understand a problem before attempting to solve the problem. The same concept applies in this context such that groups attempt to have a collective understanding of the task either through verbal discussion or via card exchanges. As indicated in hypothesis 6, it is expected that the verbal discussion would be a more efficient way of understanding the problem structure and would thus lead to better performance and lower behavioural variety. The relative use of verbal discussion and card exchange to obtain an understanding of the problem is measured using the average rank order measure.

The effect of strategy was computed using the rank order performance measure described in section 3.14.1.1. A lower rank average indicates that a group spent its longest periods of discussion at the beginning of the problem solving activity while higher rank averages imply that most discussion occurred later in the process (e.g., somewhere in the middle for long duration tasks or at the end for short duration tasks).

To test hypothesis 6 the average rank order measure was correlated with behavioural and performance measures (difference in balance area and number of card exchanges) for each experimental condition. The results (in appendix E) using Pearson's Correlation test show significant correlations at 0.01 level for both the behavioural and performance measures in six out of the eight stimuli sets. In stimuli B (basic) and C (detour) there was no significant correlation with either of the measures. This is probably due to the fact that these tasks were the least complicated and required the least information exchange to solve. Consequently, most groups solved these tasks efficiently and the choice of strategy made little or no difference. However, with the more complicated tasks (stimuli D to K) the adopted strategy significantly influenced performance and behavioural variety as shown in the results. These results support the hypothesis that behavioural strategy influences task performance on a given stimulus with earlier discussion producing better results than card exchanges as a means of understanding problem structure.

The table below reports the rank average results, as well as balance area difference and the number of card exchanges for experimental condition D. Spearman rank correlations among these three variables are reported in Table 22 below for this experimental condition. The average rank order is significantly correlated with both performance (card exchanges) and behavioural variety (area difference) indicating that early discussion is more effective than card exchanges for understanding problem structure.

Table 21: Stimulus D average rank order and performance measures

Task ID	Stimulus	Rank Average	Area Difference	Card Exchanges
060104	1D	3	11	12
090203	1D	2	3	12
100103	1D	5	16	14
100303	1D	3	7	12
010104	2D	6	55	18
010304	2D	3	18	14
040104	2D	2	6	12
040204	2D	3	4	14
040304	2D	6	26	18
040404	2D	3	2	12
070103	3D	4	20	13
080103	3D	12	53	40
120103	3D	12	42	33
090404	3D	8	26	22

Table 22: Spearman's rank correlation

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.903**	.930**
		Sig. (2-tailed)	.	.000	.000
		N	14	14	14
	Area Difference	Correlation Coefficient	.903**	1.000	.841**
		Sig. (2-tailed)	.000	.	.000
		N	14	14	14
	Card Exchanges	Correlation Coefficient	.930**	.841**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	14	14	14

**. Correlation is significant at the 0.01 level (2-tailed).

6 DISCUSSION AND CONCLUSIONS

This thesis investigated the influence of problem structure on performance and behaviour in group problem solving. Previous studies have focused more on personality factors and communication effects of individuals in group problem solving There has been limited research on effect of problem structure in group problem solving due to 1) the unavailability of a fundamental theory of task structure and 2) the complexity associated with understanding groups and group processes. This thesis addresses both concerns by defining and exploring fundamental characteristics of task structure and their effects on group problem solving behaviour and performance.

This thesis has also examined the effect of strategy on performance and behavioural outcomes in problem solving. Two general kinds of problem solving strategy were considered, that which differ in the extent to which group members worked to understand the problem structure individually or collectively: immediate card trading driven by each individual’s attempts to collect four of a kind, versus substantial early discussion to establish a collective understanding of the problem structure.

Finally, Heider’s balance theory was used in two ways in this thesis. First, it provided a representational system used to model different problem structures, which made it possible to compare problem solving behaviour and performance across different problem structure conditions. Second, problem solving is viewed as a process of moving from an initial state of imbalance towards a state of cognitive balance corresponding with the solution. Thus, balance theory provides a way of conceptualizing the problem solving process that combines both incremental search (i.e., information processing) and gestalt restructuring approaches typical of insight problems in a single framework. It highlights the dynamic process of problem solving contrary to most problem solving research approaches that examine the process using snapshots of events.

The sections below discuss the major conclusions and research findings based on the assumptions, hypotheses, theoretical constructs and final analysis. It also discusses the limitations of the study and areas of future research.

6.1 Major Findings

Below are the major findings of the experimental study:

First, the results have shown that task structure has an effect on performance and behaviour in group problem solving. Specifically, Hypothesis 1 predicted that task complexity would vary inversely with performance. This hypothesis received support from the t-tests and the correlation results. The t-tests showed substantial support for performance differences between stimuli B (when outliers were removed) and D while the comparison between stimuli C and E received partial support. The reason for the partial support may be due to the similarity in card exchanges between these stimuli types. The main characteristic that influence wrong exchanges is the detour dimension, which is present in both stimuli. And as a result, the additional characteristic in stimulus E (i.e. redundancy) may not result in substantial difference in performance between both stimuli type thus resulting in insignificant result. This notwithstanding, the p value in both insignificant measures are 0.06 and 0.077 for card exchange and card exchange rate respectively which are quite close to the acceptable range of 0.05. An alternate explanation could be the relatively small sample size, which was due to the unavailability of additional participants and number of experiments that were conducted within the allowed time frame. However, the correlation result showed consistent inverse relationship between the different structural and the performance measures. Although, the correlation results are not specific to each stimulus type, it consistently shows the inverse relationship between structural complexity and performance measures.

Second, hypothesis 2a predicted that performance would reduce with the introduction of restructuring and vice versa. This was strongly supported by both the t-tests and the correlation tests. In the t-test, all the performance measures declined significantly (except rate of card exchange in restructuring redundancy) with the introduction of restructuring, indicating that restructured tasks were perceived as more difficult than their non-restructured counterparts.

The results from the tests of hypothesis 2b however proved contrary to expectations; where homogeneity did not appear to result in more complex tasks. Although, the descriptive statistics indicated that performance varied in the expected direction (though differences were insignificant) for restructured redundancy stimuli, the restructured detour stimuli did not support the hypothesis in any analysis. In retrospect, the homogeneous detour may have been perceived as easier for two reasons: 1) the participants' can quite easily perceive that the identical items go together and 2) the sets of identical items are incomplete (i.e., 3 items per set) and thus require the addition of different extra items to form complete sets. However, in the heterogeneous detour condition, the groups probably had to make multiple decisions of which objects go together since there are different degrees of similarity among items and choices in forming possible categories. In addition, the insignificant result in the restructured redundancy condition may have been due to the relatively small sample size.

Third, consistent with Ashby's law of requisite variety (Ashby, 1958), hypothesis 3 predicted higher behavioural variety with more complex structures. The results partially supported Hypothesis 3 for both test conditions (i.e., difference in objective and subjective area and card exchange reversals) with verbal measures resulting in insignificant measures. Also, the correlation results provided support for this hypothesis showing that behavioural variety increases with task complexity.

The partial support for verbal measures may be due to the similarity in verbal categories across the two stimuli. The possible types of verbal categories are basically the same since they are based on correct and incorrect categories, which are likely to be more differentiated between restructured and non-restructured groups. The shape of the graph in Figure 23 also support this argument where the non-restructured tasks seem to be within the same range in terms of number of verbal categories and a greater differential when compared with their restructured counterparts. In addition, the partial support may also have been due to the small sample size.

Fourth, the t-tests support the prediction in hypothesis 4a that the introduction of restructuring results in higher behavioural variety. The t-tests were highly significant with $p < 0.05$ for most behavioural measures (i.e., card exchange reversals, verbal reversals and number of verbal categories) except in difference in balance area between non-restructured and restructured detour.

In tests of hypothesis 4b, the results were insignificant for all measures of behavioural variety. These results are complex and confusing. One interpretation is that the hypothesis was simply wrong, and homogeneity in fact makes tasks easier than their heterogeneous counterpart. The lack of support may also have been due to interaction effects among the various characteristics of task structure, which is beyond the scope of this research. Finally, the experimental tasks for this characteristic may have been designed inappropriately such that the signals being picked from the measures are inaccurate. Whichever reason, it is obvious that further research is required to understand the effect of homogeneity on restructured tasks.

Fifth, Hypothesis 5 received substantial support from the Chi–square analysis showing that problem solving is in the direction of increased perceived structural balance. The analysis indicated that most of the card exchanges were based on the greatest number of similar cards held by each individual. This is somewhat in line with the overall goal, which indicates that groups should come up with four of a kind. There is the natural tendency to collect only similar cards in order to come up with their four of a kind.

Finally, Hypothesis 6 showed that earlier discussion time during problem solving resulted in better performance and lower behavioural variety. The rank order correlations strongly support this for stimuli D – K. The effects were not very strong in stimuli B and C since the tasks are easy enough and the choice of strategy on performance isn't likely to be substantially different across stimuli sets.

The effect of problem structure on performance and behavioural outcomes has been predicted using various hypotheses. Data and analyzed results support the hypothesis that performance reduces and behavioural variety increases with perceived increase in structural complexity and abstraction. It also supports the hypothesis that strategy affects performance with preference for early discussion in problem solving.

While complexity and abstraction appear to have a direct relationship with behavioural outcome, degrees of homogeneity seem to portray an inverse relationship. Although the importance of the task structure has been identified in prior existing literatures, no known literature has attempted to test and confirm these effects. Even though the journey to a fundamental and encompassing theory of problem structure is still far away, this thesis serves as a step in the right direction. It provides a first attempt to examine the effects of problem structure on group problem solving behaviour and performance.

6.2 Research Contributions

This thesis contributes to the group problem solving literature in several ways:

6.2.1 Experimental Results

The experimental results provide evidence that problem structure affects performance and behavioural variety as discussed above. Different structural characteristics were varied to understand their effects on problem solving performance and behaviour. The data showed substantial support for most of the hypotheses with a few requiring further investigation.

6.2.2 Structural Effects

The difficulty associated with understanding and representing task structure has been discussed in several literatures (Roby and Lanzetta, 1956; Shaw, 1954; Zajonc and Taylor, 1963). The inability to represent task structures based on fundamental characteristics has led researchers to classify task structures in fairly general terms, such as simple and complex tasks, routine deductive and non-routine inferential tasks, or to treat task structure based on expected behaviour (Faucheux and Mackenzie, 1966; Shaw, 1954). This has impeded research on task structure and its influence in problem solving. The dimensions of task structure adopted in the present study are more generic and fundamental, and are applicable to various problem situations. These dimensions also reflect the objective properties of the task in the current experiment, independent of any other variable. This work provides an initial basis from which a general structural theory may be developed and by which task structure can be manipulated as an independent variable. The different structural dimensions introduced and examined in this thesis may provide insight into studies of structural effects in other organizational task situations.

6.2.3 Stimulus Design

The design of the experimental stimulus is another major contribution of this research. Although originally inspired by Bavelas' five squares problem (Bavelas, 1973), it provides a more robust and easily modifiable design. In the original five squares problem the structure was static and could only be used as given. The stimulus design developed in this study provides a means by which all structural dimensions can be modified to address diverse research questions and to suit different situations. For instance, for more in depth understanding of the effects of a structural dimension like redundancy, the stimulus could be redesigned to vary the degree of redundancy. Also, the number of items on each card could be increased from two and the total number of cards could be varied from the current sixteen card design. This dynamic property of the experimental stimulus provides a means by which the method developed here could be easily adapted to other scenarios and supports the call for generalizable designs (Zajonc, 1965).

6.2.4 Theoretical Contribution

This thesis has proposed a new conceptualization of the problem solving process as a progression towards structural balance based on Heider's balance theory (Heider, 1946b). Although intuitively the applicability of this theory to problem solving seems obvious, no one has studied problem solving using this theory.

Prior research considers problem solving either in terms of information processing and incremental search through a problem space or in terms of a sudden cognitive restructuring associated with the experience of insight approach, but the two have not been well integrated. The use of Heider's balance theory has the ability to capture both forms of problem types. Each move during card exchanges in a balanced network corresponds restructuring and incremental changes in total degree of balance.

Also, the balance theory framework permitted the development of unique and useful measures of the problem solving process. For example, it was possible to measure both the group's subjective and the experimenter's objective view of the problem solving situation, and to treat the difference between them as a behavioural measure of the problem solving process. The reversals measure was also used to track incremental search and periods of restructuring.

6.2.5 Construct Measures

In most studies of problem solving, performance and behaviour are assessed using simple quantitative measures such as time required to solve the problem or success or failure to reach solution. In the present study behavioural measures were adopted that reflected both the structural dimensions of the task design and the context in which the variables were used within the laboratory experiment. Furthermore, measures such as difference in balance area and the number of reversals operationalize the theoretical view of problem solving in terms of a progression towards structural balance.

6.3 Limitations of Study

Every study has limitations, which introduce unexpected sources of variability. In experimental studies these sources range from individual differences across participants to experimental design and the methods used to analyze results. Although the study participants were students and the task was conducted in the laboratory, efforts were made as explained earlier to minimize environmental and knowledge bias. However, regardless of our attempts to reduce potential sources of bias, some factors were uncontrollable and resulted in outlier effects.

One source of error resulted from the use of graduate students as subjects on a few occasions when undergraduate students failed to arrive for the experiment at the appointed time. This affected the design of the experiment since an alterative compensation scheme was required for the graduate students. As explained earlier, graduate students were promised three dollars for each completed set of four of a kind while undergraduate students were promised 0.5 course bonus marks. If all group members were able to form a set of four of a kind, the compensation would be doubled. On a couple of the occasions when graduate students were used as subjects we noticed that the undergraduate students tried to negotiate with the graduate students when the group was unable to solve the problem. The three undergraduates in the group offered to pay the graduate student to take an incomplete set of cards so the undergraduates would have full sets and thereby receive 0.5 marks. This is both an interesting observation and a potential limitation of the study. The fact that undergraduate students were trying to negotiate with the graduate students in order to forgo their goal is interesting because it raises questions about the relative influence of using either marks or money as compensation. It is also interesting because the negotiation strategy amounts to an alternative path taken by the group to maximize goal attainment in a sense, a form of problem restructuring that was unanticipated in the experimental design. It is also a limitation of the study because the experimental design assumed that the two forms of compensation were equivalent in terms of their relative influence on individual and group problem solving behaviour. However, in these mixed groups we observed that the graduate students were relatively quick to give up on the task due to the alternative payment arrangements made with the undergraduates. This scenario and the scenario where some graduate or graduating undergraduate students genuinely had very little interest in the tasks could account for some bias effects experienced in some of the experiments.

There were two additional methodological limitations that affected the comparison of results across stimuli conditions. First, the initial states of perceived balance for different stimuli types (e.g., detour versus redundancy) were slightly different but these differences were not accounted for in the analysis. This difference consequently affected the difference in balance area for these stimuli types at the initial state of problem resolution. The difference had a nominal value of 2 and thus did not result in any statistically significant effects in any of the stimuli comparisons (e.g. Detour vs. Redundancy). In addition, the t-test analyses did not compare stimuli types with potentially orthogonal relationships among structural dimensions, but rather focused on more linear relationships such as Detour vs. Detour with Redundancy.

Second, the stimuli that included the redundancy dimension required only ten card exchanges to solve the problem while other stimulus conditions required twelve card exchanges. Again, the resulting data were not normalized for this difference. However, as indicated earlier, this did not affect the results since the analyses were more along linear relationships comparing stimuli types with similar characteristics.

Finally, time and the scope of the design were major constraints in the experimental study. Since the approach and aspect of task structure addressed in this study are relatively new, the understanding and design of the problem was relatively difficult and time consuming. The time spent on background research, collection and analysis of data was over a year and considering that this is a masters program, which is a substantial amount of time within the program. This affected the amount of data that was collected and the aspects of task structure that was investigated.

6.3.1 Dependent Variables

The dependent variables in this study are performance and behavioural variety. Although the use of behavioural measures has been discussed in different literature, the measures have not been standardized. Behavioural measures can be identified to be anything from the type of facial features to the number of individual comments. This makes the measure a bit vague and difficult to generalize. Although the measures are unique for this experimental design, they were chosen based on the balance theoretical framework (i.e., balance area, no of verbal reversals, card exchange reversals etc). They were operationalized with the theoretical view that problem solving amounts to a progression towards structural balance as balance is defined in this particular design. Thus, other potential experimental designs based on a balance framework would need to measure balance effects in other ways.

6.3.2 Independent Variable

The main independent variable is problem structure and while the notion of structure is well known, it is very difficult to describe and measure. Another challenge with the adopted independent variable in this study is the relationships among the various structural dimensions defined for the study (i.e., detour, redundancy, restructuring, homogeneity), which were difficult to conceptualize and to account for in terms of quantitative measures. For instance, the "number of item pairs" measure was able to discriminate amongst detour, redundancy, but was less able to capture differences introduced with the restructuring dimension, and unable to discriminate at all between homogeneous and heterogeneous stimuli. Thus, homogeneity can be considered to be orthogonally related to the other dimensions with respect to this variable. The "number of structural dimensions" measure simply counted how many of the four dimensions were included in a particular stimulus condition, ignoring any psychologically relevant differences among these four dimensions. These difficulties of measuring problem structure in a psychologically meaningful way made it difficult to relate outcome measures back to the input structural characteristics. They also highlight the need for a general theory of structure that is applicable in diverse psychological situations.

However, regardless of the limitations identified above, it would difficult to come up with plausible alternative explanations to account for the main results of the study. The results demonstrate the effect of structure on performance and behavioural variety. In specific instances, there may be alternate measures that may provide more concise results such as interacting effects among variables but that is left to future work and beyond the scope of this research.

6.4 Areas of Future Research

This thesis raises a variety of questions for consideration in future research.

First, this research identifies various dimensions of problem structure that can be investigated further and applied to real world problems. Using variations of the stimulus design developed here, researchers could examine the effects of varying problem structure along any of these dimensions. For example, problems with two detours, three levels of restructuring, or varying amounts of redundancy could be examined.

Second, the relationships among the different variables considered here could be examined more explicitly in terms of the psychological forces operating in the problem solving situation, as discussed briefly in relation to Hypothesis 5. For example, different dimensions of problem structure introduce different psychological forces acting on the individuals involved. Further theoretical and experimental work is needed to properly formulate such a view and it has only been introduced and investigated here in a very limited sense. Using forces as a means of characterizing the effects of structure would enable the dynamic relationships among behavioural and performance variables to be examined.

Third, various social properties of the experimental situation were held constant in the present study, which could also be manipulated in future studies. For instance, performance incentives could be varied between individuals and groups to investigate the effects of different compensation schemes on group performance.

Fourth, the test of hypothesis 6 highlights the difference between problem solving strategies aimed at acquiring an understanding of problem structure prior to attempting to solve a problem, and acquiring understanding while attempting to solve the problem. This difference was captured in the present study by examining when groups spent most of their time in discussion during the card exchange process. We have demonstrated that early discussion was associated with better performance and lower behavioural variety for our experimental stimuli, but much work remains to identify the generalizability of this result and the theoretical processes involved. For instance, many problems are too complex to consider all relevant structural dimensions in advance, so groups must begin trying to solve these problems in order to identify the relevant structural properties.

Finally, as noted above, this thesis highlights the importance and absence of a general theoretical understanding of task structure. A general theory of structure would help resolve the measurement issues mentioned above. More generally it would provide a framework for future research by establishing the fundamental properties of structure to be considered by researchers in specific domains. The structural dimensions considered in the current study were defined in a somewhat ad hoc manner, partly based on prior research. While the dimensions are plausible and interesting, they do not exhaust the range of potential dimensions that could be defined, nor do they fit cleanly into an integrated theory of structure. A suitable theory would characterize structural phenomena in terms of more fundamental constructs and the dimensions considered here would represent specific instances associated with particular combinations and values of these constructs. As suggested above, one potential way forward might be to examine structure explicitly in terms of the distribution of psychological forces (Lewin, 1936) acting in a given situation, since different structural dimensions impose different psychological forces on the individuals and groups involved.

7 REFERENCES

Bavelas, A., 1950. Communication Patterns in Task-oriented Groups. Journal of the Acoustical Society of America, 22, 725-730.

Bavelas, A., 1973. The Five Squares Problem: An Instructional Aid in Group Cooperation. Studies in Personnel Psychology, 5(2), 29-38.

Bonner, B.L., 2004. Expertise in Group Problem Solving: Recognition, Social Combination and Performance. Group Dynamics: Theory, Research & Practice, 8(4), 277-290.

Bottger, Preston C. and Yetton, Phillip W., 1988. An Integration of Process and Decision Scheme Explanations of Group Problem Solving Performance. Organizational behaviour and Human decision processes, 42, 234-249.

Cartwright, D and Harary, F, 1956. A Generalization of Heider's Theory. The Psychological Review, 63, 277-293.

Cartwright, D and Zander, A, 1960, GROUP DYNAMICS: Research and Theory. Row Peterson.

Chiu, M, 2000. Effects of Status on Solutions, Leadership and Evaluations during Group Problem Solving. Sociology of Education, 73(3), 175-195.

Cohen, Susan G. and Bailey, Diane E., 1997. What Makes Teams Work: Group Effectiveness Research from the Shop Floor to the Executive Suite. Journal of Management, 23(3), 239-290.

Davis, J.A., 1967. Clustering and structural balance in graphs. Human Relations, 20, 181-187.

Derbentseva, Natalia, 2002, Study of Path Construction in Human Problem Solving. University of Waterloo, pp. 158.

Dunbar, K, 1988, Problem Solving. In: W Bechtel and G Graham (Eds.), A Companion to Cognitive Science. Blackwell, London, pp. 289-298.

Faucheux, Claude and Mackenzie, Kenneth D., 1966. Task Dependency of Organizational Centrality: Its Behavioural Consequences. Journal of Experimental Social Psychology, 2, 361-375.

Festinger, L, 1950. Informal Social Communication. Psychological Review, 57, 271-282.

Hackman, J.R, 1987, The design of work teams. Prentice Hall.

Hackman, J.R and Oldham, G.R, 1980, Work redesign. Addison-Wesley, Reading, MA.

Harary, F, 1959. On the measurement of structural balance. Behavioural Science, 4, 316-323.

Harary, F, Norman, R.Z. and Cartwright, D, 1965, Structural models: an introduction to the theory of directed graphs. Wiley, New York.

Harshbarger, D, 1971. An Investigation of a Structural Model of a Small Group Problem Solving. Human Relations, 24(1), 43-63.

Heider, F, 1946a. Attitudes and cognitive organization. Journal of Psychology, 21, 107-111.

Heider, F, 1960, The gestalt theory of motivation. In: MR Jones (Ed.), Nebraska Symposium of Motivation, pp. 145-172.

Heider, Fritz, 1946b. Attitudes and Cognitive Organization. The Journal of Psychology, 21, 107-112.

Hendrick, H, 1979. Differences in Group Problem-Solving behaviour and effectiveness as a function of abstractness. Journal of Applied Psychology, 64, 518-525.

Hoffman, LR, 1965, Group Problem Solving, Advances in Experimental Psychology, pp. 99-132.

Holloman, CR and Hendrick, HW, 1971. Problem Solving in different sized groups. Personnel Psychology, 24(3), 489 - 500.

Kershaw, TC and Ohlsson, S, 2001, Training for insight: The case of the nine-dot problem. In: JD Moore and K Stenning (Eds.), Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 489-493.

Kohler, W, 1969, The Task of Gestalt Psychology. Princeton University Press, NJ.

Larson, JR and Christensen, C, 1993. Group as Problem Solving Units - Towards a new meaning of social cognition British Journal of Social Psychology, 32, 5-30.

Laughlin, PR, Zander, ML, Knievel, EM and Tan, TK, 2003. Groups Perform Better than the Best Individuals on Letters-to-Numbers Problems: Informative Equations and Effective Strategies. Journal of Personality and Social Psychology, 85(4), 684-694.

Leavitt, HJ, 1951. Some effects of certain communication patterns on group performance. Journal of Abnormal and Social Psychology, 46, 38-50.

Lewin, K, 1936, Principles of Topological Psychology. McGraw-Hill, NY.

Lewin, K, 1951, Field Theory in social science: Selected theoretical papers. In: D Cartwright (Ed.). Harper, New York.

Maier, NRF, 1930. Reasoning in humans: I. On direction. Journal of Comparative Psychology, 12, 144-155.

McGrath, J.E., 1984, Groups: Interaction and performance. Prentice-Hall, Englewood Cliffs, NJ.

Newcomb, TM, 1968, Interpersonal balance. In: R.P. Abelson et al. (Ed.), Theories of Cognitive Consistency: A sourcebook. Rand-McNally, Chicago, pp. 28-51.

Newell, Allen and Simon, H.A., 1972a, Human Problem Solving, Englewood Cliffs, J.J., Prentice-Hall (C1972).

Newell, Allen and Simon, Herbert Alexander Joint Author, 1972b, Human Problem Solving, Englewood Cliffs, J.J., Prentice-Hall (C1972).

Osgood, CE and Tannenbaum, PH, 1955. The Principal of Congruity in the Prediction of Attitudinal Change. Psychological Review, 62, 42-55.

Regolin, L, Vallortigara, G and Zanforlin, M, 1994. Perceptual and motivational aspects of detour behaviour in young chicks. Animal Behaviour, 47, 123-131.

Roby, T B and Lanzetta, J T, 1958. Considerations in the analysis of group tasks. Psychological Bulletin, 55, 88-101.

Rosch, E, 1975. Cognitive reference points. Cognitive Psychology, 7, 532-547.

Rosch, E and Mervis, C, 1975. Family Resemblances: Studies in the Internal Structure of Categories. Cognitive Psychology, 7, 573-605.

Rosenberg, MJ and Abelson, RP, 1960, An analysis of cognitive balancing. In: MJ Rosenberg, CI Hovland, WJ McGuire, RP Abelson and JW Brehm (Eds.), Attitude, organization and change: An analysis of consistency among attitude components. Yale University Press, Oxford, England.

Shannon, Claude E, 1948. A mathematical theory of communication. The Bell System Technical Journal, 27(3), 379-656.

Shaw, M, 1954. Some effects of Problem Complexity upon Problem Solution Efficiency in different communication nets. Journal of Experimental Psychology, 48, 211-217.

Simon, HA and Kaplan, CA, 1990. In Search of Insight. Cognitive Psychology, 22, 374-419.

Simon, HA and Newell, Allen, 1972, Human Problem Solving. Prince-Hall Inc., NJ.

Swinth, RL and Tuggle, FD, 1971. A Complete Dyadic Process Model of Four Man Group Problem-Solving. Organizational behaviour and Human Performance, 6, 517-549.

Wertheimer, M, 1982, Productive Thinking. The University of Chicago Press, Chicago.

Zajonc, RB, 1965. The Requirements and Design of a Standard Group Task. Journal of Experimental Psychology, 1, 71-88.

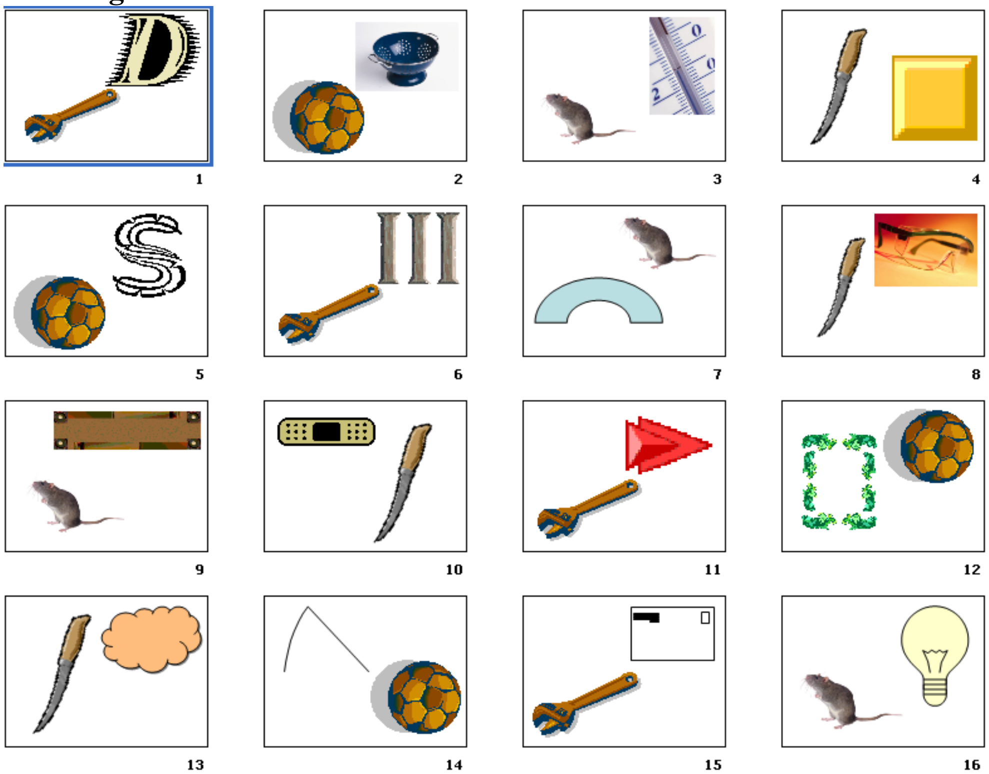
Zajonc, RB and Taylor, J, 1963. The effect of two methods of varying Group Task-difficulty on Individual and Group Performance. Human Relations, 16, 359-368.

Zhong, Zhehui, 2005, A balance theoretic view of problem solving. Department of Management Sciences, University of Waterloo.

Zucca, P, Antonelli, F and Vallortigara, G, 2005. Detour behaviour in three species of birds: quails (Coturnix sp.), herring gulls (Larus cachinnans) and canaries (Serinus canaria). Animal Cognition, 8, 122-128.

APPENDIX A: STIMULUS SET FOR EXPERIMENT I

Training 1



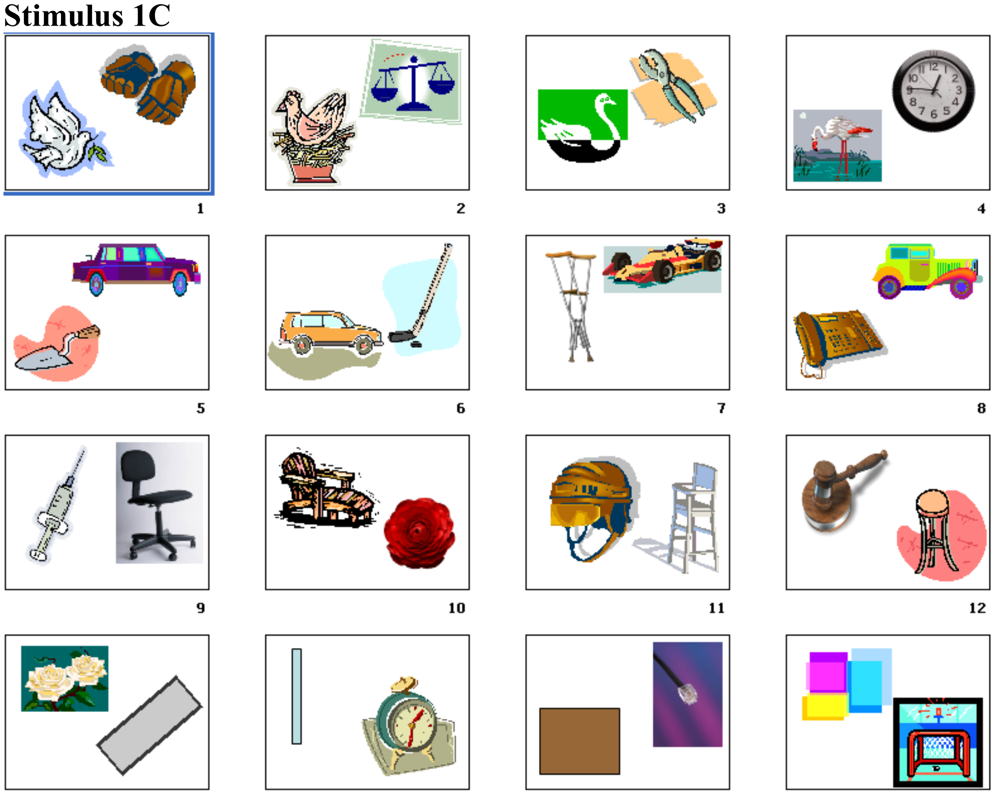
Training 2

Appendix A: Experiment I: Continued

STIMULI SET 1

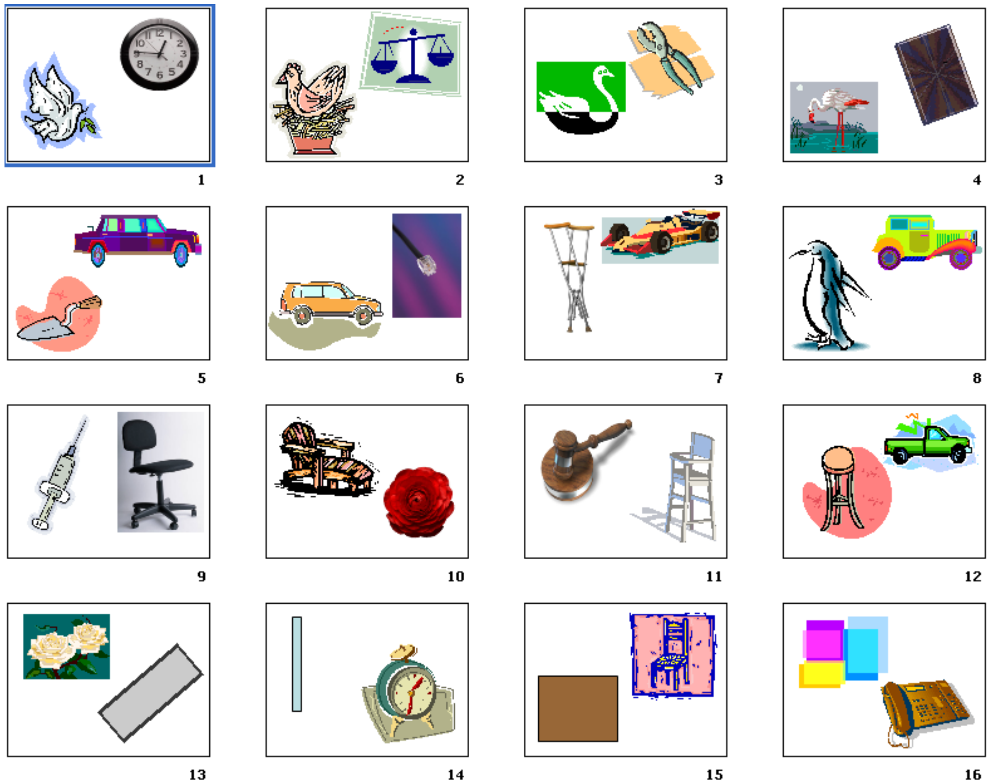
Stimulus 1B

Stimulus 1C

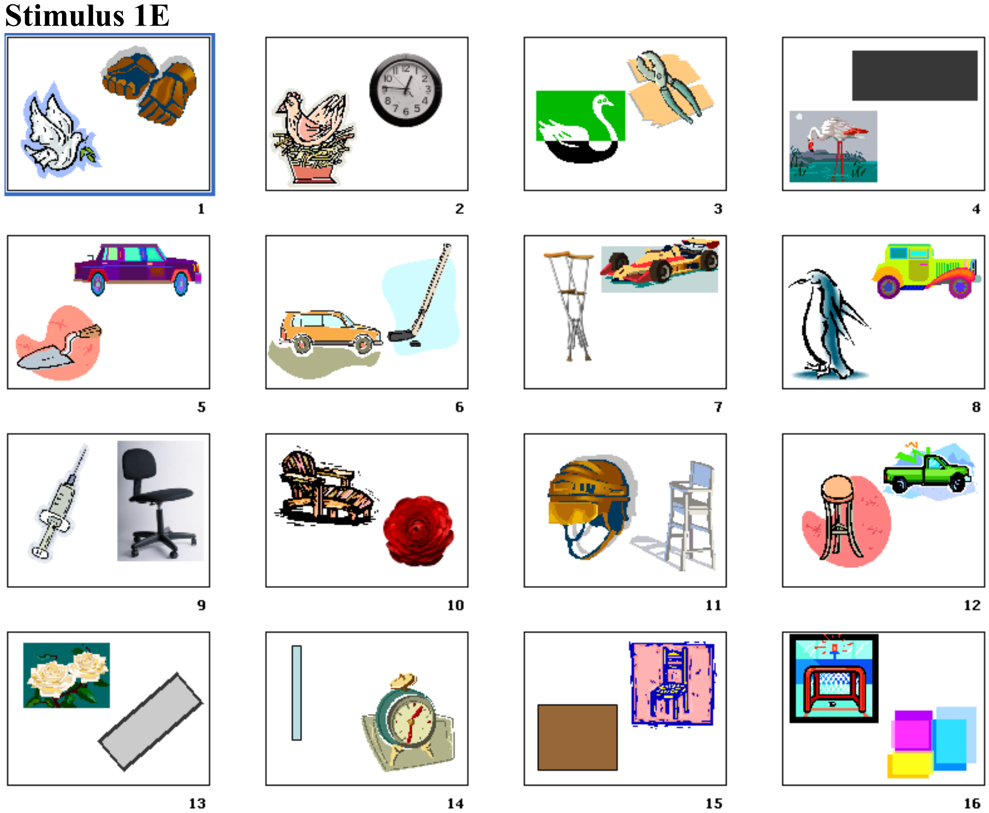


Appendix A: Experiment I: Continued

Stimulus 1D

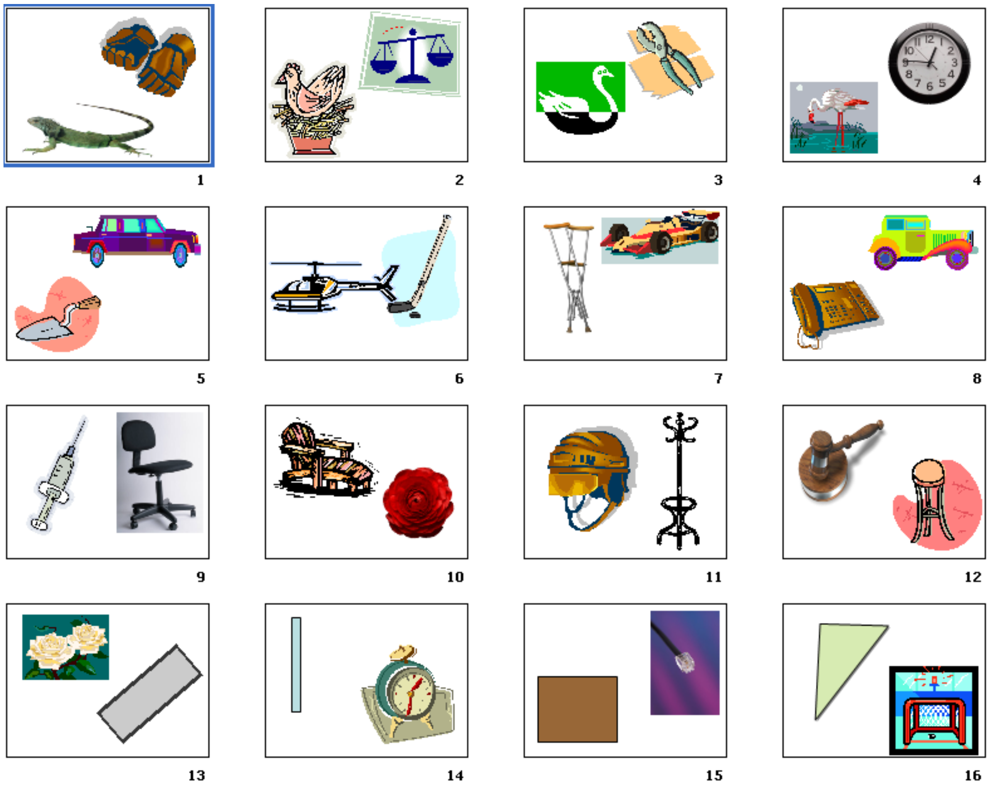


Stimulus 1E

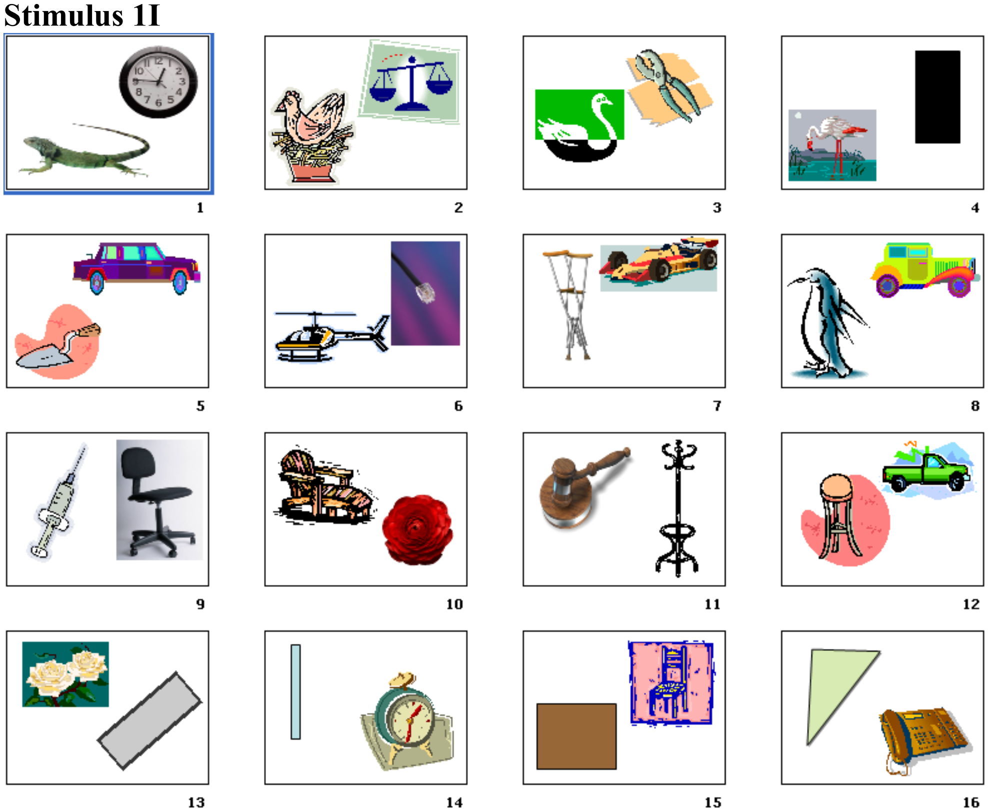


Appendix A: Experiment I: Continued

Stimulus 1H

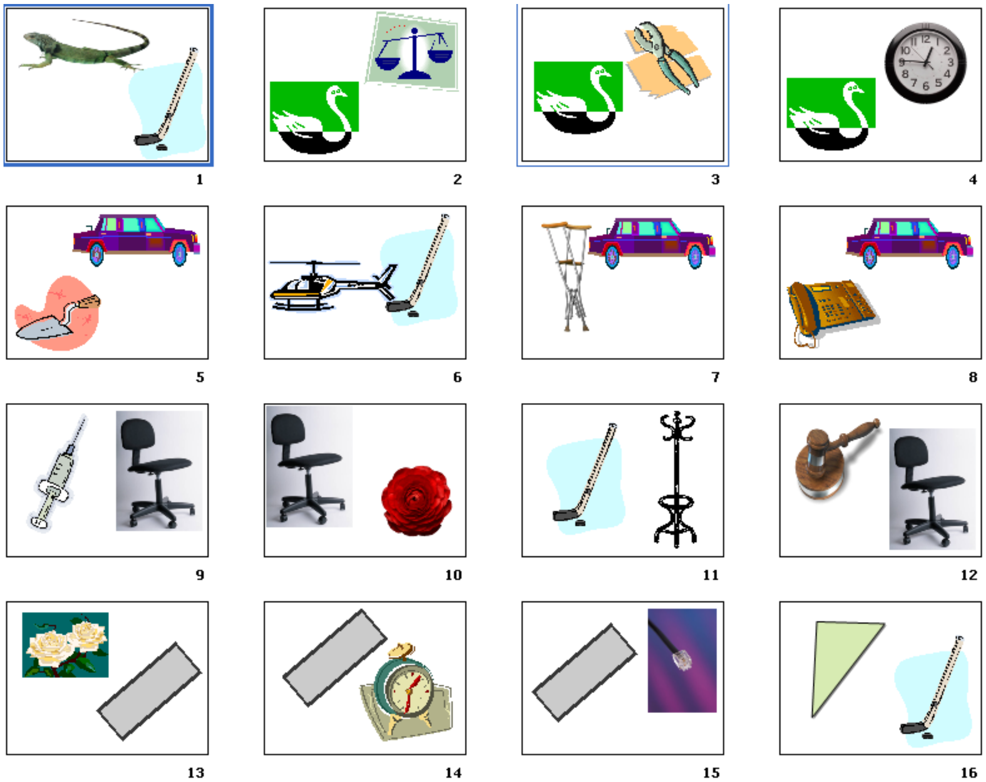


Stimulus 1I

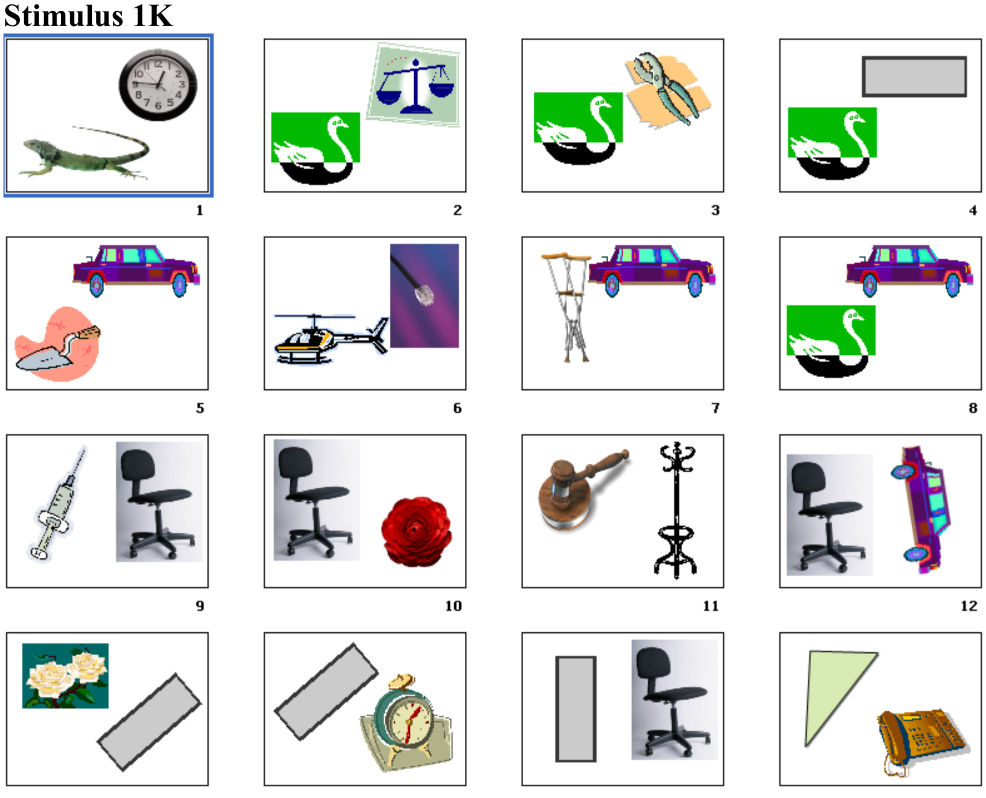


Appendix A: Experiment I: Continued

Stimulus 1J



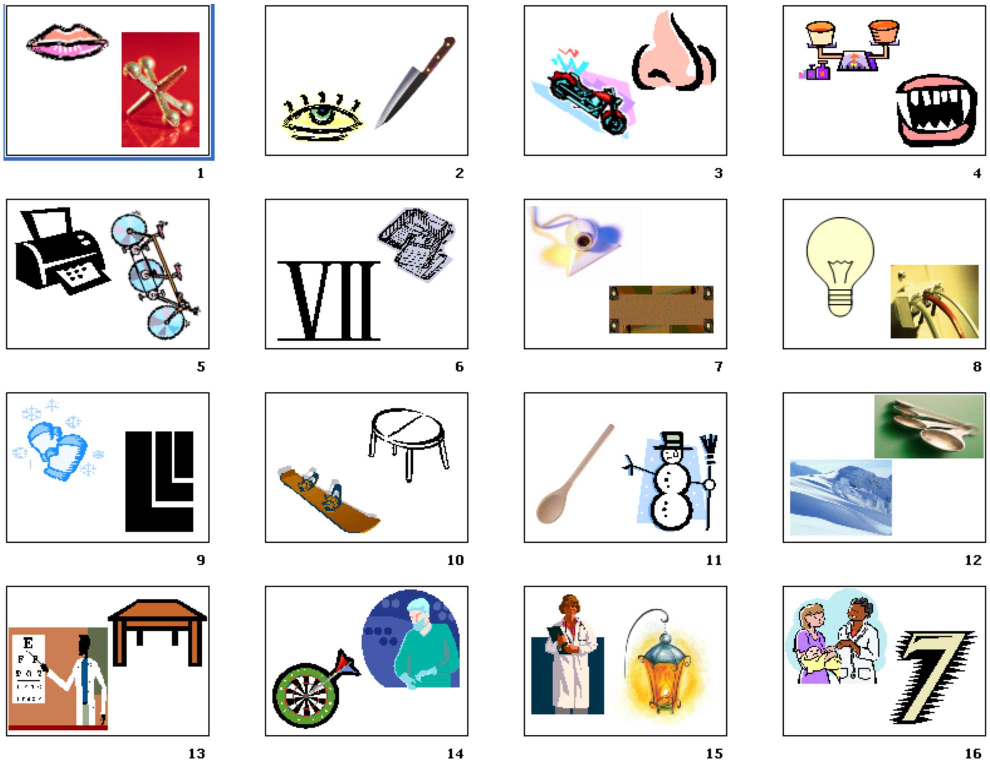
Stimulus 1K



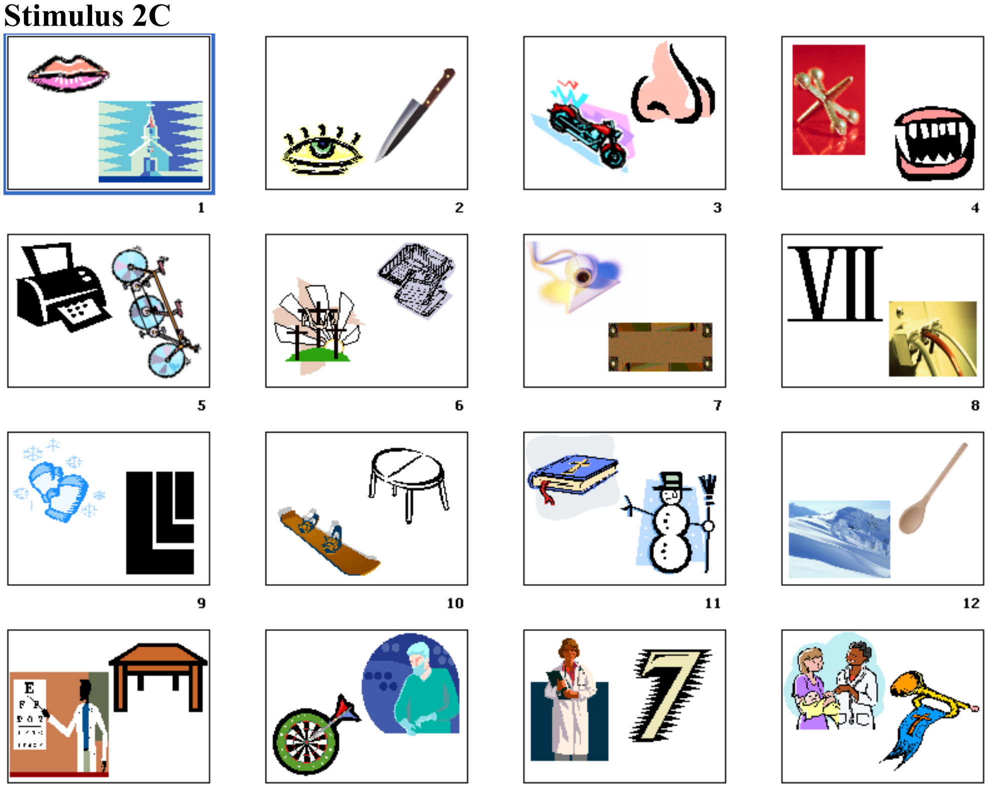
Appendix A: Experiment I: Continued

STIMULI SET 2

Stimulus 2B

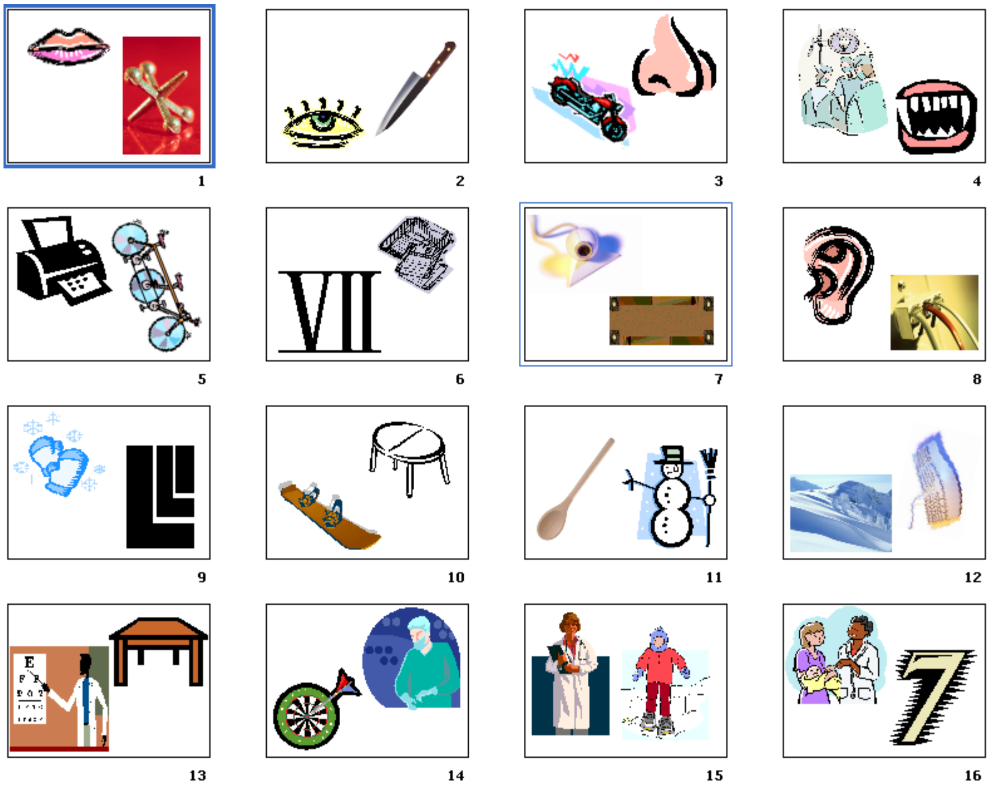


Stimulus 2C

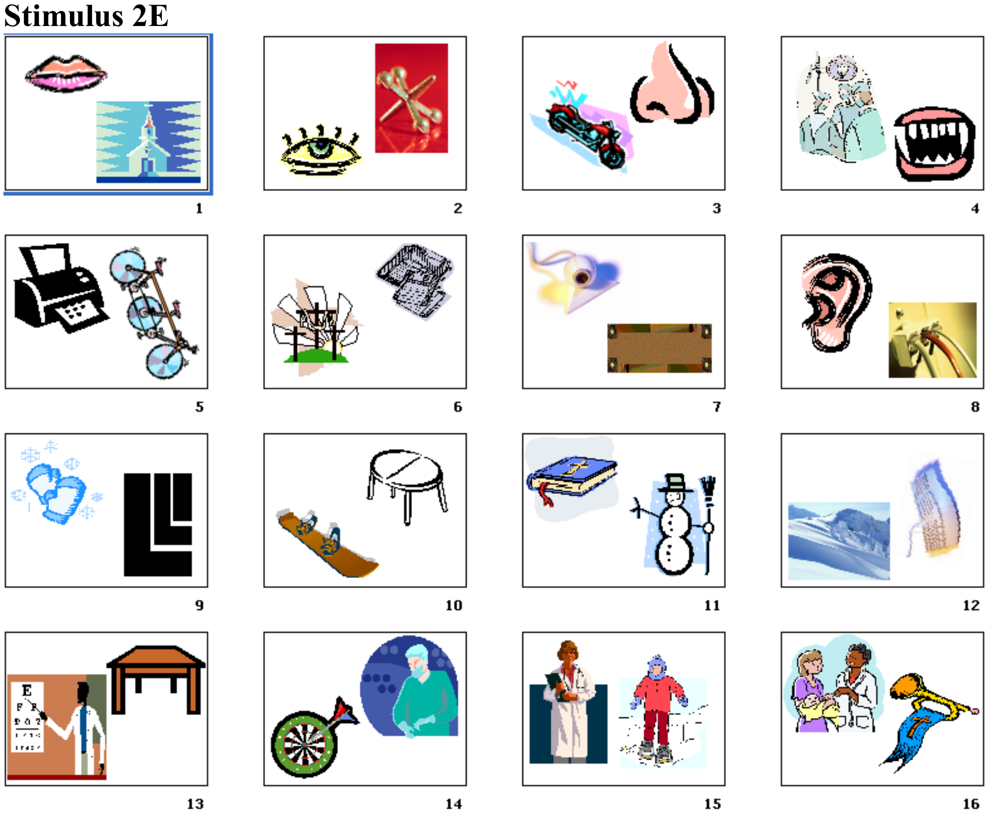


Appendix A: Experiment I: Continued

Stimulus 2D

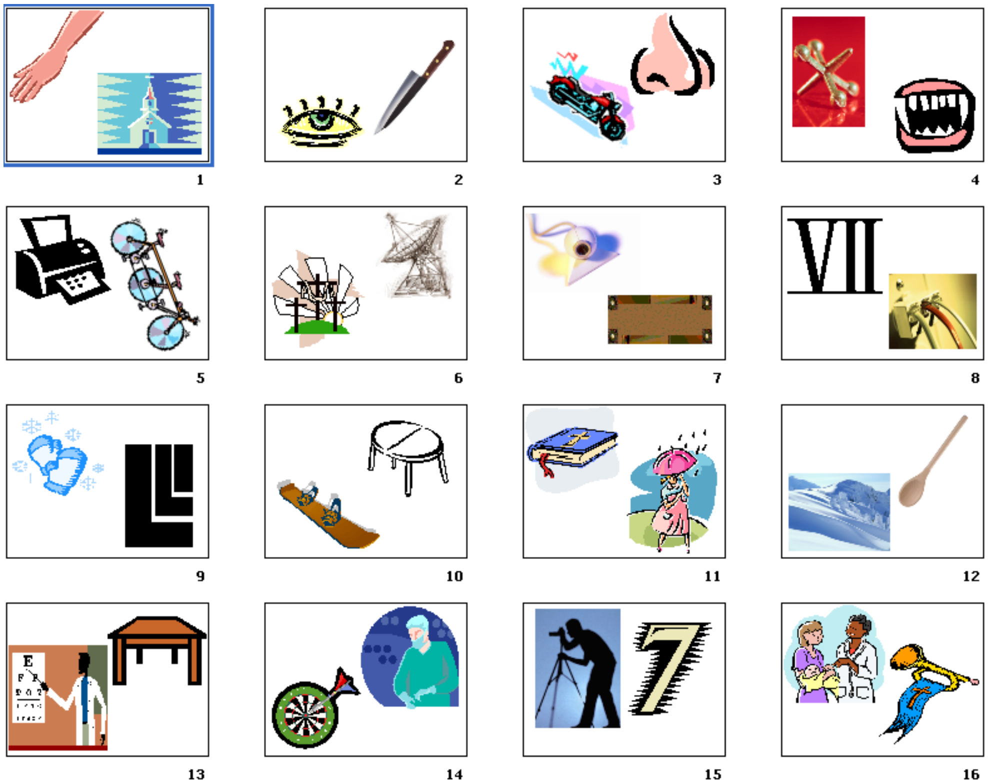


Stimulus 2E

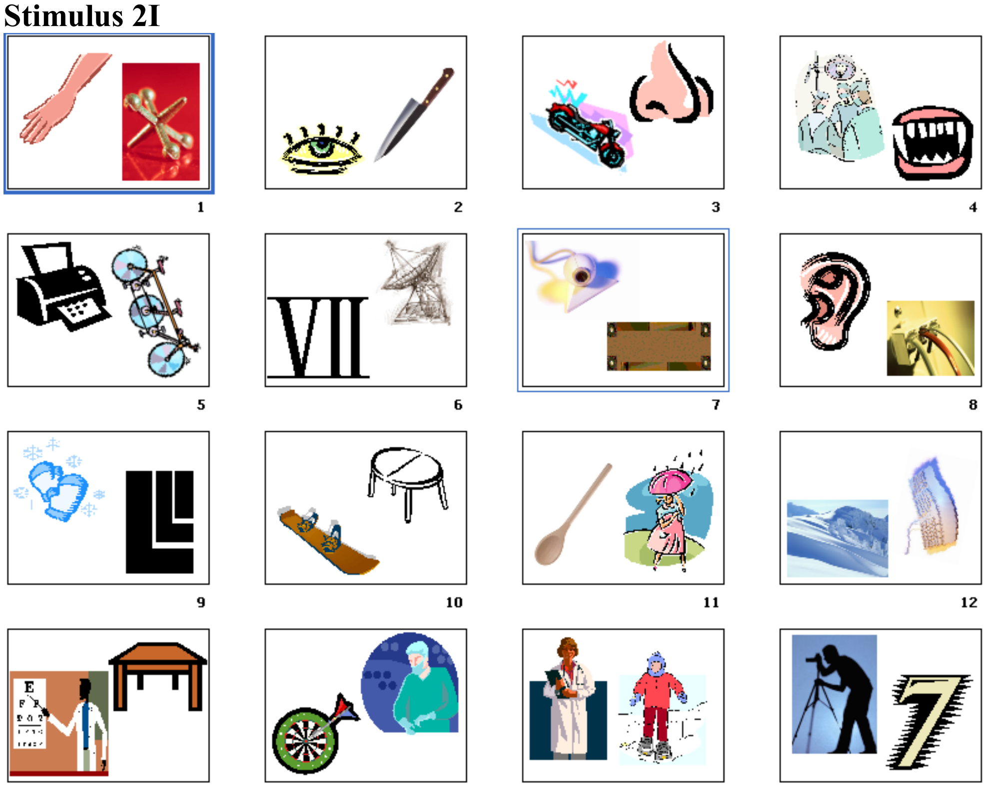


Appendix A: Experiment I: Continued

Stimulus 2H

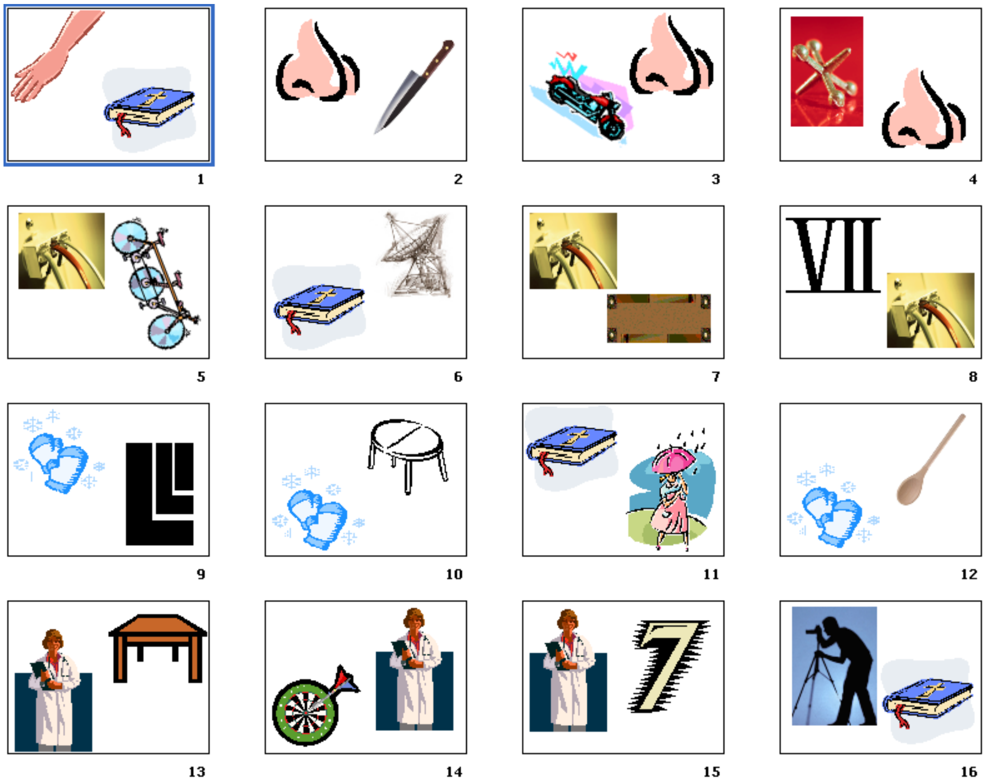


Stimulus 2I

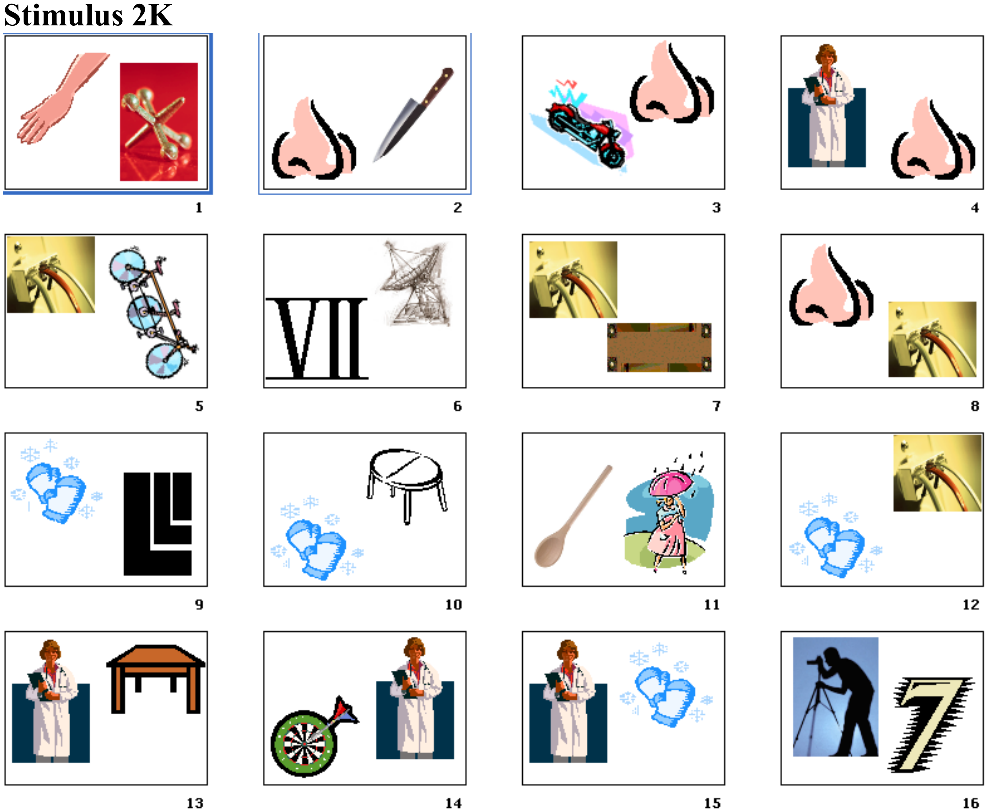


Appendix A: Experiment I: Continued

Stimulus 2J



Stimulus 2K



Appendix A: Experiment I: Continued

STIMULI SET 3

Stimulus 3B

Stimulus 3C

Appendix A: Experiment I: Continued

Stimulus 3D

Stimulus 3E

Appendix A: Experiment I: Continued

Stimulus 3H

Stimulus 3I

Appendix A: Experiment I: Continued

Stimulus 3J

Stimulus 3K

APPENDIX B: INSTRUCTIONS TO SUBJECTS OF EXPERIMENT I

Instructions

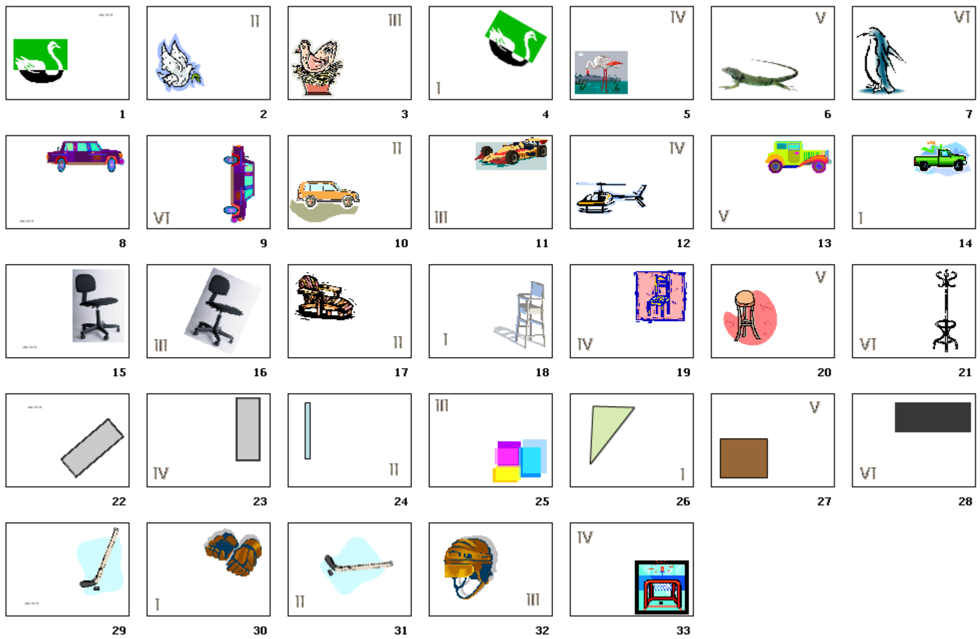
1. You will be performing three experiments and each completed experiment earns a bonus mark of one point.
2. These experiments comply with the UW ethics and research guidelines. All data collected will be used solely for research purposes. The films and videotapes will only be used in support of data analysis and during paper presentation(s) at research conferences.
3. 16 cards will be distributed equally among the 4 participants
4. The objective of this experiment is for each of you to come up with 4 cards of a kind. I will demonstrate this using the sample cards in front of you. The solution will be four cards of the same color. This is a much simpler version of the actual experiments to demonstrate the basic task.
5. You get half a mark if you get a group of 4 and you get a full mark if everyone on the group forms a group of 4 each.
6. There would be time limits for each task, 5 minutes for the first, 10 minutes for the second and 15 minutes for the third task.
7. Each participant will be required to fill a short questionnaire after each task.
8. Pls. exchange/trade cards but do not take over other subjects cards.

Thank you for your time and participation.

Good Luck

APPENDIX C: STIMULUS SET FOR EXPERIMENT II

Stimuli 1 Set



Stimuli 2 Set

Appendix B: Experiment II: Continued

Stimuli 3 Set

APPENDIX D: QUESTIONNAIRE FOR EXPERIMENT II

Group Research Questionnaire – Coding Systems

Rating Questionnaire

Name: ____

Student ID: ____

The experimenter will display 6 cards (numbered I – VI) each time with approximately 90 cards in total. You are required to rate the similarity or salience of each card in comparison with card indicated as the anchor for each set of cards.

1. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..…IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

2. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

3. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

4. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

Appendix D: Questionnaire for Experiment II: Continued

5. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

6. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

7. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

8. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

9. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

10. How similar are these cards in comparison with the anchor?

VERY DIFFERENT……………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10

II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

11. How similar are these cards in comparison with the anchor?

VERY DIFFERENT………………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

12. How similar are these cards in comparison with the anchor?

VERY DIFFERENT………………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

13. How similar are these cards in comparison with the anchor?

VERY DIFFERENT………………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

14. How similar are these cards in comparison with the anchor?

VERY DIFFERENT………………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

15. How similar are these cards in comparison with the anchor?

VERY DIFFERENT………………………………………………………..….IDENTICAL

I.	1	2	3	4	5	6	7	8	9	10
II.	1	2	3	4	5	6	7	8	9	10
III.	1	2	3	4	5	6	7	8	9	10
IV.	1	2	3	4	5	6	7	8	9	10
V.	1	2	3	4	5	6	7	8	9	10
VI.	1	2	3	4	5	6	7	8	9	10

APPENDIX E : SPEARMANS RANK CORRELATION TESTS

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.362	.379
		Sig. (2-tailed)	.	.203	.181
		N	14	14	14
	Area Difference	Correlation Coefficient	.362	1.000	.506
		Sig. (2-tailed)	.203	.	.065
		N	14	14	14
	Card Exchanges	Correlation Coefficient	.379	.506	1.000
		Sig. (2-tailed)	.181	.065	.
		N	14	14	14

STIMULUS B

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.362	.379
		Sig. (2-tailed)	.	.203	.181
		N	14	14	14
	Area Difference	Correlation Coefficient	.362	1.000	.506
		Sig. (2-tailed)	.203	.	.065
		N	14	14	14
	Card Exchanges	Correlation Coefficient	.379	.506	1.000
		Sig. (2-tailed)	.181	.065	.
		N	14	14	14

STIMULUS C

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.903**	.930**
		Sig. (2-tailed)	.	.000	.000
		N	14	14	14
	Area Difference	Correlation Coefficient	.903**	1.000	.841**
		Sig. (2-tailed)	.000	.	.000
		N	14	14	14
	Card Exchanges	Correlation Coefficient	.930**	.841**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	14	14	14

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix E: Spearman's Rank Correlation Test: Continued

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.809**	.826**
		Sig. (2-tailed)	.	.000	.000
		N	16	16	16
	Area Difference	Correlation Coefficient	.809**	1.000	.889**
		Sig. (2-tailed)	.000	.	.000
		N	16	16	16
	Card Exchanges	Correlation Coefficient	.826**	.889**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	16	16	16

**. Correlation is significant at the 0.01 level (2-tailed).

STIMULUS E

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.711**	.855**
		Sig. (2-tailed)	.	.006	.000
		N	13	13	13
	Area Difference	Correlation Coefficient	.711**	1.000	.850**
		Sig. (2-tailed)	.006	.	.000
		N	13	13	13
	Card Exchanges	Correlation Coefficient	.855**	.850**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	13	13	13

**. Correlation is significant at the 0.01 level (2-tailed).

STIMULUS H

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.860**	.740**
		Sig. (2-tailed)	.	.000	.006
		N	12	12	12
	Area Difference	Correlation Coefficient	.860**	1.000	.881**
		Sig. (2-tailed)	.000	.	.000
		N	12	12	12
	Card Exchanges	Correlation Coefficient	.740**	.881**	1.000
		Sig. (2-tailed)	.006	.000	.
		N	12	12	12

**. Correlation is significant at the 0.01 level (2-tailed).

STIMULUS I

Appendix D: Spearman's Rank Correlation Test: Continued

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.805**	.873**
		Sig. (2-tailed)	.	.003	.000
		N	11	11	11
	Area Difference	Correlation Coefficient	.805**	1.000	.863**
		Sig. (2-tailed)	.003	.	.001
		N	11	11	11
	Card Exchanges	Correlation Coefficient	.873**	.863**	1.000
		Sig. (2-tailed)	.000	.001	.
		N	11	11	11

**. Correlation is significant at the 0.01 level (2-tailed).

STIMULUS J

Correlations

			Rank Average	Area Difference	Card Exchanges
Spearman's rho	Rank Average	Correlation Coefficient	1.000	.753**	.823**
		Sig. (2-tailed)	.	.003	.001
		N	13	13	13
	Area Difference	Correlation Coefficient	.753**	1.000	.506
		Sig. (2-tailed)	.003	.	.078
		N	13	13	13
	Card Exchanges	Correlation Coefficient	.823**	.506	1.000
		Sig. (2-tailed)	.001	.078	.
		N	13	13	13

**. Correlation is significant at the 0.01 level (2-tailed).

STIMULUS K

APPENDIX F: High Frequency Test using Chi-Square Goodness of Fit Test

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Decision	122	.61	.489	0	1

Decision

	Observed N	Expected N	Residual
0	47	61.0	-14.0
1	75	61.0	14.0
Total	122		

Test Statistics

			Decision
Chi-Square[a]			6.426
df			1
Asymp. Sig.			.011
Monte Carlo Sig.	Sig.		.016[b]
	99% Confidence Interval	Lower Bound	.012
		Upper Bound	.019

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 61.0.

b. Based on 10000 sampled tables with starting seed 803356894.