Open-Ended Problem Solving in Groups

by

Hanan Alattas

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Management Sciences

Waterloo, Ontario, Canada, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Simon Taggar Professor, Lazaridis School of Business and Economics, Wilfrid Laurier University
Supervisor(s)	Rob Duimering Associate Professor, Dept. of Management Sciences, University of Waterloo
Internal Member	Frank Safayeni Professor, Dept. of Management Sciences, University of Waterloo
Internal Member	Ada Hurst Continuing Lecturer, Dept. of Management Sciences, University of Waterloo
Internal-external Member	Vanessa Schweizer Associate Professor, Dept. of Knowledge Integration, University of Waterloo

Author's Declaration

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

Previous experimental research on problem-solving has predominantly investigated well-structured problems with predefined solutions. Studies of ill-structured, open-ended problem-solving have primarily employed observational and case study methods. This study used a controlled experiment in which groups solved ill-structured categorization problems to investigate effects of problem open-endedness on problem-solving behaviors and solution outcomes. The experimental design enables precise measurement and tracking of open-ended problem-solving behaviors.

In the experiment, N=48 four-person groups solved three categorization problems, in which they grouped 16 randomly selected pictures into 4 categories of 4 pictures each. Task goals and participant beliefs were varied to create three levels of problem open-endedness. In two tasks, participants grouped pictures based on similarity, and their open-endedness beliefs were altered based on instructions suggesting either that a single best solution identified by experts should be found ("Expert"; least open-ended), or that multiple solutions were available, and a "good" solution should be found ("Good"; more open-ended). In a third task, participants grouped pictures by creating 4 simple stories involving the items ("Story"; most open-ended). The experiment investigated effects of the degree of problem open-endedness on several indicators of problem-solving behavior and properties of the solution, including problem-solving difficulty, the variability of solutions produced by different problem-solving groups, the influence of initial conditions on solutions (path dependency), the strength of concept association in solutions, structural moves toward solutions, and the variability of problem-solving search behavior.

ANOVA results across the three levels of open-endedness confirmed hypothesized negative effects of problem open-endedness on task difficulty and variability in problem-solving behavior, as well as positive effects on solution variability, path dependency, and the strength of solution association. The results also provided evidence that solutions to open-ended problems are non-random. Post-hoc pairwise comparisons between open-endedness levels partially supported our hypotheses. Differences between the similarity and story tasks strongly supported hypotheses; however, differences between the two (least open-ended vs. more open-ended) similarity tasks were mainly non-significant although the distribution means varied in the predicted directions. Regarding structural progress towards a solution, participants in the least open-ended "Expert" condition first formed categories based on the strongest associations between items, then moved to progressively weaker associations. This effect

was less prominent in the more open-ended “Good” condition and absent in the most open-ended “Story” condition.

A verbal protocol analysis conducted on nine experimental tasks provided further insights into the problem-solving process across three conditions. A prominent pattern of behavior observed in all conditions was iterative conflict recognition and resolution until groups reached a satisfactory solution. In the “Expert” condition, groups exhibited more conflict recognition and resolution iterations, more emphasis on the logic behind requested picture exchanges and more resistance to accepting proposed resolutions, compared to the Good and Story conditions. Individual group members tended to develop partial solutions independently and simultaneously in the Story condition, whereas partial solutions were developed collectively in a sequential manner in the similarity conditions.

Acknowledgements

I would like to express my sincere gratitude to Professor Rob Duimering for his guidance, insightful feedback, and continuous support during my Ph.D. studies. I am tremendously grateful that he agreed to take on this demanding role. Without Rob by my side, helping to illuminate the path, the journey would have been much more challenging. His expertise and motivation have been invaluable in shaping the direction of my research and making my Ph.D. journey a truly enjoyable and fulfilling experience. I am honored to have collaborated with him for several years.

I would also like to extend my sincere thanks to my thesis committee members: Professor Frank Safayeni, Professor Ada Hurst, and Professor Vanessa Schweizer from the University of Waterloo, and Professor Simon Taggar from Wilfrid Laurier University, for their helpful feedback on my thesis.

Furthermore, I would like to thank my colleagues and friends at the University of Waterloo. A special thanks to Aliaa Alnaggar, Remzyie Kitablty, Haya Almutairi, and Muna Alateibi. Thank you for your endless support and encouragement during the highs and lows of my Ph.D. journey; thank you for becoming a second family to me. Many thanks to my colleagues in the Department of Management Sciences: Ahmad Tanehkar, Fuat Can Beylunioglu, Arielle Grinberg, Gbemi Abimbola, Sofiia Doroshenko, and Dare Adeyeri. I have learned from each of you. Also, my gratitude to all the participants who took part in the study and made this research possible.

To my beloved parents, Saleh Alattas and Maryam Balfakhr, thank you for supporting me in all aspects of life, providing me with emotional strength, and giving me the means to start my academic path. Your unwavering belief in me and your prayers have been a constant source of strength. I will always be indebted to you. To my siblings Mohamad, Arwa, Abdullah, Shahad, and Albatoul, thank you for your unwavering support and ongoing love. Being away from you has been the most challenging aspect, but your faith in me is the cornerstone of all my achievements.

To my beloved husband, Mohsen, your support and encouragement have been indispensable on this journey. Your unwavering belief in me and the steadfast backing you provided during challenging times were pivotal in reaching this point. Your motivation was a guiding light when my enthusiasm waned, and I contemplated giving up. I am immensely grateful for everything you've contributed to the success of this journey. I express my deepest gratitude for standing by my side.

To my wonderful children, Abdullah, Al-Hussain, and Saleh, you've been my best companions on this journey. Your support, love, hugs, laughter, and jokes have made tough times easier. A special thanks to you for your patience and assistance during the pandemic while I conducted experiments from home. Over the years of my Ph.D., you've grown, and so have I, learning so much from you. You are the light of my life.

To everyone who has played a part in my academic journey, I extend my heartfelt thanks.

Finally, I would like to acknowledge the support of the Ministry of Education in Saudi Arabia.

Dedication

To my dear parents and siblings, my beloved husband Mohsen and my lovely kids Abdullah, Al-Hussain, and Saleh.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Problems arise when individuals perceive a difference between their present state and a desired goal state, without possessing a clear understanding of the means to bridge this discrepancy (Duncker, 1945). Consequently, individuals engage in searching for operators or strategies to move from the initial state to the intended goal state (Dunbar, 1998; Newell & Simon, 1972). However, certain problems are ill-structured (Reitman 1964), and introduce an element of ambiguity, where individuals recognize the existence of a problem but encounter uncertainty pertaining to how the goal state should look as well as the type of operations or method required to reach this goal. Such problem instances can manifest in various contexts, such as writing a story or designing a product, where inherent ambiguity dominates regarding how the story should end or how the product would look. These, and many other real-world problems, can be characterized as open-ended, where there is no single correct solution (Goel, 1995), but multiple feasible or satisfactory solutions (Dorst & Cross, 2001).

Previous research on problem solving has predominantly investigated the cognitive processes involved in solving close-ended problems with a single correct solution, or problems with clear-cut solution alternatives. Notable examples used in experimental studies include chess problems (Newell & Simon, 1972), the Towers of Hanoi (Dunbar, 1998), and the nine-dot problem (Maier, 1930). However, minimal focus has been given to the cognitive processes associated with open-ended problem solving. This might be due to methodological challenges associated with the complexity of open-ended problems. From a methodological point of view, the variables that affect open-ended problem-solving behavior are complicated and interrelated, which in turn leads to difficulties controlling them experimentally and identifying which variable is causing the effect.

Design problems are a prominent category of ill-structured open-ended problems that has received considerable research attention, but studies of design problem-solving have mostly taken an observational approach without employing controlled experiments. This further highlights the methodological limitations in comprehensively understanding the cognitive processes involved in tackling open-ended problems.

To address this gap in the literature, the current study experimentally investigates the behavior and process involved in solving open-ended, ill-structured problems. Through its unique experimental

design, this study specifically seeks to examine the effects of task open-endedness on problem-solving behavior and the types of solutions generated.

In our study, groups of four participants were given a sequence of three open-ended categorization problems. In each problem, participants had to group 16 pictures into four sets of four pictures each, based on specific relationships between the pictures. In each task, the pictures were randomly selected so the problems were open-ended and there was no correct solution. We further manipulated the degree of open-endedness experimentally in two ways. One manipulation used different categorization goals, asking participants to group items based on similarity, or by creating simple stories about the pictures. We also manipulated participants' beliefs about the open-endedness of the problem, suggesting either that they should seek a single best solution selected by experts, or that there was no correct solution and participants should seek a good solution. Using these manipulations, we established three levels of problem open-endedness. In two tasks, participants grouped items based on the similarity of the pictures, aiming to find either the "Expert" solution (the least open-ended) or a "Good" solution (more open-ended). In the third task, participants created four simple stories, each incorporating four of the pictures, as the basis for grouping ("Story" task; the most open-ended).

The experiment investigated the effects of problem open-endedness on various aspects of problem-solving behavior and the solutions obtained. Specifically, we tested hypotheses related to the effects of problem open-endedness on task difficulty, the variability of solutions produced by different problem-solving groups, the influence of initial conditions on solutions (i.e., path dependency), the strength of concept association in solutions, structural moves toward solutions, and the variability of problem-solving search behavior. We also used qualitative verbal protocol analysis to explore the cognitive processes involved in solving open-ended problems.

The remainder of this thesis is organized as follows:

Chapter 2 reviews literature on problem-solving, ill-structured open-ended problems, group problem solving, and categorization.

Chapter 3 introduces the study and explains the theoretical framework and the hypotheses.

Chapter 4 explains the experimental methodology.

Chapter 5 presents the results of hypothesis testing.

Chapter 6 presents the verbal protocol analysis.

Chapter 7 discusses the overall findings of the study, addresses its limitations, and proposes potential avenues for future research.

Chapter 2
Literature Review

This chapter provides a comprehensive literature review that encompasses various theories of problem-solving. The literature on two major theories of problem solving, namely Gestalt and information processing, is discussed. Furthermore, the literature comparing ill-structured problems and well-structured problems is reviewed, with a particular emphasis on design problems as a prominent example of ill-structured problems. In the context of group problem-solving, some representative literature is examined, including recent studies of the effects of problem structure on problem solving behavior, which influenced the design of the current experiment (Abimbola, 2006; Adejumo et al., 2008; Chen, 2010). The problem task used in the experiment is a categorization task, and therefore, relevant literature on categorization is also reviewed.

2.1 Overview of Problem-Solving

The literature on problem-solving is presented in terms of two main domains: Gestalt theory and information processing theory. Each approach views problem-solving from a different angle. The Gestalt approach emphasizes the importance of having a proper problem representation in problem-solving and the role of perceptual restructuring. The information processing approach views problem-solving as searching through a problem space using operations to reach a solution.

2.1.1 Problem-solving from the Gestaltist Perspective

Gestalt is a German word that means a configuration or pattern. Gestalt psychology was introduced in Germany in the early 1900s. This theory emphasizes how humans perceive the world in terms of holistic patterns organized as figure against a background, and the perceived whole is different from the sum of its parts (Duncker, 1945; Köhler, 1969; Wertheimer, 1982). Perceptual organization follows the Gestalt laws of perception, where larger wholes are perceived based on similarity, continuity, closure and proximity.

Wertheimer (1982) explained problem-solving as starting with an unsuitable holistic view that an individual holds for a certain situation, which initially prevents him/her from solving the resulting problem. The solution to the problem is eventually achieved when the individual understands the

structure and constraints of the situation. This theory emphasizes the role of perceptual restructuring and understanding the structure of the problem to be able to solve it. It places emphasis on how the problem solver perceives the problem in the first place and how this initial perception leads to difficulties in finding a solution. It also highlights the experience of insight characterized by the sudden restructuring of the problem and achievement of the solution, known as the “aha!” moment.

A classic example of an insight problem is the nine-dot problem (Maier, 1930). In this task, a problem solver is presented with a picture of nine dots that are exhibited as a set of three rows of three. The problem is to connect all the dots with four straight lines without taking the pencil off the page.

The mechanism explained by Gestalt theory to solve this problem is based on perceptual restructuring. That is, individuals initially perceive the structure of the problem as a 3x3 square, so they draw lines within the boundaries of that square, which makes it impossible to reach a solution. Thus, a restructuring of the situation is needed, where the person recognizes that they can draw lines outside the boundaries of the square to reach the solution, as shown in Figure 1.

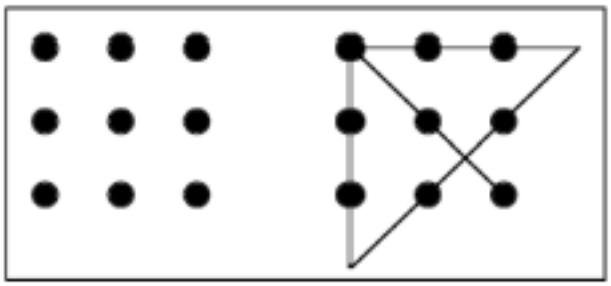


Figure 1: The nine-dot problem

Gestaltists emphasized the importance of restructuring in problem-solving and examined specific difficulties that arise during this process, such as functional fixedness, which refers to the tendency of perceiving an object as having only its traditional or usual function, making it difficult to see its potential for serving different or alternative purposes. Functional fixedness was presented in Duncker’s (1945) classic candle experiment, in which participants were given a candle, a box of thumbtacks, and a book of matches, and asked to fix the lit candle to the wall in such a way that the candle would not drip wax onto the floor. One group of participants was presented with each of the items separated from one another, while the other group received the thumbtacks inside their box. The

first group was able to solve the problem by tacking the box to the wall and using a few drops of wax to attach the candle to the inside of the box. In contrast, most of the participants in the other group were unable to solve the problem since they viewed the box of thumbtacks as a whole and were unable to restructure their perception of the box as a container for tacks, to see the box separately from the tacks as a potential holder for the candle.

In terms of the effect of prior knowledge on how to solve a problem, another concept that Gestaltists discuss is that of Einstellung, which refers to applying a procedure that is previously learned when there exists a simpler way of solving the task. An example of this is the water jar experiment by Luchins and Luchins (1950). In this experiment, participants were asked to solve a series of water jar problems that required similar steps to be solved. After solving the problems, participants were presented with easier problems. However, they applied the same steps that they used in the former problems to solve the latter even though simpler steps to solve these latter problems existed. In both cases, it can be observed that there is a structural relation built into our knowledge of parts which causes the parts to be combined into a whole in a way that is not suitable for the problem in hand.

Gestalt theory emphasizes the holistic view and the correct understanding of the task structure to solve a problem. However, this theory is subject to several limitations. One shortcoming of this approach is its inability to explain the cognitive process through which restructuring takes place. The theory describes the sudden insight or the “aha” moment without providing a mechanism or clear measurements of how this insight process works (Kohler, 2015). Furthermore, the Gestalt approach is mainly emphasizing insight problems, where the solution can be found through a sudden change in the representation of the problem structure. Thus, this raises another limitation of this approach; insight does not always characterize all types of problems. As an example, a problem that is solved through incremental steps, such as an algebra problem, cannot be properly explained through Gestalt theory.

2.1.2 Problem-solving from the Perspective of Information Processing Theory

Information processing theory was developed by Newell and Simon (1972) outlining three main components for explaining human problem solving. The first component involves the human being seen as an information processing system, which involves a physical manipulation of symbols, memory storage, a processor, sensory receptors, and motor effecters. The second component is the

task environment that contains the problem, goal, and other external factors that are objectively defined by the experimenter. Finally, the third component is the problem space, which is subjectively defined by the problem solver, and is shaped by the interaction of subjective and objective constraints. This problem space can be expressed through state space, operators, evaluation functions, and search strategies. The problem space can be viewed as containing many nodes, or knowledge states, reflecting the knowledge that the problem solver attains at a specific moment in time. The problem solver connects one node in the problem space to another until they reach a knowledge state that includes the solution, i.e., reaches the answer.

A classic problem investigated from the information processing perspective is the Towers of Hanoi task (Dunbar, 1998). In this task, an individual is given a board with three pegs and several disks of decreasing size that are placed on the left-most peg. The problem solver is asked to move all the disks onto the right-most peg, where disks can be moved one at a time, and larger disks cannot be placed on top of smaller ones.

This task can be explained in terms of information processing theory (Dunbar, 1998; Öllinger & Goel, 2010). The task environment can be defined by the given problem with objective constraints (instructions inherent in the problem). The problem space contains different states as well as operators. The operators in this task are the actions of moving the disks between pegs, while considering the disk movement instructions provided above, and applying the operators to the initial state results in changing the state to an intermediate state. In this intermediate state, evaluation takes place, where the problem solver tests if the path selected leads closer to the goal state. If not, they should go backward several steps. To reach the final state, the problem solver may use one or more of the several strategies or heuristics discussed in the literature. Examples of such strategies are means-ends analysis and the hill climbing technique (Simon & Newell, 1971).

Information processing theory provides a clear description of the process of human problem-solving. However, its proposed model emphasizes the information-processing aspect of the solution process and ignores other aspects of the process that are associated with the nature of human thinking. For instance, the dynamics of tension and frustration associated with encountering impasses or blind alleys during the process of problem-solving are not well explained by the theory (Derbentseva, 2007). Another aspect of human problem-solving that creates a challenge for information processing theory is the cognitive restructuring embodied in sudden insight or the “aha” moment (Öllinger &

Goel, 2010). Thus, in the information processing approach, there is an implicit assumption that the problem structure is already correctly perceived by the problem solver; that is, the problem solver is searching in the correct problem space. However, an individual who has the wrong conception of a problem will search in the wrong space, hence be unable to find a solution. Thus, cognitive restructuring is needed to formulate the correct problem space. This process of changing from the wrong problem space to the correct space is not properly accounted for by the theory.

2.1.3 Comparison between Information Processing and Gestalt Theories

In comparing information processing theory to the Gestalt theory, it could be noticed that the information processing theory postulates a constructed problem space, while the Gestalt theory emphasizes the way individuals create and restructure their representation of the problem itself in order to obtain insightful solutions (Chen, 2010). Moreover, as previously mentioned in Section 2.1.1, not all types of problems can be explained clearly by either theory. As an example, algebra problems can be well explained by the information processing approach, whereas Gestalt theory can better illustrate other types of problems, such as insight problems exemplified by the nine-dot task. Finally, while the information processing system focuses on the incremental movement toward the solution, the Gestalt approach seems to depend highly on the visual and mental representation of the problem.

Regarding commonalities between the two theories, both agree that problem-solving is composed of subjective and objective representations of a given problem. However, how the theories represent those subjective and objective components differs. Information processing theory explains the problem's objective representation in terms of the task environment, while the subjective representation is presented in terms of problem space. Gestalt theory emphasizes the relationship between the problem solver's subjective representation of the problem structure and the objective problem situation. Problem solving difficulty arises when the perceived problem structure does not reflect the objective structure needed to reach the solution. Only after perceptual restructuring is the problem solver able to reach the objective solution.

Having highlighted the two dominant theoretical views in the literature of problem solving, the following section discusses the distinction between well and ill-structured problems as they relate to the open-ended problems investigated in the present work.

2.2 Ill-Structured Open-ended Problems

Most research in the literature investigates well-structured problems with a single solution. However, problems experienced by individuals in everyday life usually have various possible solutions and multiple possible operations can be used to reach a solution. Such problems are classified as ill-structured, open-ended problems. In this section, we will review ill-structured problems as they relate to the open-ended problems in the present work. We will highlight the distinction between ill- and well-structured problems, then will review design problems as a relevant example of ill-structured open-ended problems.

2.2.1 Ill-Structured Versus Well-Structured Problems

Reitman (1964) classified problems as ranging from ill-structured to well-structured. This classification is based on the availability of information in each of three components of the problem: the start state, the goal state and the transformation function describing how the problem solver moves from the start to the goal (Reitman 1964). When the information content of each of the three components is completely specified, the problem is defined as well-structured. A relevant example is the Towers of Hanoi problem, where the start and goal state are clearly defined (i.e., when all the disks are stacked in descending order on the right-most peg for the start state, and on the left-most peg for the goal state) and the operations are specified in terms of all the constraints given in the problem instructions.

When information content in one of the three components is incomplete or unavailable, the problem is categorized as an ill-structured problem. This can be exemplified by the task of preparing dinner for guests (Öllinger & Goel, 2010). In terms of the start state, some of the relevant information, such as how hungry the guests are, cannot be clearly specified. In terms of the goal state, it is difficult to articulate the success of the meal and how satisfied the guest will be. Finally, the transformation functions, such as whether the food will be ordered or cooked at home, are not specified. Such incomplete information in the components of the problem classifies it as an ill-structured problem

Whereas close-ended problems have one specific solution, ill-structured open-ended problems have multiple possible solutions, with some being better than others. In terms of information processing theory, it can be assumed that the problem space of a well-structured problem would be smaller and have limited paths leading to the solution. In contrast, the problem space in an ill-structured problem

would be more complicated as it allows for a diverse set of solutions and, consequently, wider possibilities of paths than well-structured problems.

Research in various fields has examined ill-structured problems for different purposes. In the field of education, for example, studies have investigated the skills required to solve such ill-structured problems (Becker & Shimada, 1997), as well as the differences in problem-solving approaches between experts and novices (Schunn, McGregor & Saner, 2005). In public health, a study by Sarsfield (2014) explored how novices and expert nurses tackled ill-structured problems, revealing that possessing domain-specific knowledge facilitated superior solutions. On the other hand, functional knowledge can inhibit creativity by introducing constraints that prevent individuals from thinking outside the boundaries of that knowledge (Wieth and Francis, 2018).

Ill-structured problems have been discussed in the literature of "complex problem solving". Complex problems, as defined by Fischer, Greiff & Funke, (2012), are problems characterized by the dynamics of the situation, the complex structure, interrelated variables involved, and ambiguity in the way to solve the problems. An iterative two-stage process of solving such problems was proposed by these authors. The first phase is knowledge acquisition, where the problem solver gains knowledge through the exploration of the problem. The second phase is goal-oriented knowledge application, when the problem solver uses the acquired knowledge to predict the behavior of the problem and evaluate the selected solution from amongst the available alternatives.

The following section focuses on literature devoted to design problem-solving, which is a relevant example of open-ended problems.

2.2.2 Design Problems as Open-ended Problems

Design problems are well-known examples of open-ended and ill-structured problems. This section considers some of the distinctive features of design problems and provides a review of relevant research into the process of solving these problems.

Scholars have described design problems as open-ended problems with certain characteristics. One of the distinctive features in design problems is the adaptable nature of constraints. For example, Goel (1995) discussed two types of constraints that influence design problem solving: rules that are negotiable and natural laws that are never negotiable. Similarly, in studies of mechanical engineering design, Ullman and Dietterich (1988) noted that some constraints are given with the problem, while

others are introduced by the domain knowledge of the designer or derived by the designer while exploring solutions. Another unique feature in design problem solving is the continuous redefinition of new task goals. Akin (1978) reported that architects exercise the freedom to change both goals and constraints as they gradually develop their comprehension of the problem and definition of a solution. The situated nature of constraints implies that solutions are often domain-specific and not easily transferred to other settings (Jonassen, 2011). As new constraints are encountered during the solution process, designers may need to reframe and restructure the problem to resolve trade-offs between conflicting goals and requirements (Nickel, Duimering & Hurst, 2022), leading to the co-evolution of problem and solution over time (Maher, 1994; Maher & Poon, 1996, and Dorst & Cross, 2001)

Protocol analysis (Ericsson & Simon, 1984) is a methodology that has been widely used to examine the design process where participants are instructed to think aloud while solving design problems providing potential insight into the cognitive processes involved in design problem solving. This method has been introduced in design studies by Eastman (1970), and since then, it has been used by other researchers to study different design phenomena (Cross, 2001; Christensen & Ball, 2014).

In his book "Sketches of Thought" (1995), Goel discussed the process of solving design problems based on a verbal protocol analysis involving three designers: an architect, a mechanical engineer, and an instructional designer. Goel described the process of solving such problems as a gradual move from an ill- to a well-structured problem. Goel defined four stages through which the designers progress while solving the design problem: problem structuring; preliminary solutions; refinement; and detailing of solution. As the problem solver proceeds from the preliminary to the detailing phases, the problem becomes more structured. Goel observed that the problem solver conducts an extensive structuring of the problem at the beginning stage of problem solving. The structuring stage is necessary due to the lack of information in ill-structured problems. The preliminary design stage subsequently takes place where the problem solver generates a few core ideas and performs a lateral transformation between ideas. In other words, the problem solver moves from one idea to a slightly different idea rather than a more detailed version of the same idea, which results in widening and exploring the problem space. In this stage, there is a lower degree of commitment to the generated ideas. In the later refinement and detailing stages, the design becomes more constrained and there is a higher degree of commitment to the generated ideas. Also, a large number of vertical transformations between ideas were observed in the later stages where the problem solver moves from one idea to a more detailed version of the same idea. This in turns results in narrowing down the problem space

and eventually reaching a final design. Goel noticed that problem restructuring constitutes as much as 30% of the overall task time. In this protocol analysis, an extensive problem structuring was observed during the first stages of the task, but the restructuring constantly reoccurs until almost the last phases of the task.

Other studies also explained the process of solving design problems by analyzing the think aloud protocols into different behavioral categories. For example, in an investigation of an individual designer and a team of three designers, Gunther et al. (1996) coded the design episodes by analyzing the designers' statements into three categories corresponding to different stages of design. The first stage is clarifying the task, where the problem solvers understand the task and collect information about its requirements. In the second stage, searching for concepts, the designers search for different principal solutions for the subfunction of the design problem. In the third stage, they first evaluate, then select from these solutions and combine them to achieve an overall concept for the design problem. In the final fourth stage, fixing the concept, the problem solvers develop and optimize the concept while considering the technical and economic criteria. At this stage, the designers produce a hand-drawn sketch of the design. Gunther noticed considerable iteration between different stages throughout the design session, especially in the first half where designers extensively iterate between the first and second stages of clarifying the task and searching for concepts. In a study on how different professions solve ill-structured problems, Fernandes and Simon (1999) coded think-aloud protocol transcriptions into basic cognitive chunks. These chunks included actions such as recall, read, infer, evaluate and recommend.

Design problems are ill-structured problems characterized by ambiguity and incomplete information. Thus, the designer may not fully understand the constraints affecting design problems, especially in the initial stages of solving the problem. They need to develop a better understanding of the problem and the operations needed to reach a satisfactory solution. One means of developing an understanding of such ill-structured problems is to start solving the problem by considering initial solutions (Kolodner & Wills, 1996). In an attempt to offer a more comprehensive view of the design process, Maher (1994) proposed a co-evolutionary model of the design process based on the biological concept of genetic evolution (Maher, 1994; Maher & Poon, 1996). Specifically, Maher modelled the design process in terms of an interaction between the design problem space (requirements of the design) and the solution space (potential structural combinations of solutions that

constitute a design). In other words, the co-evolutionary model views problem and solution spaces as evolving separately while mutually affecting one another.

Dorst and Cross (2001) applied Maher's model in a protocol study of nine industrial designers devising a litter system for a train. Based on their observations of protocols, the authors described the design process as an iterative movement between the problem space and solution space. The problem solvers started by exploring the problem space, where they discovered a partial structure by using the information provided in the design brief and asking questions to the experimenter. This partial structure of the problem helped the problem solvers to develop a partial structure of the solution space in which the designers started to generate initial design ideas. This developed structure of the solution space was then fed back to the problem space as the designer considered the implications of the partial solution and again extended the structure of the problem space. This process of creating a matching problem-solution pair continued until a complete solution was reached.

Ulrich (2011) proposed a four-stage product design process that starts by sensing a gap in the user's experience, followed by defining the problem where the designer articulates the gap and the reasons it is experienced by users. Designers then explore alternatives and finally evaluate those alternatives and select a plan. The problem solver determines if the gap has been closed and, if not, the process will be repeated.

Overall, these various studies suggest certain general processes involved in solving open-ended design problems. For example, there is usually an early stage of exploring the problem and trying to understand its requirements and constraints, followed by later stages of generating different solution ideas, evaluating them, and selecting those that meet the constraints. Additionally, all of the studies refer to some process of structuring, i.e., the process starts with an ill-structured situation and designers/problem solvers make decisions/actions that introduce a structure, until a final solution state is reached. The process involves ongoing structuring and restructuring, whereby the designer iteratively moves between the various stages of solving the design problem until settling on a solution (Seidel & Fixson, 2013). Regardless of the different views of structuring (e.g., moving from the general to the specific; changing constraints and goals; lateral and vertical transformations), structuring seems to be an important aspect in solving design problems.

2.2.3 Limitations

The methodologies used to analyze design behavior are based on case studies or observational lab studies of design teams (e.g. Gunther et al., 1996) or individual designers (e.g. Goel, 1995; Gunther et al., 1996; Fernandes & Simon, 1999; Dorst & Cross, 2001; Ulrich, 2011) solving design problems. In a review of the design studies literature, Hernández, Cooper, Tether and Murphy (2018) showed that much of the published data are gathered and analysed based on case studies, observations, and focus groups, or based on the authors' own experiences, often using protocol analysis methods to analyze think aloud, video or sketching data (e.g., Eastman,1970; Cross, 2001; and Ulrich, 2011). Such methods provide valuable insights into design and other open-ended problem-solving domains, but sample sizes are often too small to test hypotheses about the effects of variables like open-endedness on cognition and behavior.

Furthermore, from the above review, one can observe that the literature on design used different theories to explain the process of solving design problems. This may be due to the fact that design is a broad concept that refers to various kinds of human activities in different domains, such as fashion, business, architecture and engineering. Therefore, the process is interpreted by design researchers using different terminologies in their different domains. For example, problem definition is called programming in architecture, identifying customers' needs in product design and establishing specifications in engineering design (Duerk, 1993; Ulrich, 2011).

The current study addresses a type of ill-structured problem, where there is restructuring that must be ongoing in order to find a solution. In other words, the problem solver introduces some kind of structure as part of the problem-solving process. However, this process is not a matter of finding a good structure, which both theories seem to emphasize; rather it is a matter of flexibility in the choice of the structure that is introduced. The present study addresses the methodological limitation of prior work by providing an experimental method that enables the controlled manipulation of aspects of problem structure such as degree of open-endedness. Though the experimental task of the current study is not a completely open-ended task as is the case with some real-world design problems, it operationalizes the idea of open-endedness in a controlled way that can be experimentally manipulated and measured.

The present study uses a group problem solving task as an alternative to think-aloud methods. Although think aloud methods have provided many insights into problem solving processes in design

and other domains, it is nonetheless an indirect and potentially incomplete method for accessing internal thinking processes. By contrast, in a group problem-solving situation, the group members are forced to communicate and actively engage with the information at hand, and their communications play a real part in the problem-solving process. Thus, the utilization of a group setting in our study is not primarily driven by the intention to investigate social properties of groups. Instead, we have intentionally chosen the group context as it provides an opportunity to access internal cognitive processes that are usually hidden in individual studies. Using a group setting offers an alternative and potentially better approach compared to think aloud protocols.

Given that the experiment in this study was conducted in a group setting, the following section review relevant literature on group problem solving.

2.3 Problem-Solving in a Group Context

In many real-life situations, problem-solving takes place in group rather than individual settings (Dunbar, 1998). In these cases, a group of individuals work together by sharing information and opinions to collectively come to a solution. Studying group problem-solving in the literature involves exploring various aspects of group behavior and performance. For example, different studies have focused on comparing group to individual performance. Some arrived at a general conclusion of the superiority of groups over individuals (e.g., Shaw, 1932; Kelley & Thibaut, 1954; Davis, 1969; Baron & Kerr, 2003), while others reached the opposite conclusion (Bouchard, 1969; Barron, 2003).

Approaching group problem-solving from a social psychology perspective, researchers have investigated the impact of individual characteristics on group dynamics. For instance, collective efficacy, which refers to a group's shared belief in their ability to successfully accomplish a specific task, has been shown to enhance team motivation and performance (e.g., Bandura, 1997; Tasa, Taggar, & Seijts, 2007). Studies showed that certain factors, such as group composition, cohesiveness, and motivation can influence team effectiveness (e.g., Guzzo & Dickson, 1996). Research by Taggar (2001) found that having the number of highly creative individuals in a group can influence overall creative performance. Other work has shown that team member conscientiousness can have both positive and negative effects on team creativity (e.g., Robert & Cheung, 2010; Taggar, 2021). In one study, the achievement striving aspect of conscientiousness, characterized by high aspirations and diligent effort towards personal goals, was found to predict team creativity positively, but the dependability aspect, associated with individuals being reliable,

organized, and responsible, had negative effects (Taggar, 2021). Other research highlighted the importance of person-group fit, referring to the extent to which individual's personality, values, and goals align with those of their coworkers, showing its positive impacts on job satisfaction (Kristof-Brown et al., 2005).

While these studies have primarily focused on individual characteristics, such as conscientiousness and creativity, and their impact on group outcomes like performance and group creativity, the current work takes a different perspective. This study places more emphasis on the structural properties of the task and their influence on internal group processes.

Laughlin and Adamopoulos (1980) classified group tasks as ranging from intellective to judgmental tasks. Intellective tasks have a demonstrably correct answer, while judgmental tasks are evaluative, behavioral, or artistic and do not have a demonstrably correct answer. The task of the current study can be considered more on the judgmental side than on the intellective side, although the participants will have to create a solution that they feel to be good rather than merely share opinions and select from among given alternatives. Within the literature of group problem-solving, there are two main bodies of research that investigate the internal processes of groups solving judgmental decision problems. These will be discussed in the following section.

2.3.1 Group Decision Making and Information Sharing

The literature has discussed two prominent approaches to investigate group problem solving and decision making. The Social Decision Scheme approach describes how individual group member preferences are combined into collective group decision outcomes. The Hidden Profile approach considers the effects of information sharing and exchange between members on group decisions.

The Social Decision Scheme approach has been widely used in the field of group problem-solving and decision-making research. It assumes that the group formulates one collective response through combining members' preferences by certain processes. The study of Social Decision Scheme stemmed from the studies of Arrow (1951) and Black (1948, 1958) on social choice theory, which deals with how individual preferences are combined to make collective decisions through voting systems and parliamentary procedures. Studies developed through similar work (e.g., Lorge and Solomon, 1955; Smoke & Zajonc, 1962; Davis, 1969a, 1969b), and in 1973, Davis introduced the social decision scheme, which is a mathematical model that assigns probabilities to individual preferences for a solution alternative, and based on that distribution of member preferences, it

calculates the collective group's potential choice. The Social Decision Scheme approach emphasizes group communication and consensus-building, as it suggests that groups like juries or committees should communicate until they reach a united decision regardless of whether it is a correct one or not (Davis, 1969; Arrow, 1951; Black, 1948, 1958; Lorge & Solomon, 1955; Smoke & Zajonc, 1962).

The Hidden Profile approach focuses on the effect of information distribution and sharing among team members on the optimality of the solution in group decision problems. The Hidden Profile effect was discovered in a study by Stasser and Titus in 1985. In this study, a group of university students were asked to vote for a student body president from a pool of three candidates A, B, and C. Participants were given information profiles about the three candidates in a way that the order of the best candidate should be perceived to be person A, B, then C. Participants first indicated their personal preferences after reading the profile information, and then formed a group and conducted a discussion before indicating their preferences as a group. Two conditions were tested in the experiment: equal distribution of profile information among group members in one condition, and unequal distribution with different subsets of information supporting candidate A distributed among different members in the other condition. The results showed that in the first condition, with equal information sharing, 67% of participants preferred candidate A before and 83% after the discussion. In the second condition, with unequal information distribution, pre-discussion preference for candidate A was 23%, and post-discussion preference was 18%. These results demonstrate the Hidden Profile effect: when information is held by a subset of members, the group may choose a suboptimal decision despite having the necessary collective information for an optimal choice (Stasser & Titus, 1985).

Subsequent research explored factors influencing information sharing, such as leadership (Henningsen, Henningsen, Jakobsen & Borton, 2004), proportion of unique information (Cruz, Boster & Rodriguez, 1997), group size (Mennecke, 1997), and the use of technology (Hollingshead, 1996). The Hidden Profile effect highlights the importance of effective information sharing in group tasks and how unshared information may have a limited impact on group decisions. However, in judgmental tasks like voting based on candidate attributes, the value of information is ambiguous. Different voters may perceive the value of information differently, leading to varying assessments of the group decision's optimality. As a result, there are no clear specifications for the value of the information nor, consequently, for the definition of an optimal decision.

In the Social Decision Scheme and Hidden Profiles literature, the main task is transitioning from individual preferences to a group decision. The distinction lies in the availability of information: Social Decision Schemes involve equal access to all necessary information, while Hidden Profile starts with unequal distribution of information. Consequently, shared information receives more emphasis than unshared information. Both approaches focus on decision-type problems with clear alternatives and unambiguous solution states, such as the election candidates or the guilty/innocent outcome in jury decisions.

The current research problem differs from Hidden Profiles and Social Decision Schemes. In our open-ended task, there are no predetermined alternatives, leading to dominant uncertainty about possible solution states. Instead of combining initial preferences, participants devise a new state collectively. While Hidden Profile literature emphasizes the impact of unequal information distribution on group outcomes, our research focuses on the varying levels of attention members pay to available information. However, it is expected that the lack of optimal information sharing is not mainly the result of the initial distribution of information, but more likely due to more attention being paid to some items of information than others.

The following section reviews relevant studies in group problem-solving, with a focus on the influential works of Adejumo et al. (2008) and Chen (2010). These studies have played a significant role in shaping the direction and design of the current study.

2.3.2 Group Problem-Solving and Task Structural Difficulties

The studies by Duimering and his students (Abimbola, 2006; Adejumo et al., 2008; Chen, 2010) investigate group problem-solving in terms of the effects of structural difficulties and type of incentives on group performance.

The experimental design in these studies was based on a categorization task where 16 cards, with two pictorial items on each, are distributed equally among a group of four participants. They are asked to exchange cards until each possesses a set of four of a kind, by using one picture from each card as the basis for their categorization. In these experiments, participants were seated at a round table separated by T-barrier, which allowed them to communicate and exchange cards but prevented them from seeing each other's cards (see Figure 2). The card movements and discussion were recorded by installed cameras. The time and number of card exchanges required to solve the problem were used to evaluate group performance. The first study (Abimbola, 2006; Adejumo et al., 2008)

compared group behavior and performance across different problems of increasing structural complexity. The findings show that as problem complexity increased, there was a corresponding increase in the number of card exchanges and the time required for solution, more complex search behaviors and greater deviations between perceived and objective problem structure.

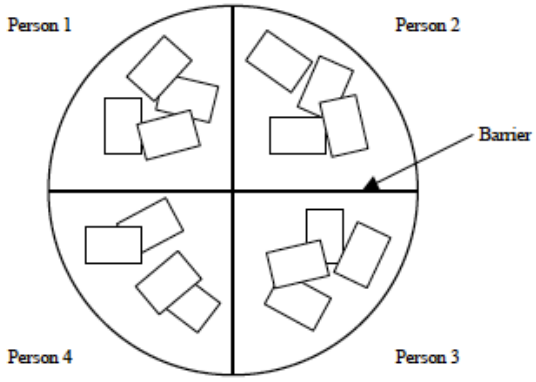


Figure 2: Overhead view of the experimental setup (Adejumo et al., 2008)

Using a similar experimental method, a second study by Chen (2010) investigated the effects of group versus individual incentives on the behavior and performance of the group. Under group incentives, the study found that groups exhibited free exploration behavior, associated with a higher number of card exchanges to reach solutions, as well as a tendency for the group to get stuck in blind alleys. In contrast, under individual incentives, group members exhibited risk-averse behavior and made careful card exchanges when a clear path to a satisfying solution was perceived, resulting in a lower likelihood of getting stuck in blind alleys.

The problem-solving tasks used in these studies combined aspects of information processing and insight problem-solving. Information processing involved the exchange of cards as participants incrementally searched the problem space for the correct solution categories; insight problem-solving was represented by problem structures that required groups to navigate detour paths and restructure their initial perceptions of the problem. Furthermore, through tracking of group discussions and card exchanges, the methodological design provided means of measuring aspects of problem solving that are not accessible in most other designs. For instance, the design makes it easier to detect when

participants become stuck in blind alleys and the time it takes them to consider a detour and reach the solution.

The tasks employed in both studies, along with other problem-solving processes in the literature, primarily involve well-structured, close-ended problems with a single correct solution. Therefore, both the performance and the process of solving the problem are measured by analyzing the path that the group takes while solving the problem, which is mostly predictable in terms of interacting with a detour and approaching the solution. However, the task design used in these studies is flexible and can be modified to encompass various problem structures, ranging from well- to ill-structured, and from closed- to open-ended problems. In the present study, we adapt the experimental design used in these studies to investigate effects of problem open-endedness. This will involve utilizing a modified version of the categorization problem introduced by Abimbola (2006), Adejumo et al. (2008), and Chen (2010), to create tasks with varying degrees of open-endedness. The task design will be further discussed in Section 4.2.

Since the task to be implemented here is an open-ended categorization problem, a brief review of relevant concepts from the categorization literature will be provided next.

2.4 Categorization

Categorization is one of the basic concepts that has been studied throughout history. Ancient Greek philosophers such as Aristotle introduced the classical view of categorization, in which groups of objects are considered to be members of the same rigidly defined categories based on similar properties or features shared by its members. In the 20th century, this view was challenged when Wittgenstein (1953) pointed out that some categories, such as games, do not have common properties shared by all of the members. Some games involve only amusement, others involve competition or no competition, and some, such as snakes and ladders, involve luck. The concept of category then shifted to focus on cognitive psychological aspects. One of the influential names in this field is Eleanor Rosch, who developed an experimental paradigm, formalizing previous findings in the area and exploring how people classify objects based on a mental image that represent a category. Therefore, this brief review will focus on the categorization literature based on Rosch's work.

According to Mervis and Rosch (1981), people spontaneously categorize objects and events which helps them to mentally organize their thoughts. Theories of categorization argue that a category exists when people treat two or more distinct events equally.

Rosch, Mervis et al. (1976) defined the main principles of forming a system of categories as follows:

1. Cognitive economy, which refers to the function of the category system as providing the maximum amount of information with the minimum cognitive effort.

2. Perceived world structure, which argues that the perceived world is viewed as structured information rather than random or unpredictable attributes. Therefore, achieving a maximum level of information with the least cognitive effort occurs when categories map the perceived world structure as closely as possible.

Rosch viewed category systems in terms of two dimensions: a vertical dimension that concerns the level of inclusiveness of a category; and a horizontal dimension that focuses on the differentiation between categories at the same level of inclusiveness (Rosch & Lloyd, 1978).

Based on the vertical dimension view, each category within a taxonomy is entirely included within a higher-level category but is not exhaustive of this higher level. Accordingly, Rosch defined the term "level of abstraction", which is the level of inclusiveness within a taxonomy, where the greater the inclusiveness of the category within a taxonomy, the higher its level of abstraction. Figure 3 offers an example of taxonomies from Rosch and Mervis's 1975 empirical study.

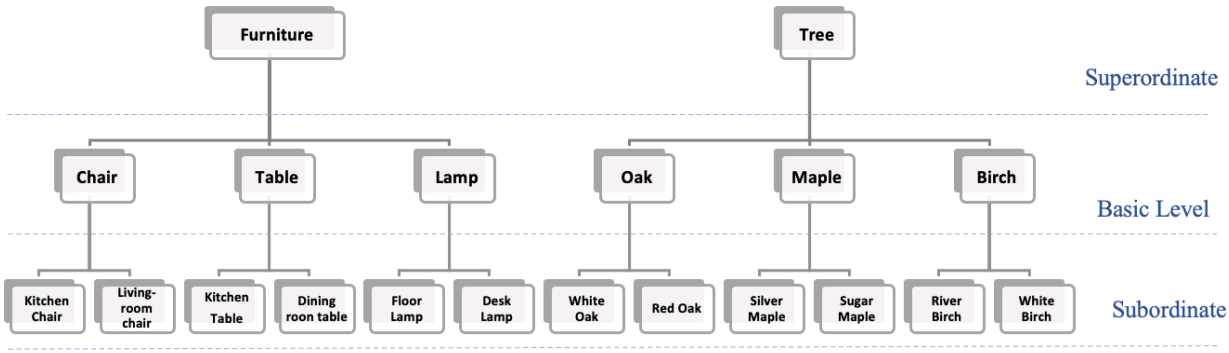


Figure 3: Example of two taxonomies used in Rosch and Mervis's study (1975)

Basic level categories are those that contain the optimal number of distinct attributes among their members. Categories that are one level more abstract will be superordinate categories (i.e., furniture and tree) as their members share only a few attributes. Categories below the basic level will have

common predictable attributes, but they will contain many attributes that overlap with the other categories, such as kitchen table and dining table.

In a 1975 study, Rosch and Mervis asked participants to list as many attributes as possible for items that are taken from nine common taxonomies (An example of the taxonomy used in the study is shown in Figure 3). The results showed that participants listed a significantly greater number of attributes in common for the basic level objects than for superordinate level objects. The results also indicated that basic level categories were the most inclusive level of classification, where objects have a number of attributes in common.

Through a series of other experiments, Rosch, together with fellow researchers also postulated that the basic level of abstraction is the most generic level, where members can be highly differentiated from one another and carry the maximum information. It was also found that a basic level category is the first category to be formed during perception of an environment, and the earliest category to be sorted and named by children (Rosch et al., 1976).

In contrast, the horizontal dimension concerns the segmentation of categories at the same level of abstraction. Rosch defined prototype theory, which argues that membership of items in a category is structured in a graded mode, where the clearest case of category membership, as defined by people's judgment, is the center of the category. For instance, a robin is a clearer, or more prototypical, member of the bird category than a penguin or an ostrich, i.e., a robin is the center of the bird category (Rosch & Lloyd, 1978).

A study by Rosch, Simpson and Miller (1976) showed that, when participants were asked to list the members of the category, the most prototypical items were the first and most frequently produced items. Other studies conducted by Rosch and others from 1972-1977 demonstrated the effect of prototypicality on major psychological measures such as speed of learning and speed of processing.

The above-mentioned literature on categorization is related to the current study from different aspects. Firstly, it highlights how categories are structured in terms of relations among concepts. This relation is built based on the perceptual view of the world around us. Ultimately, this leads us to a main assumption, which will be used in the present study, that knowledge is organized in the mind in terms of a network of concepts that are related to one another.

Secondly, Rosch's view of categorization provides a strong basis for predicting problem-solving behavior and how structuring would work in an open-ended categorization task. It can be observed that both basic level and prototypical categories are the most recognizable when viewing a category system from vertical and horizontal dimensions.

2.5 Summary of the Literature Review

A review of the literature sheds light on the two broadly defined trends in the literature of problem-solving, namely Gestalt theory and the information processing approach. The two views do not appear to be well integrated in terms of the problems that are classified as ill-structured, which is the focus of the current research. Therefore, this work combines, to some degree, the two theories that are needed to understand ill-structured problems. It is further noted that studies on the process of solving ill-structured design problems are mostly observational, with no controlled experiments. The problem-solving task in the current study will be performed in a group setting for the sake of better accessing internal cognitive processes. While different studies of groups focused on individuals' characteristics and their impact on group outcomes, the current study places more emphasis on the structural properties of the task and their influence on internal group processes. Two main bodies of research explored the internal processes of groups: Social Decision Scheme and Hidden Profile. In contrast to Social Decision Schemes and Hidden Profile tasks, the open-ended task in this work is characterized by ambiguous potential solutions alternatives, while decision schemes and Hidden Profile tasks feature clear-cut alternatives without any ambiguity. The task developed by Adejumo et al. (2008) is also considered, since it provides an appropriate environment that offers the opportunity to explore these kinds of open-ended problems. This is because it is a structured task that enables careful tracking of group behavior and is also flexible in a way that allows different sorts of problems to be created, including this study's ill-structured open-ended task.

Conceptually, the ill-structured problem-solving task in this study involves structuring by way of the process of creating a system of categories. Therefore, in this work, basic ideas from Rosch and others' categorization studies are used as a basis for forming our hypotheses, as well as for understanding how this structuring process might work and what factors may affect it.

Chapter 3
The Present Work

Most problems studied in academic literature are close-ended, even though many real-world problems are characterized by open-endedness. In an open-ended problem, the specificity of the goal state can affect the variability of the solution. As an example, let us imagine an open-ended task in which two groups of people are told to go to a certain destination. Members in the first group are told to "go to the other half of the campus", while the other group members are given a more specific location, such as "go to the south campus". The final location of each group member, and the paths they will use to arrive there, are unlikely to be the same. That is, there would be diversity in the exact location that individuals reach as well as the paths they use to reach them. However, this variability will be higher among the first group than the second.

The purpose of this study is to experimentally investigate the effects of the degree of problem open-endedness on problem-solving behavior, process and solution outcomes. Specifically, we focus on quantitative analysis by utilizing an experimental design similar to the categorization problems outlined by Abimbola (2006), Adejumo et al. (2008) and Chen (2010). They used well-structured problems with pre-defined solutions to study effects of specific problem structures on group problem solving behavior. In the present study, we modified the task to create ill-structured open-ended problems with no predefined solutions. We also manipulated the task instructions given to participants to vary the degree of problem open-endedness.

The main experimental task involved groups of four participants solving a sequence of three open-ended categorization problems, in which they grouped 16 pictures into subsets of four pictures each, based on specific relations between the pictures. The 16 pictures were randomly selected from collections of images, such that there was no right or wrong way to categorize the items. The instructions were varied to create different levels of open-endedness, by utilizing two different categorization goals (similarity-based versus story-based categories) and by manipulating the participants' beliefs about the open-endedness of the task. The levels of open-endedness are as follows:

1. **Similarity-based categories; belief in one correct solution (*Expert).*** This is the least open-ended task, where participants were instructed that the best categorization solution had been identified by experts and they should seek this "expert solution".
2. **Similarity-based categories; belief in multiple correct solutions (*Good)***. This is a higher open-endedness level, where participants were instructed that there exist multiple acceptable solutions to the problem, and they should find a good solution.
3. **Story-based categories; belief in multiple correct solutions (*Story).*** This is the highest level of open-endedness, where participants grouped items by coming up with 4 simple stories that link the four pictures within each group.

The degree of open-endedness serves as the main independent variable in the experiment, allowing us to explore the variations in problem-solving behavior and solution outcomes as the level of open-endedness changes. Further details regarding the experimental method are discussed in Section 4.2

3.1 Assumptions of the Study

We hypothesized effects of open-endedness on problem solving behavior based on some general assumptions.

First, we assume problem solvers have bounded rationality and are likely to exhibit satisficing behavior in their search for solutions (Miller, 1956; Simon,1997). That is, once they discover a satisfactory solution to the problem, they are unlikely to continue searching for an optimal, or best possible, solution if this is not a requirement of the task. For well-structured problems with one pre-defined solution, there may be little or no difference between satisficing and optimizing behavior. However, for ill-structured, open-ended problems, there may be substantial difference, particularly when the problem space is large and complex.

Second, we assume that all problems have inherent constraints that influence search behavior and potential solutions. Constraints are introduced by properties of the task situation and by the problem solvers. With respect to the task situation, well-structured close-ended problems generally introduce more constraints on problem-solving behavior than ill-structured open-ended problems. As noted above, different instructions varied the degree of problem open-endedness, using either similarity-based or story-based categorization goals, and by encouraging participants to identify either "good" or "expert" solutions. Problem solvers introduce further constraints due their perception of the task

situation and their background knowledge, which influence the perceived structure of the problem space and the information processing operations available to search that space. In our experiment, participants form categories of picture items based on certain perceived relations between them. In the two similarity tasks participants group items based on either direct visual perception of similar picture features (e.g., two round items perceived as similar based on shape) or their conceptual knowledge of similarity relations associated with the pictures (e.g., a plant and an animal grouped together based on participants' knowledge of the items as living things). In the story task participants group items into simple stories, drawing on their conceptual knowledge of a much wider variety of potential relations than just similarity. For example, stories might group items together based on various functional or logical relations between them, on different ways the pictured objects might interact with one another, and so on.

The third assumption is that conceptual knowledge can be represented as a network of concepts that are associated with one another. The experimental task in this work is basically a categorization problem, where participants are asked to group together items based on specific relations. These relations, according to Rosch's view (Rosch & Lloyd, 1978), are established based on the participants' understanding of the world. For example, a category of birds is constructed of items that are related to each other by shared attributes associated with birds, such as feathers or wings. This assumption was adopted by Ran (2007) in a study that examined how the meaning of conceptual combinations is constructed through an interaction of associated concepts. By implementing experimental problems that involve basic processes of categorization, our tasks operationalize the idea of problem structure in a way that enables measurement of various aspects of problem-solving search behavior and various properties of solutions.

Additionally, this study adopts the view of Newell and Simon (1972) of the problem space in which problem solvers navigate using operators to find a solution. This view considers a problem space as containing many nodes of knowledge states through which the problem solver moves during the problem-solving process until reaching a knowledge state that contains a solution. This study adopts Reitman's (1964) definition of an ill-structured problem, which entails the absence of information in one or more components of the problem, including the initial state, operation, and goal state. However, we specifically consider the problem as ill-structured due to the absence of information in the goal state.

The assumptions adopted in this study will be used as a basis to manipulate the degree of open-endedness of a given problem, and to explain the hypotheses.

3.2 Problem Open-Endedness

Solving an open-ended problem is associated with behavioral variety and a wide range of solution alternatives. The goal of this study is to investigate the effect of problem open-endedness on problem-solving behavior and the kind of solutions that the groups devise. Thus, problem open-endedness is the main independent variable in this study. The open-endedness of the problem is manipulated through the problem's goal state and the belief of the problem solver about the problem open-endedness.

3.2.1 Structure of the Goal State

Unspecified information in the goal state of the problem classifies a problem as ill-structured (Reitman, 1964). In this study, the specific experimental problem can be considered as ill-structured because information about the goal state is incomplete. It is not clearly defined in terms of how items must be sorted or if there is any basis, such as color, shape, or other shared attributes, for categorizing them. Consequently, various solutions will be acceptable, and there will also be various paths to reach those solutions.

To understand the behaviors associated with solving open-ended problems, the study manipulates the degree of open-endedness in the goal state component of the problem. Since the degree of the problem open-endedness cannot be explicitly defined in terms of the number of constraints and solutions, this variable is implicitly manipulated by defining two different categorization goals that vary in their degree of flexibility.

1. **Similarity-based categories.** Participants are asked to exchange pictures until each person obtains four pictures that are similar to one another, but different from the pictures in other categories.
2. **Story-based categories.** Participants are asked to exchange pictures until each person obtains four pictures that make up a story.

Based on the knowledge network assumption discussed in Section 3.1, it is assumed that similarity-based categorization is less open-ended than its story-based counterpart. In addition to manipulating

the degree of open-endedness, the story versus similarity conditions provides two different kinds of problem-solving goals using the same stimulus. Because these two different categorization problems draw on different associations in participants' knowledge networks, we expect participants to follow different moves as they progress through the problem space and to develop solutions that have different structures.

For illustration purposes we will use a hypothetical example of 16 pictures randomly distributed among four participants as shown in Figure 4. Based on this example, one possible partial solution that participants will reach in the similarity condition might be the category of 'living things': e.g., flower, tree, monkey, and bird. However, in the story condition, a partial solution might group the items: hammer, person, monkey, and banana , forming a category based on a story of 'a person is a carpenter who uses a hammer to build a cage for a monkey that eats a banana'. The presumed partial solution for similarity and story conditions is shown in the bottom right half of Figures 5 and 6, respectively.

Figure 4: A hypothetical example of 16 cards distributed among four participants in the experimental design.

Figure 5: A possible partial solution for the similarity condition

Figure 6: A possible partial solution for the story condition

3.2.2 Problem Solvers' Belief of Open-endedness of the Problem

Variability in search behavior when solving an open-ended problem can also be determined based on the problem solver's belief about the existence of one versus many solutions for the problem. Believing in the existence of a unique solution may increase exploratory search behavior or create a feeling that the problem is unsolvable or difficult. However, if the participant believes that the problem has many solution alternatives, a decrease in the exploratory search behavior may be exhibited. The participant may also be more motivated to reach a satisfactory justifiable solution.

The beliefs of the problem solver regarding the open-endedness of the problem are used as another way to manipulate open-endedness. The proposed experimental task manipulates the problem solver's belief about the problem open-endedness by introducing two conditions:

1. **Belief in one correct solution**. The instructions indicate that the problem has a best solution (Expert solution).
2. **Belief in multiple correct solutions**. The instructions indicate that the problem could have many correct solutions (Good solution).

In brief, the problem's degree of open-endedness is manipulated in two ways: goal structure (similarity-based versus story-based) and the problem solver's open-endedness beliefs (there is a

unique correct solution versus multiple correct solutions). Combining these variables will yield three different conditions as follows:

1. Similarity-based and a belief in one correct solution
2. Similarity-based and a belief in more than one correct solution
3. Story-based and a belief in more than one correct solution[1]

Following are hypothesized effects of open-endedness on several dependent variables related to problem-solving behavior and solution outcome.

3.3 Effect of Open-endedness on Task Difficulty

3.3.1 General Hypotheses

Constraints introduced to a task may create a source of difficulty while solving it. In this study's task, some constraints are common across all task conditions, including the restrictions of the task being completed in the context of a group and for which participants must devise four categories, each containing four items that have something in common but are distinct from items in other categories. On the other hand, some constraints are different across the three task conditions due to the degree of problem open-endedness that is manipulated through the goal structure and belief of the problem solver regarding the availability of a unique versus multiple solutions. In general, the more constrained this task, the less open-ended it is and the more difficult the problem solver may perceive it to be.

Given our three experimental conditions, it is expected that the level of difficulty in an open-ended task varies inversely with the degree of the open-endedness of the problem.

H1: Problem open-endedness has a negative effect on task difficulty.

[1] We excluded a fourth possible condition of "Story-based and a belief in one correct solution" since a story-based category of a problem implies open-endedness; that is, it is counterintuitive to tell participants that a story has only one unique solution.

3.3.2 Effects of Similarity versus Story Goals on Task Difficulty

The experimental task of this work is open-ended, without a right or wrong solution. However, participants may face some difficulties in solving the problem in both the story and similarity-based conditions due to the different knowledge association structure in each condition. The network of potential knowledge associations in the similarity-based condition are constrained to include only similarity relations, while the story condition allows for a wider range of conceptual relations from which participants could make stories. In other words, the set of associations from which participants can choose to create a solution category in the similarity condition is narrower than it is in the story condition. Therefore, it is expected that participants will experience more difficulty, in the similarity than the story condition.

H1a: Groups that form similarity-based categories experience a higher level of task difficulty than groups that form story-based categories.

3.3.3 Effects of Open-endedness Beliefs on Task Difficulty

The difficulty of a problem can be affected by the problem solver's beliefs about the open-endedness of the task. That is, believing in the availability of different acceptable solutions reduces the need for exhaustive exploration of the problem space, compared to believing that a single best solution must be found. Therefore, within the similarity condition, problem solvers are likely to indicate less difficulty, exhibit fewer picture exchanges, and spend less time to solve the problem if they believe the problem is open-ended than if they believe it is not.

H1b: Groups that believe there is only one correct solution experience a higher level of difficulty with the task than groups that believe there are multiple solutions.

3.4 Effects of Open-endedness on Solution Variability

Given that this task is open-ended, it is improbable that different problem-solving groups will devise identical solutions. However, patterns of similarity may show up in the different solutions across groups. The degree of variability in the solution indicates the potential frequency of common solution patterns across the different groups. A certain pattern of grouping items that appears more regularly in the solution categories of the different groups may be identified.

3.4.1 General Hypotheses

Although the categorization task involves participants grouping random items, it is expected that their solution categories will exhibit similarities for the following reasons:

Firstly, based on the items' different shared attributes, one or more relationships can be formed between two or more items in a set. Accordingly, it can be assumed that participants will perceive some kind of common relationship linking together different items. That is, certain relations might be more dominant and appear more frequently in the solution categories across different groups. Referring back to the hypothetical example in Figure 4, in the similarity condition, common relations in the solution category across different groups could be observed, such as the relation between 'flower' and 'tree' as 'plants'. Similarly, common relations such as 'monkey' eats 'banana' may be noticed in the story condition. Also, participants may categorize items through direct visual perception of similar picture features. For example, they might group together a cup and a balloon based on their shape as two round items.

Secondly, it is expected that the solution categories will be roughly similar since the proposed experimental task is a general knowledge task that does not require participants to possess any specialized knowledge to solve it. Additionally, it is assumed that participants in this study are not a totally heterogeneous group. Despite potential differences in cultural background, or other demographic characteristics, they are likely to share similar concepts and ways of perceiving the environment. Based on the earlier hypothetical example, they all will presumably recognize items such as 'bird' and 'monkey' as animals and 'car' and 'airplanes' as transport. Therefore, all participants are assumed to have relatively similar general background knowledge about the concepts of different pictures in the picture set.

Although different participant groups are expected to arrive at relatively similar solutions, the degree of variability across their solutions is expected to differ in response to the degree of problem open-endedness. The more open-ended the problem, the more flexibility in forming associations between items, and therefore the more variability in the solutions (and vice versa for the less open-ended problem). The following hypothesis is accordingly made:

H2: Problem open-endedness has a positive effect on solution variability.

3.4.2 Effects of Similarity versus Story Goals on Solution Variability

Based on the assumption that knowledge is organized as a network of associated concepts, this work views the set of potential associations for a given concept as being larger in the story condition than in the similarity condition. In other words, the relations between items in the solution network of the similarity goal are mainly limited to similarity relations, while in the story goal, these relations are wider and might include all potential associations such as similarity or functional relations. Thus, more variability in the solution categories is expected in the story condition than in the similarity condition.

H2a: Groups that form story-based categories exhibit higher variability in their solutions than groups that form similarity-based categories.

3.4.3 Effects of Open-endedness Beliefs on Solution Variability

As previously discussed, the picture items used in our task are general knowledge items that should be similarly perceived by different participants. Therefore, it is proposed that groups will devise similar ways of categorizing these items and, consequently, exhibit moderate variability in their solution categories. However, it is expected that this solution variability will be affected by the problem solvers' belief in the open-endedness of the problem. That is, within the similarity condition, if the problem solvers believe that the problem has a unique correct solution, they are likely to analyze the situation more carefully and draw on their background knowledge to make sure they identify the correct similarity relations. In other words, they would try to maximize the similarity relations between items within the category and minimize the similarity relations between items in different categories. This will result in solutions that represent good similarity relations. Specifically, participants will strive to identify categories containing items that they assume others, such as the experimenter, would agree are the most similar to one another (and the most dissimilar from items in other categories).

In contrast, considering bounded rationality and cognitive limits to information processing in human problem solving (Miller, 1956; Simon, 1997), it is expected that groups that believe there are multiple solutions will accept solutions that they perceive to be satisfactory, rather than seek the best possible solutions. A range of solutions that includes those that are acceptable is wider than that which only includes the best similarity relations. Therefore, higher variability of the solutions is

expected if the problem solvers believe the problem has multiple solutions than if they believe it has only a single solution.

H2b: Groups that believe there are multiple solutions exhibit greater solution variability than groups that believe there is one correct solution.

Although this task is open-ended, there are different constraints in the task that can lead participants to arrive at similar types of solutions. Such constraints include that all groups must come up with four categories, that participants share similar background knowledge, and that pictures in this task contain general knowledge items whose relations are likely be perceived similarly by the participants. These constraints ultimately are expected to lead to similar solutions for different participants groups. Stating this differently, when considering all possible solution alternatives for the current task, which amounts to a total of 1820 possible categories (16 choose 4), participants are expected to select only a fraction of all the possible solution categories due to the constraints operating in this task. Therefore, it is assumed such solutions to be far from a random sampling of that solution space.

H2c: The variability of solutions of this problem will be far less than the theoretical variability, assuming random sampling of the solution space.

3.5 Effects of Open-endedness on Path Dependency

3.5.1 General Hypotheses

Path dependency refers to the general idea that past states can affect current and future states. In the present context, problem solvers are likely to start searching the problem space in the vicinity of the initial problem state, and gradually search more widely until a solution is found. For close-ended problems with a single correct solution, the initial state cannot alter the solution but may influence other factors such as problem-solving difficulty or search path complexity. For open-ended problems, multiple potential solutions may be acceptable, so cognitive limits to information processing imply that problem solvers should engage in satisficing behavior, accepting a satisfactory solution found near the initial state, rather than rigorously searching the entire problem space for the optimal solution. Thus, the initial state of an open-ended problem should influence the solution state.

H3: Problem open-endedness has a positive effect on the path dependency of solutions.

3.5.2 Effects of Similarity versus Story Goals on Path Dependency

In the similarity condition, problem solvers search in the problem space for associations that are limited to similarity relations, whereas the story condition allows for a much wider set of potential associations and conceptual relations from which to create stories. This implies that the set of potential acceptable solutions is narrower for the similarity condition than the story condition. Thus, to find a solution from a smaller set of acceptable solution, problem solvers may need to search rigorously in the problem space and consequently move further from the initial problem state. On the other hand, in the story condition, problem solvers may settle on an acceptable solution as soon as they find it without the need to exhaustively search in the problem space or move too far from the initial state. Thus, the initial state of an open-ended problem should have a greater influence on the solution state in the story condition than in the similarity condition.

H3a: Path dependency is greater for groups that form story-based categories than groups that form similarity-based categories.

3.5.3 Effects of Open-endedness Beliefs on Path Dependency

When problem solvers believe that there are multiple acceptable solutions to a problem, they are more likely to accept a satisfactory solution instead of extensively searching the entire problem space for the optimal solution. This belief in the availability of different acceptable solutions reduces the motivation to explore further and move further away from the initial problem state. In contrast, if problem solvers believe that the problem has only one correct solution, they are more likely to engage in a rigorous search for the optimal solution, potentially leading them to explore the problem space more extensively and move further from the initial problem state. Thus, the initial state of the problem should have a greater influence on the solution state if the problem solvers believe in multiple possible solutions than if they believe there is only one correct solution.

H3b: Path dependency is greater for groups that believe there are multiple solutions than groups that believe there is only one correct solution.

3.6 Effects of Open-endedness on Variability of Category Association Strength

The solution to the experimental problem involves creating categories of associated items based on participants' knowledge of conceptual relations between items or their direct perception of item

similarity. These relations vary in strength, resulting in variability in the association strength observed in solution categories.

3.6.1 General Hypothesis

The strength of association between concepts in the solutions is assumed to be the basis of creating solution categories. As will be explained in in more detail in Section 3.7, participants are expected to start tackling the problem by initially focusing on highly related items or items with strong associations based on their conceptual knowledge of relations between the items or their direct perception of item similarity (e.g., colour, shape, etc.) . However, if they are not able to create four complete categories based on these identified relations, participants may break those strong relations and adopt less obvious, weaker ones. This behavior is likely to be different in each of the conditions, leading to differences in the variability of concept association strength in the solutions. Accordingly, we hypothesize an inverse relationship between the variability of the strength of association in the solution categories and the level of problem open-endedness.

H4: Problem open-endedness has a positive effect on the variability of solution association strength.

3.6.2 Effects of Similarity versus Story Goals on Variability of Category Association Strength

As group members progress towards a solution involving four items, they establish relations between the items based on their existing knowledge network. Initially, their attention is directed towards the perceived higher strength associations between items that is influenced by their knowledge of conceptual relations, such as similarity. However, they must also deal with lower strength relations to find a solution that includes all items and accommodates all constraints of the task situation. The behavior of creating solution categories is expected to be different in each of the two goal conditions, which should lead to different degrees of variability in solution association strength. In the similarity goal of the hypothetical example mentioned earlier (Figure 5), participants might initially perceive a strong relation between 'bird' and 'monkey' due to their association as 'animals'. However, when considering the constraint that they must find sets of four-of-a-kind with the remaining items in the problem space, they would be forced to give up this ‘animal’ association and might create a more

general category of ‘living things’ that include (flower, tree, monkey, and bird), based on somewhat weaker association between these items.

On the other hand, in the story condition (Figure 6), participants may identify a strongly associated pair, such as (monkey and banana), and link it to other items, such as ‘person’ and ‘hammer’ collectively. When considering strength of associations between items in this story-based category, some items may exhibit a high association strength such as ‘monkey’ and ‘banana’, while others exhibit a very low association strength, such as ‘monkey’ and ‘hammer’. Overall, higher variability is expected in the association strength of the story-based solution categories than in the similarity-based versions.

H4a: Solutions in the story-based categories exhibit higher variability in the strength of association than in the similarity-based category.

3.6.3 Effects of Open-endedness Beliefs on Variability of Category Association Strength

Within the similarity conditions, variability in the strength of solution associations is expected to be different according to whether the problem solvers believe that there is a single solution to the problem or multiple possible solutions. If the problem solvers believe the problem has one solution, they will search more carefully for similarity relations that ensure all the categories are of homogeneous association strength. In other words, their aim is to identify solution categories with the highest possible strength of concept association. However, they may need to compromise and settle for an average association strength due to task constraints, such as a scarcity of strongly related items that can form complete sets of four similar items. In contrast, believing in multiple solutions may lead problem solvers to settle for a satisfactory solution (Simon, 1956) whose association represents a higher variability in the strength of concept network association.

H4b: Groups that believe there is a single correct solution develop solutions that exhibit less variability in the strength of associations than groups that believe there are multiple possible solutions.

3.7 Structural Moves across Levels of Open-endedness

Solving the current problem amounts to creating a structure out of random items. As previously mentioned in the literature on information processing, Newell and Simon (1972) view the problem

space as containing many nodes of knowledge states through which the problem solver moves during the problem-solving process. Reaching a solution occurs by connecting these nodes until a knowledge state that contains the solution is achieved. The structural moves toward a solution in our open-ended categorization problem can be viewed as moving through a knowledge network. The structural moves in this task take place by exchanging pictures between categories until groups find four sets of four items that satisfy a similarity or story solution. By exchanging pictures, participants move in the problem space by creating links between items based on their conceptual knowledge and modifying those links until a solution state is reached.

Creating and breaking links is assumed to follow a notion similar to the findings of Rosch & Lloyd (1978) in terms of the horizontal and vertical dimensions of categories. They noted that both basic level and prototypical categories are the most distinctive when viewing and forming category systems. Accordingly, it is assumed that forming categories in our open-ended problem starts with the most distinctive items or the items with the strongest relationships to one another.

In brief, the process of solving our open-ended problem can be considered as following a path in a knowledge network where participants create a structure of a solution by linking items that are based on their knowledge network. The process of forming links is assumed to be initiated with the strongest perceived associations between items and gradually to move to weaker associations as problem solvers progress toward a solution. As an example from the hypothetical case mentioned earlier, in the similarity category of (birds, monkey, flower, tree), participants may start by grouping ‘bird’ and ‘monkey’ based on the prominent relation that they are both ‘animals’. However, when considering other items available in the problem, they would relax this ‘animal’ association and shift their attention to weaker associations between these items and ‘tree’ and ‘flower’, thereby creating a more abstract category of ‘living things’. Association strength seems to be relatively higher in the categories of ‘animal’ or ‘plants’ than ‘living things’. In other words, people are more likely to categorize a ‘tree’ as ‘plants’ and call a ‘monkey’ an ‘animal’ than to call either of them a ‘living thing’. Nonetheless, since participants will still have three groups of miscellaneous items, they will have to break these links and make new, weaker links to satisfy the four-of-a-kind solution. In the story condition, a similar outcome is expected to occur in terms of starting from the perceived strongest relation, although with different types of relations. As an example, participants may start grouping ‘monkey’ and ‘banana’ due to the relation ‘monkey eats banana’, and move on from there.

In general, it is expected that in both the story and similarity conditions, participants will start with the strongest associations and gradually move to weaker associations as they structure a solution.

H5: The process of solving an open-ended problem is represented by moving from high to low strength concept associations.

3.8 Effects of Open-endedness on the Variability of Problem-solving Search Behavior

In order to move toward a solution to this open-ended problem, participants are exchanging pictures and eventually creating and breaking category relations between pictures till reaching solutions. This behavior of creating and breaking category relations between items is assumed to be affected by the open-endedness of the problem. In more open-ended problems, multiple potential solutions may be acceptable, and due to cognitive limits to information processing, problem solvers might accept a satisfactory solution without investing much effort to break initially created links between items and creating new ones to improve the solution. In less open-ended problems, they would strive for a correct solution and thus exhibit higher variability in the problem-solving behavior presented by higher rates of breaking initial category relations and creating new ones.

H6: Problem open-endedness has a negative effect on the variability in the problem-solving search behavior.

3.8.1 Effects of Similarity versus Story Goals on Variability of Problem-solving Search Behavior

In the similarity conditions, when participants become stuck with two or three categories and cannot find the fourth, they might break the current links between items and create completely different new links. Meanwhile, in the story condition, participants will be less likely to break the initial strong relations they have created. Rather, they would build upon the strong association already created, such as (banana, monkey), by introducing additional weak associations like (person, hammer) to complete a story. Therefore, they are less likely to totally break the links and create new links in the story condition than in the similarity condition.

H6a: Groups that form story-based categories exhibit lower variability in the problem-solving behavior than groups that form similarity-based categories.

3.8.2 Effects of Open-endedness Beliefs on Variability of Problem-solving Behavior

In the two similarity conditions as participants are building solution categories, they are searching for similarity relations starting with the strongest association, but eventually reach a point where they have items with weaker relations. Believing in the existence of a single correct solution would motivate them to break the strong links they created initially between items to improve the weakly connected solution categories. On the other hand, believing in the existence of multiple solution would encourage a tendency towards satisficing behavior, accepting a satisfactory solution without breaking initial strong relation and creating new links.

H6b: Groups that believe that there is a single right answer exhibit higher variability in the problem-solving search behavior than groups that believe there are multiple solutions.

3.9 Summary

This chapter provided a general description of our experimental task and explained the main assumptions and hypotheses for the study. To investigate effects of problem open-endedness on problem-solving behavior and outcomes, we used an adaptation of the experimental design of Abimbola (2006), Adejumo et al. (2008), and Chen (2010). In the main task of our study, groups of four participants solved open-ended categorization problems, where they grouped 16 pictures in four categories based on specific relations. Instructions were varied to create different levels of increasing open-endedness: Expert (belief in one correct solution), Good (belief in multiple correct solutions), and Story (story-based categories). The degree of problem open-endedness serves as the main independent variable for the study, and the effects of open-endedness on dependent variables related to problem-solving behavior and solution are examined.

Several assumptions were adopted in this study as the basis of manipulating degree of open-endedness and hypothesizing its effects on problem-solving behavior. Firstly, due to bounded rationality, problem solvers may exhibit satisficing behavior when solving open-ended problems and may settle for satisfactory solutions rather than searching for optimal ones. Secondly, we considered the influence of constraints on problem-solving behavior, which can be introduced by the task situation and the problem solvers themselves. The third assumption is that conceptual knowledge is represented as a network of associated concepts. Furthermore, we adopted the view of the problem space as a network of knowledge states, where problem solvers navigate using operators to reach a solution.

We hypothesized that problem open-endedness has a negative effect on task difficulty and variability in problem-solving search behavior. On the other hand, we hypothesized that problem open-endedness has a positive effect on solution variability, path dependency, and on the variability of solution association strength. We also hypothesized that the process of solving an open-ended problem is represented by moving from high to low strength concept associations.

The next chapter describes the experimental methods of this study in more details and explains the measures we used to assess the behavioral and solution outcome variables.

Chapter 4
Methodology

The present study's methodology comprises two integral components: a main experimental procedure and a concept association survey. The survey was created to construct a measure used to assess the strength of association between the pictures used in the main experiment. This chapter will first provide a description of the concept relation survey, followed by a detailed explanation of the primary experiment.

4.1 Concept Association Survey

The concept association survey was designed to provide an independent measure of the degree of association between the pictures used as visual stimuli in the main experiment. This measure is necessary to test the hypotheses related to variability in the strength of solution association, variability in problem-solving search behavior, and the structural moves toward the solution.

4.1.1 Participants

Forty-eight undergraduate student participants were recruited from an undergraduate Organizational Behavior course and received extra course credit (i.e., bonus marks) as compensation for their participation. Students who opted not to participate in the survey were given the option to complete an alternative coursework for the same extra credit. Participation was voluntary and none of the survey respondents also participated in the main experiment.

4.1.2 Procedure and Survey Design

In the main experiment, we used three versions of visual stimulus with 16 pictures each, which will be referred to as “Clipart”, “Icon” and “Walmart”. (A detailed description of the three stimulus versions will be provided in Section 4.2). We designed an online survey using the Qualtrics platform (Qualtrics, n.d.) to measure the strength of association between the 48 pictures used in the experimental task (i.e., between the 16 pictures for each of the three stimulus versions). In each survey question, participants were presented with a target picture from one of the three stimulus versions, along with the other 15 pictures from the same stimulus, and asked to select five of the 15 pictures that they thought were most similar to the target picture. Thus, we had 48 questions pertaining to each of the 48 pictures from the three stimulus versions. To avoid participants losing

their focus and to ensure more accurate similarity ratings, we created two separate versions of the survey, each with 24 questions, by randomly selecting eight pictures from each stimulus version.

Participants were invited to complete the survey through an email invitation sent to their university email addresses. Upon clicking the survey link, participants were directed to the Qualtrics online platform to complete the questionnaire. Participants were randomly assigned to one of the two versions of the survey. The survey questions were presented in a randomized order for every participant to control for possible order effects. Appendix A provides a complete sample of one of the survey versions.

4.1.3 Measuring Concept Association Strength

Table 1 presents the survey results for the "Clipart" stimulus version. The table provides frequency counts indicating the number of survey respondents who selected a given picture (each row in Table 1) as one of the 5 pictures that were most similar to each of the other pictures (each column in Table 1). For instance, the "tennis racket" was rated as most similar to the "soccer ball" by 21 participants, and the "soccer ball" was rated as most similar to the "tennis racket" by 20 participants. To compute the strength of association scores between each pair of pictures, we added the frequency of ratings from both directions. Thus, the similarity score for "soccer ball" and "tennis racket" was 41 (i.e., 21+20). The combined pair-wise similarity scores for all pairs of Clipart pictures are presented in Table 2. Twenty-four respondents completed each version of the survey, so the maximum frequency was 24 for each one-way rating, and the maximum pair-wise similarity score was 48. Corresponding frequency counts and pair-wise similarity scores for the Icon and Walmart stimulus versions are provided in Appendix B.

Table 1: One-way concept association scores from the survey for the Clipart stimuli type

From \ to	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	(M)	(N)	(O)	(P)
(A) Bird	0	14	9	4	6	11	18	3	4	6	4	2	7	20	2	10
(B) Campfire	3	0	13	3	14	9	16	2	17	2	15	1	3	16	4	2
(C) Winter-House	6	19	0	3	4	20	7	4	6	3	17	0	8	15	8	0
(D) Tennis-racket	1	3	0	0	9	12	4	15	4	21	1	10	16	4	10	10
(E) Waves	11	17	12	5	0	10	11	3	2	3	11	4	9	6	9	7
(F) House	3	8	21	2	1	0	8	10	2	8	12	5	16	3	19	2
(G) Vegetables	8	15	2	2	5	11	0	12	20	3	14	2	4	20	2	0
(H) Backpack	1	2	3	16	0	13	8	0	5	22	3	21	11	5	5	5
(I) Pizza	1	17	4	3	2	6	22	8	0	5	16	5	8	17	0	6
(J) Soccer-ball	5	4	2	20	1	13	7	19	11	0	1	13	4	2	9	9
(K) Coffee	2	18	13	1	5	13	13	7	18	3	0	4	15	7	0	1
(L) Grad-hat	1	0	3	11	2	17	4	21	4	9	5	0	19	2	12	10
(M) Man	4	5	3	13	2	19	4	3	7	12	8	14	0	6	19	1
(N) Hay	18	20	13	1	2	6	22	7	14	0	10	1	6	0	0	0
(O) Car	3	3	12	8	7	18	1	12	1	11	4	14	21	2	0	3
(P) Balloon	9	7	3	17	6	10	2	15	8	13	8	11	6	1	4	0

Table 2: Combined pair-wise concept association scores for Clipart

	(A)	**(B)**	**(C)**	**(D)**	**(E)**	**(F)**	**(G)**	**(H)**	**(I)**	**(J)**	**(K)**	**(L)**	**(M)**	**(N)**	**(O)**	**(P)**
(A) Bird	0															
(B) Campfire	17	0														
(C) Winter-House	15	32	0													
(D) Tennis-racket	5	6	3	0												
(E) Waves	17	31	16	14	0											
(F) House	14	17	41	14	11	0										
(G) Vegetables	26	31	9	6	16	19	0									
(H) Backpack	4	4	7	31	3	23	20	0								
(I) Pizza	5	34	10	7	4	8	42	13	0							
(J) Soccer-ball	11	6	5	41	4	21	10	41	16	0						
(K) Coffee	6	33	30	2	16	25	27	10	34	4	0					
(L) Grad-hat	3	1	3	21	6	22	6	42	9	22	9	0				
(M) Man	11	8	11	29	11	35	8	14	15	16	23	33	0			
(N) Hay	38	36	28	5	8	9	42	12	31	2	17	3	12	0		
(O) Car	5	7	20	18	16	37	3	17	1	20	4	26	40	2	0	
(P) Balloon	19	9	3	27	13	12	2	20	14	22	9	21	7	1	7	0

4.2 Main Experiment

4.2.1 Participants and Experimental Design

To examine the effects of problem open-endedness on problem solving behavior and solution outcome, we used a 3x2 within-by-between repeated measures design. 192 participants (90 female) completed the experiment, in N=48 four-person problem-solving groups. Each group was given a sequence of the three (within group) open-endedness task conditions (Expert, Good, Story), and was randomly assigned to one of two (between group) initial conditions, representing different initial distributions of the pictures to group members (see Appendix C). Of the 48 groups, 30 began with one initial condition, 18 with the other initial condition. The study and experimental procedures were approved by the University of Waterloo Research Ethics Committee under protocol number 42550.

Participants were undergraduates who received extra course credit (i.e., bonus marks) for their participation. Participating in the experiment was voluntary and students who did not want to participate in the experiment had the option of completing alternative course work for the same extra credit. The experiment was completed online using video-conferencing software. In addition to the four group members, we recruited an extra fifth participant per group to reduce the risk of cancelation due to no-shows. When all five arrived, the fifth person was given an observer role, instructed to mute microphone and take notes on the difficulties the group faced. When only four participants showed up, to be consistent, a confederate played the role of the fifth participant and was assigned the observer role. Observers are not included in the above sample size since their data were not used in the analysis of results.

4.2.2 Experimental Conditions and Stimuli Design

Problem open-endedness was varied by manipulating the goal of the categorization problem (similarity-based versus story-based) and participants' beliefs about the solution (i.e., that there was a unique correct "expert" solution to the problem versus multiple correct "good" solutions). Based on these manipulations, we created three experimental conditions corresponding with three levels of problem open-endedness:

1. Similarity-based categories; belief in one correct solution (*Expert*; least open-ended). Participants were instructed to categorize 16 pictures into four categories based on similarity "by exchanging pictures with other members until each of you has four pictures that belong to

the same category". They were also told that "this task was given to groups of students in a previous experiment and a panel of experts determined the best solutions. Your job is to try to find the best solution, as judged by a panel of experts".

2. Similarity-based categories; belief in multiple correct solutions (*Good*; more open-ended). This task was similar to Expert; however, participants were told that "there is no single correct solution to this problem, but your job is to try to come up with a solution that you think is a good one".

3. Story-based categories; belief in multiple correct solutions (*Story*; most open-ended). Participants were instructed to categorize pictures "by exchanging pictures with other members until each of you has four pictures that make up one story…your group will have a total of four different stories".

Each group solved three problems, one per condition. Three different sets of pictures were used to create three different versions of the stimulus: a Microsoft Word Clipart version (Microsoft Office 365, 2020), a Microsoft Word Icons version (Microsoft Office 365, 2020), and a version of Walmart product photos (Walmart, 2020), drawn from publicly available Microsoft and Walmart Canada websites respectively. Items in all three versions rely on general knowledge; thus, we assume that undergraduate participants likely perceive them in similar ways regardless of their demographic background. The selection of the three stimulus versions was intended to control for potential effects due to differences in participant knowledge and to get evidence of the generalizability of the results. To control for any potential bias due to the types of pictures, we balanced the assignment of the three versions to the three experimental conditions across the groups. Each group initially started with the Clipart task, followed by Icons, and then Walmart; however, the order of the three experimental conditions (Expert, Good, Story) was balanced across the groups (i.e., all sequence permutations were balanced across the groups). These resulted in six unique orders in which the task was presented, labelled A-F in Table 3. We used a within-participant (repeated-measures) design to increase data collection efficiency where each group completed three tasks, allowing to obtain three data points per participant instead of one.

Table 3: The six different orders of conditions across the three versions of the stimulus.

Condition orders / Stimuli version	A	B	C	D	E	F
Task 1 (Clipart)	Expert	Expert	Good	Good	Story	Story
Task 2 (Icon)	Good	Story	Story	Expert	Good	Expert
Task 3 (Walmart)	Story	Good	Expert	Story	Expert	Good

The following procedure was used to create the three picture sets. Since pictures on the Microsoft and Walmart websites can be revised, any future replication of the experiment necessitated creating a pool of categories and their picture content to have a fixed set of pictures from which to choose. Each of the three sources provide many different categories of images. To create the pools, 16 of the main picture categories were randomly selected from each source, and one picture was randomly selected from each category. Since product pictures on the Walmart website are organized in a hierarchical classification of categories and sub-categories, random selection proceeded from top level categories down to the lowest subcategory level until reaching the product pictures used in the stimuli (see Appendix D). After creating the picture sets, we did a trial run to make sure that the pictures were clear for the participants; three of the pictures were ambiguous and difficult to identify for participants, so we replaced them with other randomly drawn pictures from the same source categories. We coded the 16 pictures using letters from A to P. The three versions of the stimuli with their letter codes are presented in Figure 7. Table 1 and Appendix B provide corresponding labels for each picture item (e.g., Clipart picture A “Bird”), based on the labels most frequently used by participants to refer to the items during the experiment.

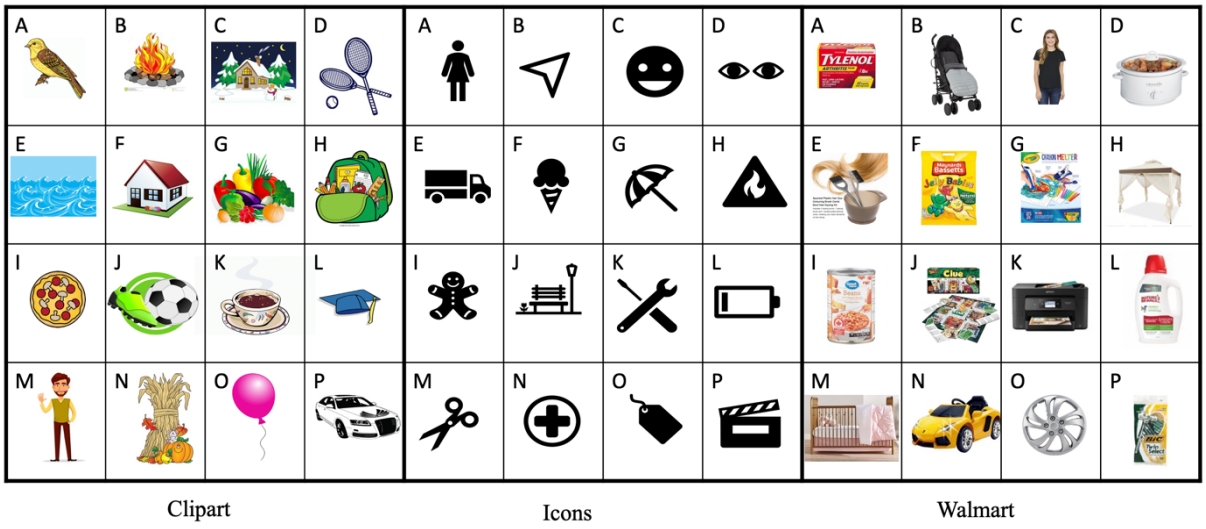


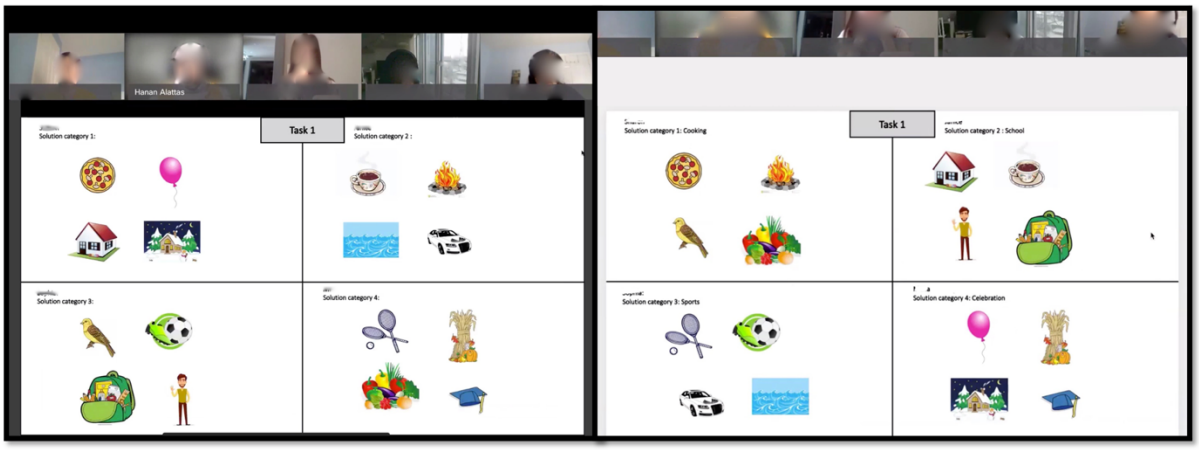
Figure 7: The three versions of the stimuli

4.2.3 Procedure

The experiment was conducted online using WebEx video conferencing software (Cisco WebEx, 2020), which allows screen recording and transcription of the groups' conversations. Participants were informed in advance that the video call would be recorded, and prior to their participation, they provided their informed consent to take part in the research. Each group comprised four participants randomly assigned to each session. Each group member received an invitation to the video call. Upon starting the call, participants were introduced to each other, then presented with a screen sharing showing the tasks and their instructions.

Every categorization problem contained 16 pictures that were distributed equally among the four participants. Participants were presented with a screen that was divided into four quadrants with four pictures per quadrant; each participant was randomly assigned to one of the quadrants. Group members were instructed to exchange pictures with one another until each member possessed a set of four pictures belonging to a category based on either a similarity or story goal. The experimenter served as facilitator, moved the pictures as requested by group members and wrote the names of the solution categories that participants identified after exchanging the pictures. Figure 8 displays an

example of the experimental setting, showing the initial problem state and solution developed by the experimental groups.



Initial state Solution state

Figure 8: Experimental setting for a Clipart similarity-based task, showing initial problem state and sample solution state

Participants began with two training tasks before completing the three experimental tasks. In the first training task, participants solved a similarity-based categorization problem with straightforward solution categories: tools, animals, people and fruit (See Appendix E). In the second training task, participants were given a demonstration of a possible story-based solution using the same pictures from the first training task. Specifically, the pictures of (hammer, man, monkey, and banana) were used to create a simple story of 'a man who is a carpenter uses a hammer to build a cage for a monkey that eats a banana'. The purpose of the training tasks was to familiarize participants with the basic requirements to solve the actual experimental tasks.

After completing the training tasks, participants were given a sequence of the three experimental tasks (Clipart, Icon, Walmart) solving one problem from each of the three experimental conditions as shown in Table 3. Upon solving them, each group member was emailed screenshots of their three solutions with a link to an online questionnaire (Qualtrics, n.d.) related to the three tasks they solved. The survey included questions about the difficulty of each task and other aspects of their problem-solving experience. The questionnaire is provided in Appendix F. A complete version of the problems of an experiment with the instructions is provided in Appendix E.

4.2.4 Measures

4.2.4.1 Task Difficulty

To examine the level of difficulty across the different levels of open-endedness we defined two types of measures: perceptual and behavioral observational measures.

The perceptual measure is presented by perceived task difficulty that was directly assessed through the post-experiment survey questionnaire (see Appendix F). In the survey, participants rated the difficulty of each task on a scale from 1 (very easy) to 7 (very difficult).

We defined two behavioral observational measures as indicators of task difficulty: the time taken to solve the problem, and the number of picture exchanges required to reach a solution.

Time refers to the duration, expressed by seconds, that a group takes to finish a specific task. Time can be considered as an objective indicator of the difficulty of each task. The less difficult the task, the less time it should take to complete, while the more difficult the task, the longer it should take. We consider the time of the task based on the difference between the time when participants first saw the pictures of the task and the time when they did their last picture exchange.

Picture exchange is the basic means to progress toward a desirable solution in the task of this study. We considered the number of picture exchanges needed to reach a solution as an objective indicator of task difficulty. Namely, the higher the number of pictures exchanges the more difficult the task.

4.2.4.2 Solution Variability

We compared solution variability across the three levels of open-endedness (Expert, Good, Story), and relative to a fourth hypothetical random condition in which we generated 48 random solutions (i.e., random distributions of pictures to categories). To measure solution variability across these four conditions, we used the Rand index (RI) similarity measure (Rand, 1971), which computes the pair-wise similarity between two different partitions of a set of items into disjoint subsets. Specifically, in the experimental tasks, a solution corresponds with a particular distribution of the 16 pictures to the 4 group members, with members holding 4 pictures each. Thus, the RI was used to compare different solutions corresponding to different partitions of the 16 pictures into 4 subsets of 4 pictures each. For our categorization task, RI ranges from 0.6 when two solutions are maximally dissimilar, to 1 when two solutions are identical. Referring to Table 3 in Section 4.2.2, there are nine unique comparable combinations of stimulus version and experimental conditions as follows:

1. Clipart\Expert (task 1, order A+B)
2. Clipart\Good (task 1, order C+D)
3. Clipart\Story (task 1, order E+F)
4. Icon\Good (task 2, order A+E)
5. Icon\Story (task 2, order B+C)
6. Icon\Expert (task 2, order D+F)
7. Walmart\Story (task 3, order A+D)
8. Walmart\Good (task 3, order B+F)
9. Walmart\Expert (task 3, order C+E)

For each of the 9 combinations of stimulus version and experimental condition above, one group's solution was randomly selected as a baseline for comparison, and the RI was measured for the other groups' solutions relative to these baselines. For the hypothetical random condition, we replicated the same structure as the actual conditions, with 48 randomly generated solutions to match the number of solutions in each of actual conditions. For a fair comparison, we divided the 48 random solutions into three subsets (based on the three different stimuli) and randomly selected one solution from each subset as baseline.

4.2.4.3 Path Dependency

To measure path dependency for each of the three experimental tasks, we used two different initial distributions of the 16 pictures to the four group members. These two distributions were maximally different from one another, such that any items assigned to the same category in one initial distribution were assigned to different categories in the other initial distribution, and the RI similarity score between them was the minimum possible value of 0.6 (Appendix C shows an example of the two different versions). To examine the effect of initial conditions on solutions, we used the RI to measure the similarity of each solution to the initial picture distribution used for that task, and we also measured the RI similarity of the solution to the opposite initial picture distribution. The difference between the two RIs for each solution was used as an indicator of path dependency. The larger the difference, the more path dependency and the smaller the difference, the less path dependency. That is, if the RI difference is large, this would indicate that a group's solution is closer to their initial

condition than to the opposite initial condition (i.e., more path dependent). If the RI difference is small, this would indicate that their solution is approximately equal distance from both initial conditions (i.e., less path dependent).

4.2.4.4 Variability of Category Association Strength

The concept association scores acquired from the survey were used to measure this variable. For each group's solution, we first assigned a score for each of the six pairwise relations between the four items in each of the four solution categories. To measure the variability in the strength of solution association, we computed the range of the relation scores for each of the four final solution categories and analyzed the average of the four ranges. We used the following formula to calculate the average range for all the four solution categories in each participant group j :

$$\boldsymbol{Average\ Range_j} = \frac{\sum_i(\max_C(\boldsymbol{Association\ Score_{(i,j)}}) - \min_C(\boldsymbol{Association\ Score_{(i,j)}}))}{|\boldsymbol{I}|} \quad \textbf{(1)}$$

where i is the index of a solution category (i.e. each individual category) in group j, $i \in I = \{1, 2, 3, 4\}$, $|I|$ represents the number of solution categories in an experiment j, C is the set pairwise pictures of solution category i, $C = \{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$.

4.2.4.5 Structural Moves

We initially planned to test the structural moves hypothesis by analyzing the association score for the sequence of picture exchange (moves) during the problem-solving process. However, due to pandemic restrictions, we had to change the experiment to an online format. As a result, testing the hypothesis at the level of every picture exchange was challenging. Therefore, we opted for an alternative approach to test this hypothesis. The rationale for the modification in plan will be explained first, followed by a detailed description of the methodology utilized for measuring the structural moves.

4.2.4.5.1 Change in Plan

The strength of concept association during problem-solving can be investigated at two levels: the progress of items within categories and the progress of final solution categories. Originally, our

experiment was planned to be conducted in person, similar to previous studies of Adejumo et al. (2006) and Chen (2010). However, due to Covid-19, we had to redesign it for an online setting. In the physical experiment, partitions were used to prevent participants from seeing each other's pictures, leading to more picture exchanges as a primary means of sharing information. However, in the online experiment, participants could see all the pictures, resulting in more verbal exchanges, where they discussed the potential associations between different pictures, and fewer actual picture exchanges.

To track and quantify the progress of solutions, we needed to create a dataset including both verbal and actual picture exchanges. Analyzing the verbal protocol of all 144 tasks to extract categories and items mentioned would have been challenging. Instead, we focused on the dataset based on actual picture exchanges. Therefore, we explored the hypothesis at the level of the evolution of final solution categories. We expected the first category completed by the group to have the highest average concept association scores, followed by the second and then third and fourth categories.

4.2.4.5.2 Structural Moves Measure

Similar to the previous measure, we used concept association scores obtained from the survey to measure this variable by assigning a score to each of the six pairwise relations between the four items in each completed solution category. The following formula was used to compute the average associations score for each solution category:

$$Average\ Association_i = \frac{\sum_{(k,l)\in C} Association\ Score_{(k,l)}}{|C|} \quad \textbf{(2)}$$

Where i and C are as defined in equation 1, and $|C|$ represents the number of elements in C.

For each group, we also coded their four final solution categories, based on the order in which the group completed each category during the problem-solving process. Due to the exchange nature of the task, every picture must be traded with another picture. Therefore, the groups typically completed their first and second solution categories sequentially, followed by the third and fourth solution categories, which are always completed at the same time. For this measure, we compared the average association scores between categories, based on three possible category completion orders: first, second, and the average of 3rd and 4th together. In a few cases, groups completed their first and

second solution categories simultaneously (in one Expert task, two Good tasks, and eight Story tasks). In these cases, we determined the order of the first and second solutions through random selection.

Appendix G provides an illustrative example of the order of the four solutions with the corresponding average association scores obtained from one of the experimental tasks.

4.2.4.6 Variability of the Problem-solving Search Behavior

Again, we used the concept association scores obtained from the survey as a basis to measure the variability in problem-solving search behavior. We computed the total average concept association score across the four solution categories repeatedly at every move (i.e., every picture exchange) during the problem solving process using the following formula:

$$Overall\ Group\ Average\ Association = \frac{\sum_i Average\ Association_i}{|I|} \quad \textbf{(3)}$$

Where the Average Association is as defined as in equation 2, and $|I|$ represents the number of solution categories.

The variability in problem-solving search behavior was measured through the number of direction reversals in the sequence (i.e., the time-series) of intermediate states representing partial solutions. We calculated the number of direction reversals based on the directional change in the overall average association score. A direction reversal happens when the overall average association score changes from an increasing trend to decreasing trend, or from a decreasing trend to increasing trend. A higher number of directional reversals corresponds with high variability in the problem-solving search behavior. Examples of similarity score trajectories for each problem structure condition are presented in Appendix H.

Chapter 5
Results

5.1 Task Difficulty

5.1.1 Perceived Task Difficulty

These results are based on participants' rating of the task difficulty taken from the survey that was given at the end of the experiment. Distributions of perceived difficulty scores deviated significantly from normality for all three task conditions (Shapiro's test $p < 0.05$). Therefore, a non-parametric Friedman test was used to test the perceived difficulty hypotheses. Results showed significant differences between the three conditions, $\chi2(2) = 121.09$, $p < 0.0001$. To further explore the main effect, a post-hoc analysis using pair-wise Wilcoxon signed-rank testing was conducted with a Bonferroni correction. The post-hoc tests revealed that difficulty scores for the Expert condition were significantly higher than both the Good and Story conditions ($p < 0.0001$), and difficulty scores for the Good condition were significantly higher than the Story condition ($p < 0.0001$) (see Figure 9).

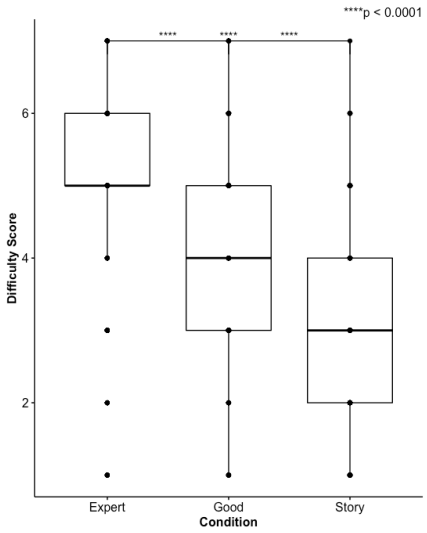


Figure 9: The effect of problem open-endedness on perceived task difficulty.

These results provide strong support for H1. Increased open-endedness reduced perceived difficulty for both the Story vs. Similarity (H1a) and Good vs. Expert (H1b) manipulations.

5.1.2 Time to Solve

The solution time data deviated from normality (Shapiro's test $p < 0.05$), so a non-parametric Friedman test was performed to analyze the effect of open-endedness on task difficulty represented by time. The results showed significant differences between the three conditions, $\chi2(2) = 40.17$, $p < 0.0001$. A post-hoc analysis using pair-wise Wilcoxon signed-rank testing with a Bonferroni correction revealed that time spent solving the Expert condition was significantly longer than the time spent on both the Good ($p < 0.001$) and Story conditions ($p < 0.0001$); and time spent on the Good condition was significantly higher than the time spent on the Story condition ($p < 0.05$) (see Figure 10). These results provide strong support for H1. Increased open-endedness reduced the level of difficulty, as reflected by the shorter time spent on both the Story vs. Similarity (H1a) and Good vs. Expert (H1b) manipulations.

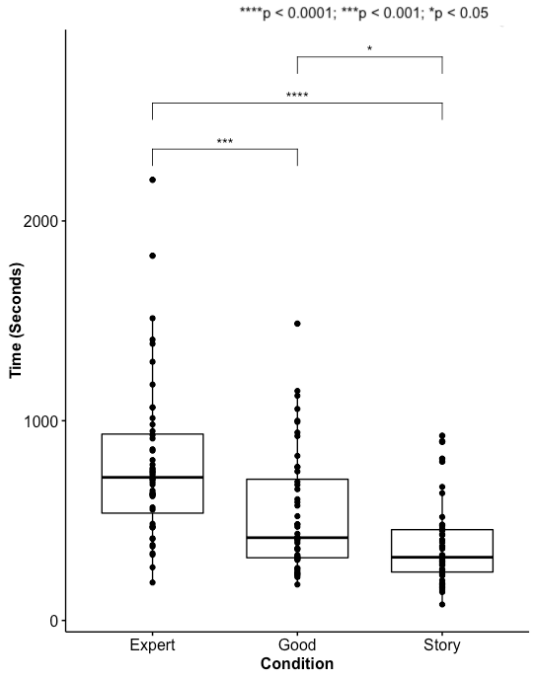


Figure 10: The effect of problem open-endedness on time spent solving the problems.

5.1.3 Number of Picture Exchanges

To further test the effect of problem open-endedness on task difficulty, a one-way repeated measures ANOVA was conducted on the number of picture exchanges across the Expert, Good, and Story conditions. The results indicated significant statistical differences in the average number of picture exchanges across the three conditions, $F(2, 94) = 21.21$, $p < 0.0001$, generalized eta squared = 0.21. Mauchly's test for sphericity was non-significant ($p > 0.05$). Shapiro's test for normality indicated $p < 0.05$ for the Good condition only ($p > 0.05$ for Expert and Story), but q-q plots indicated relatively small deviations from normality, so we chose to use standard ANOVA methods. As shown in Figure 11, post-hoc analyses using pairwise t-tests with a Bonferroni adjustment revealed that the number of picture exchanges in both the Expert ($p < 0.0001$) and Good ($p < 0.001$) conditions was significantly higher than in the Story condition. However, the number of picture exchanges were not significantly different between the Expert and Good conditions ($p > 0.05$).

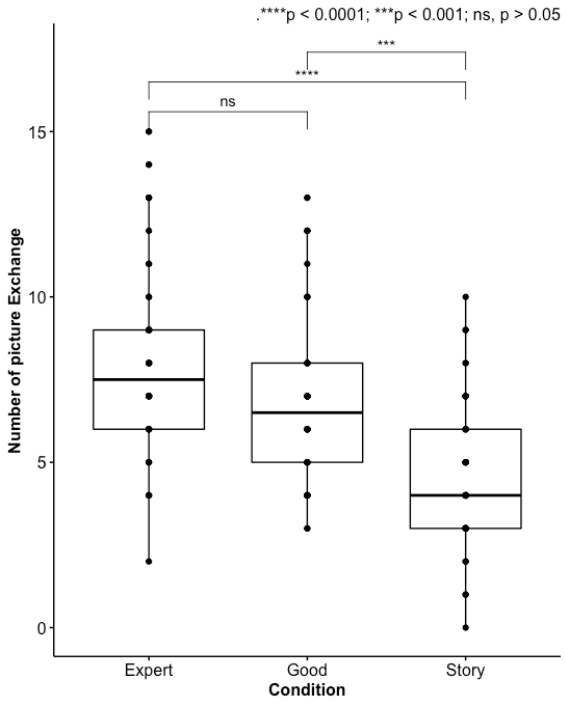


Figure 11: The effect of problem open-endedness on task difficulty represented by the number of picture exchanges.

These results provide mixed support for H1. Level of difficulty represented by the number of picture exchanges was significantly lower in the most open-ended Story condition than both the Good and Expert similarity-based conditions (H1a). However, in the similarity-based conditions, the belief in one unique solution instead of multiple solutions did not significantly affect task difficulty as represented by the number of picture exchanges (H1b).

5.2 Solution Variability

One-way repeated measures ANOVA was conducted to test the effect of problem open-endedness on the variability of solutions between the Expert, Good, Story, and Random conditions. The results indicated significant statistical differences in the mean Rand Index (RI) similarity scores across the four conditions, $F(3, 114) = 49.98$, $p < 0.0001$, generalized eta squared = 0.49. Mauchly's test for sphericity was non-significant ($p > 0.05$). Shapiro's test for normality indicated $p < 0.05$ for the Good condition only ($p > 0.05$ for Expert, Story, and Random conditions), but q-q plots indicated relatively small deviations from normality, so we chose to use standard ANOVA methods. As shown in Figure 12, post-hoc analyses using pairwise t-tests with a Bonferroni adjustment revealed that both the Expert and Good conditions had significantly higher similarity scores than the Story condition ($p < 0.0001$), but differences between the Expert and Good conditions were not significant ($p > 0.05$). On the other hand, RI scores for the Random condition were significantly lower than RI scores in all three conditions of open-endedness ($p < 0.0001$ for Expert and Good conditions; $p < 0.001$ for Story condition).

The results provide mixed support for H2. Solution variability was significantly higher in the most open-ended Story condition than both the Good and Expert similarity-based conditions (H2a). However, in the similarity-based conditions the belief in one unique solution instead of multiple solutions did not significantly affect solution variability (H2b). Additionally, solution variability in the three open ended problems was significantly less than the theoretical variability of a Random solution (H2c)

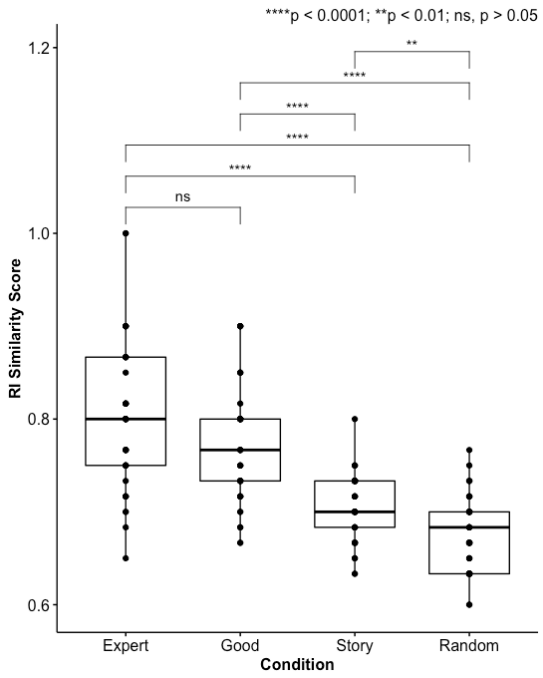


Figure 12: The effect of problem open-endedness on solution variability

5.3 Path Dependency

To test the effect of the initial picture distribution on the final solution under the three conditions, we conducted repeated measures ANOVA on the difference between the RI comparing each solution to the initial distribution for the problem and the RI comparing each solution to the opposite initial distribution. We found that the RI difference scores were significantly different across the three conditions, $F(2, 94) = 8.629$, $p < .001$, generalized eta squared = 0.112. Mauchly's sphericity test was non-significant ($p > 0.05$). Shapiro's test for normality indicated $p < 0.05$ for the Story condition only ($p > 0.05$ for Expert and Good), but q-q plots indicated relatively small deviations from normality, so we chose to use standard ANOVA methods. Post-hoc analyses using pairwise t-tests with a Bonferroni adjustment revealed that RI difference scores were significantly lower in the Expert than Story condition ($p < 0.01$), and in Good compared to the Story condition ($p < 0.05$), but there was no significant difference between the Expert and Good conditions ($p > 0.05$) (see Figure 13).

These results provide mixed support for H3. Solutions to the most open-ended Story problem exhibited significantly more path dependency than solutions to the less open-ended Good and Expert similarity problems (H3a). However, Good versus Expert open-endedness beliefs did not significantly affect path-dependency (H3b).

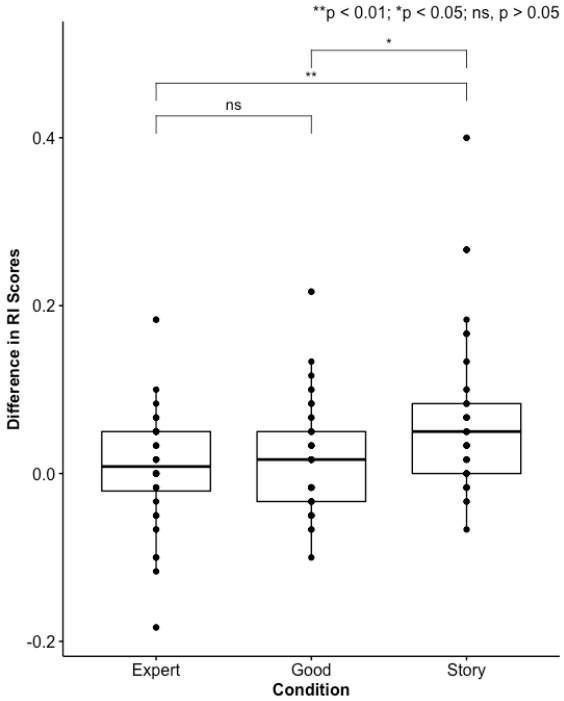


Figure 13: Path dependency effect of the initial condition on solution

5.4 Variability of Category Association Strength

To test the effect of problem open-endedness on the variability in the strength of solution association, we conducted one-way repeated measures ANOVA on the average range of solution association scores across the three conditions. The results indicate significant statistical differences in the average range of solution association scores across the three conditions, $F(2, 94) = 6.458$, $p < 0.01$, generalized eta squared = 0.094. Mauchly's test for sphericity was non-significant ($p > 0.05$). Shapiro's test for normality indicated $p < 0.05$ for the Good condition only ($p > 0.05$ for Expert and Story), but q-q plots indicated relatively small deviations from normality, so we chose to use standard ANOVA methods. As shown in Figure 14, post-hoc analyses using pairwise t-tests with a Bonferroni adjustment revealed that both the Expert and Good conditions had significantly lower ranges in

association scores than the Story condition ($p < 0.01$), but differences between the Expert and Good conditions were not significant ($p > 0.05$).

These results provide mixed support for H4. Variability in the strength of solution associations in the most open-ended Story condition was significantly higher than both the Good and Expert similarity-based conditions (H4a). However, in the similarity-based conditions the belief in one unique solution instead of multiple solutions did not significantly affect variability in the strength of solution association (H4b).

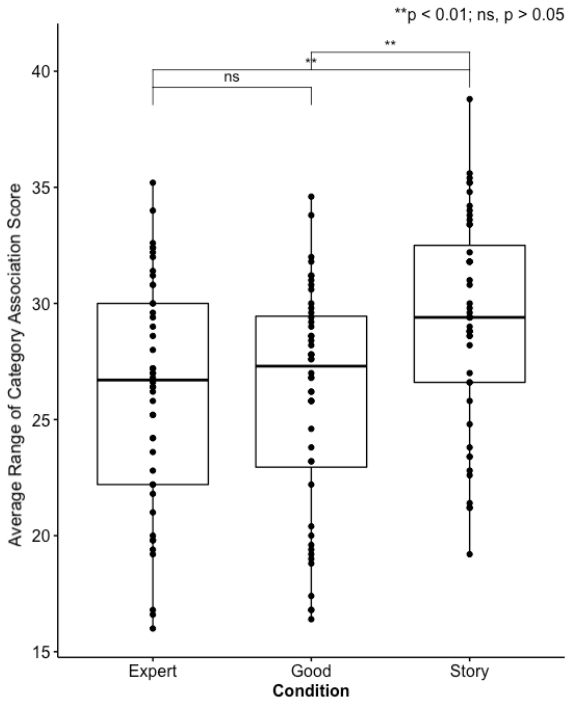


Figure 14: Effect of problem open-endedness on variability in strength of category association.

5.5 Structural Moves

We examined the structural progress of final solution categories within each experimental condition (i.e., at each level of problem open-endedness) separately. As explained in Section 4.2.4.5.1, we compared the average solution association scores between groups' four solution categories, based on the order in which groups completed each category during the problem-solving process. Since the third and fourth solution categories are completed simultaneously, the statistical analysis compared

three orders of solution category completion: first, second, and the average of the third and fourth solution categories.

5.5.1 Structural Moves in the Expert Condition

The average solution association scores in the Expert task were not normally distributed (Shapiro's test $p < 0.05$). Thus, we conducted a non-parametric Friedman test to compare the average solution association scores across the three category completion orders in the Expert task. The results showed significant differences in average solution association scores across the three orders of the Expert

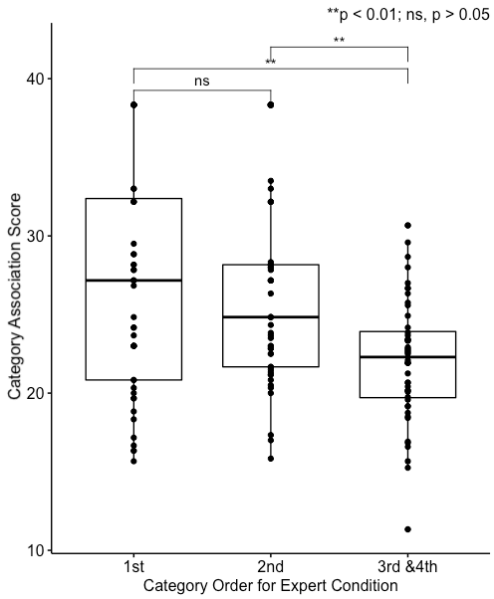


Figure 15: Average solution association scores across the orders of solution in the Expert condition

solutions, $\chi 2(2) = 12.67$, $p < 0.001$. To further explore the main effect, we performed a post-hoc analysis using pair-wise Wilcoxon signed-rank testing with a Bonferroni correction. The results, as shown in Figure 15, revealed that average concept association scores in the first and second solution categories were significantly higher than in the third and fourth ($p < 0.01$). However, average concept association scores were not significantly different between the first and second solution categories. These results partially support H5; the first and second solution categories had stronger association scores than the third and fourth categories, but association scores were not significantly different between the first and second categories.

5.5.2 Structural Moves in the Good condition

To compare the average solution association scores across the three category completion orders in the Good tasks, we performed one-way repeated measures ANOVA. The results indicated significant statistical differences in average solution association scores between the three orders of the Good solutions, $F(1.73, 81.52) = 4.31$, $p < 0.05$, generalized eta squared = 0.06. Mauchly's test for sphericity was non-significant ($p > 0.05$). Shapiro's test for normality indicated $p < 0.05$ for the first solution category only ($p > 0.05$ for second and third with fourth), but q-q plots indicated relatively small deviations from normality, so we chose to use standard ANOVA methods. As shown in Figure 16, Post-hoc analyses using pairwise t-tests with a Bonferroni adjustment revealed that the average concept association scores were significantly higher in the first solution category than in the third and fourth solution categories ($p < 0.05$), but average concept association scores were not significantly different between the first and second categories or between the second and third with fourth categories ($p > 0.05$). These results partially support H5. That is, the first solution category had a stronger average association score than the second and third with fourth solution categories.

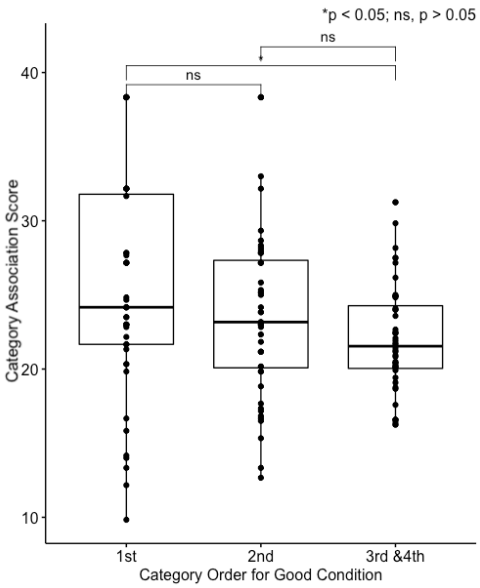


Figure 16: Average solution association scores across orders of solution in the Good condition

5.5.3 Structural Moves in the Story Condition

One-way repeated measures ANOVA was performed to compare the average solution association scores across the three category completion orders in the Story tasks. The results indicated non-significant statistical differences in average concept association scores across the three orders of the Story solutions, $F(2, 94) = 0.23$, $p > 0.05$, generalized eta squared = 0.003. Mauchly's test for sphericity and Shapiro's test for normality were non-significant ($p > 0.05$). As shown in Figure 17, post-hoc analyses using pairwise t-tests with a Bonferroni adjustment revealed that the concept association scores were not significantly different between any of the three solution category completion orders ($p > 0.05$).

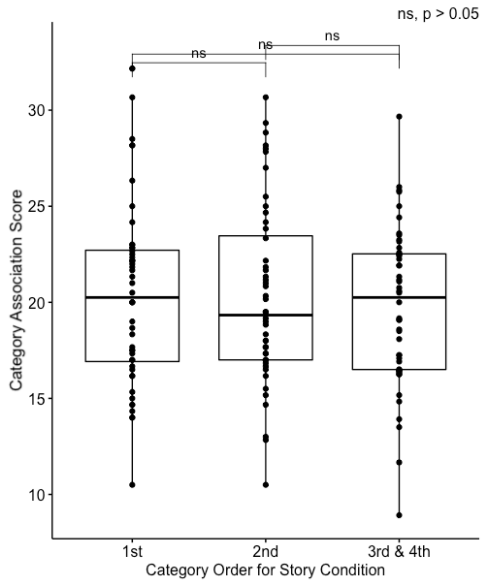


Figure 17: Average solution association scores across orders of solution in the Story condition.

These results may not support the main hypothesis H5, that is, the first solution has a stronger association score than the second, third and fourth solutions. However, collectively the preceding results for all three conditions of open-endedness (Expert, Good, Story) suggests differences in category association strength by solution category completion order varies depending on the degree of problem open-endedness. The tendency to focus on the strongest associations for the first category and weaker associations in the second, third and fourth categories is stronger for less open-ended problems (Expert) but reduces as open-endedness increases.

5.5.4 Interaction Between Problem Open-Endedness and Solution Order Effects on Structural Moves

The above one-way repeated measures ANOVA results show that participants in the least open-ended Expert condition exhibited a trend of moving from the strongest to the weakest solution association. However, this pattern was less distinct in the more open-ended Good condition, and it did not manifest at all in the most open-ended Story condition. These results indicate a possible interaction between levels of problem open-endedness and solution category order, in terms of their effects on structural moves. To explore this interaction, a two-way repeated-measures ANOVA was performed, analyzing the effect of problem open-endedness and solution category order on average solution association scores. The results revealed a significant main effect of problem open-endedness on average solution association scores, $F(2, 94) = 48.65$, $p < 0.0001$, generalized eta squared = 0.134. A significant main effect of solution category order on average solution association scores was also observed, $F(1.72, 80.64) = 9.94$, $p < 0.001$, generalized eta squared = 0.055. However, the interaction between problem open-endedness and solution category order was not quite statistically significant at

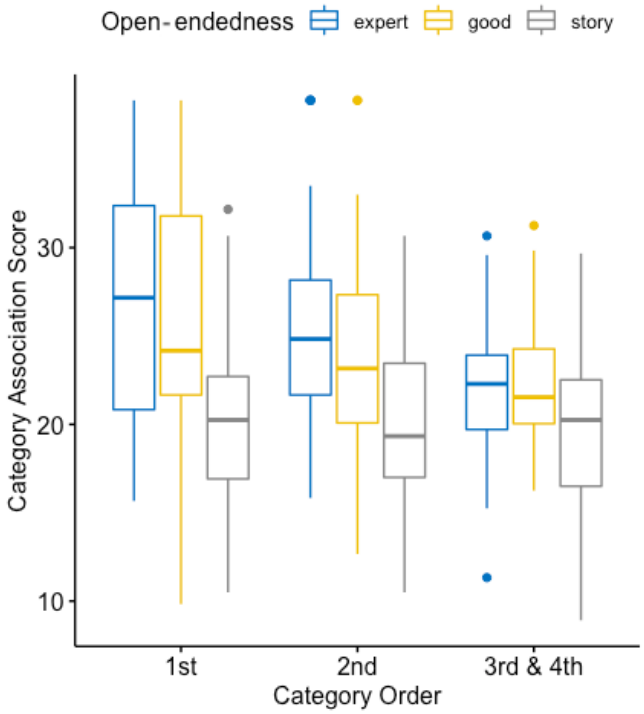


Figure 18: Average solution association scores across orders of solution in the three conditions

the 0.05 level, $F(4, 188) = 2.394$, $p = 0.052$, generalized eta squared = 0.024. Figure 18 shows solution association scores across orders of solution in the three conditions of open-endedness.

5.6 Variability in Problem-solving Search Behavior

Distributions of number of direction reversals in the time-series of average association scores during the problem-solving process deviated significantly from normality for all three task conditions (Shapiro's test $p < 0.05$). Thus, to examine the variability in problem-solving search behavior across the three conditions of open-endedness, we performed a non-parametric Friedman test on the number of picture exchange reversals. The results showed a significant difference in the number of exchange reversals across the three conditions, $\chi2(2) = 6.83$, $p < 0.05$. Post-hoc analyses using pair-wise Wilcoxon signed-rank testing was conducted with a Bonferroni correction, revealing that the average number of reversals in the Expert was significantly higher than the Story condition ($p < 0.05$), and in Good compared to the Story condition ($p < 0.05$), but there was no significant difference between the Expert and Good conditions ($p > 0.05$) (see Figure 19).

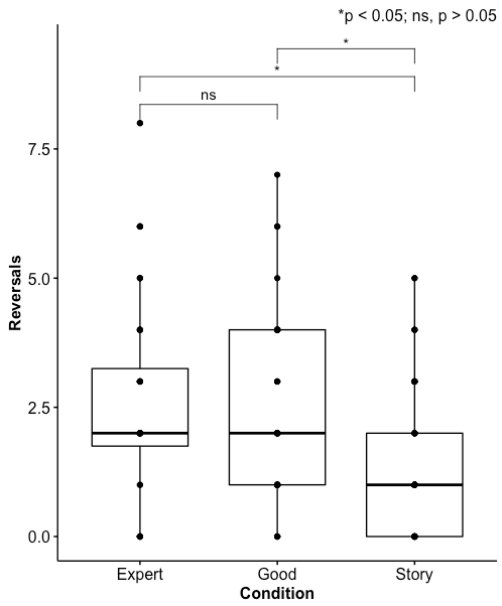


Figure 19: Number of direction reversals in the average association score

These results provide mixed support for H6. The variability in the search behavior, indicated by the number of direction reversals in the time-series of average association scores, was significantly lower

for the most open-ended Story problem than in the less open-ended Good and Expert similarity problems (H6a). However, the variability of search behavior was not significantly affected by Good versus Expert open-endedness beliefs (H6b).

5.7 Differences in the Effects of Open-endedness Across Stimulus Types

The preceding repeated measures ANOVA results were based on combined data from the three different stimulus versions (Clipart, Icons, Walmart). To investigate potential differences in the results that may occur across the different stimuli, we conducted statistical analysis on each stimulus separately. ANOVA and Bonferroni-corrected pair-wise t-tests were used when applicable; when normality assumptions were violated, non-parametric Kruskal-Wallis and Bonferroni-corrected Wilcoxon signed-rank tests were used instead. Tables 4 and 5 summarize the stimulus-specific ANOVA and Kruskal-Wallis test results and the relevant post-hoc pairwise tests. The results in Table 4 are generally consistent with the combined results reported earlier, with the Walmart results showing the strongest statistical effects, followed by Icon and Clipart. Specifically, the Walmart results showed significant effects of open-endedness for all dependent variables. The Icon results showed significant effects for most of the dependent variables, except for path dependency and variability in the problem-solving behavior. Lastly, the Clipart results showed significant effects of open-endedness for solution variability, perceived task difficulty, solution time, and the number of picture exchanges, but not for path dependency, variability in the problem-solving behavior, and the variability in solution association.

Table 5 presents the results of the structural moves towards the solution for each stimulus type in the three conditions of open-endedness. In the Expert condition, a significant relationship between category completion order and picture association scores was found for the Icon and Walmart stimuli, specifically between the first and last two completion orders, and between the second and last two completion orders for the Icon stimulus, and between the first and last two completion orders for the Walmart stimulus. In the Good condition, the Walmart stimulus showed a similar pattern to the main results, with a significant positive relationship between the first and last two completion orders. In the Story condition, there was no statistical relationship between category completion order and picture association scores, consistent with the combined stimulus results.

This mixed support to our main findings could be attributed to two possible factors: the sample size for non-repeated measure tests and the nature of the stimulus types we used in the study. Firstly, the

sample size that we used for a non-repeated measure tests is relatively small with low statistical power. That is, the sample size for the non-repeated measure tests we conducted is basically one third the number of observations that we used in our repeated measure tests. Secondly, the nature of stimulus types themselves might have affected the result. For example, the complexity of the pictures could have an influence on the results. Even though the study did not measure or control for the complexity of the pictures, it seems that Walmart version is the most complex and it was the one that most strongly demonstrated the effect of open-endedness that was observed in the main results.

Table 4: Summary of the stimulus-specific ANOVA and Kruskal-Wallis test results with relevant post-hoc pairwise tests[2]

Dependent Variable / Stimuli	Task difficulty (Perceived task difficulty) n = 64	Task difficulty (Time to solve) n = 16	Task difficulty (Number of picture exchanges) n = 16	Solution variability n = 15	Path dependency n = 16	Variability of category association strength (Range) n = 16	Variability of problem-solving behavior (Reversals) n = 16
Clipart	$\chi^2(2) = 52.64$, $p < .001$. Expert > Good**** Expert > Story**** Good > Story**	$\chi^2(2) = 15.414$, $p < 0.001$ Expert > Good** Expert > Story*** Good = Story	$\chi^2(2) = 8.4952$, $p < 0.05$ Expert = Good (0.07) Expert > Story* Good = Story	$\chi^2(3) = 29.218$, $p < 0.0001$ Expert = Good Expert > Random*** Expert > Story** Good > Random ** Good = Story Story > Random*	$F(2, 45) = 1.734$, $p = 0.188$ Expert = Good Expert = Story Good = Story	$\chi^2(2) = 0.87188$, $p = 0.6467$ Expert = Good Expert = Story Good = Story	$\chi^2(2) = 3.0871$, $p = 0.2136$. Expert = Good Expert = Story Good = Story
Icons	$\chi^2(2) = 51.26$, $p < .001$ Expert > Good**** Expert > Story**** Good > Story****	$\chi^2(2) = 12.497$, $p < 0.01$ Expert > Good** Expert > Story** Good = Story	$\chi^2(2) = 8.6001$, $p < 0.05$ Expert = Good Expert > Story* Good = Story	$\chi^2(3) = 36.561$, $p < 0.0001$ Expert = Good Expert > Random**** Expert > Story**** Good > Random ** Good > Story** Story = Random	$F(2, 45) = 3.005$, $p = 0.0596$ Expert = Good Expert < Story * Good = Story	$\chi^2(2) = 20.333$, $p < 0.0001$ Expert = Good Expert < Story**** Good < Story**	$\chi^2(2) = 0.67583$, $p = 0.7133$ Expert = Good Expert = Story Good = Story
Walmart	$\chi^2(2) = 68.948$, $p < 0.001$ Expert > Good** Expert > Story**** Good > Story****	$\chi^2(2) = 17.839$, $p < 0.001$ Expert = Good Expert >Story*** Good > Story***	$F(2, 45) = 13.95$, $p < .001$ Expert = Good Expert > Story **** Good >Story ***	$\chi2(3) = 28.412$, $p < .001$ Expert = Good Expert > Random *** Expert = Story Good > Random *** Good > Story** Story > Random *	$\chi^2(2) = 11.49$, $p < 0.01$ Expert = Good Expert = Story Good < Story**	$\chi^2(2) = 15.193$, , $p < 0.001$ Expert = Good Expert < Story** Good < Story**	$\chi^2(2) = 9.6111$, $p < 0.01$ Expert = Good Expert > Story * Good > Story*

[2] ****$p < 0.0001$; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; ns, $p > 0.05$

Table 5: Stimulus-specific ANOVA and Kruskal-Wallis test results with relevant post-hoc pairwise tests for structural moves

Variable / Stimuli	**Structural Moves for Expert condition n = 16**	**Structural Moves for Good condition n = 16**	**Structural Moves for Story condition n = 16**
Clipart	$F(2, 45) = 1.289$, $p = 0.285$ 1st = 2nd 1st = 3rd &4th 2nd =3rd &4th	$\chi^2(2) = 1.7477$, $p = 0.4173$ 1st = 2nd 1st = 3rd &4th 2nd =3rd &4th	$F(2, 45) = 0.461$, $p = 0.633$ 1st = 2nd 1st = 3rd &4th 2nd =3rd &4th
Icons	$\chi^2(2) = 13.779$, $p = 0.001019$ 1st = 2nd 1st > 3rd &4th ** 2nd > 3rd &4th **	$\chi^2(2) = 1.7108$, $p = 0.4251$ 1st = 2nd 1st = 3rd &4th 2nd =3rd &4th	$F(2, 45) = 0.014$, $p = 0.986$ 1st = 2nd 1st = 3rd &4th 2nd =3rd &4th
Walmart	$F(2, 45) = 2.602$, $p = 0.0853$ 1st = 2nd 1st > 3rd &4th * 2nd =3rd &4th	$\chi^2(2) = 10.237$, $p = 0.005984$ 1st = 2nd 1st > 3rd &4th** 2nd =3rd &4th	$F(2, 45) = 0.032$, $p = 0.969$ 1st = 2nd 1st = 3rd &4th 2nd =3rd &4th

5.8 Summary of the Statistical Results

In line with our hypothesis, the overall results of the ANOVA or non-parametric Friedman tests shows significant main effect of problem open-endedness on the behavioral and solution outcome variables. The post hoc pairwise tests revealed a general pattern of results on most cases: the Story condition and both similarity conditions constantly showed statistically significant differences, which aligns with our predictions, whereas in several cases, differences between the two similarity conditions (Expert and Good) were non-significant.

Problem open-endedness was hypothesized to have negative effects on problem-solving difficulty. Difficulty was measured using perceptual ratings on the post-experiment survey, solution times, and the number of picture exchanges. Our findings revealed a significant main effect of problem open-endedness on perceived task difficulty. Participants perceived the two similarity problems as significantly more difficult than the Story one, and the Expert task was perceived as significantly more difficult than the Good task. The time taken to complete the task followed a similar pattern. Participants took significantly longer to solve the Expert condition, followed by the Good, and then the Story condition. In terms of task difficulty indicated by the number of picture exchanges, we observed significantly more exchanges in the two similarity conditions compared to the Story condition; however, there were no significant differences between the Expert and Good conditions.

Problem open-endedness was hypothesized to have a positive effect on solution variability, but the variability of solutions for all three levels of open-endedness (Expert, Good, Story) was predicted to be substantially lower than the variability of a random solution. We used the Rand index (RI) similarity measure to assess solution variability. Consistent with our hypothesis, we found a significant main effect of problem open-endedness on solution variability. The post hoc results revealed significant differences in solution variability between Story- and similarity-based categorization problems, but not between the Good and Expert similarity tasks. As hypothesized, solution variability for Expert, Good and Story were all significantly less than the random condition.

A positive effect of problem open-endedness on path dependency was hypothesized. The difference between the RIs of the solution to their initial picture distribution and to the opposite initial picture distribution was used as an indicator of path dependency. Consistent with our hypothesis, we found a significant main effect of problem open-endedness on solution variability. The post-hoc analysis revealed that solutions to Story problems exhibited significantly greater path dependency than both

Good and Expert similarity problems, but path dependency was not statistically different between Good and Expert.

The variability in the solution category association strength was hypothesized to be positively affected by problem open-endedness. We used scores obtained from the concept association survey to assess strength of solution association, and measured variability in the solution association strength using the average range of the strength of solution association. The results showed a statistically significant main effect of problem open-endedness on variability in the solution association strength. The post-hoc test revealed that the range of the strength of solution association was not significantly different between the Expert and Good condition, but significantly higher in the Story condition compared to the two similarity conditions.

We hypothesized that solving an open-ended problem involves moving from strong associations to weak associations in a knowledge network. To measure that, we compared the average solution association scores between groups' four solution categories for every condition separately. In the least open-ended Expert condition, we found significant statistical differences in average solution association scores across the orders of the solution categories. The post-hoc test showed that average solution association for the first and second solution categories identified by participants had significantly stronger association scores than the average of the third and fourth solution categories. Similarly, a significant main effect was found in the Good condition with the post-hoc test showing that the first solution categories had significantly higher scores than the average of third and fourth solution categories, but scores for the second solution categories were not significantly different from the first, nor from the average of the third and fourth solution categories. In the Story condition, association scores were not significantly different regardless of the orders of the solution category completion. Subsequent two-way repeated measures ANOVA analyses demonstrated significant main effects of problem open-endedness and solution category order on average solution association scores; however, the interaction effect was not quite significant at the 0.05 level.

The variability in the problem-solving behavior, measured in terms of the number of picture exchange reversals, was hypothesized to be negatively affected by problem open-endedness. We found a main effect of problem open-endedness on the number of exchange reversals across the three conditions with post-hoc results showing significantly fewer reversals in the Story condition than in

the two similarity-based conditions, but no significant difference between the two similarity conditions.

Finally, we conducted separate stimulus-specific hypothesis testing to explore generalizability and the sensitivity of our results to different stimuli. The results of this analysis generally supported the main results, but also revealed differences across our three stimulus types, with the Walmart results most consistent with our main results, followed by Icon and then Clipart.

These hypothesis testing results provide insight into the effects of problem open-endedness on solution outcomes and group problem-solving behavior. However, to gain further insights into the cognitive, communication and behavioral processes involved in open-ended problem solving, we analyzed the recorded verbal protocols of the tasks and conducted a qualitative analysis on a select number of them. This analysis will be presented in the following chapter.

Chapter 6
Verbal Protocol Analysis

Chapter 5 presented quantitative analysis of certain hypothesized effects of open-endedness on problem-solving. This statistical analysis provides insight into the outcomes and behaviors associated with solving problems at different levels of open-endedness. Nonetheless, the analysis is limited in terms of providing a comprehensive view of the underlying behavioral and cognitive processes involved in solving such open-ended problems. For instance, what is the process undertaken by participants, so the problem was considered as more difficult in the less open-ended problems compared to the more open-ended problems? What processes contributed to path dependency effects, which were more prominent in more open-ended than less open-ended problems? This raises the question about the underlying cognitive processes that might have led to such effects on problem-solving behavior. To address this limitation and get a better understanding of the underlying problem-solving processes, we conducted a qualitative analysis of the verbal protocols of nine problem-solving task sessions, including three from each of the experimental conditions (Expert, Good and Story)[3]. This analysis was based on the video recordings and transcripts of these tasks. All nine task sessions used the Clipart stimulus, to enable some degree of comparison across conditions. The task sessions analyzed were selected randomly from both versions of pictures' initial distributions.

During the problem-solving process, we observed a general pattern in the behavior of the group while solving the three problem conditions. At the start of each task, there is typically a period of silence, during which participants look over the 16 picture items and seem to be thinking of possible solutions. This is followed by a discussion in which group members suggest possible solution

[3] The nine tasks are as follow:
Expert 1 (Experiment A5 / Clipart/ Expert / Version 1)
Expert 2 (Experiment B2 / Clipart/ Expert / Version 1)
Expert 3 (Experiment A'1 / Clipart/ Expert / Version 2)
Good 1 (Experiment D5 / Clipart/ Good / Version 1)
Good 2 (Experiment C3/ Clipart/ Good / Version 1)
Good 3 (Experiment D1/ Clipart/ Good / Version 1)
Story 1 (Experiment E3/ Clipart/ Story / Version 1)
Story 2 (Experiment F'3/ Clipart/ Story / Version 2)
Story 3 (Experiment F4/ Clipart/ Story / Version 1)

categories, and list items in each of the categories they identified. Then, because these initial categories may be incomplete or in conflict with one another, the group goes through repeated cycles of conflict recognition and resolution until they reach a final solution. During the analysis of the nine task sessions, conflicts that emerged within the tasks and the subsequent resolution strategies employed were noted. Through this analysis, certain consistent types of conflicts and resolution strategies were identified across all the tasks. Tables 6 and 7 present the types of conflicts identified, and the resolution strategies employed to address them along with their corresponding definitions.

Table 6: Conflict types observed in the problem-solving tasks

Conflict Type	**Definition**
Scarcity	Identifying a category but noticing missing items that are needed to complete this category (e.g., noticing only three "food" items or two "sport" items)
Redundancy	Noticing that more than four items fit as potential members of an identified category (e.g., after identifying a "human made" category, they notice that other items are also human made). Alternatively, sometimes there are two items available to complete a category that only requires one additional item.
An item fits better in another category	Noticing that an item initially classified in one solution category would be more suitable in another category, or that an item from someone else's category would fit better in their own category. For instance, "waves" may be better suited in an "element" category than a "sport" category.
Weak category	Uncertainty about a category because there is no clear membership relation between the category and some or all items inside it. For example, asking a question about what the category is, or identifying an item (e.g., "balloon") that does not seem to be strong member of a category (e.g., "activity")
Random item	Identifying a random item that does not seem to belong to any of the defined categories. For example, noticing an item (e.g., "bird") does not belong to any of the categories. In the Story condition, a random item is defined when a person offers an item that is not needed for her/his solution.

Table 7: Conflict resolution strategies observed in the problem-solving tasks

Resolution Strategy	**Definition**
Affordance-based assumption	Making a flexible interpretation of an item such that it fits a predefined solution category (e.g., the "haybale with the pumpkin" affords potential interpretations of "food" or "Halloween").
Relabeling the category	Changing the name of a category to make it more specific (narrow) to include fewer items, or more abstract (broader) to encompass more items. An example of narrowing is calling a category that has (backpack, tennis, soccer ball, house) a "family house". Broadening a category of "sport" to "activity "is an example of relabeling for abstracting proposes.
Compromising	Suggesting or making an exchange based on what item fits better into other categories. For example, one group combined "hay bale" with "vegetables" to represent the earth in a category of "elements", but then narrowed the category to "vegetables" when it was decided that "hay bales" better represented "Fall" or "Thanksgiving" for another category.
New solution category	Proposing a new solution category that has not been previously considered, such as grouping seemingly unrelated items like (fire, snow, water, and bird) together into a new category called "things belonging to nature".
Discard the conflict	Discarding a conflict that has been raised about a specific category and accepting the category without making exchanges. For example, one person proposed replacing "man" with "haybale" to complete the "education" and "seasons" categories, but another person argued to keep the original items and provided a justification for doing so.
Discard the solution idea	Participants abandon the category that has a conflict by not dealing with the conflicts and switching to another topic; such as giving up trying to find a fourth item to complete a category and changing the topic.

To further explain how the problem-solving process unfolds and how the solution emerges, we will go through the transcript of an Expert task (Expert 1), followed by a Good task (Good 1), and then a Story task (Story 1). These three tasks were selected from the same version of the initial picture distribution (version 1) as shown in Figure 20.

As explained in the Method section, participants were presented with the 16 pictures on a computer screen divided into four quadrants, with four pictures per quadrant, and each participant assigned to one of the quadrants. They solved the problems by exchanging pictures with one another until each had four pictures from the same category (Expert and Good conditions), or four pictures making up a story (Story condition). Group discussions were recorded and automatically transcribed and time-

stamped using the Webex video conferencing software (Cisco WebEx, 2020). Transcript text was later manually corrected for errors by the researcher. In the discussion transcripts presented below, we will use the letters A to D to denote each of the four participants, and use "EXP" to denote the experimenter.

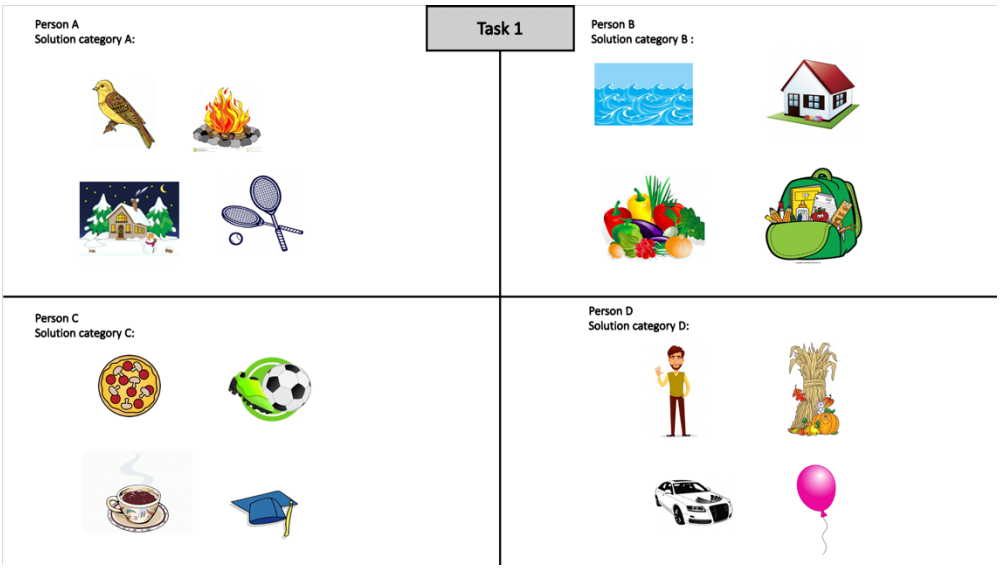


Figure 20: The initial picture distribution of a Clipart task

6.1 Verbal Protocol Analysis for an Expert Condition Task

In this section we analyze the verbal protocol of a group completing an Expert task (Expert 1) by going through chunks of the transcript of the group discussion. Each chunk contains the dialogue and corresponding timestamp.

After being presented with the 16 pictures, the participants spent about 17 seconds looking at the pictures without speaking. Then they started a discussion by generating ideas of possible solution categories.

287 00:35:36.599 → 00:35:39.898 B: Okay, so I think there's a food category. 288 00:35:39.898 --> 00:35:43.559 B: Like school. 289 00:35:43.559 --> 00:35:47.849 A: Adulthood? B: Adulthood. 290 00:35:52.708 --> 00:35:59.338 A: why don't we say sport D: what what is the food category? 291 00:35:59.338 --> 00:36:04.228 B: There's like the pizza vegetables, vegetables, coffee. D: ok ok ok 292 00:36:04.228 --> 00:36:09.119 B: Is there 1 more?	293 00:36:09.119 --> 00:36:16.018 D: Does that count it uh? 294 00:36:16.018 --> 00:36:20.548 D: The pumpkin thing does that count as a food?. 295 00:36:20.548 --> 00:36:24.838 C: Yeah, B: yeah, I think that too. I think the…. 296 00:36:26.458 --> 00:36:31.108 B: And then there's like elements like water fire. 297 00:36:31.108 --> 00:36:35.699 D: Right, water fire. B: snow, Maybe 298 00:36:35.699 --> 00:36:39.208 A: Snow is water I think.

As shown in the above transcript, participants identified potential categories of *food, school, adulthood*, and *sport* (lines 287-290). Then, they started asking questions about the items in these categories. They listed three items in the *food* category, which are *pizza*, *vegetables*, and *coffee* (line 291). After 20 seconds (line 294), they identified another possible item that might also fit into the *food* category: the *haybale with the pumpkin*. It seems that the group identified a conflict of scarcity (a missing item), so to deal with this conflict the group made an affordance-based assumption that the *haybale with the pumpkin* affords a potential interpretation of food to be a fourth *food* item. Then, they identified another category of *elements* that includes *water*, *fire*, and *snow* (lines 296-298).

299 00:36:40.438 --> 00:36:43.648 B: Tennis and football. 300 00:36:47.489 --> 00:36:52.079 A: If you could do activities like, and and'it's. 301 00:36:52.079 --> 00:36:55.469 A: tennis and football, driving a car and. 302 00:36:55.469 --> 00:37:04.018 A: A birthday party such a balloon, or, I guess graduation, hmm I feel like the book bag and the hat. 303 00:37:04.018 --> 00:37:07.018 A: Can go together, um. 304 00:37:07.018 --> 00:37:14.579	A: You could talk about, like a school house where you have a man going to school with a book that he graduates at a house. 305 00:37:14.579 --> 00:37:19.648 A: That should be 4 things I count. D: Yes. 309 00:37:42.358 --> 00:37:46.798 A: Bird watching is an activity. 310 00:37:48.929 --> 00:37:53.518 A: I think I have activities. 311 00:37:54.778 --> 00:38:00.088 B: Okay' I'll take the food stuff again. 312 00:38:00.088 --> 00:38:04.829 C: I'll take the elements, I guess.

In line 299 they grouped the *tennis* and *soccer ball* as a tentative category, initially without labeling it. Then, after a silence of 6 seconds, they called that group *activities* and added the *car* as *driving a car* and the *balloon* as *birthday party* or *graduation* (lines 301-302). In this case, they grouped what seems to be the most obvious *sport* items together, but they seem to identified another scarcity conflict as there are only two sport items (*tennis racket and soccer ball*). They then relabeled the category to be *activity* as a way of abstracting or broadening it to include more non-sport items such as "*car*", "*balloon*", and "*graduation*".

While discussing the *activity* category, another solution category emerged simultaneously. Person A noticed that the *graduation hat* and the *backpack* could go together and added the *house* and the *man* to them as one category: "*schoolhouse where you have a man going to school with a book that he graduates at a house*" (line 304). Then three participants settled on the category names they wanted to take, and they exchanged the pictures accordingly (lines 305-312). These are the categories they had after exchanging:

A: (bird, soccer ball, car, tennis) > Activities

B: (Pizza, coffee, vegetable, pumpkin) > Food

C: (water, Fire, house, winter house) > Elements

D: (man , backpack, grad-hat, balloon) > Unnamed, but this seems similar to the school house scenario given above

331 00:40:07.199 --> 00:40:10.349 A: So, what's everyone's category? just to be clear. 332 00:40:11.820 --> 00:40:18.449 B: Well, I think that "C" and my categories pretty distinct the other 2 are a bit like. 333 00:40:18.449 --> 00:40:21.809 B: it's hard to differentiate between them. 334 00:40:21.809 --> 00:40:30.329 D: graduation, … D: Like school, backpack is graduating. 335 00:40:30.329 --> 00:40:34.650 D: Graduation hat and then balloon.	336 00:40:34.650 --> 00:40:39.269 A: But I feel like "C" has a weaker. 337 00:40:39.269 --> 00:40:46.079 A: Um, category, because the house doesn't may be fit in with elements, but if you switch balloon and house, so it can be like, balloon is like, air. 338 00:40:46.079 --> 00:40:51.269 A: And house, if the house goes to you, "D", then becomes like, school. 339 00:40:51.269 --> 00:40:54.599 A: being the category where person goes to school and graduates 340 00:40:54.599 --> 00:40:58.440 D: I see. C: Yeah

After they had exchanged pictures and created some solution categories, they started another round of discussion, asking questions about the existing categories, identifying conflicts in them, and using some strategies to deal with these conflicts (lines 331-339). One conflict emerged when person B argued that A's and D's categories were weaker ones. It seems that, as a way of dealing with the conflict, person D justified their category indicating that all items in their quadrant belong to a category of *graduation*. Then, it seems that another conflict of an item that fits better into another category was identified when person A said that category C was a weak one because the *balloon* fit better into the *element* category (representing air) than the *winter house* (as snow). As a resolution, person A made a compromise by suggesting that the *house* go to category D as a school in exchange for the *balloon*. So they made the exchange, accordingly, resulting in the following categories:

A: (bird, soccer ball, car, tennis), > Activities

B: (Pizza, coffee, vegetable, pumpkin) > Food

C: (water, Fire, balloon, winter house)> Elements

D: (man, backpack, grad-hat, house)> School

341 00:40:58.440 --> 00:41:06.030 B: But then the bird is kind of random in category 1 A: Bird watching is an activity. B: Oh, yeah, that's true. 347 00:41:41.460 --> 00:41:44.789 Silence. 348 00:41:48.389 --> 00:41:56.489 A: "C", What's your category? 349 00:41:56.489 --> 00:42:01.079 C: umm like elements I guess 350 00:42:02.159 --> 00:42:12.329 D: you can umm like a season.	351 00:42:13.710 --> 00:42:18.630 D: Like, fall, I don't know, I'm just throwing ideas. 352 00:42:19.710 --> 00:42:24.480 D: "B", Can you have, like, fall category where you're like, your harvest food? 353 00:42:24.480 --> 00:42:28.679 B: Yeah, yeah, I was thinking that too. Cause it's like the pumpkin. : 364 00:43:41.909 --> 00:43:46.800 B: I don't know. I'm not going to do seasons, we don't have seasons here. {Not clear} 365 00:43:46.800 --> 00:43:50.610 D: yeah might be wrong

Then, another conflict of weak category arose when person B identified the *bird* as a "kind of random" item (line 341) that did not seem to belong in the *activities* category. To resolve this conflict, person A justified it by relating it to the activity of "bird watching". It seems that person A made an affordance-based assumption considering the *bird* as affording a potential interpretation of "bird

watching activity" to complete the category of *activity*. After 15 seconds of silence, person A asked C about their category; person C answered "elements I guess", suggesting they were unsure. This could be classified as recognition of a conflict that a category was weak or not clear. So to deal with this conflict, person D tried to suggest a new solution category of *seasons*, where *pumpkin* and *vegetables* in B represent *fall* and *fire* goes with *winter house*. They discussed it for around a minute and a half, but it seemed not to work, so the group discarded this new category idea. Suggesting new categories seems to be a strategy that person D used to resolve the conflict of a weak category though this attempt to resolve the conflict was not implemented.

367 00:43:57.059 --> 00:44:06.000 A: Okay, so we have these categories right now. 368 00:44:06.000 --> 00:44:09.269 A: Um, the often there are any categories that are 369 00:44:09.269 --> 00:44:15.090 A: a little weak, or we can move towards to make our overall. 370 00:44:15.090 --> 00:44:19.440 A: task 1 stronger? 372 00:44:28.019 --> 00:44:31.409 D: I think it's, it's. 373 00:44:31.409 --> 00:44:39.900 D: The balloon from "C" seems a little awkward, but, except for that. 374 00:44:39.900 --> 00:44:44.610 D: I think it makes sense. 375 00:44:47.039 --> 00:44:52.050 D: Yeah, I can be school ,"C" can be elements. 376 00:44:52.050 --> 00:44:56.969 D: Yeah. 377 00:44:56.969 --> 00:45:01.920 D: I think this is fine for now. I got no objection. B: I think this is fine too 378 00:45:01.920 --> 00:45:04.920 C: Sure. Okay.	379 00:45:04.920 --> 00:45:10.559 EXP: So, do you want to give names or are you still thinking. 380 00:45:11.610 --> 00:45:15.239 EXP: Do you want to give names to the categories or not yet. 381 00:45:15.239 --> 00:45:23.099 A: 1, last thing … ... {Not clear} 382 00:45:23.099 --> 00:45:27.030 A: Maybe. 383 00:45:27.030 --> 00:45:30.690 EXP: Um, we couldn't hear you actually. C: Yeah. 384 00:45:30.690 --> 00:45:34.769 A: Sorry, "C", do you think it makes sense to swap the bird with the balloon maybe. 385 00:45:34.769 --> 00:45:42.059 A: Does it make sense. I'm asking what what you think. 386 00:45:42.059 --> 00:45:46.139 C: Oh, sure. And then, I guess I'll be like natural elements. 387 00:45:46.139 --> 00:45:49.199 A: Yeah, like nature. That I was thinking. 388 00:45:49.199 --> 00:45:54.960 But is a balloon an activity?. 389 00:45:54.960 --> 00:45:58.170 D: birthday party A: a balloon is like a birthday party.

390 00:45:58.170 --> 00:46:03.269 A: Kind of activity. Yeah, you read my mind here we go. 391 00:46:03.269 --> 00:46:11.070 D: it might make more sense than. 392 00:46:11.070 --> 00:46:16.949 Bird watch. Well, would they really put bird watching as an Activity.	393 00:46:16.949 --> 00:46:27.119 A: These are great questions. I don't think they would… 394 00:46:27.119 --> 00:46:35.429 A: so, yeah, let's switch "C" and that gives you a more concrete category. And activities is already very generalizes so D: right A: technically anything could could be with that.

Another example of a weak category with an item that did not seem to belong occurred in line 373, when person D suggested the balloon was an "awkward" item in the *element* category (category C). Initially, the group seemed to discard this conflict by disagreeing with person D's suggestion and almost accepting *balloon* as part of the *element* category. However, person A reconsidered this conflict and suggested switching the *balloon* (to be in the *activities* category) with the *bird* (to be in the *elements* category), and changing the name of the *elements* category to *nature* (lines 384-390). They accepted *balloon* in the *activity* category as a party thing, and to accommodate the *bird* in category C, they adjusted the name of the category from *element* to *nature*. They spent some time justifying this last suggestion of switching the *bird* and the *balloon*, asking questions to make sure that this was an acceptable solution. It seems that they resolved the conflict through compromising which item fits better into a category (*balloon* an as *activity* rather than *bird*), and by adjusting the name of the category (from *element* to *nature*) to accommodate the newly added item (*bird*). After they all agreed on swapping the *balloon* with the *bird*, they gave names to the final solution categories they settled on:

A: Activities: (balloon, soccer ball, car, tennis)

B: Food: (pizza, coffee, vegetable, pumpkin)

C: Nature: (water, Fire, bird, winter house)

D: School: (man, backpack, grad-hat, house)

The task was completed when the experimenter asked the participants if they were satisfied with the solution, to which they answered "yes". The complete transcript for this task and a summary of the conflicts and resolutions in that task are available in Appendices I and J (Task: Expert 1).

6.2 Verbal Protocol Analysis for a Good Condition task

Problem solving in the Good condition followed a process that was similar to that in the Expert condition, with some differences. Therefore, for the Good condition task, we will conduct a brief analysis of the verbal protocol to identify the key behaviors observed during the process using examples from the transcript. The full transcript of this task and a summary of the observed conflicts and resolutions can be found in Appendices I and J (Task: Good 1).

After 30 seconds of looking at the pictures in silence, participants started generating ideas of possible solution categories. Person A noticed *buildings* (*Winter house* "calling it *cottage*" and *house*); and then person B and C identified other possible categories, including *seasons* (*hot chocolate* and *winter house* representing *winter*; *pumpkin* and *vegetables* representing *fall harvest*; *waves* representing *summer*, *spring*). Then person B identified *education* as a category (*backpack and graduation hat*), and Person B then suggested combining the *backpack* and *graduation hat* with the *fall theme* "because it's back to school".

They then started with the *winter* theme idea for category A by switching the *tennis racket* (they called it *badminton*) with the *coffee* (they called it *hot chocolate*). The categories they have after exchanging are as follows:

A: (bird, fire, winter house, coffee) > Winter Category

B: (backpack, waves, vegetables, house,) > Unnamed

C: (Pizza, soccer ball, tennis, grad hat) > Unnamed

D: (man, hay, balloon, car) > Unnamed

202-203 00:30:29.189 --> 00:30:41.009 B: I think that bird kind of looks like a winter bird that would be in there in winter. So I think maybe Just a category A is complete. What's your senses?

Person B suggested the *bird*, which was an existing item in their quadrant, could belong to the *winter* category, so this category was complete. It seems that they completed the category by using an affordance-based assumption that the *bird* was a *winter bird*.

204-206
00:30:41.009 --> 00:30:55.348
C: I think I can change the the graduation hat with the with the wave from B.

207
00:30:55.348 --> 00:30:58.348
B: No, I think the waves are part of a cottage theme
C: Yeah

Then, in what seems to be an attempt to complete the *Fall/back to school* category discussed earlier, person C wanted to exchange the *graduation hat* with the *waves* from person B; however, person B thought this would conflict with a *cottage* theme (though they had not discussed the cottage theme earlier). This seems to be a conflict of an item (*waves*) that fits better in another category (*cottage*), so they discarded this solution idea and switched to a *cottage* theme category. Working toward the *cottage* theme, they exchanged the *man* with the *vegetables*, the *car* with the *backpack*. The resulting categories were as follows:

A: (bird, Fire, winter house, coffee) > Winter Category

B: (car, wave, man, house) > Cottage (presumably)

C: (pizza, soccer ball, tennis, grad hat) > Unnamed

D: (vegetables, hay, balloon backpack) > Unnamed

They then switched the *pizza* with the *backpack* without giving an explanation; however, this exchange seems to be a way of combining the school items together (*backpack* and *graduation hat*). The resulting categories were as follow:

A: (bird, Fire, winter house, coffee) > Winter Category

B: (car, wave, man, house) > Cottage (presumably)

C: (backpack, soccer ball, tennis, grad hat) > School (presumably)

D: (vegetables, hay, balloon, pizza) > Unnamed

221
00:32:14.788 --> 00:32:19.169
B: why is there a balloon there.

222
00:32:19.169 --> 00:32:24.419
C: I can see the balloon with the graduation hat.

223
00:32:24.419 --> 00:32:32.009
B: That's what I was thinking as well. Yeah. Maybe what sport do you play in like, an August September.

224
00:32:32.009 --> 00:32:37.288
B: Soccer badminton maybe soccer.

225
00:32:38.848 --> 00:32:43.709
c: With the with
B: with the balloon

226
00:32:45.659 --> 00:32:48.959
C: Okay, yeah, that works may be

Then they recognized a conflict of a weak category where the *balloon* seemed to be an unrelated item in category D. To deal with the conflict, person C proposed that the *balloon* could go with the *graduation hat*, and person B made an affordance-based assumption that *soccer* was a game that is played in August and September (around Fall time); therefore, they made a compromise by swapping the *balloon* with the *soccer ball* in category C. The resulted categories are:

A: (bird, Fire, winter house, coffee) > Winter Category

B: (car, wave, man, house) > Cottage (presumably)

C: (backpack, balloon, tennis, grad hat) > School/Fall (presumably)

D: (vegetables, hay, soccer ball, pizza) > Unnamed

228
00:32:59.939 --> 00:33:07.828
A: What would category B be though?
B: This would be like a cottage time when you take your car to that cottage and there's like, waves there and that there's like a pool not pool, but like, maybe a river or like a lake there
C: maybe a vacation
B: vacation type. Yeah

As seen in the above discussion, the final exchange appears to have caused some confusion. A conflict of a weak category was recognized by person A, who expressed uncertainty about Category B. As a means of resolving the conflict, person B justified this category by giving a scenario that combined all items and relabeled it to *cottages* or *vacation*.

230
00:33:19.348 --> 00:33:23.429
C: What about D, What would it be?

232
00:33:30.298 --> 00:33:36.298
B: Like, I guess you can make pizza with these fruits of the top that you can put, I don't know.

Then, in the above discussion, a similar conflict of weak category was raised when person C asked about category D. To deal with the conflict, person B suggested a possible scenario for that category, but seemed unsure. This conflict of weak category seemed to be due to uncertainty about the category. It was followed by a suggestion of a potential way of labeling the category such that the items in it make sense as a group.

233
00:33:43.138 --> 00:33:51.628
B: Maybe we should keep the balloon where it is. I don't know where it was, like, in category D.

234
00:33:53.278 --> 00:33:57.269
B:Yeah, because A and B feel like it's Don, I think.

235
00:33:57.269 --> 00:34:02.159
C, and D don't look as good as.

After a few seconds of silence, they decided to swap back the *balloon* with the *soccer ball* because they thought that categories A and B were complete but not C, and D. The categories they had after exchanging were:

A: (bird, fire, winter house, coffee) > Winter

B: (car, wave, man, house) > Cottage/Vacation

C: (backpack, soccer ball, tennis, grad hat) > School/Fall (presumably)

D: (vegetables, hay, balloon, pizza) > Unnamed

239
00:34:22.168 --> 00:34:25.978
C: I am thinking what D is.

242
00:34:41.369 --> 00:34:47.818
C: There is vegetables, pizza, pumpkin and a balloon, right?

243
00:34:47.818 --> 00:34:57.449
B: The Halloween party may be. I don't know, I guess
A: yeah,
B: because you need, because you need because you need vegetables to make pizza and then.

244
00:34:57.449 --> 00:35:01.739
B: You can have you need a pumpkin to have a Halloween and balloon is just for a party kind of thing.

245
00:35:01.739 --> 00:35:04.768
C: That works actually. [laughing]

As seen in the above transcript, they were still not sure about what category D should be called and after a 30 second period of silence, person B suggested the name "*Halloween party*" and provided a justification. They laughed and agreed on that name. There were no further switches but here they decided on the name of the categories as follows:

A: (bird, fire, winter house, coffee) > Winter

B: (car, wave, man, house,) > Vacation/ Cottage

C: (backpack, soccer ball, tennis, grad hat) > School activities

D: (vegetables, hay, balloon, pizza) > Halloween party

When giving names for the categories at the end, they seemed unsure about the name of category C; they used the word "maybe", and there was a questioning tone when they were giving the name of the category.

6.3 Verbal Protocol Analysis for a Story Condition Task

Following is a partial analysis of the verbal protocol of a Story condition task (Story 1), highlighting significant behaviors observed during the process. For a complete record of this task and a summary of the conflicts and resolutions in that task, please refer to Story 1 in Appendices I and J, respectively.

Participants started the discussion after around 40 seconds of looking at the pictures in silence. They proposed a strategy that one of them start a story and then the rest of the group go from there to make their stories from the rest of the pictures. They considered pictures that might go together like *pizza* and the *beach* (*waves*)

233-237
00:31:26.489 --> 00:32:09.298
D: I think for mine for category D. Like, I can honestly just swap out that pumpkin thing for almost anything else here and that would be a story. So it'd be like this guy start uh has a birthday, he needs to go to so many shops in his car and goes and buys a balloon, boom That's like a story right there. I mean, to swap out that pumpkin thing but, I mean if you guys are down with that?
A: Yeah. Okay. What can you swap the pumpkin with ?
D: Ah I don't know

238
00:31:57.929 --> 00:32:09.298
A: the pizza like pizza party
D: Yeah, it could be a pizza pizza.
C: Sure
A: Yeah. That's good.
D: Yeah, so that's one.

Person D started offering other members the items that they did not need (*pumpkin*) proposing a *birthday* story based on 3 out of the 4 items they had. This can be considered as a conflict of a random item where the *pumpkin* did not seem to belong to the story person D had in mind. They then compromised by switching the *pumpkin* with the *pizza* from person C since *pizza* fit the theme of *birthday party*.

A: (bird, fire, winter house, tennis) > Unnamed

B: (waves, house, vegetable, backpack) > Unnamed

C: (pumpkin, soccer ball, coffee, grad-hat) > Unnamed

D: (man, pizza, car, balloon) > Birthday story

240-243
00:32:15.659 --> 00:32:40.108
C: I feel like I have too much food I'm not sure how to come up with a story with the pumpkin.
A: Yeah,
C: yeah, like the pumpkin just seems out of place here I think.
C: umm let me See.

244-245
00:32:43.169 --> 00:33:02.939
D: You could, like, stretch the truth and be like, there's that house there, "B", in category B. There's that house there that's that's on a farm that grows crops like pumpkins and vegetables, and then throw like a 4th thing in there. So that would use the house the, and the vegetables and then something else.

Person C was not sure how to come up with a story with the *pumpkin*. This seems to be a conflict of weak category due to uncertainty on how to link all items in a story. To deal with this conflict, person D proposed a solution idea of a *farm story* for category B that incorporated the *pumpkin* with the *house* and *vegetables*. They agreed on that story for category B, and to complete it, they switched the *pumpkin* from C's category with the *backpack* from B's, and the *bird* from A's with the *waves* (calling it *beach*) from B's. At the same time, an idea of a *graduation story* had emerged for category C, which included the *backpack* and the *grad-hat*. The resulting categories were as follows:

A: (waves, fire, winter house, tennis) > Unnamed

B: (bird, house, vegetable, pumpkin) > Farm story

C: (backpack, soccer ball, coffee, grad-hat) > Graduation story

D: (man, pizza, car, balloon) > Birthday story

255-257
00:33:55.288 --> 00:34:06.659
B: Like I went to a beach house, or something Had a bonfire and played some tennis; that works.
D: Yeah,
B: yeah. [Laugh]

After they swapped those pictures, person B proposed a solution for category A of a *beach house* story that incorporated all items existing in that category. They agreed and laughed.

258-260
00:34:06.659 --> 00:34:33.958
A: Oh, wait.
D: Oh, wait, no, I would swap the I would swap the, the tennis. whatever that badminton thing for the coffee, right?
A: Yeah so it's like a cold winter day. coz, it's like that that house,
D: yes. Yeah. That's what I think cold winter day for category A and then category C, you'd be like, there's this scholar that graduated in, like, high academics with, like, a love for sports or something. I don't know.

However, person A and person C seemed to recognize a conflict of an item (*coffee*) that fit better into category A, and suggested switching the *tennis-racket* from A with the *coffee* from C, such that the *coffee* would fit better in category A as a story of *cold winter day* and the *tennis* fits in category C as a *graduation* story.

263-265
00:34:45.449 -->00:35:20.009
A: but like the, I don't think the wave, like, the beach fits into that then.
D: Oh, but like, I mean, that you could argue, that's like, you know, how like, beaches on Airbnb they sell during the winter time, right? It's like a cozy like winter stay, it could still be on the beach. It's just that you know, yeah, like a winter potty or something.

Person A recognized a conflict of a weak story for the proposed story of category A, such that the *beach* (*waves*) did not fit the theme of *winter house*. Then, they discarded that as a conflict by making a justification that the *waves* represent the beach house on a winter day. They all agree on the assumption and laughed. Accordingly, they switched the *tennis* from A with the *coffee* from C and started writing their stories as follows:

A: (waves, fire, winter house, coffee): Someone rents a condo for a winter vacation close to the beach where they enjoy a warm fire and warm tea indoors

B: (bird, house, vegetable, pumpkin): A bird flies by a farmhouse that grows crops like vegetables and pumpkins.

C: (backpack, soccer ball, tennis, grad-hat): A graduate student packs his cleats and racquets in his backpack.

D: (man, pizza, car, balloon): There's a man who has to attend a pizza party. So he goes out in his nice car and grabs a pizza and a balloon to look festive.

6.4 General Observations on the Process

The verbal protocol analyses provide further insight into the processes involved in solving the open-ended problems used in our experiment. The above analyses were part of the nine verbal protocol analyses that we conducted. Overall, in the nine tasks we analyzed, some common behaviors were observed in the problem-solving process across the three conditions. After participants are presented with the pictures in each of the tasks, they first spend some time looking at the pictures in silence (from 17 seconds to a minute and a half). During this time, participants seem to be individually

exploring the problem and thinking of possible solutions. After that, they start a group discussion in which they identify pictures that go together, generate possible solution categories from the pictures they have seen, inquire about the categories they have generated and list the items that belong to each category. Then they start executing the potential solutions and building categories by exchanging pictures based on their initial discussions. While attempting to implement potential solutions, the groups engage in a series of further discussions, identifying and resolving conflicts with the categories they have created until they eventually reach a final solution.

We view the iteration of conflict recognition and resolution as a main element in this process. Essentially, recognizing and resolving conflicts is an iterative process that the group undergoes during the problem-solving task until they reach a satisfactory solution. The transcripts analyzed in sections 6.1-6.3 include examples of most of the conflicts the group recognized and resolved while solving the problems listed in Table 6. The conflict of redundancy, however, has not been shown in these transcripts. An example of redundancy conflict was observed in the Expert 2 protocol, when one participant suggested a solution category of *elements* that included air (*balloon*), water (*waves*), fire (*campfire*) and earth, but there were two items that could represent earth, either *haybale* or *vegetables*. In this case the redundancy was that more than one item (*haybale* and *vegetables*) was available to fulfill the category of elements but only one item was required for completion. Table 8 presents an overview of the conflict types observed in the nine tasks, along with their corresponding resolution strategies. For a more detailed description of each conflict and its respective resolution in the nine tasks, refer to Appendix J.

Table 8: Conflict observed with their corresponding resolution strategy in the nine tasks[4]

Resolution / Conflict	Affordance-based assumption	Relabeling the category	Compromising	New solution category	Discard the conflict	Discard the solution idea
Scarcity	X1C1; X2C1; X2C2a, b, c; X3C1; X3C2, X3C8, G1C1; G2C1; G3C1	X1C2; G3C1				X3C5
Redundancy		X2C6	X3C7	X3C6	X3C6	
An item fits better in another category		G2C5	X1C4; X3C3; G3C2; S1C3; S2C1; S3C3; S3C4	X3C4	X3C4; G2C5	G1C2
Weak category	X2C4; X2C5a; X3C9; G1C3; G2C3, S1C2; S2C3	X1C7b, X2C5b, X2C7, X2C8, G1C4; G1C5a, c; G3C4b	X1C7b, X2C4, G1C3; G1C5b	X1C6; X2C3	X1C3 X1C7a; G3C3; G3C4a; G3C4b; S1C4	
Random item	X1C5; G2C2; S2C4	G2C4	S1C1; S2C2; S3C1; S3C2;	G2C2		

At the beginning of the task, when participants have their initial discussion, they generate some possible solution categories without fully recognizing all the potential conflicts related to them. However, as they exchange pictures, they seem to become more aware of the constraints limiting the current solution and gradually develop a better understanding of the problem situation. As a result, they engage in more constructive cycles of conflict recognition and resolution after exchanging

[4] The conflicts have been coded such that the first letter represents the condition (X: Expert, G: Good, S: Story), followed by the task number. The letter "C" and subsequent number indicate the conflict order, while the lowercase letter represents the attempt to resolve it. For example, “X2C4” denotes the fourth conflict in the second Expert task, resolved in a single attempt, while “G3C4b” signifies the second attempt to resolve the third conflict in the fourth Good task.

pictures compared to the initial discussion. In the tasks we analyzed, we observed that the solution categories generated by the group at the beginning of the task appeared to be incomplete or contained more than four items. For instance, during the initial discussion of the Expert task 1, the group identified an *element* category with only three items (*waves* representing water, *fire*, *winter house* representing snow), and an *activity* category with more than four items (*tennis racket*, *soccer ball*, car, *balloon*, *graduation hat*, *bird*). However, after exchanging pictures and going through several rounds of conflict recognition and resolution, they eventually arrived at an activities category consisting of (*balloon*, *soccer ball*, *car*, *tennis rackets*). The *element* category was relabeled as *nature* and included (*water*, *fire*, *bird, winter house*). This might suggest that the participants were able to visualize the conflicts in the solution more effectively and actively engage in conflict recognition and resolution during the later stages after exchanging pictures.

One possible explanation for this is that there is a substantial amount of information to consider, which makes it difficult to pay attention to everything or predict all the potential conflicts associated with every potential solution category they initially identify. The conflicts become apparent only when they attempt to implement potential solutions by exchanging pictures. In other words, as the group tries to create categories, they introduce new constraints, leading to conflicts that need to be resolved. Thus, the participants not only fail to recognize these conflicts initially, but some conflicts arise directly from the decisions made by the group itself, based on incomplete information.

A general observation at the level of category and the items is that when building categories, participants seem to first pay attention to categories that seem to have a larger number of highly connected items followed by categories that have a smaller number of highly connected items. Then they work with the rest of the items by using conflict resolution strategies, such as making affordance-based assumptions for the missing items, abstracting the categories to include extra items, or adjusting the category name to be more representative to its members. For example, in several cases we observed participants start by mentioning the *food* category and identifying three items that seemed obviously related to food: the *cup of coffee*, *vegetables*, and *pizza*. To complete the food category, they often included the *pumpkin* with the *hay bale* picture as a fourth item, based on the affordance-based assumption that the picture of *pumpkin with the hay bale* can be interpreted as a kind of food. On the other hand, several groups identified categories with fewer obvious members, such as *sports*, including the *tennis racket* and *soccer ball*, or *school*, including the *backpack* and *graduation hat*. Because they could not find additional sports or school items, they were forced to

include extra items that broadened these categories, and consequently also relabelled the categories to better reflect the broader contents (e.g., *activities*).

In the later stages of the problem-solving process, participants seemed to be willing to accept weaker categories without making further exchanges. More specifically, after finalizing the first two or three solution categories, they end up with one or two categories that might not be coherent or clear for them, including poorly fitting items that do not go well with the others. Instead of restructuring the previously completed categories, they dealt with the conflict by making affordance-based assumptions about the poorly fitting items, interpreting them in ways that better fit as members of the existing categories, relabeling the categories in broader ways that could accommodate the poorly fitting items, or by discarding a conflict and accepting the solution as is and giving justifications for the weaker categories. Following is an example taken from the end of the Expert 3 task where the categories the group had at that point were:

A: (pizza, coffee, tennis, soccer ball) > Round

B: (balloon, fire, waves, vegetable) > Elements

C: (bird, winter house, house, hay) > Seasons (partially complete)

D: (backpack, man, car, graduation hat) > Education (partially complete)

242-243
00:32:54.413 --> 00:33:13.169
C: Yeah, categories A and B are pretty good. I think C and D just need a few adjustments. Maybe there's the man, the backpack, the car and the graduation cap.
B: he goes to school.
B: He goes to university.
C: Yeah, it could be a teacher going to school right?

244-245
00:33:13.169 --> 00:33:25.138
B: yeah,
C: may be that house with the green grass represent summer, and the house with winter. Snow represents winter and then the bird is the spring

246-247
00:33:26.699 --> 00:33:44.548
B: That might actually, right
D: yeah,
C: it's green grass versus snow so, maybe it's just trying to say summer versus winter.
A: And the car could be like, either commuting to work or like, oh you going to school so you can afford a car now, that a bit of a stretch but {laughing}

248
00:33:44.548 --> 00:33:58.523
B: he goes to School he graduates he gets a car it's career or education.
C: I yeah, I personally think this is as good as we're gonna get it unless anyone wants to make any changes.

In the above example, after completing the first two categories (A and B), the group found themselves with the two remaining categories (C, D), which were not yet finalized. Instead of doing further exchanges to improve the coherence of these categories, they reframed their interpretations of poorly fitting items and adjusted category labels to make things fit.

They called category C (bird, winter house, house, hay) *seasons* with the *bird* representing spring, the *winter house* representing winter, and the *hay* representing fall. Although they did not find something that corresponded well with summer, they made an assumption of an existing item in the category, *house*, to be a summer item, made no further exchanges and accepted this category as "*seasons*". They called category D (backpack, man, car, graduation hat) *education*, and although the *car* was incongruent with this category label, they did not make any further exchanges. Instead, they interpreted the *car* as a way of commuting to school and relabelled the category as *going to school*.

In this case participants used the affordance-based assumption to accommodate the conflict of weak categories they encountered. Something similar occurred towards the end of Expert Task 1 (line 373), when the group encountered a conflict where the *balloon* was considered an "awkward" item within the *element* category. Instead of restructuring other categories, they seemed to ignore this conflict and accepted the *balloon* as part of the *element* category. It seems that participants satisfice at the end of the task by trying to accommodate the poorly fitted items using resolution strategies that do not require further exchanges, such as making an affordance-based assumption, relabelling a category, or ignoring the conflict.

A prominent behavior observed in the Story tasks was that participants seemed to start by working independently, each trying to create a story in their own head using the items in their own quadrant, and then eventually offering other participants any items that they did not need or asking for items from other participants to complete their story. This behavior was observed at the beginning of the above Story 1 task (Section 6.3), when participant D offered the *pumpkin* to other participants proposing a *birthday* story based on their other 3 items. However, even though they are offering other members pictures they don't need, there is still some kind of tension in terms of accepting exchanges

requested by other members. That is, when a participant wanted an item from another participant's category, the latter participant seemed to be hesitant to accept the exchange, and asked questions such as "why do you need it?" or "what is your story?". The group member requesting the item would then share their story idea, and the other member might then accept the request or offer an alternative item for exchange. It seems that by asking such questions, participants wanted to help the other group members make their stories, but they also wanted to make sure the exchange would help them make their own stories.

6.5 Differences in the Process between Good and Expert Conditions

The problem-solving processes were highly similar under the Good and the Expert conditions; however, some differences were observed. The first difference we noticed is that groups exhibited fewer conflict recognition and resolution iterations in the Good condition than in the Expert condition. We consider an iteration to include the recognition of a conflict being raised in discussion, and a resolution strategy being used to try to solve this conflict. In the three Expert tasks we analysed, we observed from seven to nine conflicts; two of these conflicts were followed by two attempts to resolve them and one was flowed by three attempts. In the three Good condition tasks, we observed from 4 to 5 conflicts; out of these conflicts, one was followed by two resolution attempts and another one was followed by 3 resolution attempts (see Appendix J).

Another difference we noticed was more discussion and disagreement about the appropriateness of conflict resolution strategies in the Expert than the Good condition. On the other hand, when a resolution was proposed in the Good condition, it seemed to be quickly accepted with little resistance or questioning. For example, in the Good condition task (Good 1) described in Section 6.2, the assumptions that the bird was a "winter bird" (line 202-303), and that soccer was played in the "fall" (line 223-226) were accepted with no objection from other team members. It seems that the only objection in that task was when proposing the exchange of the graduation hat with the waves (line 204-207).

Another difference was that there was less discussion about the logic behind exchanging pictures in the Good condition compared to the Expert one. For example, in the Good 1 task we analyzed (Section 6.2), the group made several exchanges without explicitly saying the reasons or asking any questions about these exchanges. However, in the Expert task, the logic behind exchanges seemed to

be always stated. In the Good condition task, there seemed to be less justification or attempts to convince others of the correctness of a proposed solution category compared to the Expert task.

Lastly, groups differed between Expert and Good in their behavior at the last stages of their tasks, when they were finalizing their solutions. In both conditions, groups were likely to satisfice, and accept the last solution categories even if they seemed to not be entirely satisfied, and some items were not seen as obvious members of the chosen categories. However, in the Expert condition, there were more efforts by participants at the end to try to make the solution look better. For example, in the last conflict of the Expert condition analyzed in Section 6.1 (Expert 1, line 373-375), after participants identified a weak category due to an unrelated item (*balloon*), they first gave justification for keeping the *balloon* where it was without making an exchange. However, they then suggested an exchange that would improve the category and relabeled it as well. Thus, it seems that there was a kind of last-minute effort to improve the overall solution.

6.6 Differences in the Process between Similarity and Story Conditions

There were also some noticeable differences in the problem-solving process between the two similarity conditions (Expert and Good) and the Story condition. In terms of the number of conflicts observed, the Story task exhibited the fewest overall (four conflicts in each of the three-story tasks, versus four to five in the Good tasks, and seven to nine in the Expert tasks). Each of the conflicts in the three Story tasks was followed only by one resolution attempt, whereas the similarity tasks sometimes required several iterations of conflict recognition and resolution. In other words, there seemed to be less resistance against proposed resolution strategies in the Story condition than in the similarity conditions and participants seemed to accept exchanges suggested by others. Though in the Story condition participants asked questions regarding the reason for exchanging pictures to or from their own quadrant, they did not ask questions regarding the clarity of other members' categories.

In the Story task we observed that the four stories were mostly developed independently by the four individuals in a kind of simultaneous manner. That is, they created stories in their heads out of the items they already had and then made some exchanges, if necessary, to complete the story. Thus, stories were mostly developed at the same time. On the other hand, solutions of the similarity conditions were developed in a kind of sequential style, where the group members worked together starting with one solution category and then moving on to work on the others.

6.7 Alignment of Verbal Protocol Analysis and Statistical Results

The verbal protocol analysis provides complementary evidence to the statistical analysis and further insight into how problem open-endedness affected our dependent variables. Specifically, the observation from the verbal protocol analysis indicates a greater number of conflict recognition and resolution iterations in the Expert than in the Good than in the Story condition. Notably, participants tackling less open-ended problems were observed to strive for an optimal solution and explore various alternative solutions to enhance the overall group solution. On the other hand, participants approaching more open-ended problems knew that any solution could be satisfactory; thus, they put less effort to improve the overall solution. These observations are consistent with the statistical results showing greater perceived task difficulty, longer time spent solving problems, and more picture exchanges and path reversals in less open-ended problems than in more open-ended problems. Additionally, the higher number of conflict recognition and resolution cycles observed in less open-ended problems suggests a more exhaustive search within the problem space. This aligns with the findings on path dependency, indicating a more rigorous search in the problem space and moving further from the initial problem state to seek better solutions. The observation that participants in the Story condition start making stories using items in their own quadrants and then offering items that they don't need to others, is also consistent with the path dependency results and suggests that the solution in the Story condition is highly dependent on the initial state of the problem.

It has been observed from the verbal protocol analysis that participants in the similarity tasks started creating categories by initially paying attention to categories with a higher number of strongly related items and then moving to items that did not easily fit into obvious categories. This observation provides support to our statistical analysis of the structural moves, showing that progress toward the solution starts with categories with the strongest association scores followed by those with weaker associations. Also, the observation that participants keep trying to improve the categories until the end of the task in the Expert condition compared to the Good or Story conditions supports the statistical findings of structural moves in the sense of the gradual decrease in the strength of solution association across the order of the categories, which was more evident in the Expert than in the Good condition. On the other hand, the observation that solutions developed more sequentially in the similarity problems (i.e., starting with one category and then moving to the next one) but in a parallel manner in the Story condition, is consistent with the structural moves findings of no differences in the strength of solution association between categories in the Story condition.

6.8 Summary

This chapter summarized results of the verbal protocol analysis of nine experimental tasks and provided a detailed discussion of three specific tasks, including one from each experimental condition. The results shed light on some commonalities and differences that have been observed in the problem-solving process across the three conditions. Generally, participants spend some time in silence after being presented with the task, then start a discussion in which they identify pictures that go together, generate possible solution categories, inquire about these categories, list items that belong to each category, and exchange pictures based on their discussion. A prominent behavior that was observed in all task conditions is that the groups engage in iterations of conflict recognition and resolution until reaching a solution.

Some differences in the problem-solving behavior between the Expert and Good condition were observed. Compared to the Good condition, participants in the Expert condition exhibit more iterations of conflict recognition and resolution, and more discussion and disagreement before accepting resolutions. Furthermore, participants in the Expert condition seem to put more emphasis on the logic behind exchanges compared to the Good condition. In both conditions participants tend to accept the last solution categories without making exchanges, but in the Expert condition they seem to put in more last-minute efforts to improve the solution through category relabeling and other strategies.

The problem-solving process differed between the similarity conditions and the Story condition. Compared to similarity conditions, the Story tasks had fewer conflicts that were each followed by a single resolution attempt. Also, there was less resistance to proposed resolution strategies in the Story condition, as participants readily accepted suggested exchanges without extensive conflict recognition and resolution iterations. While participants in the Story condition tend to question about the reasons for requesting to exchange pictures from their own quadrant, they did not inquire about the clarity of other members' categories as was observed in the similarity conditions. In the Story tasks, participants developed solutions independently and simultaneously, creating stories based on items in their quadrants and making exchanges as needed to complete the narrative. Conversely, the similarity conditions followed a sequential approach, with the group starting with one solution category before moving on to the next.

Chapter 7
General Discussion and Future Research Directions

Many real life problems can be classified as ill-structured, open-ended problems, involving various possible solution outcomes rather than clear-cut alternatives. Psychological experiments on problem-solving have generally emphasized well-structured problems with predefined solution states, rather than open-ended, ill-structured problems. Previous research has explored ill-structured open-ended problems using qualitative observational methods, such as those employed by Goel (1995), Gunther et al. (1996), Fernandes and Simon (1999), Dorst and Cross (2001), and Ulrich (2011). However, this study goes beyond traditional observational methods by incorporating controlled experiments, allowing for more precise measurement and tracking of problem-solving behaviors. Furthermore, it aims to investigate effects specific to open-ended problems that have not been previously examined.

In this thesis, effects of problem open-endedness were examined by employing a flexible categorization task adapted from Adejumo et al. (2008). The experiment of our study involved generating ill-structured open-ended problems, with no predetermined solutions, by asking participants to create categories from a set of randomly selecting pictures. By adjusting the task instructions, we created three levels of problem open-endedness, ranging from least to most open-ended: (1) similarity/belief in a best solution (Expert); (2) similarity/belief in multiple solutions (Good); and (3) story/belief in multiple solutions (Story). Specifically, we investigated the effects of the degree of problem open-endedness on: 1) task difficulty; 2) the variability of solutions developed by different problem-solving groups; 3) the path-dependency of solutions (i.e., the degree to which solutions depend on initial conditions); 4) variability of solution association strength; 5) structural moves toward solutions; and 6) variability of problem-solving search behavior. We also conducted a verbal protocol analysis to further explore the cognitive processes involved in solving open-ended problems.

The study makes both methodological and substantive contributions to the open-ended ill-structured problem-solving literature. Methodologically, the study addresses the lack of controlled

experiments in prior literature, contributing a method that allows for precise measurement of various aspects of problem-solving behaviour and solution outcomes. The experimental method introduced in this study enables the controlled manipulation of open-endedness and other structural properties of complex problems, providing insights into cognitive and behavioral processes relevant to understanding different aspects of ill-structured open-ended problems. The use of group problem-solving tasks offers a valuable opportunity to explore the cognitive processes involved in moving from the initial problem state to a solution state. The group setting forces participants to articulate their thoughts and engage in explicit picture exchanges, which provide insights into internal cognitive processes like information processing and search that are challenging to observe using think-aloud or other observation methods.

In terms of substantive contributions, the results of our study provide quantitative evidence supporting several hypothesized effects of the degree of problem open-endedness on problem-solving behavior and solution outcomes. The qualitative protocol analysis results provide additional insights into these processes, highlighting the roles of conflict recognition and resolution in the solution of complex open-ended problems. Collectively, the results enhance our understanding of the effects of problem open-endedness on problem-solving, providing new insights into the cognitive and behavioral processes involved in solving ill-structured open-ended problems.

7.1 Summary of Findings

The findings from the ANOVA or non-parametric Friedman tests provide support for our hypotheses, indicating significant main effects of problem open-endedness on various behavioral and solution outcome variables. Specifically, the results support our hypothesized negative effects of problem open-endedness on task difficulty and variability in problem-solving search behavior, as well as hypothesized positive effects on solution variability, path dependency and variability in the strength of solution association. Also, the results provide evidence that solutions of open-ended problems are far from random. The pairwise comparisons between different open-endedness experimental conditions provided partial support for our hypothesis for most of the dependent variables that we tested. Specifically, differences between the similarity and Story conditions consistently supported our hypotheses. On the other hand, differences between the two similarity conditions (Expert vs. Good) were not statistically significant in several cases, although the distribution means for the two similarity conditions were in the predicted direction.

In terms of structural progress toward a solution, participants in the least open-ended Expert condition followed a progression from the strongest to the weakest solution association. This effect was less prominent in the more open-ended Good condition and absent in the most open-ended Story condition.

Separate stimulus-specific hypothesis testing provided further support for our main findings but also highlighted variation across the three stimulus types, with the Walmart results most consistent with our main results, followed by Icon and then Clipart.

The verbal protocol analysis conducted on nine experimental tasks provided further insights into the problem-solving process across the three experimental conditions. Participants initially spent time in silence and then engaged in discussions to identify related pictures, generate solution categories, inquire about them, list items, and exchange pictures. A prominent pattern of behavior for all conditions was the iteration of conflict recognition and resolution until groups reach a solution. Several different types of conflicts and corresponding resolution strategies were identified in the analysis.

Although the small sample size prevented statistical hypothesis testing, the protocol analysis indicated some potential differences between experimental conditions that could be investigated in further research. Regarding the differences between the Expert and Good conditions, participants in the Expert condition exhibited more conflict recognition and resolution iterations. Expert participants also placed greater emphasis on the logic behind exchanges and exhibited greater resistance to accepting resolutions proposed by other group members compared to participants in the Good condition. In both conditions, participants seemed to accept weaker solution categories towards the end of their task session without making further picture exchanges, but participants in the Expert condition seemed to exert more last-minute efforts to improve or justify the solution more than those in the Good condition.

In terms of differences between the similarity conditions and the Story condition, fewer conflict recognition and resolution iterations were observed in the Story than in the similarity conditions. Story participants typically applied only a single resolution attempt for each conflict and provided less resistance to proposed resolution strategies, whereas participants often tried two or more resolution strategies for conflicts and often resisted strategies proposed by others in the similarity tasks. In the Story condition, participants tended to create their own narratives independently and

simultaneously, developing partial solutions using their own initial picture distributions and then making a few additional exchanges as necessary. In contrast, the similarity conditions followed a more collective and sequential approach, with group members working together to complete one solution category before moving to the next.

7.2 Limitations and Potential Alternative Explanations

7.2.1 Influence of Good vs. Expert Manipulation

For most of the outcome variables, the results comparing Story vs. Similarity tasks strongly supported our hypotheses, but not the results comparing the Expert vs. Good similarity tasks. The significant differences for perceived difficulty as well as in the time spent solving problems imply that the Good vs. Expert manipulation did in fact alter participants' open-endedness beliefs, but perhaps not enough to significantly change their problem-solving behavior with respect to outcome variables such as the number of picture exchange, solution variability, path dependency, variability in the strength of solution association, and variability in problem-solving behavior.

There could be several potential explanations for the non-significant differences between Good and Expert in terms of these variables. Given that the distribution means were in the predicted direction, one possible explanation is that the study's sample size was too small to provide the statistical power needed to detect significant differences between the two groups. Problem space complexity might have also played a role, in that the 16-item categorization problem may not have been complex enough to induce detectable differences between satisficing and optimizing behavior under the Good vs. Expert tasks respectively.

Another potential explanation is that task order effects in the study's repeated-measures design could have influenced the observed behavior. Despite randomizing the order of tasks, it is possible that search behavior norms established by groups under prior tasks influenced their subsequent behavior, thereby reducing the magnitude of differences between Good and Expert conditions. The use of three different stimulus versions (Icons, Clipart, and Walmart images) may have also contributed to variability in the repeated-measures data. Even though the purpose of using different stimulus versions was to control for potential effects of specific types of images on problem-solving behavior, this may also have introduced additional variability, reducing statistical differences between task conditions. For example, the stimulus-specific analyses conducted in Section 5.7 indicated

varying problem open-endedness effects, with Walmart and Icon demonstrating greater consistency with our main results compared to Clipart. Nevertheless, this analysis might not be sufficient since it is based on smaller sample sizes than the repeated measure analyses used for our main results.

7.2.2 Similarity vs. Story Comparison

To introduce different levels of problem open-endedness, we created two categorization goals: Similarity and Story. Additionally, we manipulated problem solvers' beliefs about problem open-endedness in the similarity task. The story vs. similarity manipulation allowed us to create a spectrum of problem open-endedness, wherein the potential for multiple solutions could be larger in the story category due to the broader range of associations that problem solvers could draw from when creating a story category, compared to a similarity category. However, a potential limitation may arise from the comparison between Story and Similarity conditions, which introduces the possibility of unfair comparisons and potential ambiguity in result interpretation. It could be argued that making a story is a different activity from making similarity judgments, so the manipulation changes more than just the open-endedness of the problem. For example, the behavior of solving a story task can be different than the one for the similarity tasks; thus, comparing the Story versus similarity conditions might be biased relative to comparing the two similarity conditions.

7.2.3 Online Experiment

Due the COVID-19 global pandemic, it was necessary to change from our proposed in-person experiment to an online setting. Changing to the online version of the experiment reduced our ability to capture certain aspects of problem-solving behavior because participants did not rely on picture exchanges to process information during the problem-solving process. As discussed in Section 4.2.4.5.1, since participants could see all 16 pictures in the online experiment, so groups could search the problem space verbally by proposing and discussing possible picture exchanges without taking action and actually exchanging pictures. This has limited our ability to quantitively capture the thinking processes involved in solving this open-ended problem.

7.2.4 Ecological Validity

A potential limitation in this study pertains to ecological validity, particularly concerning the extent to which the student participants are representative of the real population, in terms of their knowledge or motivation to complete the experimental tasks. With respect to participant knowledge, it is important

to note that the experimental task is a general knowledge task, which implies that students' responses are likely to reflect those of non-student individuals.

With respect to potential influences of participant motivation, it is possible that students are motivated solely by bonus marks and did not take the task seriously. However, the presence of significant differences in the study's results across the different conditions challenges this idea (i.e., why would student motivation differ between different experimental conditions?) Moreover, the experimental task is not a complex or a tedious task where participants could easily lose motivation. Instead, it resembles a game or a puzzle-like challenge, where participants seemed to be engaged in solving it. As such, the likelihood that student participation or motivation greatly affected the results, considering ecological validity, seems low.

7.2.5 Subjectivity of the Verbal Protocol Analysis

The verbal protocol analysis was developed to gain a better understanding of the problem-solving process though observing the videos of the tasks and carefully detecting common behavior patterns. Despite the efforts made to improve consistency and accuracy of interpretations in our verbal protocol analysis through revisiting the initial observations and considering alternative explanations, it is essential to acknowledge the inherent limitations associated with the qualitative nature of this methodology. These limitations arise from potential researcher subjectivity when interpreting the problem-solving process. Different researchers might give alternative explanations of the behaviors we observed. For example, our identification and classification of conflicts and resolutions may differ from how other researchers might perceive and define them. This might lead to variations in the identifying and characterizing conflicts and their resolutions, potentially influencing subsequent analysis and conclusions drawn from the verbal protocol analysis.

7.3 Future Research

Future research could address the limitations of this study, answer question that arise from this work in relation to other literature and propose possible extensions to the current research.

To address some of the limitations of this research, future work could replicate our experiment using a larger sample size. That would increase the power of the study and might reveal significant differences between the two similarity conditions. Another possible research direction could further

investigate the potential influence of task order effects on the observed behavior. This could be done by replicating this study using a between-subject design to avoid any potential order effects on the behavior. Another potential direction is to investigate how the complexity of the problem space affects problem-solving behavior, considering factors such as the number of categories, number of items per category, and the interpretative flexibility of the stimulus items. It is possible that such method changes could strengthen the open-endedness belief manipulation.

With respect to the story vs. similarity manipulation it has been noticed that Story and similarity relations are not discrete and overlap in some cases. For example, in one similarity task, man, graduation hat, car and balloon were grouped together as a category named "Graduation"; in another story task, the same items were grouped together as a story of "A person got a car for his graduation gift". Another example includes grouping together tennis racket, house, soccer ball and backpack in a similarity task as a category of "Family Home" and in a Story task as a story that "After school, Dana goes home with her backpack and plays soccer and badminton to get at least an hour of physical activity". Further work is needed to investigate potential overlap between the cognitive processes involved in making similarity judgments and making stories. For example, with respect to the similarity tasks in our experiment it is possible that overtime, participants may switch from forming categories based on similarity perceptions to forming story-like categories based on conceptual relations between items. As participants progress toward a solution, they may have a harder time identifying similarity relations with the remaining items; thus, they might start relaxing the meaning of similarity and eventually, get to kind of very general categories that seem like stories.

Future work could investigate the psychological experience of the problem solver in terms of the build-up of tension while solving an open-ended problem, and the release of that tension after finding a satisfactory solution. A study by Derbentseva (2007) using closed-ended insight problems showed that the more tension accumulates, the more intense the experience of insight and the greater intensity of release after recognizing the solution. Such experiences reflect the structure of close-ended problems, where the direction of the psychological force is toward one specific solution, and any barrier that blocks progress toward this solution creates an opposing force on the individual to search for a detour to reach this solution. This creates a buildup of tension that is released when the person restructures the problem and/or identifies a detour path around the barrier that leads to the solution. However, in our study the problem is open ended, with different satisfactory solutions, so the experience of tension build-up and release should be different. Specifically, the problem solver may

be attracted to different alternative solutions, corresponding with psychological forces of different strengths in different directions. When barriers are encountered that prevent progress towards one solution, the availability of other satisfactory options suggests we should expect to observe less tension build-up and subsequently less tension release when settling on a solution. Our experimental evidence of perceived difficulty, solution time, and the number of moves suggest that the level of tension experienced should vary inversely with the level of open-endedness, highlighting the need for future research in this area.

A related aspect concerns the psychological experience associated with satisficing when solving open-ended problems. From the perspective of information processing theory (Miller, 1956; Simon,1997), satisficing is viewed as a strategy of selecting an acceptable alternative rather than an optimal one. However, in open-ended problems, the presence of psychological forces attracting the person toward different solutions in different directions, implies that opting for one solution alternative rather than another one should result in some remaining tension or dissonance (Festinger, 1957) after solving the problem. That is, problem solvers may not be completely satisfied with the selected solution as they are still attracted to some of the alternatives they did not choose. Future research could investigate factors that influence the level of dissonance associated with satisficing solutions to open-ended problems, and how problem-solvers reduce that dissonance. For example, given equally satisfactory solution alternatives, problem-solvers forced to choose just one of them might engage in post-solution dissonance reduction strategies that justify their chosen solution while discrediting other rejected alternatives.

Our analysis of structural moves indicated that in the least open-ended (Expert) problem, groups moved toward a solution in a kind of sequential approach, starting with solution categories of higher similarity scores and ending with categories of lower similarity scores. In more open-ended (Good) task, this sequential manner was less obvious, and even less so in the most open-ended Story problem where group members tended to work in parallel to form their stories simultaneously using the initial distribution of pictures in their quadrants. These findings raise questions about the factors that influence the extent to which complex problems can be decomposed into separate sub-problems that can be completed in parallel, versus completed collectively by the group in a sequential manner. In the language of the architecture of complexity (Simon, 1962), the Story task can be viewed as a more modular task in the sense that individual group members could work on each part of the problem (each story) independently and there is relatively little interdependence between the sub-tasks. The

similarity problems were more integral, with more interdependence between the sub-tasks (the four categories), necessitating close collaboration between the group members. Future research could experimentally manipulate problem architectures to investigate how it interacts with problem open-endedness to influence group problem solving behavior. For example, the Story task could be made more integral by requiring participants to come up with four inter-related stories that tie together into some overall story (e.g., like the chapters of a book). Under such conditions, we would expect to observe more collaborative problem-solving behavior with closer coordination between members.

Future research can also investigate further the effects of path-dependency on solution outcomes. The path dependency results indicated a significant effect of the initial picture distribution on the solution state. Effects of the initial solution categories proposed by group members on the path followed and on the final solution are also of potential interest. For example, we noticed that some of the categories identified by group members at early stages of the problem-solving process seemed to persist throughout the duration of the task and eventually became one of their final solution categories. Given that some pictures could be labelled and categorized in quite different ways (e.g., *pumpkin/haybale* as *Halloween, food, harvest*, etc.), it is possible that such initial labels/categories introduce constraints that limit later search processes. Such effects could be investigated by manipulating the interpretive flexibility of items, for example by including labels with the pictures that encourage participants to perceive an item as a member of one category rather than another. Different labels might lead different groups to different solution categories.

Finally, this study has primarily used the group context as a methodological device to provide insight into problem solving behavior and cognitive processes that are difficult to access using think-aloud and related methods in individual-based experiments. Nonetheless, the group context provides many opportunities for further research on the effects of various group properties or social psychological variables on group behavior and performance in open-ended problem-solving. For example, future studies could incorporate varying incentives, communication patterns, conflicting individual goals, or group member characteristics such as expertise, creativity, or other individuals' differences, to investigate how they influence problem-solving behavior and performance. In addition, it is possible to implement our task in an individual-based experiment to investigate differences in performance or solutions outcomes when the open-ended problems are solved by individuals versus groups.

References

Abimbola, G. (2006). EFFECTS OF TASK STRUCTURE ON GROUP PROBLEM SOLVING [UWSpace]. http://hdl.handle.net/10012/824

Adejumo, G., Duimering, P. R., & Zhong, Z. (2008). A balance theory approach to group problem solving. Social Networks, 30(1), 83–99.

Akin, O. (1978). How do architects design? In J. C. Latombe (Ed.), Artificial intelligence andpattern recognition in computer aided design (pp.65-104). New York, NY: IFIP/North-Holland Publishing.

Arrow, K. J. (1951). Social Choice and Individual Values. Wiley: New York.

Bandura, A. (1997). Self-efficacy: The exercise of control. In Self-efficacy: The exercise of control. W H Freeman/Times Books/ Henry Holt & Co.

Baron, R. S., & Kerr, N. L. (2003). Group process, group decision, group action (2nd ed.). Maidenhead, UK: Open University Press.

Becker, J. P., & Shimada, S. (1997). The open-ended approach: A new proposal for teaching mathematics. Reston, VA: National Council of Teachers of Mathematics.

Black, D. (1948). On the Rationale of Group Decision-making. *Journal of Political Economy, 56*, 23 - 34.

Black, D. (1958). The Theory of Committees and Elections. Cambridge, At the University Press.

Bouchard Jr., T. J. (1969). Personality, problem-solving procedure, and performance in small groups. Journal of Applied Psychology, 53(1, Pt.2), 1–29. https://doi.org/10.1037/h0026747

Chen, L. (2010). Effects of Individual versus Group Incentives on Group Problem Solving [UWSpace]. http://hdl.handle.net/10012/5128

Christensen, B. T., & Ball, L. J. (2014). Studying design cognition in the real world using the 'in vivo' methodology. In P. Rodgers & J. Yee (Eds.), The Routledge companion to design research (pp.317-328). New York, NY: Taylor & Francis Ltd. Routledge.

Cisco Webex Meetings. Cisco (2020), Cisco Systems, Inc.

Cross, N. (2001). Design cognition: Results from protocol and other empirical studies of design activity. In C. Eastman, W. Newstatter, & M. McCracken (Eds.), Design knowing and learning: Cognition in design education (pp. 79-103). Oxford, UK: Elsevier Science.

Cruz, M. G., Boster, F. J., & Rodríguez, J. I. (1997). The impact of group size and proportion of shared information on the exchange and integration of information in groups. Communication Research, 24(3), 291-313.

Davis, J. H. (1969). Group performance. Reading, MA: Addison-Wesley.

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. Psychological Review, 80(2), 97-125.

Derbentseva, N. (2007). The Intensity of the Insight Experience in Problem Solving: Structural and Dynamic Properties [UWSpace]. http://hdl.handle.net/10012/2666

Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem–solution. Design Studies, 22(5), 425-437. doi: 10.1016/s0142-694x(01)00009-6

Duerk, D. P. (1993). Architectural programming: Information management for design. New York, NY: Wiley.

Dunbar, K. (1998). Problem solving. In W. Bechtel and G. Graham (Eds.), A Companion to Cognitive Science (pp. 289-298). London, UK: Blackwell.

Duncker, K. (1945). On problem solving. Psychological Monographs, 58, 5, i-113.Retrieved from https://doi.org/10.1037/h0093599

Eastman C. M., (1970). On the analysis of intuitive design processes. In G. T. Moore (Ed.), Emerging methods in environmental design and planning (pp. 21-37). Cambridge, MA: MIT Press.

Ericsson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. In Protocol analysis: Verbal reports as data. The MIT Press.

Fernandes, R., & Simon, H. A. (1999). A study of how individuals solve complex and ill-structured problems. Policy Sciences, 32(3), 225-245.

Festinger, L. (1957). A theory of cognitive dissonance. In A theory of cognitive dissonance. Stanford University Press.

Fischer, A., Greiff, S., & Funke, J. (2012). The Process of Solving Complex Problems. The Journal Of Problem Solving, 4(1). doi: 10.7771/1932-6246.1118

Goel, V. (1995). Sketches of thought. Cambridge, MA: MIT Press.

Gunther, J., Frankenberger, E., & Auer, P. (1996). Investigation of individual and team design processes. In N. Cross, H. Christiaans & K. Dorst (Eds.), Analysing design activity (pp.117-132). New York, NY: Wiley.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. In Annual Review of Psychology (Vol. 47, pp. 307–338). Annual Reviews. https://doi.org/10.1146/annurev.psych.47.1.307

Henningsen, D. D., Henningsen, M. L. M., Jakobsen, L., & Borton, I. (2004). It's good to be leader: The influence of randomly and systematically selected leaders on decision-making groups. Group Dynamics: Theory, Research, and Practice, 8(1), 62-76.

Hernández, R. J., Cooper, R., Tether, B., & Murphy, E. (2018). Design, the language of innovation: A review of the design studies literature. She Ji: The Journal of Design, Economics, and Innovation, 4(3), 249-274.

Hollingshead, A. B. (1996). Information suppression and status persistence in group decision making: The effects of communication media. Human Communication Research, 23(2), 193– 219.

Jonassen, D. H. (2011). Learning to solve problems: A handbook for designing problem-solving learning environments. Routledge.

Kelley, H. H., & Thibaut, J. W. (1954). Experimental studies of group problem solving and process. In G. Lindzey (Ed.), Handbook of Social Psychology, (pp, 735-785). Cambridge, MA: Addison-Wesley

Kohler, W. (2015). The task of Gestalt psychology. Princeton, NJ: Princeton University Press.

Kolodner, J. L., & Wills L. M. (1996). Powers of observation in creative design. DesignStudies, 17(4), 385-416.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individual's fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. Personnel Psychology, 58(2), 281–342. https://doi.org/10.1111/j.1744-6570.2005.00672.x

Laughlin, P. R. (2011). Group problem solving. Princeton, NJ: Princeton University Press.

Laughlin, P. R., & Adamopoulos, J. (1980). Social combination processes and individual learning for six-person cooperative groups on an intellective task. Journal of Personality and Social Psychology, 38(6), 941-947.

Lorge, I. & Solomon, H. (1955). Two models of group behavior in the solution of Eureka-type problems. Psychometrika, 20, 139-148. Retrieved from https://doi.org/10.1007/BF02288986

Luchins, A. S. & Luchins, E. H. (1950). New experimental attempts at preventing mechanization in problem solving. The Journal of General Psychology, 42, pp. 279 – 297.

Maher, M. L. (1994). Creative design using a genetic algorithm. Computing in Civil Engineering, 2, 2014-2021.

Maher, M. L., & Poon, J. (1996). Modeling design exploration as co-evolution. Computer-Aided Civil and Infrastructure Engineering, 11, 195-209.

Maier, N.R.F. (1930). Reasoning in humans. I. On direction. Journal of Comparative Psychology 10(2):115–143. Retrieved from https://doi.org/10.1037/h0073232

Mennecke, B. E. (1997). Using group support systems to discover hidden profiles: An examination of the influence of group size and meeting structures on information sharing and decision quality. International Journal of Human Computer Studies, 47(3), 387–405.

Mervis, C. and Rosch, E. (1981). Categorization of Natural Objects. Annual Review of Psychology, 32(1), 89-115. doi: 10.1146/annurev.ps.32.020181.000513

Microsoft Word Clipart, Microsoft Office 365, Microsoft 2020. Accessed 20 May 2020

Microsoft Word Icons, Microsoft Office 365, Microsoft 2020. Accessed 20 May 2020

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review, 63(2), 81-97. doi: 10.1037/h0043158

Newell, A. & Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Nickel, J., Duimering, P. R., & Hurst, A. (2022). Manipulating the design space to resolve trade-offs: Theory and evidence. Design Studies, 79, 101095. https://doi.org/https://doi.org/10.1016/j.destud.2022.101095

Öllinger M., & Goel V. (2010). Problem Solving. In B. Glatzeder, V. Goel, & A. Müller (Eds.), Towards a Theory of Thinking. Building Blocks for a Conceptual Framework (pp.3-21). Springer, Berlin- Heidelberg, Germany: Springer.

Qualtrics. (n.d.). Qualtrics. Retrieved from https://uwaterloo.qualtrics.com

Ran, B. (2007). Meaning construction: Cognitive processes of conceptual interaction. (Unpublished doctoral thesis). University of Waterloo, Ontario, Canada. Retrieved from http://hdl.handle.net/10012/3125

Rand, W. M. (1971). Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association, 66(336), 846–850. https://doi.org/10.2307/2284239

Reitman, W.R. (1964). Heuristic decision procedures, open constraints, and the structure of ill-defined problems. In M.W. Shelley & G.L. Bryan (Eds.), Human Judgments and Optimality (pp 282–315). New York, NY: Wiley.

Robert, C., & Cheung, Y. H. (2010). An examination of the relationship between conscientiousness and group performance on a creative task. Journal of Research in Personality, 44(2), 222–231. https://doi.org/https://doi.org/10.1016/j.jrp.2010.01.005

Rosch, E., & Lloyd, B. (1978). Cognition and categorization. Hillsdale, N.J: Lawrence Erlbaum.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7(4), 573-605. doi: 10.1016/0010-0285(75)90024-9

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive psychology, 8(3), 382-439.

Rosch, E., Simpson, C. and Miller, R.S. (1976). Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 2(4), pp. 491-502. doi: 10.1037//0096-1523.2.4.491.(b)

Sarsfield, E. (2014). Differences between novices' and experts' solving ill-structured problems. Public Health Nursing, 31(5), 444-453.

Schunn, C. D., McGregor, M. U., & Saner, L. D. (2005). Expertise in ill-defined problem-solving domains as effective strategy use. Memory & Cognition, 33(8), 1377-1387.

Seidel, V., & Fixson, S. (2013). Adopting design thinking in novice multidisciplinary teams: The application and limits of design methods and reflexive practices. Journal of Product Innovation Management, 30(S1): 19–33.

Shaw, M. E. (1932). A comparison of individuals and small groups in the rational solution of complex problems. The American Journal of Psychology, 44, 491-504.

Simon, H. (1956). Rational choice and the structure of the environment. Psychological Review, 63(2), 129-138. doi: 10.1037/h0042769

Simon, H. A. (1962). The Architecture of Complexity. Proceedings of the American Philosophical Society, 106(6), 467–482. http://www.jstor.org/stable/985254

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. American Psychologist, 26(2), 145.

Simon, H. (1997). Administrative behavior (4th ed.). New York, NY: Free Press.

Smoke, W. H., & Zajonc, R. B. (1962). On the reliability of group judgments and decisions. In J.H.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. Journal of Personality and Social Psychology, 48(6), 1467.

Taggar, S. (2001). Group composition, creative synergy, and group performance. The Journal of Creative Behavior, 35(4), 261–286. https://doi.org/10.1002/j.2162-6057.2001.tb01050.x

Taggar, S. (2021). Conscientiousness in Teams Completing Creative Tasks: Does it Predict? The Journal of Creative Behavior, 55(2), 319–336. https://doi.org/https://doi.org/10.1002/jocb.453

Tasa, K., Taggar, S., & Seijts, G. H. (2007). The development of collective efficacy in teams: a multilevel and longitudinal perspective. Journal of applied psychology, 92(1), 17.

Ullman, D.G., Dietterich, T.G., & Stauffer, L.A.(1988), A model of the mechanical design process based on empirical data. AI in Engineering Design and Manufacturing, 2(1), 33-52.

Ulrich, K. T. (2011). Design: Creation of artifacts in society. SSRN Electronic Journal. doi: 10.2139/ssrn.1951106

Walmart. https://www.walmart.ca/en. Accessed 12 May 2020.

Wertheimer, M., (1982). Productive thinking. Chicago, IL: The University of Chicago Press.

Wieth, M. B., & Francis, A. P. (2018). Conflicts and consistencies in creativity research and teaching. Teaching of Psychology, 45(4), 363-370.

Wittgenstein, Ludwig. 1953. Philosophical Investigations. New York: Macmillan.

Appendix A

Concept Relation Survey[5]

This survey is designed to help us better understand how people group things together into categories. In each task, you will be presented with one item and you will be asked to pick 5 items, out of a list of 15 other items, that you think are most similar to the target item. Please take your time to think about your answers and be as precise as possible when completing the survey. Your response will be confidential.

Participant's name:

Student ID number:

Department:

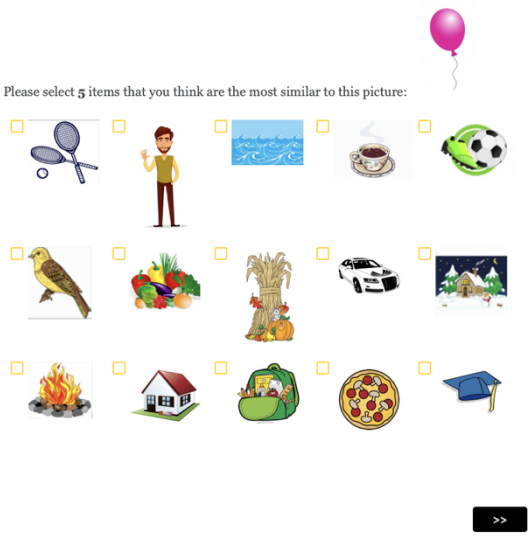
Please select your age group

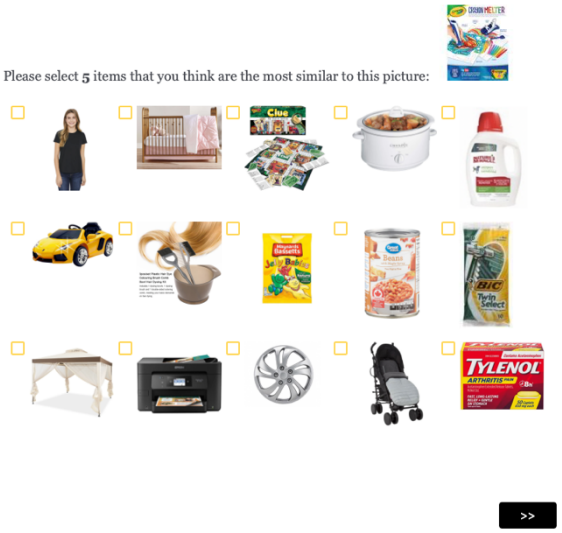
- ○ Below 17
- ○ 18-20
- ○ 21-25
- ○ 26-30
- ○ 31-35
- ○ Above 35

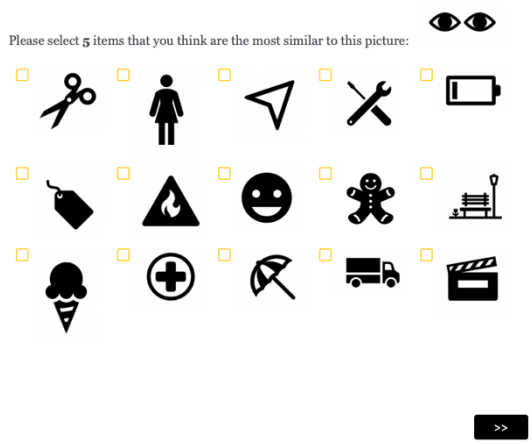
Please select your Gender.

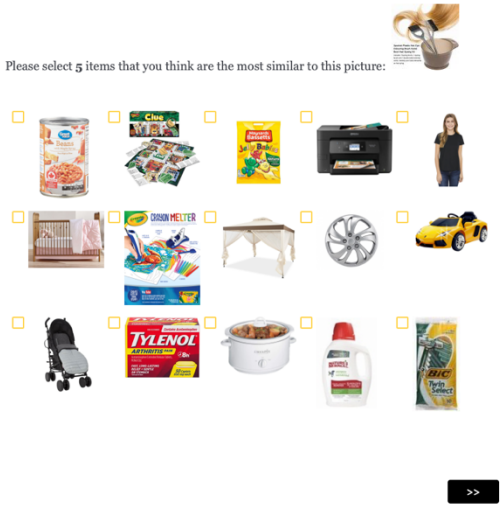
- ☐ Man/Male
- ☐ Woman/Female
- ☐ Trans* / Non-binary;
- ☐ Prefer to self-identify (specify)
- ☐ Prefer to not answer

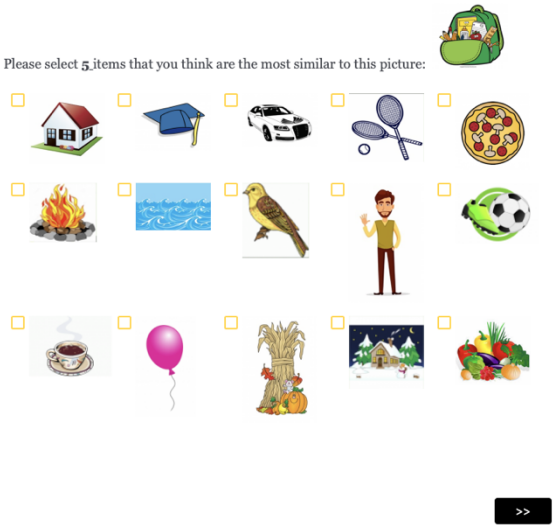
[5] Each question was presented in a separate page

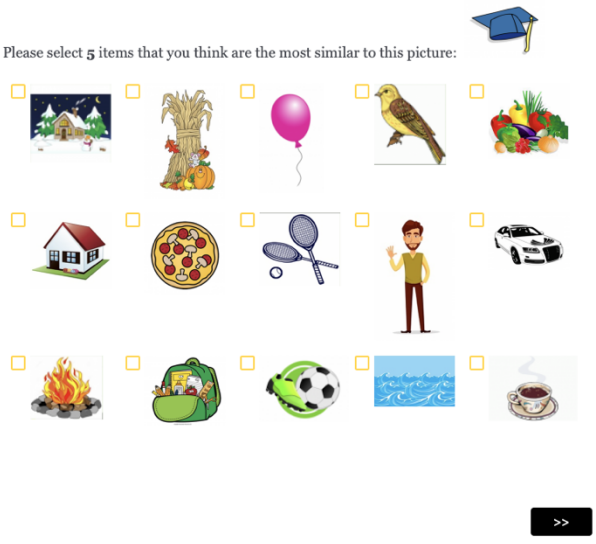

Please select 5 items that you think are the most similar to this picture:
>>


Please select 5 items that you think are the most similar to this picture:
Clue
TYLENOL
ARTHRITIS
>>

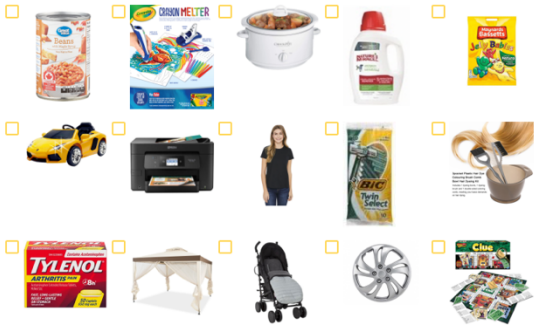

Please select 5 items that you think are the most similar to this picture:
>>


Please select 5 items that you think are the most similar to this picture:
Clue
CRAYON MELTER
TYLENOL
ARTHRITIS
>>


Please select 5 items that you think are the most similar to this picture:
>>

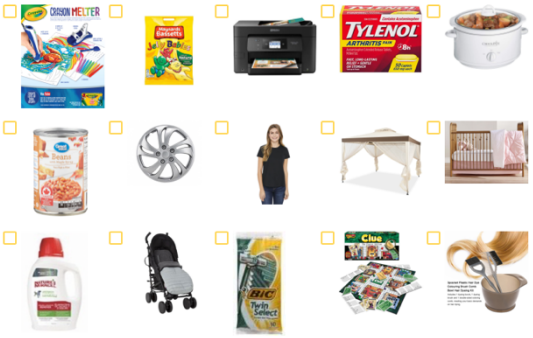

Please select 5 items that you think are the most similar to this picture:
>>

Please select 5 items that you think are the most similar to this picture:



>>

Please select 5 items that you think are the most similar to this picture:

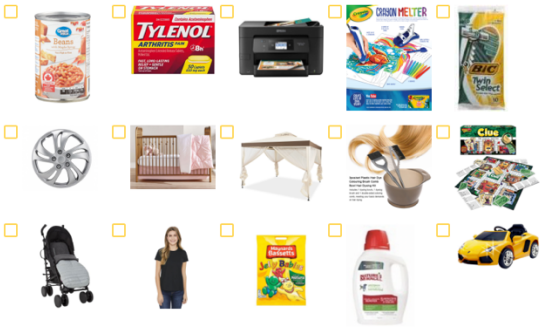


>>

Please select 5 items that you think are the most similar to this picture:

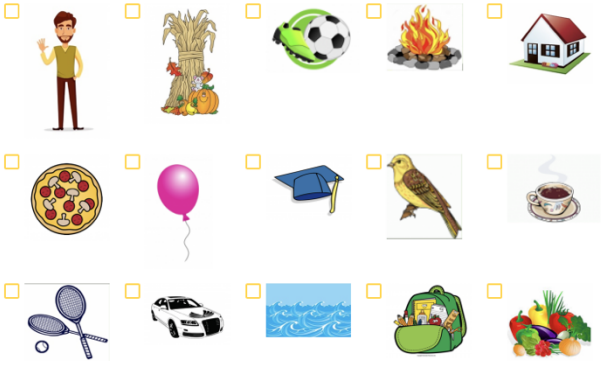
>>

Please select 5 items that you think are the most similar to this picture:



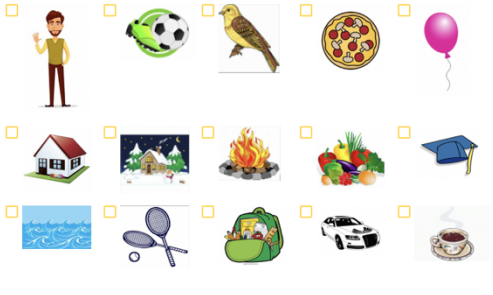
>>

Please select 5 items that you think are the most similar to this picture:



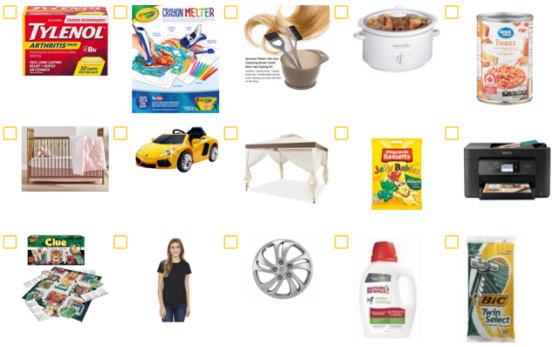
>>

Please select 5 items that you think are the most similar to this picture:



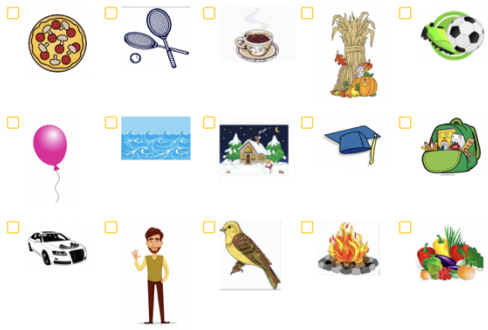
>>

Please select **5** items that you think are the most similar to this picture:



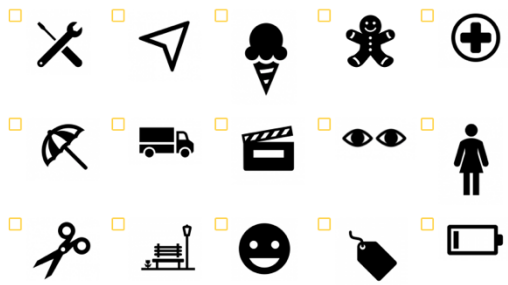
>>

Please select **5** items that you think are the most similar to this picture:



>>

Please select **5** items that you think are the most similar to this picture:



>>

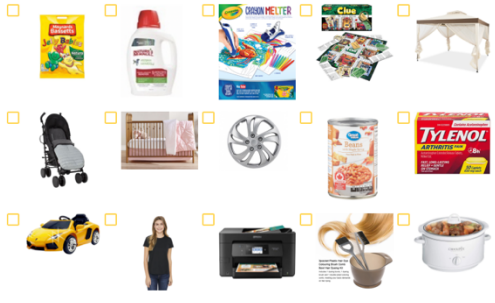
Please select **5** items that you think are the most similar to this picture:

>>

Please select **5** items that you think are the most similar to this picture:

>>

Please select **5** items that you think are the most similar to this picture:

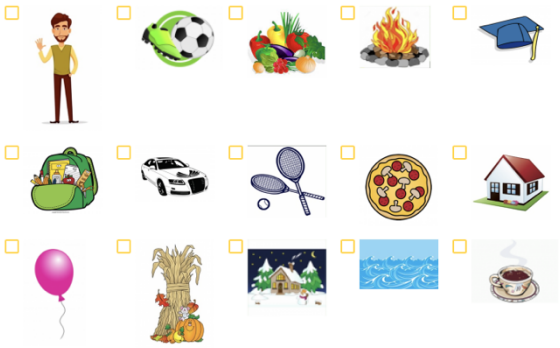


>>

Please select **5** items that you think are the most similar to this picture:

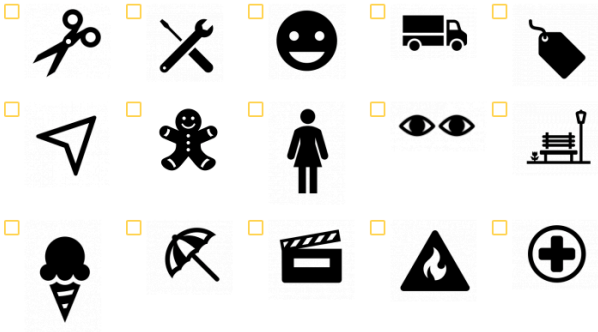
>>

Please select **5** items that you think are the most similar to this picture :



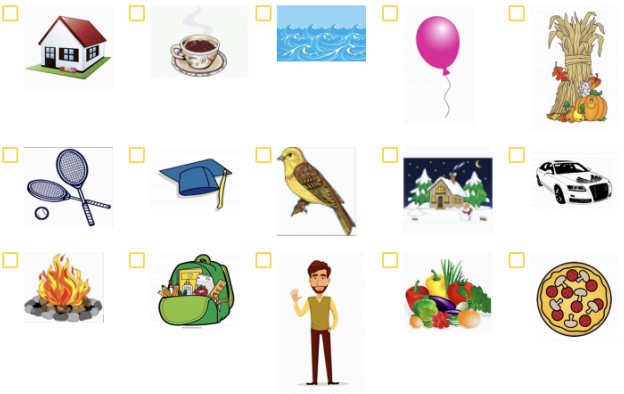
>>

Please select **5** items that you think are the most similar to this picture:



>>

Please select **5** items that you think are the most similar to this picture:



>>

Please select **5** items that you think are the most similar to this picture:

>>

Please select **5** items that you think are the most similar to this picture:

>>

Appendix B

Frequency Counts and Pair-wise Similarity Scores for the Icon and Walmart Stimulus

Concept association scores from the survey for the Icon stimuli type

From\to	Woman (A)	Arrow (B)	Face (C)	Eyes (D)	Truck (E)	Ice-cream (F)	Umbrella (G)	Fire-sign (H)	Gingerbread (I)	Bench (J)	Tools (K)	Battery (L)	Scissors (M)	First-aid (N)	Tag (O)	Movie (P)
Woman (A)	0	3	21	17	1	18	8	1	18	10	2	2	4	5	7	3
Arrow (B)	8	0	6	7	13	4	9	8	0	12	9	13	8	12	6	5
Face (C)	21	2	0	15	0	17	7	1	21	11	2	1	3	8	3	8
Eyes (D)	17	6	21	0	5	1	2	3	15	6	4	2	8	10	4	16
Truck (E)	8	13	3	6	0	5	2	14	0	7	18	8	6	14	8	8
Ice-cream (F)	16	4	20	1	7	0	18	2	21	18	0	0	0	3	6	4
Umbrella (G)	17	9	14	13	6	11	0	5	4	15	5	3	1	7	5	5
Fire-sign (H)	6	7	2	8	11	4	5	0	4	1	16	12	12	20	5	7
Gingerbread (I)	21	0	24	19	3	22	4	2	0	5	0	0	5	3	6	6
Bench (J)	16	10	13	10	15	12	14	2	4	0	4	5	2	2	3	8
Tools (K)	4	6	1	4	18	1	6	15	1	4	0	13	22	12	7	6
Battery (L)	6	14	6	8	7	1	3	10	1	9	15	0	6	16	6	12
Scissors (M)	10	8	2	4	8	2	7	5	6	2	21	5	0	14	18	8
First-aid (N)	11	9	14	9	4	4	2	13	2	7	13	15	14	0	3	0
Tag (O)	11	8	9	9	9	10	8	3	9	2	6	7	13	4	0	12
Movie (P)	17	8	12	16	9	9	5	3	1	8	8	10	4	0	10	0

Combinations of concept relation scores for Icon

Picture	Woman (A)	Arrow (B)	Face (C)	Eyes (D)	Truck (E)	Ice-cream (F)	Umbrella (G)	Fire-sign (H)	Gingerbread (I)	Bench (J)	Tools (K)	Battery (L)	Scissors (M)	First-aid (N)	Tag (O)	Movie (P)
Woman (A)	0															
Arrow (B)	11	0														
Face (C)	42	8	0													
Eyes (D)	34	13	36	0												
Truck (E)	9	26	3	11	0											
Ice-cream (F)	34	8	37	2	12	0										
Umbrella (G)	25	18	21	15	8	29	0									
Fire-sign (H)	7	15	3	11	25	6	10	0								
Gingerbread (I)	39	0	45	34	3	43	8	6	0							
Bench (J)	26	22	24	16	22	30	29	3	9	0						
Tools (K)	6	15	3	8	36	1	11	31	1	8	0					
Battery (L)	8	27	7	10	15	1	6	22	1	14	28	0				
Scissors (M)	14	16	5	12	14	2	8	17	11	4	43	11	0			
First-aid (N)	16	21	22	19	18	7	9	33	5	9	25	31	28	0		
Tag (O)	18	14	12	13	17	16	13	8	15	5	13	13	31	7	0	
Movie (P)	20	13	20	32	17	13	10	10	7	16	14	22	12	0	22	0

Concept association scores from the survey for the Walmart stimuli type

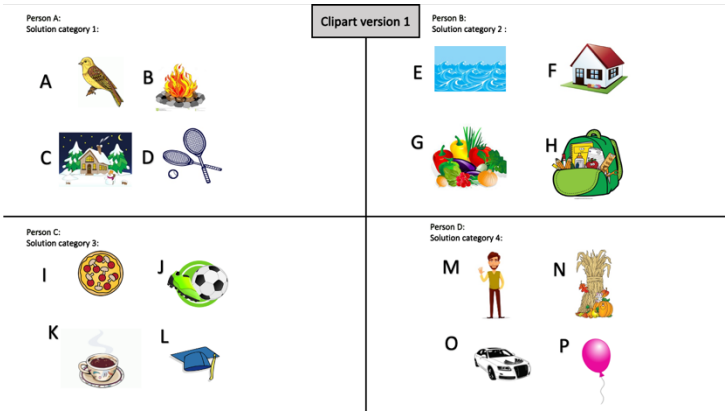
From\to	Tylenol (A)	Stroller (B)	Lady (C)	Crock-pot (D)	Hair-dye (E)	Crayon_melter (F)	Jelly-babies (G)	Canopy (H)	Canned-beans (I)	Clue (J)	Printer (K)	Stain-remover (L)	Crib (M)	Toy-car (N)	Rim (O)	Razors (P)
Tylenol (A)	0	8	11	10	7	8	12	4	15	7	2	13	5	3	3	12
Stroller (B)	1	0	17	5	1	15	15	6	1	7	1	2	23	16	9	1
Lady (C)	7	17	0	10	20	2	5	10	5	3	7	5	18	5	2	4
Crock-pot (D)	4	6	18	0	8	3	12	11	23	1	8	7	8	5	4	2
Hair-dye (E)	7	5	19	7	0	8	5	11	6	4	8	13	9	6	2	10
Crayon_melter (F)	8	12	7	3	8	0	20	1	3	19	5	6	10	6	0	12
Jelly-babies (G)	11	9	7	7	2	17	0	2	15	16	1	4	9	7	0	13
Canopy (H)	4	18	16	12	8	2	4	0	6	4	5	6	16	10	6	3
Canned-beans (I)	14	5	12	23	6	3	17	5	0	7	1	12	7	0	3	5
Clue (J)	9	8	7	3	3	21	17	2	5	0	11	5	8	9	1	11
Printer (K)	10	5	7	6	7	18	7	4	3	14	0	7	5	6	6	15
Stain-remover (L)	11	8	10	10	14	5	5	5	14	5	4	0	7	5	9	8
Crib (M)	7	23	16	7	6	7	9	15	6	5	4	3	0	12	0	0
Toy-car (N)	2	21	10	2	2	10	11	6	3	10	8	1	12	0	17	5
Rim (O)	4	20	10	8	5	1	1	10	5	2	10	7	6	22	0	9
Razors (P)	15	4	14	1	17	11	8	5	6	8	5	11	5	6	4	0

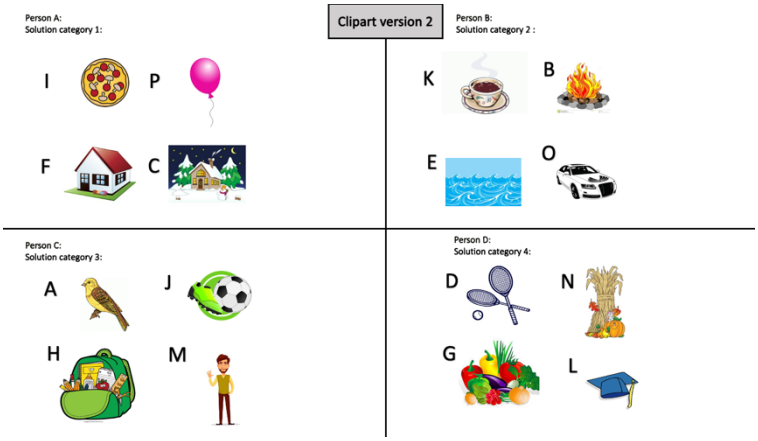
Combinations of concept relation scores for Walmart

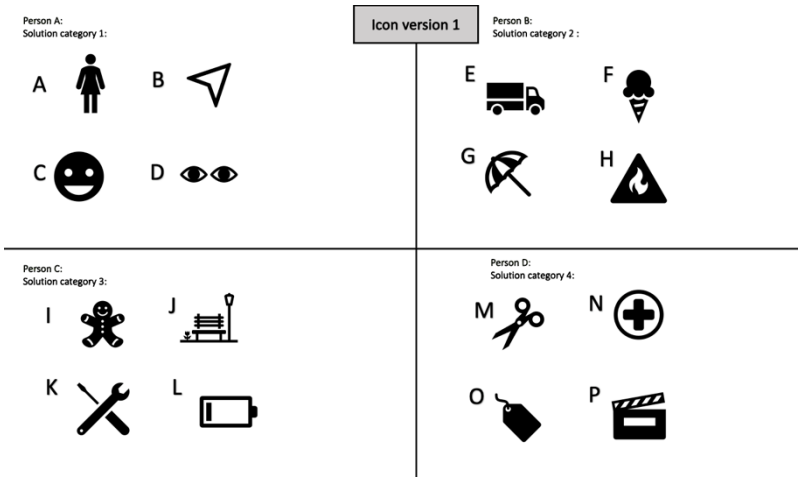
Picture	Tylenol (A)	Stroller (B)	Lady (C)	Crock-pot (D)	Hair-dye (E)	Crayon_melter (F)	Jelly-babies (G)	Canopy (H)	Canned-beans (I)	Clue (J)	Printer (K)	Stain-remover (L)	Crib (M)	Toy-car (N)	Rim (O)	Razors (P)
Tylenol (A)	0															
Stroller (B)	9	0														
Lady (C)	18	34	0													
Crock-pot (D)	14	11	28	0												
Hair-dye (E)	14	6	39	15	0											
Crayon_melter (F)	16	27	9	6	16	0										
Jelly-babies (G)	23	24	12	19	7	37	0									
Canopy (H)	8	24	26	23	19	3	6	0								
Canned-beans (I)	29	6	17	46	12	6	32	11	0							
Clue (J)	16	15	10	4	7	40	33	6	12	0						
Printer (K)	12	6	14	14	15	23	8	9	4	25	0					
Stain-remover (L)	24	10	15	17	27	11	9	11	26	10	11	0				
Crib (M)	12	46	34	15	15	17	18	31	13	13	9	10	0			
Toy-car (N)	5	37	15	7	8	16	18	16	3	19	14	6	24	0		
Rim (O)	7	29	12	12	7	1	1	16	8	3	16	16	6	39	0	
Razors (P)	27	5	18	3	27	23	21	8	11	19	20	19	5	11	13	0

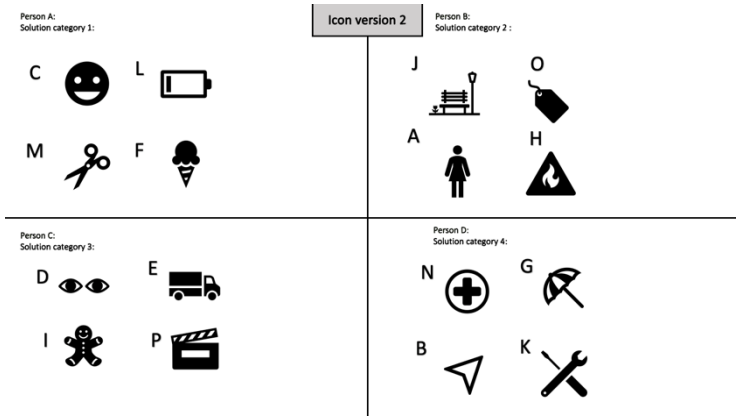
Appendix C

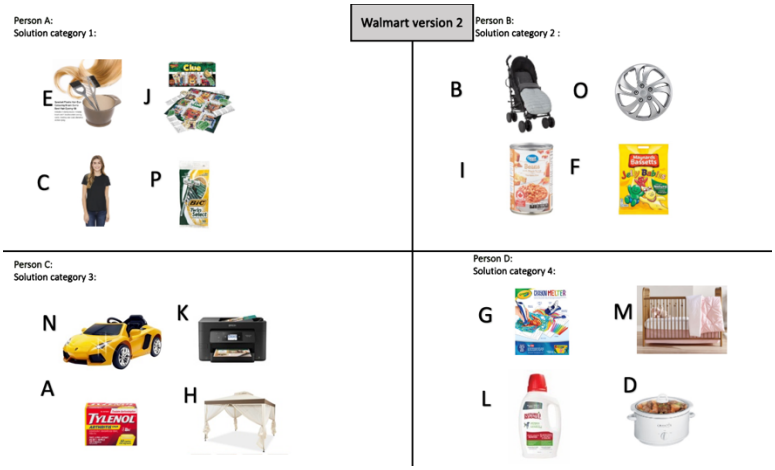
The Two Different Initial Distributions of Pictures to Group Members









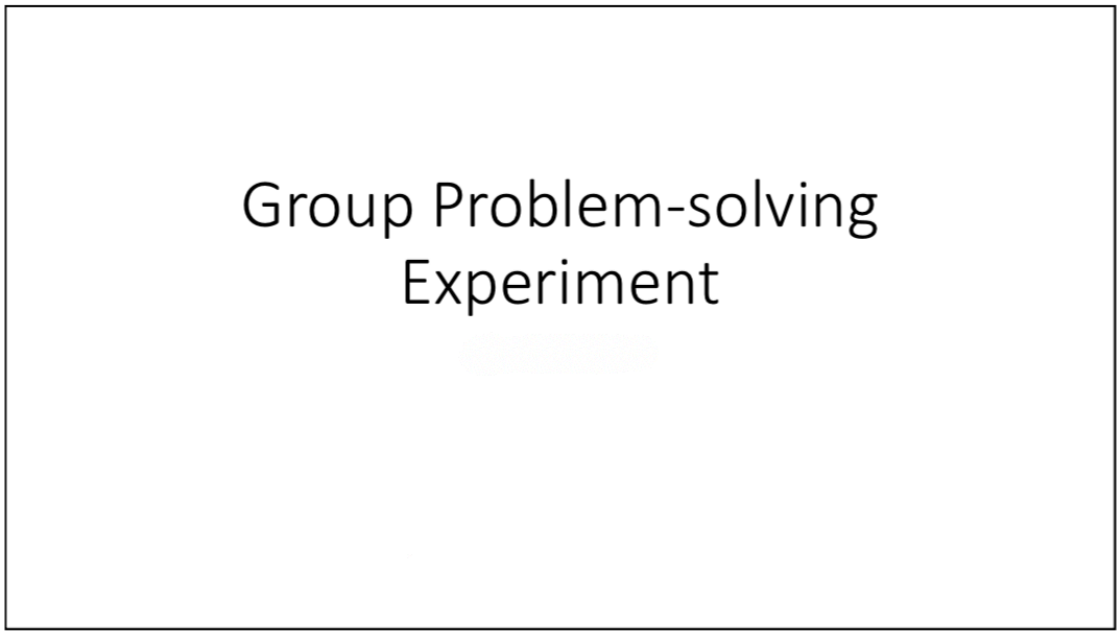


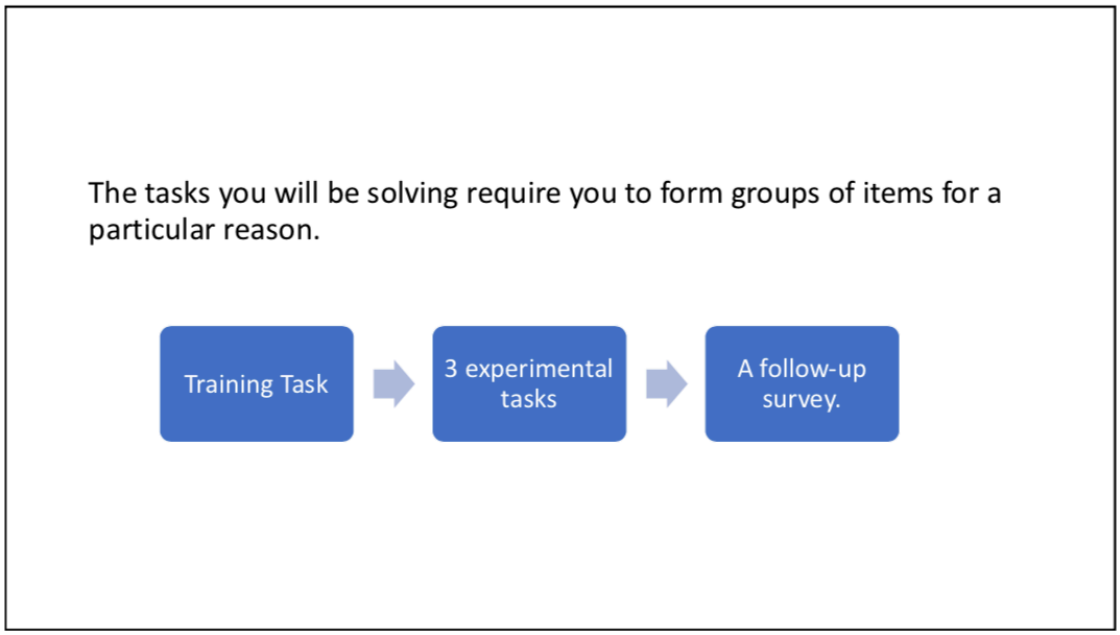
Appendix D

An Example of the Procedure for Selecting Walmart Pictures

Appendix E

A complete experiment with the instructions (condition order B)[6]





[6] Participants are presented with one slide at a time

General instructions:

- A respectful dialogue to discuss and justify the categories you are making.
- I will be moving the pictures for you as you tell me.
- Always keep 4 pictures in your slot.
- In case you are disconnected please enter the link of meeting again, in case I am disconnected stay in the room until I connect back again.
- Each problem should take about 15 minutes to solve.

Instructions to the observer:

Your role is to observe and take notes as the other group members solve the problems.

- Please do not interfere or influence the solution the other 4 participants come up with.
- Please keep your microphone on mute until the group is finished solving the three problems.
- Please make notes about:
  - The kind of difficulties you notice the group experience while solving the problem,
  - How they resolve these difficulties.

  (We will ask you to submit your notes afterwards)

Training task 1

- Your group will be given 16 pictures (i.e. each of you will have 4 pictures).
- You need to find the solution by exchanging pictures with other members until each of you has 4 pictures that belong to the same category.
- Each category must contain 4 pictures that are similar to one another but different from pictures in the other categories. i.e. your group will have a total of 4 different categories

Training 1

Participant 1
Solution category 1:




Participant 2
Solution category 2 :




Participant 3
Solution category 3:




Participant 3
Solution category 4:




Training task 2 *(story)*

- Exchange pictures with other members until each of you has 4 pictures that make up one story.
- Each category must contain four pictures that make up a story; i.e., your group will have a total of 4 different stories.

Training 2 Story

Participant 1
Solution category 1:




Participant 2
Solution category 2 : A man is a carpenter who uses a hammer to build a cage for a monkey that eats a banana




Participant 3
Solution category 3:




Participant 4
Solution category 4:




Task 1

- Your group will be given 16 pictures (i.e. each of you will have 4 pictures).
- You need to find the solution by exchanging pictures with other members until each of you has 4 pictures that belong to the same category.
- Each category must contain 4 pictures that are similar to one another but different from pictures in the other categories. i.e. your group will have a total of 4 different categories

Note, this task was given to groups of students in a previous experiment and a panel of experts determined the best solutions. Your job is to try to find the best solution, as judged by a panel of experts.

Task 1

Sophia:
Solution category 1:




Sophia:
Solution category 2 :




Sophia:
Solution category 3:




Sophia:
Solution category 4:

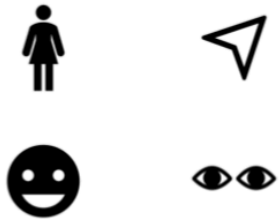



Task 2

- Your group will be given 16 pictures (i.e. each of you will have 4 pictures).
- You need to find a solution by exchanging pictures with other members until each of you has 4 pictures that make up one story.
- Each category must contain four pictures that make up a story. In other words, your group will have a total of 4 different stories.

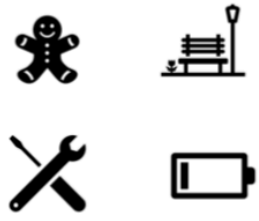
Task 2

Sophia:
Solution category 1:

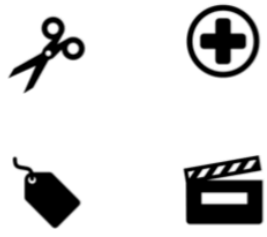


Sophia:
Solution category 2 :

Sophia:
Solution category 3:



Sophia:
Solution category 4:



Task 3

- Your group will be given 16 pictures (i.e. each of you will have 4 pictures).
- You need to find the solution by exchanging pictures with other members until each of you has 4 pictures that belong to the same category.
- Each category must contain 4 pictures that are similar to one another but different from pictures in the other categories. i.e. your group will have a total of 4 different categories

Note, there is no single correct solution to this problem, but your job is to try to come up with a solution that you think is a good one.

Task 3

Sophia:
Solution category 1:



Sophia:
Solution category 2 :




Sophia:
Solution category 3:



Sophia:
Solution category 4:




You will receive an email contains a screenshot of your solution and a link of a survey. Please complete the survey.

Group Code: B

Thank you for your participation

- Hanan Alattas, Department of Management Sciences, email: h4alatta@uwaterloo.ca, or
  Rob Duimering, Department of Management Sciences, email: rduimeri@uwaterloo.ca
- Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

Appendix F

Post-experiment Questionnaire following Completion of Three Tasks

Group problem solving (Solvers)

Start of Block: Default Question Block

This questionnaire is designed to help us better understand your role in the process of solving the three problems. When needed, please refer to the screen shot of your solution categories that was included in the email. To enable better understanding of your perception of the group dynamics, please be as precise as possible. Be assured that your response will be confidential. Please complete this questionnaire on your own, without consulting the other members of your group.

Page Break

Q1 Participant's name:

_____

Q2 Student ID:

_____

Q42 Department:

_____

Q4 Select the course for which you are participating in this experiment

- ○ MSCI 211 (4)
- ○ MSCI 311 (5)
- ○ Other (specify) (6) _____

Q5 Please select your age group

- ○ Below 17 (1)
- ○ 17-20 (2)
- ○ 20-25 (3)
- ○ 25-30 (4)
- ○ 30-35 (5)
- ○ Above 35 (6)

Q6 Please select your group code

▼ A (1) ... F (6)

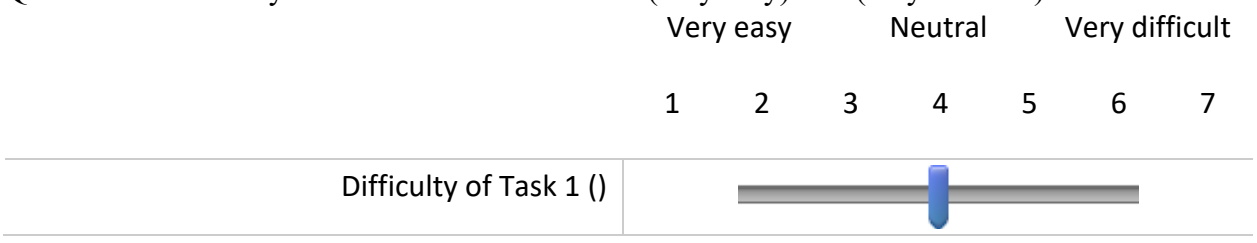
End of Block: Default Question Block

Start of Block: Block 1

The following questions relate to **Task 1**. Please refer to the email attachment to view your solution.

Page Break

Q7 Rate the difficulty of **Task 1**on a scale from 1 (very easy) to 7 (very difficult)



Q8 Can you give some examples of what made **Task 1** difficult?

_____

Q9 How did you deal with these difficulties to reach a solution for **Task 1**?

_____

Q10 Please rate your satisfaction with the overall team solution and each of the solution categories for **Task 1.**

	Extremely satisfied (1)	Somewhat satisfied (2)	Neither satisfied nor dissatisfied (3)	Somewhat dissatisfied (4)	Extremely dissatisfied (5)
The Overall Team Solution	☐	☐	☐	☐	☐
Solution Category 1	☐	☐	☐	☐	☐
Solution Category 2	☐	☐	☐	☐	☐
Solution Category 3	☐	☐	☐	☐	☐
Solution Category 4	☐	☐	☐	☐	☐

Q11 If not satisfied with either the overall team solution or any of the solution categories, can you please explain what do you think will make a better overall solution or a better solution category in **Task 1**?

_____

Page Break

The following questions relate to **Task 2**. Please refer to the email attachment to view your solution.

Page Break

Q12 Rate the difficulty of **Task 2** on a scale from 1 (very easy) to 7 (very difficult)

	Very easy 1	2	3	Neutral 4	5	6	Very difficult 7
Difficulty of Task 2 ()							

Q13 Can you give some examples of what made **Task 2** difficult?

_____

Q14 How did you deal with these difficulties to reach a solution for **Task 2**?

_____

Q15 How satisfied are you with the overall team solution for **Task 2**?
Please evaluate each of the story in the solution

	Extremely satisfied (1)	Somewhat satisfied (2)	Neither satisfied nor dissatisfied (3)	Somewhat dissatisfied (4)	Extremely dissatisfied (5)
The Overall Team Solution	○	○	○	○	○
Solution Category 1	○	○	○	○	○
Solution Category 2	○	○	○	○	○
Solution Category 3	○	○	○	○	○
Solution Category 4	○	○	○	○	○

Q16 If not satisfied with either the overall team solution or any of the solution categories, can you please explain what do you think will make a better overall solution or a better solution category in **Task 2**?

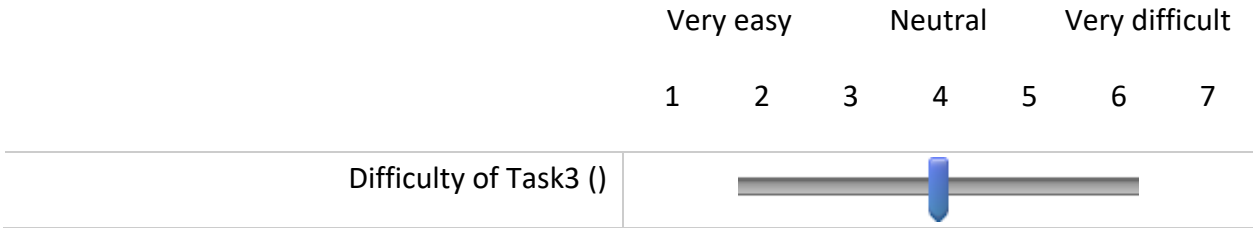
_____

Page Break

The following questions relate to **Task 3**. Please refer to the email attachment to view your solution.

Page Break

Q17 Rate the difficulty of **Task 3** on a scale from 1 (very easy) to 7 (very difficult)



Q18 Can you give some examples of what made **Task 3** difficult?

_____

Q19 How did you deal with these difficulties to reach a solution for **Task 3**?

_____

Q20 How satisfied are you with the overall team solution for **Task 3**?
Please evaluate each of the story in the solution

	Extremely satisfied (1)	Somewhat satisfied (2)	Neither satisfied nor dissatisfied (3)	Somewhat dissatisfied (4)	Extremely dissatisfied (5)
The Overall Team Solution	○	○	○	○	○
Solution Category1	○	○	○	○	○
Solution Category2	○	○	○	○	○
Solution Category3	○	○	○	○	○
Solution Category4	○	○	○	○	○

Q21 If not satisfied with either the overall team solution or any of the solution categories, can you please explain what do you think will make a better overall solution or a better solution category in **Task 3**?

_____

Page Break

The rest of the questions refer to **all three tasks** and to your overall problem-solving experience.

Page Break

Q22 Please rank the relative difficulty of the **three tasks** from 1 (very easy) to 7 (very difficult):

	Very easy			Neutral			Very difficult
	1	2	3	4	5	6	7
Task 1 ()							
Task 2 ()							
Task 3 ()							

Q23 Do you have any comments about the three tasks in general?

_____

End of Block: Block 1

Start of Block: Appreciation

Thank you for participating in our Group problem solving survey! Your feedback is extremely valuable.

If you are interested in viewing the results of this survey, they will be posted at UWSPACE
This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#XXXX). If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca

For all other questions or if you have general comments or questions related to this study, please contact:

Hanan Alattas, Department of Management Sciences, email: h4alatta@uwaterloo.ca**, or**
Rob Duimering, Department of Management Sciences, email: rduimeri@uwaterloo.ca

End of Block: Appreciation

Appendix G

Order of the Four Solutions with the Average Association Scores in an Expert Clipart Task7

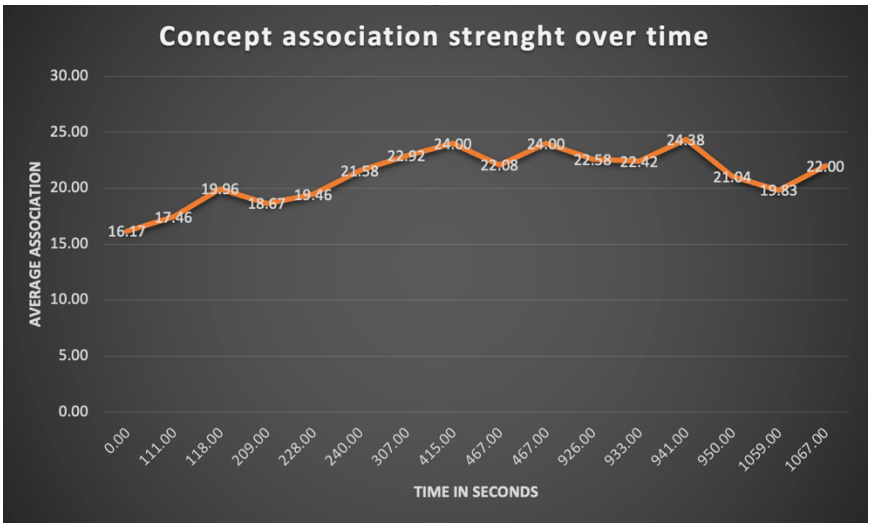
	Task ID:Experiment 1_A5_EXPERT										Design Set: ClipArt			
Time in seconds	Person A											Person B	Person C	Person D
	Items in solution category				Association score for each pair in the solution						average	average	average	average
	1	2	3	4	1_2	1_3	1_4	2_3	2_4	3_4	Association	Association	Association	Association
0.00	A	B	C	D	17	15	5	32	6	3	**13.00**	**15.33**	15.67	11.50
194.00	A	B	C	D	17	15	5	32	6	3	**13.00**	**20.83**	10.17	11.50
205.00	A	B	C	D	17	15	5	32	6	3	**13.00**	**24.33**	14.33	11.50
230.00	A	J	C	D	11	15	5	5	41	3	**13.33**	**24.33**	14.67	11.50
240.00	A	J	L	D	11	3	5	22	41	21	**17.17**	**24.33**	24.67	11.50
256.00	A	J	O	D	11	5	5	20	41	18	**16.67**	**24.33**	24.67	12.83
278.00	A	J	O	D	11	5	5	20	41	18	**16.67**	**32.17**	24.67	22.83
369.00	A	J	O	D	11	5	5	20	41	18	**16.67**	**32.17**	17.33	**28.17**
689.00	P	J	O	D	22	7	27	20	41	18	**22.50**	**32.17**	**21.33**	28.17

[7] Note that columns related to items in solution category and their pairwise association score has been hidden for person B, C, D. In this case the order of solution categories is: B, D, and A and C together.

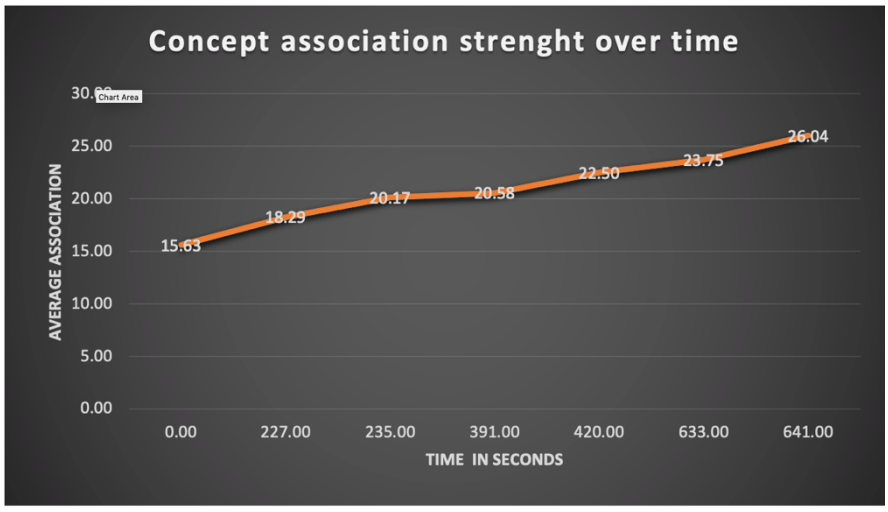
Appendix H

Examples of Concept Relation Trajectories During Problem Solving Process

Concept relation trajectory for a problem structure with 8 reversals



Concept association trajectory for a problem structure with no direction reversals



Appendix I

Verbal Protocol Transcripts

Expert 1 (Experiment A5 / Clipart/ Expert/ Version 1 / Solving Time = 692 sec)

Transcript	Interpretation
285 00:35:18.059 --> 00:35:22.079 EXP: Here are the pictures you can see your names, right? 286 00:35:22.079 --> 00:35:25.708 D: Yup. 287 00:35:36.599 --> 00:35:39.898 B: Okay, so I think there's a food category. 288 00:35:39.898 --> 00:35:43.559 B: Like school. 289 00:35:43.559 --> 00:35:47.849 A: Adulthood? B: Adulthood. 290 00:35:52.708 --> 00:35:59.338 A: Why don't we Say sport,	After around 17 seconds of seeing the pictures, participants started generating ideas of possible solution categories. The categories they identified are: 1. food 2. school 3. Adulthood 4. sport
D: what what is the food category? 291 00:35:59.338 --> 00:36:04.228 B: There's like the pizza vegetables, vegetables, coffee. D: ok ok ok 292 00:36:04.228 --> 00:36:09.119 B: Is there 1 more? 293 00:36:09.119 --> 00:36:16.018 D: Does that count it uh? 294 00:36:16.018 --> 00:36:20.548 D: The pumpkin thing does that count as a food?. 295 00:36:20.548 --> 00:36:24.838 C: Yeah,	Then, they started asking questions about the items in each category they already identified: -They listed three items in the food category: pizza vegetables coffee After 20 seconds, they identified possible items that fit into a food category which is pumpkin. **Conflict recognition 1: Scarcity:** They listed three items in the food category, which are pizza, vegetables, and coffee.. **Conflict resolution 1: Affordance-based assumption** To deal with this conflict the group made an affordance-based assumption that the picture of the haybale with the pumpkin is the fourth food item.

B: yeah, I think that too. I think the…. ---------- 296 00:36:26.458 --> 00:36:31.108 B: And then there's like elements like water fire. 297 00:36:31.108 --> 00:36:35.699 D: Right, water fire. B: snow, Maybe 298 00:36:35.699 --> 00:36:39.208 A: Snow is water I think. ---------- 299 00:36:40.438 --> 00:36:43.648 B: Tennis and football. 300 00:36:47.489 --> 00:36:52.079 A: If you could do activities like, and and it's. 301 00:36:52.079 --> 00:36:55.469 A: A ball, driving a car and. 302 00:36:55.469 --> 00:37:04.018 A: A birthday party such a balloon, or, I guess graduation hmmm I feel like the book bag and the hat. 303 00:37:04.018 --> 00:37:07.018 A: Can go together, um. 304 00:37:07.018 --> 00:37:14.579 A: You could talk about, like a school house where you have a man going to school with a book that he graduates at a house. 305 00:37:14.579 --> 00:37:19.648 A: That should be 4 things I count. D: Yes. 306 00:37:19.648 --> 00:37:25.528 EXP: So, do you want to swab or you still thinking. 307 00:37:25.528 --> 00:37:32.429 B: So, are we doing for similar categories? Are we already doing the story. 308 00:37:32.429 --> 00:37:42.358 EXP: No, this is not the stories it's similar to the tools fruit and people B: Okay.	---------- They identified another category of “Elements” that includes (Water, Fire, Snow) ---------- They grouped together tennis and soccer ball then they called it activities and added to them the car as “driving a car” and a balloon as” birthday party or graduation”. Then it seems that while discussing the activity category, another solution category has emerged simultaneously when person A then noticed that the graduation hat and the backpack could go together and added the house and the man to them as one category in which “school house where you have a man going to school with a book that he graduates at a house”. Person A considers bird as a bird watching activity **Conflict recognition 2: Scarcity:** In the sport category, there are only two sport items (tennis racket and soccer ball). **Conflict resolution 2: Relabelling the category** They renamed it to “activity” as a way of abstracting or widening the category to include more non-sport items such as “car”, “balloon”, and “graduation”

309 00:37:42.358 --> 00:37:46.798 A: Bird watching is an activity.	
310 00:37:48.929 --> 00:37:53.518 A: I think I have activities. 311 00:37:54.778 --> 00:38:00.088 B: Okay, I'll take the food stuff again. 312 00:38:00.088 --> 00:38:04.829 C: I'll take the elements, I guess. 313 00:38:04.829 --> 00:38:08.969 EXP: Okay, so who wants to start swapping. 314 00:38:08.969 --> 00:38:13.860 B: Um, "C", can I have your pizza and your coffee. 315 00:38:13.860 --> 00:38:17.519 C: Yeah B: and I'll just give you. 316 00:38:17.519 --> 00:38:21.900 B: Sorry, what is your category again? C: Elements. 317 00:38:21.900 --> 00:38:32.250 B: Elements I'll give you the water and the house for now. C: Okay B: the house could possibly go somewhere else. C: Yeah, 318 00:38:32.250 --> 00:38:38.039 B: the pizza and the coffee. EXP: Okay. B: With the water and the. 319 00:38:40.650 --> 00:38:44.190 B: House EXP: okay. 320 00:38:46.260 --> 00:38:51.539 EXP: great. 321 00:38:51.539 --> 00:38:56.130 C: Sorry, I missed the category 1 the activity category?	Then it seems that 3 of them settled on the categories they wanted to take, and they exchanged the pictures accordingly. **These are the categories they have after exchanging:** **A: (bird, soccer, car, tennis), > activities** **B: (Pizza, coffee, vegetable, pumpkin) > food** **C: (water, Fire, house, winter house)> elements** **D: (man , backpack, grad-hat, balloon)**

322 00:38:56.130 --> 00:39:02.940 A: Yeah, yeah, I must be doing it. C: can I get the fire and the kind of like the winter scenery. 323 00:39:02.940 --> 00:39:08.909 A: Sure, C: and then I'll switch with the soccer and the graduation hat . 324 00:39:08.909 --> 00:39:12.539 C: for now. EXP: Is that's okay. 325 00:39:12.539 --> 00:39:22.349 A: Yeah, EXP: soccer and this and the winter house with.. C: the graduation hat. EXP: Okay. 326 00:39:26.639 --> 00:39:33.210 A: Can we can I give you a graduate hat and can I give you the car?. 327 00:39:33.210 --> 00:39:39.179 D: Yeah. 328 00:39:49.530 --> 00:39:52.650 B: "D", can I give you my backpack? 329 00:39:52.650 --> 00:39:57.420 B: And take the pumpkin D: yeah. the pumpkin 330 00:40:01.079 --> 00:40:07.199 A: This.	
331 00:40:07.199 --> 00:40:10.349 A: So, what's everyone's category? just to be clear. 332 00:40:11.820 --> 00:40:18.449 B: Well, I think that "C" and my categories pretty distinct the other 2 are a bit like. 333 00:40:18.449 --> 00:40:21.809 B: It's hard to differentiate between them 334 00:40:21.809 --> 00:40:30.329 D: graduation, … D: Like school, backpack is graduating. 335 00:40:30.329 --> 00:40:34.650	**Then they start recognizing conflicts in the set of the four categories they had.** After they exchanged, they started another round of **discussion asking** questions about the excising categories and identifying conflicts in them. **Conflict recognition 3: Weak categories** Here person B thinks category A and D are weaker ones. **Conflict resolution 3: Discard the conflict** Then Person D is justifying his category This seems to be an attempt from Person D to deal with the conflict by making a justification that all items belong to a category of graduation.

D: Graduation hat and then balloon.	
336 00:40:34.650 --> 00:40:39.269 A: But I feel like "C" has a weaker. 337 00:40:39.269 --> 00:40:46.079 A: Um, category, because the house doesn't may be fit in with elements, but if you switch balloon and house, so it can be like, balloon is like, air. 338 00:40:46.079 --> 00:40:51.269 A: And house, if the house goes to you, "D", then becomes like, school. 339 00:40:51.269 --> 00:40:54.599 A: Being the category where person goes to school and graduate 340 00:40:54.599 --> 00:40:58.440 D: I see. C: Yeah.	**Conflict recognition 4: An item fits better in another category** Person A thinks that C is a weak category because balloon fits better in the element category than the house. **Conflict resolution 4: A compromise based on what fits better in other categories** An attempt to improve the weaker categories. They made a compromise based on what fits better into the other categories. (The house from C fits better in D and the Balloon from D fits better in C) so they made the exchange accordingly.
341 00:40:58.440 --> 00:41:06.030 B: But then the bird is kind of random in category 1 A: Bird watching is an activity. B: Oh, yeah, that's true. 342 00:41:07.230 --> 00:41:13.739 EXP: Do you want to swap pictures ? D: yeah. Uh.wait ammm 343 00:41:13.739 --> 00:41:18.599 A: I think it Makes sense. 344 00:41:19.800 --> 00:41:27.360 D: Okay, yeah. Can I swap the balloon with the. 345 00:41:27.360 --> 00:41:30.659 D: The red house. 346 00:41:33.659 --> 00:41:38.400 B: I guess that house could just be a school then. A: Yeah. Yeah.	**Conflict recognition 5: Identifying a random item that has no category membership** Person B asked question about an item (Bird) that seems to not belong to a given category (activity category) **Conflict resolution 5: Affordance-based assumption** But **A** justified why it would belong (as a bird watching activity) > making an interpretation of the random item (bird) to complete the category (activity)
347 00:41:41.460 --> 00:41:44.789 Silence. 348 00:41:48.389 --> 00:41:56.489 A: "C", What's your category? 349 00:41:56.489 --> 00:42:01.079 C: umm like elements I guess	**Conflict recognition 6: Weak category** After 15 seconds of silence, A asked C about her category, person C answered elements but seemed not sure. **Conflict resolution 6: Suggesting new solution category** Then D tried to suggest a category of seasons they discussed about it for a minute and a half, but it seemed to not work so the group discarded this idea of season.

350
00:42:02.159 --> 00:42:12.329
C: I guess
D: you can like a season.

351
00:42:13.710 --> 00:42:18.630
D: Like, fall, I don't know, I'm just throwing ideas.
352
00:42:19.710 --> 00:42:24.480
D: "B", Can you have, like, fall category where you're like, your harvest food?

353
00:42:24.480 --> 00:42:28.679
B: Yeah, yeah, I was thinking that too. Cause it's like the pumpkin.

354
00:42:30.869 --> 00:42:39.210
B: Oh, you mean my category how it is now?
D: No, no, no I'm just throwing ideas.
B: Yeah.

355
00:42:39.210 --> 00:42:44.039
D: Not sure if this is the right thing? Yeah.

356
00:42:44.039 --> 00:42:52.650
D: What would be in the fall category then, what are…
B: the pumpkin.

357
00:42:52.650 --> 00:42:56.489
D: And the food, uh the vegetables.

358
00:42:57.510 --> 00:43:02.940
B: Arguably the fire, but that could also be winter too. Like the bird.

359
00:43:06.360 --> 00:43:13.289
D: yeah the bird.
B: like going back to school.
D: Yeah. Or like 4 seasons.? I.

360
00:43:16.199 --> 00:43:22.769
Like, well,
B: what would the water one be though with seasons?
D: Oh, no. Oh, so the bird.

361
00:43:22.769 --> 00:43:28.650
D: Oh, no never mind. The snowing is obviously winter harvesting is fall.

362
00:43:28.650 --> 00:43:32.550
D: The pumpkin thing is fall and summer would be…
B: water?

363
00:43:33.570 --> 00:43:41.909
A: Here you go on the beach I guess a birthday party or a summer. We can use a balloon.

……
364
00:43:41.909 --> 00:43:46.800
B: I don't know, I'm not going to do seasons, we don’t have seasons here. {Not clear}

365
00:43:46.800 --> 00:43:50.610
D: Yeah, might be wrong. Yeah.

367
00:43:57.059 --> 00:44:06.000
A: Okay, so we have these categories right now.

368
00:44:06.000 --> 00:44:09.269
A: Um, the often there are any categories that are

369
00:44:09.269 --> 00:44:15.090
A: a little weak, or we can move towards to make our overall.

370
00:44:15.090 --> 00:44:19.440
A: task 1 stronger?

371
00:44:20.849 --> 00:44:28.019
EXP: Are you asking me?
A: Sorry I'm asking the team.

372
00:44:28.019 --> 00:44:31.409
D: I think it's, it's.

373
00:44:31.409 --> 00:44:39.900
D: The balloon from “C” seems a little awkward, but, except for that.

374
00:44:39.900 --> 00:44:44.610
D: I think it makes sense.

375
00:44:47.039 --> 00:44:52.050
D: Yeah, I can be school ,”C” can be elements.

376
00:44:52.050 --> 00:44:56.969
D: Yeah.

377
00:44:56.969 --> 00:45:01.920
D: I think this is fine for now. I got no objection.

Conflict recognition 7: Weak category
They tried to identify and improve weaker categories again: Person D identified the balloon as “awkward” in the element category (category C).

Conflict resolution 7a : Discard the conflict
They almost accepted it (Satisficing as strategy to deal with the Conflict), in (374-378)

Conflict resolution 7b : Compromising + Relabeling the category.
Person A suggested switching the balloon (to be in activities category) with the bird in the elements category but this time they changed its name from elements to nature.
Here they resolve the conflict through compromising which one fits better and adjusting the name of the category. They accepted balloon as an activity as a party thing and to switch the bird to the element category the adjusted the name of the category to accommodate the bird into it.

Note that in the discussion they acknowledged that bird watching activity it's not a real activity and activity is a general category that can accept many other items.
There were questions and moments of clarifications and justifications when they made the last suggestions off switching the bird and the balloon so they asking questions trying to make sure that this is an acceptable solution.

B: I think this is fine too

378
00:45:01.920 --> 00:45:04.920
C: Sure. Okay.

379
00:45:04.920 --> 00:45:10.559
EXP: So, do you want to give names or are you still thinking.

380
00:45:11.610 --> 00:45:15.239
EXP: Do you want to give names to the categories or not yet.

381
00:45:15.239 --> 00:45:23.099
A: 1, last thing … "C" ... {Not clear}

382
00:45:23.099 --> 00:45:27.030
A: Maybe.

383
00:45:27.030 --> 00:45:30.690
EXP: Um, we couldn't hear you actually.
C: Yeah.

384
00:45:30.690 --> 00:45:34.769
A: Sorry, "C", do you think it makes sense to swap the bird with the balloon maybe.

385
00:45:34.769 --> 00:45:42.059
A: Does it make sense. I'm asking what what you think.

386
00:45:42.059 --> 00:45:46.139
C: Oh, sure. And then, I guess I'll be like natural elements.

387
00:45:46.139 --> 00:45:49.199
A: Yeah, like nature. That I was thinking.

388
00:45:49.199 --> 00:45:54.960
But is a balloon an activity?.

389
00:45:54.960 --> 00:45:58.170
D: birthday party
A: a balloon is like a birthday party.

390
00:45:58.170 --> 00:46:03.269
A: Kind of activity. Yeah, you read my mind here we go.
391
00:46:03.269 --> 00:46:11.070
D: it might make more sense than.

392

00:46:11.070 --> 00:46:16.949 Bird watch. Well, would they really put bird watching as an Activity. 393 00:46:16.949 --> 00:46:27.119 A: These are great questions. I don't think they would… 394 00:46:27.119 --> 00:46:35.429 A: so, yeah, let's switch "C" and that gives you a more concrete category. And activities is already very generalizes so D: right A: technically anything could could be with that.	
399 00:47:07.619 --> 00:47:12.000 D: Yeah, I think we can give out names now, these 4 categories. 400 00:47:12.000 --> 00:47:18.900 A: So category 1 is going to be Activities. 401 00:47:22.980 --> 00:47:29.969 B: It's like food, I guess. 402 00:47:29.969 --> 00:47:34.230 EXP: sure. 403 00:47:39.480 --> 00:47:49.320 EXP: three? C: Uh the nature. 404 00:47:50.820 --> 00:47:56.969 EXP: Four? D: School.	They all agreed and swapped the balloon with the bird : The categories now: A: Activities (Balloon, soccer, car, tennis), B: Food (Pizza, coffee, vegetable, pumpkin) C: Nature (water, Fire, bird, winter house) D: School (man , backpack, grad-hat, house) A: Activities: (Balloon, soccer, car, tennis), B: Food: (Pizza, coffee, vegetable, pumpkin) C: Nature: (water, Fire, bird, winter house) D: School: (man , backpack, grad-hat, house)

Good 1 (Experiment D5 / Clipart/ Good / Version 1/ Solving Time = 361 sec)

Transcript	Interpretation
176 00:28:02.398 --> 00:28:05.519 EXP: Okay, so here are your pictures. 177 00:28:05.519 --> 00:28:09.388 EXP: The names are correct right? 178 00:28:09.388 --> 00:28:12.449 B: Yeah. 179 00:28:17.159 --> 00:28:20.999 Silence. 180 00:28:22.378 --> 00:28:27.689 EXP: You can discuss. 181 00:28:27.689 --> 00:28:37.199 EXP: You can think about it discuss and ask me whenever you want me to swap pictures for you I will be exchanging pictures ----------	They looked at the picture for 30 seconds then started the discussion ----------
182 00:28:38.429 --> 00:28:44.638 A: Okay, so 1 that I see is like, buildings, I guess, like the house and. 183 00:28:44.638 --> 00:28:53.608 A: The cottage in mine. Yeah, actually, I see like, there's also like that. 184 00:28:54.689 --> 00:29:01.709 C:Sorry A: go ahead. C: Yeah, I was just going to say, I think, in the pumpkin… 185 00:29:01.709 --> 00:29:05.669 C: hot Chocolate in mine, I kind of see, like a winter theme. 186 00:29:05.669 --> 00:29:12.269 C: The wave is the summer B: yeah, I think I see seasons. Like, you can see. 187 00:29:12.269 --> 00:29:15.479 B: There is a pumpkin, so you can maybe have, like, harvest. 188 00:29:15.479 --> 00:29:20.159 B: Something along that you have a winter them, you have like a spring.	Person A notice buildings: the cottage and the house. Person B&C noticed season themes (winter : winter house, fall harvest, summer waves, spring) They thought of educational theme (backpack, graduation hat) A think that the backpack, graduation hat can go with the fall theme because it’s like back to school.

189
00:29:20.159 --> 00:29:25.858
B: And then the other 1 might be educational because you see backpacks and you see a graduation hat.

190
00:29:25.858 --> 00:29:30.538
C:Um, yeah,
A: maybe the.

191
00:29:30.538 --> 00:29:38.519
A: Like, graduation hat and the backpack goes with, like, the fall theme, because it's back to school. Maybe
C: back to school probably.

192
00:29:38.519 --> 00:29:43.078
B: Yeah, I guess so. Yeah.
A: So should we switch.

193
00:29:43.078 --> 00:29:48.088
C: Yeah.

194
00:29:48.088 --> 00:29:51.719
silence

195
00:29:51.719 --> 00:29:58.348
B: I I think I'm going to switch that the backpack.

196
00:29:58.348 --> 00:30:03.328
B: for pizza maybe I don't know.

197
00:30:06.209 --> 00:30:10.679
B: Or, actually, you know, I, I think A should switch.

198
00:30:10.679 --> 00:30:14.068
B: That I don't know this badminton maybe.

199
00:30:14.068 --> 00:30:19.528
B: Yeah, that with chocolate
C:yeah. Yeah.
A: Yeah

201
00:30:24.269 --> 00:30:29.189
A: Yes.

202
00:30:29.189 --> 00:30:34.858
B: I think that bird kind of looks like a winter bird that would be in there in winter. So I think maybe.

203

They started the exchanges by switching the badminton with the hot chocolate.

Person B thinks category 1 is complete as winter theme and the bird can belong to the winter category.

No one answered.

A: (bird, fire, winter house, coffee) > Winter Category

B: (backpack, waves, vegetables, house,) > Unnamed

C: (Pizza, soccer, tennis, grad hat) > Unnamed

D: (man, hay, balloon, car) > Unnamed

Conflict recognition 1: Scarcity, mssing item: Person B wanted to complete the winter theme category.

Conflict resolution 1: Affordance-based assumption

00:30:34.858 --> 00:30:41.009 B: Just a category one's complete. What's your senses? 204 00:30:41.009 --> 00:30:47.519 C: I think I can change the. 205 00:30:47.519 --> 00:30:50.818 C: The graduation hat with the. 206 00:30:50.818 --> 00:30:55.348 C: With a wave from B 207 00:30:55.348 --> 00:30:58.348 B: No, I think the waves are part of a cottage theme C: Yeah 208 00:30:59.699 --> 00:31:06.628 B: Yeah, I think I think I should switch maybe this person for the fruits. 209 00:31:06.628 --> 00:31:12.269 B: Um, A: yeah, B: yeah, I, I think that might be a good option. 210 00:31:14.429 --> 00:31:22.019 B: What do you guys think do you think we should switch the fruit with the person. 211 00:31:22.019 --> 00:31:26.669 A: Yeah, I think. 212 00:31:26.669 --> 00:31:33.328 B: The fruits and vegetables yeah. EXP: Is that okay? D?. 213 00:31:33.328 --> 00:31:38.038 D: Yeah.	He assumes the bird is a winter bird and can belong to the winter so category 1 is complete as winter theme category. **Conflict recognition 2: Item better fits another category** C wanted to exchange graduation hat (in category C) with the wave (in category B) but it seems not to work because person B thinks the wave is a part of a cottage. **Conflict resolution 2: Discarding the conflict** They just discarded the idea and jumped to another switch (man with the vegetables). Note: It seems that they might be stuck with the themes that they created initially in the brainstorming of seasons. They then exchanged the man with the vegetables (they did not explain why). A(bird, Fire, winter house, coffee), Winter B: (backpack, wave, man, house,) C: (Pizza, soccer, tennis, grad hat) D: (vegetables, hey, balloon, car)
--- 214 00:31:39.929 --> 00:31:43.138 A: I think the backpack should go with the.. 215 00:31:43.138 --> 00:31:47.368 A: In in D's category. 216 00:31:47.368 --> 00:31:52.439 A: Right. 217 00:31:52.439 --> 00:31:58.919	A seemed to adhere to the "Fall back to school" category the group mentioned earlier and wanted to add the backpack to category 4. Thus, they then switch the car with the backpack. A: (bird, Fire, winter house, coffee) > Winter Category B: (car, wave, man, house) > Cottage (presumably) C: (pizza, soccer, tennis, grad hat) > Unnamed

D: Yes swhich the car with the backpack 218 00:31:58.919 --> 00:32:03.148 EXP: Okay, A: yeah, I think, B: yeah. Yeah. Great --- 219 00:32:04.528 --> 00:32:08.398 B: Let me switch the backpack with the, the pizza. 220 00:32:11.608 --> 00:32:14.788 EXP: Okay --- A: Okay. What the theme then? 221 00:32:14.788 --> 00:32:19.169 B: why is there a balloon there. 222 00:32:19.169 --> 00:32:24.419 C: I can see the balloon with the graduation hat. 223 00:32:24.419 --> 00:32:32.009 B: That's what I was thinking as well. Yeah. Maybe what sport do you play in like, an August September. 224 00:32:32.009 --> 00:32:37.288 B: Soccer badminton maybe soccer. 225 00:32:38.848 --> 00:32:43.709 c: With the with B: with the balloon 226 00:32:45.659 --> 00:32:48.959 C: Okay, yeah, that works may be 227 00:32:48.959 --> 00:32:59.939 B: Yeah, this feels --- 228 00:32:59.939 --> 00:33:07.828 A: What would category 2 to be though? B: This would be like a cottage time when you take your car. 229 00:33:07.828 --> 00:33:19.348 B: To that cottage and there's like, waves there and that there's like a pool not pool, but like, maybe a reverse or like a lake there C: maybe a vacation	D: (vegetables, hay, balloon backpack,)> Unnamed --- They then switch the pizza with the backpack (did not explain why). A: (bird, Fire, winter house, coffee) > Winter Category B: (car, wave, man, house) > Cottage (presumably) C: (backpack, soccer, tennis, grad hat) > School (presumably) D: (vegetables, hay, balloon, pizza) > Unnamed --- **Conflict recognition 3: Weak category because an item seems not related** The last exchange seems to create confusion for some, where they are not sure of the theme that category D will be (it conflicts with the category idea they defined earlier of “Fall back to school”); person A is smiling and asking what the theme for category D is then. Person B is asking why there is a balloon in category D. **Conflict resolution 3: Compromise+ affordance based assumption** Person C thinks the balloon can go with the graduation They all agreed, then person B ask what sport you play in August September he thought maybe soccer, so he switched soccer ball with the balloon: A: (bird, Fire, winter house, coffee) > Winter Category B: (car, wave, man, house) > Cottage (presumably) C: (backpack, balloon, tennis, grad hat) > School/Fall (presumably) D: (vegetables, hay, soccer, pizza) > Unnamed --- **Conflict recognition 4: Weak category:** Person A ask “what would category B be”? **Conflict resolution 4 (Relabeling):** Person B&C justified to be a cottage or vacation category

B: vacation type. Yeah that would work --- 230 00:33:19.348 --> 00:33:23.429 C: What about 4? 231 00:33:23.429 --> 00:33:28.888 C: What would it be? A lot of noise. 232 00:33:30.298 --> 00:33:36.298 B: Like, I guess you can make pizza with these fruits of the top, and that you can put I don't know. 233 00:33:43.138 --> 00:33:51.628 B: Maybe we should keep the balloon where it is. I don't know where it was, like, in category 4. 234 00:33:53.278 --> 00:33:57.269 B:Yeah, because 1 and 2 feel like it's Don, I think. 235 00:33:57.269 --> 00:34:02.159 3, and 4 don't look as good as. 236 00:34:02.159 --> 00:34:05.489 Exp: So do you want to swap B: yeah. 237 00:34:05.489 --> 00:34:15.898 A: Yeah. 238 00:34:15.898 --> 00:34:22.168 B Yeah, then this is good. 239 00:34:22.168 --> 00:34:25.978 C: I am thinking What. 240 00:34:25.978 --> 00:34:30.148 C: What D is ? 241 00:34:35.009 --> 00:34:41.369 C: And, uh, Exp: you can discuss what you think. 242 00:34:41.369 --> 00:34:47.818 C: There is vegetables, pizza, pumpkin and a balloon, right? Right. 243 00:34:47.818 --> 00:34:57.449 B: The Halloween party me I don't know I guess yeah, because you need, because you need because you need vegetables to make pizza and then. 244	--- **Conflict recognition 5: Weak category** Now they're asking about the 4th category D what is it? **Conflict resolution 5a : Relabeling** Person B thinks of a scenario that someone can make pizza with the vegetable and then can play, but he seemed not sure. **Conflict resolution 5b: Compromising** They seemed to not have a clear idea what category 4 would be. After few seconds of silence, they decided to swap back the balloon with the soccer ball because They think categories A and B are Done but not C, and D. A: (bird, Fire, winter house, coffee) > Winter Category B: (car, wave, man, house,) > Cottage/ vacation C: (backpack, soccer, tennis, grad hat) > School/Fall (presumably) D: (vegetables, hay, balloon, pizza) > Unnamed **Conflict resolution 5c :Relabeling** (adjusting the name of the category to fit the a less similar item, “pumpkin”) They were still not sure about what category D must be called and after a 30 second period of silence, Person B suggested a “Halloween party” name and gave justification of that. They laughed and agreed.

00:34:57.449 --> 00:35:01.739
B: You can have you need a pumpkin to have a Halloween and balloon is just for a party kind of thing.

245
00:35:01.739 --> 00:35:04.768
C: That works actually. [laughing]

246
00:35:04.768 --> 00:35:10.679
B: All right,
C: tell me you kind of told me on that one.

247
00:35:10.679 --> 00:35:13.798
C:I think good.

248
00:35:13.798 --> 00:35:18.268
B:Yeah, good.

EXP:So the 1st, 1 is winter.

249
00:35:19.918 --> 00:35:28.559
B: All right yeah. 2nd is a vacation Yeah vacation slash cottage site.

250
00:35:30.838 --> 00:35:37.559
A: Might just be a house we didn't know
C: on the lake or.

251
00:35:38.668 --> 00:35:41.998
EXP: Do you want to call it this way?

252
00:35:41.998 --> 00:35:45.599
All: Yeah, yeah, that.

253
00:35:48.389 --> 00:35:51.838
EXP: 3.

254
00:35:51.838 --> 00:35:55.588
B: School activities maybe.

255
00:35:55.588 --> 00:35:59.128
C: Yeah.

256
00:35:59.128 --> 00:36:02.579
B: School activities.

257
00:36:09.239 --> 00:36:13.079
Oh, okay.

258
00:36:16.889 --> 00:36:20.429
D: Halloween party party.

There were no switches but here they decided on the name of the categories:

A: (bird, Fire, winter house, coffee) > Winter

B: (car, wave, man, house,) > Vacation/cottage

C: (backpack, soccer, tennis, grad hat) > school activities

D: (vegetables, hay, balloon, pizza) > Halloween party

*They seem not sure when they are giving the names of category C; they used the word maybe and there was a questioning tone when they were giving the category names.

259 00:36:20.429 --> 00:36:26.728 EXP: So, you're happy with this solution. 260 00:36:26.728 --> 00:36:30.239 EXP: I'll take a screenshot if you are. Yeah.	

Story 1 (Experiment E3/ Clipart/ Story / Version 1/ Solving Time = 254 sec)

Transcript	Interpretation
228 00:30:29.159 --> 00:30:33.538 EXP: So here are the Pictures . 229 00:30:33.538 --> 00:30:40.769 EXP: You will be forming, like, 4 different stories. Here are the pictures and you see your name's right? 230 00:30:40.769 --> 00:30:47.878 EXP:So you can think and discuss and ask me whenever you want me to swap pictures for you. 231 00:31:04.949 --> 00:31:18.628 EXP:Do you want to talk? A: It's like 1 of us has a story then we can start making it from, like, their corner. Like like,[if one of us] can think of something. 232 00:31:18.628 --> 00:31:26.489 B: I think of pictures that would go nicely together like, maybe the pizza, the beach or something. 233 00:31:26.489 --> 00:31:31.078 D: I think for mine for category 4. 234 00:31:31.078 --> 00:31:35.909 D: Like, I can honestly just swap out that pumpkin thing for. 235 00:31:35.909 --> 00:31:40.588 D: For almost anything else here and that would be a story. So it'd be like this guy. 236 00:31:40.588 --> 00:31:52.348 D: Start uh has a birthday, he needs to go to so many hops in his car and goes and buys a balloon boom. That's like a story right there. I mean, to swap out that pumpkin thing but, I mean. 237 00:31:52.348 --> 00:31:57.929 D: If you guys are down with that? A: Yeah. Okay. What can you swap the pumpkin with ? D: Ah I don't know 238 00:31:57.929 --> 00:32:09.298	Started the discussion after around 40 seconds of seeing the pictures. They thought of strategy that one of them start a story and then the group go from their making the other stories. Some one considers pictures that can go together like the pizza and the beach (waves). **Conflict recognition 1: Random items** Person D started offering other members the items that he doesn't need (pumpkin)and Started proposing his stories. That is, D has a birthday story based on 3 out of the 4 items he has and offered the item he doesn't need, which is the pumpkin. **Conflict resolution 1: Compromise** They then switch the pumpkin with the pizza from C since pizza fits the theme of the birthday, A: (bird, fire, winter house, tennis), > unnamed B: (waves, house, vegetable, backpack) > unnamed C: (pumpkin, soccer, coffee, grad-hat) > unnamed D: (man, pizza, car, balloon)Birthday > story 1

A: the pizza like pizza party
D: Yeah, it could be a pizza pizza.
C: Sure
A: Yeah. That's good.
D: Yeah, so that's 1.

239
00:32:11.189 --> 00:32:15.659
D: Uh.

240
00:32:15.659 --> 00:32:19.259
C: I feel like I have

241
00:32:19.259 --> 00:32:25.828
C: Too much food I'm not sure how to come up with a story with the pumpkin.

242
00:32:25.828 --> 00:32:32.909
A: Yeah,
C: yeah, like the pumpkin just seems out of place here I think.

243
00:32:33.959 --> 00:32:40.108
C: umm let me See.

244
00:32:43.169 --> 00:32:49.409
D: You could, like, stretch the truth and be like, there's that house there "B" in category 2.

245
00:32:49.409 --> 00:33:02.939
D: There's that house there that's that's on a farm that grows crops like pumpkins and vegetables, and then throw like a 4th thing in there. So that would use the house the, and the vegetables and then something else.

246
00:33:02.939 --> 00:33:10.588
C: Okay, actually, can I get a house like, from either of you and I can give you Whatever you guys want I don't[..].

247
00:33:10.588 --> 00:33:23.878
B: I was going to swap the beach with the bird and just how? Sorry the yeah, that with the bird and then the bag with the pumpkin and that's the story for me.

248
00:33:23.878 --> 00:33:27.179
C: Like, do you want to? Okay. Oh, okay. Okay.

249
00:33:27.179 --> 00:33:31.648
B: Yeah,
A: like a farm?
B: Yeah.

Conflict recognition 2: Weak category

But person C seems not happy with this since he now has too many food items;
though he had 2 food items the coffee and pumpkin then,

Conflict resolution 2: Affordance-based assumption

Person D proposed a farm story for category 2 that uses the pumpkin with the house and vegetable.

They agreed but they did not stick with the same story but similar to it as they switched the pumpkin from C with the backpack from B and the bird from A with the with the water from B considering category B as a fam story and at the same time the backpack would go together with a grad-hat as a graduation story:
A: (waves, fire, winter house, tennis), > unnamed
B: (bird, house, vegetable, pumpkin) > farm story 2
C: (backpack, soccer, coffee, grad-hat) > unnamed
D: (man, pizza, car, balloon)> Birthday story 1

A: Okay. Yeah.

250
00:33:31.648 --> 00:33:36.628
C: Okay, yeah, actually I could.. the bag will be perfect.

251
00:33:36.628 --> 00:33:42.659
B: Yeah, like a graduation thing for you
C: yeah
B: ok so the pumpkin with the bag.

252
00:33:42.659 --> 00:33:46.288
C:Yeah, the pumpkin with the bag.

253
00:33:46.288 --> 00:33:50.669
B: Yeah, and then the bird with the beach is that Okay.

254
00:33:50.669 --> 00:33:55.288
A: Yeah. sure

255
00:33:55.288 --> 00:33:59.818
B: Like I went to a beach house or something.

256
00:33:59.818 --> 00:34:03.239
Had a bonfire and played some tennis that works.

257
00:34:03.239 --> 00:34:06.659
D:Yeah,
B: yeah. {Laugh}

258
00:34:06.659 --> 00:34:12.838
A: Oh, wait.
D: Oh, wait, no, I would swap the I would swap the, the tennis.

259
00:34:12.838 --> 00:34:20.278
D: The whatever that badminton thing for the coffee, right?
A: Yeah so it's like a cold winter day.

260
00:34:20.278 --> 00:34:33.958
A: coz, it's like that that house,
D: yes. Yeah. That's what I think cold winter day for category. 1 and then category 3, you'd be like, there's this scholar that graduated in, like, high academics with, like, a love for sports or something. I don't know.

261
00:34:33.958 --> 00:34:38.489
EXP: You want me to swap pictures ?
C: what did you want to swap? Sorry? I missed that.

262
00:34:38.489 --> 00:34:45.449
A: Like, swap the rackets for, like, the cup of tea.

With that last swap, ideas of stories for the last two categories (A and C) emerged simultaneously. That is , as soon as they swapped the pictures, person B proposed a story of a beach house for category A incorporating the items that already in there.

Conflict recognition 3: Item better fits another category

But immediately Person A and C suggested another idea that would fit the coffee in A "cold winter day story" and the tennis in C as a "graduating story".

Conflict resolution 3: Compromising

They propose switching the tennis from A with the coffee from C

C: Rackets for the Cup of tea OK sure .

263
00:34:45.449 --> 00:34:52.048
A: but like the, I don't think the way, like, the beach fits into that then.

264
00:34:54.449 --> 00:35:07.018
D: Oh, but like, I mean, that you could argue, that's like, you know, how like, beaches on airbnb they sell during the winter time, right? It's like a cozy like winter stay, it could still be on the beach. It's just that.

265
00:35:07.018 --> 00:35:20.009
D: You know, yeah, like a winter potty or something,
C: it could be could be like a pond or whatever too. I mean, I don't know. Instead of being a Beach {Laugh}

266
00:35:20.009 --> 00:35:23.429
C: Uh, like a graduate student.

267
00:35:24.809 --> 00:35:30.719
EXP: All right, so you think you me now, you made the stories.

268
00:35:30.719 --> 00:35:34.498
D: Yeah, and so athlete.

269
00:35:34.498 --> 00:35:40.378
EXP: Great in this case in the story scenario, I would ask you to type down.

270
00:35:40.378 --> 00:35:46.914
EXP: your story is in the chat box, and I will copy it on the screen.
B: Sure.

271
00:36:30.838 --> 00:36:37.588
D: Can you see my story that I posted.

272
00:36:37.588 --> 00:36:44.128
EXP: Yeah, I see Two stories .
D: Okay, great.

273
00:36:44.128 --> 00:36:47.489
EXP: Yes.

274
00:37:19.498 --> 00:37:28.559
EXP: Well, guys can you, okay.

275
00:37:28.559 --> 00:37:35.099
EXP: Let me take.

276
00:37:38.338 --> 00:37:42.659

Conflict recognition 4 Weak category:
Person A thought that the beach won't fit to the theme of winter house
Conflict resolution 4: Affordance-based assumption
But person D justified for it by making the assumption that the water represents the beach house in a winter day.
(They all agree on the assumption with no objection and laughed.)

Accordingly, they switched the tennis from A with the coffee from C:
A: (waves, fire, winter house, coffee) > winter day
B: (bird, house, vegetable, pumpkin) > farm story 2
C: (backpack, soccer, tennis, grad-hat) > graduating story
D: (man, pizza, car, balloon) > Birthday story 1

Participants started typing and experimenter copied and pasted the stories:

Solution category A: someone rents a condo for a winter vacation close to the beach where they enjoy a warm fire and warm tea indoors
Solution category B : A bird flies by a farmhouse that grows crops like vegetables and pumpkins.
Solution category C: A graduate student packs his cleats and racquets in his backpack.
Solution category D: There's a man who has to attend a pizza party. So he goes out in his nice car and grabs a pizza and a balloon to look festive.

EXP: We have all the stories

277
00:38:15.960 --> 00:38:20.400
EXP: This is the third one

278
00:38:35.550 --> 00:38:40.739
EXP: So, Here are the stories you can see them, right?
All: yeah

279
00:38:40.739 --> 00:38:47.579
EXP: So, what I'll be doing now, I will take a screen shot of the.

280
00:38:47.579 --> 00:38:50.699
EXP: Stories to send that to, you.

281
00:38:55.860 --> 00:39:01.769
EXP: great

Appendix J

Summary of observed conflicts and resolutions in the Nine Tasks

Expert 1 (Experiment A5 / Clipart/ Expert/ Version 1 / Solving Time = 692 sec)

Line reference	Conflict Code	Observed Conflict	Observed Resolution
290-295	**X1C1**	**Conflict recognition 1: Scarcity** They listed three items in the food category, which are pizza, vegetables, and coffee..	**Conflict resolution 1: Affordance-based assumption** To deal with this conflict the group made an affordance-based assumption that the picture of the haybale with the pumpkin is the fourth food item.
299-303	**X1C2**	**Conflict recognition 2: Scarcity** In the sport category, there are only two sport items (tennis racket and soccer ball).	**Conflict resolution 2: Relabelling the category** They renamed it to "activity" as a way of abstracting or widening the category to include more non-sport items such as "car", "balloon", "graduation", and "bird"
331-333	**X1C3**	**Conflict recognition 3: Weak category** Here person B thinks category A and D are weaker ones.	**Conflict resolution 3: Discard the conflict** An attempt from person D to deal with the conflict by making a justification that all items belong to a category of graduation
336-340	**X1C4**	**Conflict recognition 4: An item fits better in another category** Person A thinks that the balloon in category C fits better in the element category than the house	**Conflict resolution 4: Compromising** And in an attempt to improve the weaker categories. They made a compromise based on what fits better into the other categories. (The house from C fits better in D and the Balloon from D fits better in C)
341-346	**X1C5**	**Conflict recognition 5: Random item** Person B Asked question about an item (Bird) that seems to not belong to a given category (activity category)	**Conflict resolution 5: Affordance-based assumption** But **A** justified why it would belong (as a bird watching activity) > making an interpretation of the random item (bird) to complete the category (activity)
347-365	**X1C6**	**Conflict recognition 6: Weak category** After 15 seconds of silence, A asked C about her category, person C answered elements but seemed not sure	**Conflict resolution 6: Suggesting new solution category** Then D tried to suggest a category of seasons they discussed about it for a minute and a half, but it seemed to not work so the group discarded this idea of season.
373-375	**X1C7**	**Conflict recognition 7: Weak category** They tried to identify and improve weaker categories again: Person D identified the balloon as "awkward" in the element category (category C).	**Conflict resolution 7a : Discard the conflict** They almost accepted It (Satisficing as strategy to deal with the Conflict) **Conflict resolution 7b : compromising + relabeling the category**. But person A suggested switching the balloon (to be in activities category) with the bird in the elements category but this time they changed its name from elements to nature. Here they resolve the conflict through compromising which one fits better

			and adjusting the name of the category. They accepted balloon as an activity as a party thing and to switch the bird to the element category the adjusted the name of the category to accommodate the bird in it.

Expert 2 (Experiment B2 / Clipart/ Expert /Version 1 / Solving Time = sec 762)

Line reference	Conflict Code	Observed Conflict	Observed Resolution
220-221	**X2C1**	**Conflict recognition 1: Scarcity** Person C ask what the last item is that complete the food or the edible category	**Conflict resolution 1 : Affordance-based assumption** Then persons A and B suggested the hay with the pumpkin. So, they added the hay to complete the food category by exchanging the soccer with the hay.
225-229	**X2C2**	**Conflict recognition 2: Scarcity** They notice that soccer and tennis can go together however there are only two items	**Conflict resolution 2a: Affordance-based assumption** Person C identified the car as a sport item to be included in the sport category. **Conflict resolution 2b: Affordance-based assumption** Balloon is also an existing item in this quadrant, they asked if it could be considered as a fourth sport item, but they seemed to not be convinced with the idea as they laughed. **Conflict resolution 2c: Affordance-based assumption** Then person C suggested that water could be a sports item considering it as board surfing sport.
235-249	**X2C3**	**Conflict recognition 3: Weak category** The rest of the items they left with do not make sense.	**Conflict resolution 3: Suggesting a new solution category:** Person C suggested two categories: things that belong to the nature (fire, snow, water, bird) and human made things.
250-252	**X2C4**	**Conflict recognition 4: Weak category** After calling category C Human made, person C noticed that vegetables are an existing item in this category, but it is natural not human made	**Conflict resolution 4: Compromising + Affordance-based assumption** Suggested fire would work, so she came up with the idea that "this fire seems to be made by human because there are stones around it; so, we can switch the vegetable with the fire."
253-258	**X2C5**	**Conflict recognition 5: Weak category** Person B thinks that fire would not fit with the "human made" category	**Conflict resolution 5a: Affordance-based assumption** C tries to justify for the switch using the same assumption they used above that the fire is made by a human. This justification seems not to work **Conflict resolution 5 b: Relabeling the category** Person A suggested relabelling category **1** to be named nature and elements, "even if it's made by human, it's still elements".

			And person B suggested calling it “things that you do when you're in the nature” like sitting campfire.
258-266	**X2C6**	**Conflict recognition 6: Redundancy** Then person C is holding on the idea of human made category, but other things are also human made, they listed items that are human made which are (House, Car) but then noticed that other things are also human made items.	**Conflict resolution 6: Relabeling the category** Then person B relabeled category 2 as “measure of success” and justified for that.
267-270	**X2C7**	**Conflict recognition 7: Weak category** Now they're left with category D that has the following items: (man, tennis, balloon, soccer), and they're not sure what to call it.	**Conflict resolution 7 : Relabeling the category** They suggested the following category names: “miscellaneous”, “things you do in a birthday party” and agreed on the second category name though they were laughing about the category names.
271-283	**X2C8**	**Conflict recognition 8 Weak category** when they reached the last category, (D), they seemed not to be convinced with the name they created for it which is “birthday party”.	**Conflict resolution 8: Relabeling the category** They tried to find another name for it. They brainstormed other names: entertainment, playing cosplay, or listening to a stand-up comedy. They settled on entertainment.

Expert 3 (Experiment A’1 / Clipart/ Expert / Version 2 / Solving Time = 628 sec)

Line reference	Conflict Code	Observed Conflicts	Observed Resolutions
164-166	**X3C1**	**Conflict recognition 1: Scarcity** They listed 3 items that can represent food and asked if there is a fourth item?	**Conflict resolution 1: Affordance-based assumption** After few second s of discussion, they considered the pumpkin with haybale as affording the interpretation of a food item.
175-176	**X3C2**	**Conflict recognition 2: Scarcity** however, they noticed there is no more sport items to complete the sport category.	**Conflict resolution 2: Affordance-based assumption** Then they considered the car as a sport car and water (waves) as swimming.
177-178	**X3C3**	**Conflict recognition 3: Item better fits another category** Identified water as an item that is needed for the element category than in the sport	**Conflict resolution 3: Compromising** They discussed the suggestion of grouping (fire, water, balloon and haybale) in a category of “elements.”

219-220	**X3C4**	**Conflict recognition 4: Item better fits another category** Person A asked if they wanted to swap the haybale with the (man) as teacher to complete the education and seasons categories,	**Conflict resolution 4: discard the conflict +New solution idea** but then B thought to keep the man in category C and that could make human needs category including (man, house, food ,warmth) and Person D suggested human house and car would go in the driveway.
221-223	**X3C5**	**Conflict recognition 5: Scarcity** but Another person said if it's (human, house and car) there was no fourth item	**Conflict resolution 5 Discard the solution:** They discarded the idea and jump to another topic (seasons).
226-233	**X3C6**	**Conflict recognition 6: Redundancy** They're not sure of the 4th item that will complete the season category either vegetables or waves representing summer (it seems here another solution category has emerged while trying to resolve for the season category)	**Conflict resolution 6a: Discard the conflict + Affordance-based assumption** One suggested both items might be needed for a possible food category. Person A suggested campfire could afford a potential interpretation of a summer item.
234-238	**X3C7**	**Conflict recognition 7: Redundancy** Person B thought of the solution idea of elements water fire and earth they proposed two items to represent earth either haybale or vegetables	**Conflict resolution 7: Compromising** So they decided to use vegetable as a representation of earth because the haybale represent fall or Thanksgiving.
244-246	**X3C8**	**Conflict recognition 8: Scarcity** They want to complete seasons category but they're looking for something that represent summer	**Conflict resolution 8: Affordance-based assumption** They assume the house would represent summer because it has green grass.
247-251	**X3C9**	**Conflict recognition 9: Weak category** for category B the car does not seem to belong to the education	**Conflict resolution 9: Affordance-based assumption** They assume that the car is a way to commute to school.

Good 1 (Experiment D5 / Clipart/ Good / Version 1/ Solving Time = 361 sec)

Line reference	Conflict Code	Observed Conflicts	Observed Resolutions
202-203	**G1C1**	**Conflict recognition 1 Scarcity** They needed one item to complete the winter theme category * They didn't explicitly state that there is a conflict however.	**Conflict resolution 1: Affordance-based assumption** Person B thinks the bird can belong to the winter as a winter bird so category 1 is complete as winter theme category.
204-207	**G1C2**	**Conflict recognition 2: Item better fits another category** C wanted to exchange graduation hat (in category C) with the wave (in category B) but it seems not to work because person B thinks the wave is a part of a cottage theme	**Conflict resolution 2: Discard the solution** They just discarded the idea and jumped to another switch.
220-227	**G1C3**	**Conflict recognition 3: Weak category:** Category D seems to not be clear as person A is smiling and asking what the theme for category D is then Person B is asking why there is a balloon in category D.	**Conflict resolution 3: Compromising+ Affordance-based assumption** Person C thinks the balloon can go with the graduation hat and person B assumes that soccer is a sport you play in August September. So, they switched soccer ball with the balloon.
228-229	**G1C4**	**Conflict recognition 4: weak category:** Person A ask "what would category B be?"	**Conflict resolution 4 : Relabeling the category** Person B suggested a cottage or vacation scenario for that category.
230-223	**G1C5**	**Conflict recognition 5: weak category:** Now they're asking about the 4th category D what is it?	**Conflict resolution 5a : Relabeling the category** Person B thinks of a scenario that someone can make pizza with the vegetable that you can put, but he seemed not sure. **Conflict resolution 5b Compromising:** They seemed to not have a clear idea what category 4 would be. After few seconds of silence, they decided to swap back the balloon with the soccer ball. **Conflict resolution 5c:Relabeling the category** Person B suggested a "Halloween party" name and gave justification of that. They laughed and agreed.

Good 2 (Experiment C3/ Clipart/ Good / Version 1 / Solving Time = 217 sec)

Line reference	Conflict Code	Observed Conflict	Observed Resolution
134-138	**G2C1**	**Conflict recognition 1 : Scarcity** they noticed a category of sport that includes two items: tennis and soccer ball.	**Conflict resolution 1: Affordance-based assumption** The water is a sport of swimming.
139-142	**G2C2**	**Conflict recognition 2: Random Item** Person C notice that bird does not belong to a specific category such that there's no category of animal, so he asked where would the bird go.	**Conflict resolution 2: Affordance-based assumption + new solution category** So, B suggested a camping category that include: the cottage campfire bird and water as lake:
144-150	**G2C3**	**Conflict recognition 3: Weak category:** Now they are left with (house, person, car, balloon, graduation hat)	**Conflict resolution 3: Affordance-based assumption** Person D Person D thought of the balloon as part of equipment and the graduation is like a party item (graduation hat, balloon] or [balloon hat man school bag] as celebrating graduating category.
151-153	**G2C4**	**Conflict recognition 4: Random Item** They ask what to do with the car and the house	**Conflict resolution 4 (relabeling): S**omeone suggested an idea of you live in a house and drive to a sport practice and he called a category of sport related things. Person B laughed.
161-163	**G2C5**	**Conflict recognition 5 : Item better fits another category** One suggested to move the car with the house (This suggestion corresponds to the category they identified earlier of driving to a sports practice)	**Conflict resolution 5: Discard the conflict+ Relabelling** But then another person thought of keeping the house in the same category with the backpack tennis and soccer and call it a family house.

Good 3 (Experiment D1/ Clipart/ Good/ Version 1 / Solving Time = 238 sec)

Line reference	Conflict Code	Observed Conflict	Observed Resolution
182-186	**G3C1**	**Conflict recognition 1 : Scarcity** Only two sport items that include: tennis and soccer ball.	**Conflict resolution 1: Relabelling + Affordance-based assumption** The sport category can be called activity by adding driving a car to it , or may be by adding bird as a bird watching activity.
188-189	**G3C2**	**Conflict recognition 2: Item better fits another category** identifying a better way to categorize backpack with the graduation rather than in activity category	**Conflict recognition 2 Compromising:** **S**uggesting exchanging backpack with the graduation hat and the person as school items and the house as the school.
190-195	**G3C3**	**Conflict recognition 3: Weak category** Another conflict they recognized in the newly identified category is that "how the house would represent a school"(CR2)	**Conflict recognition 3 Discard the conflict** To resolve this conflict they immediately made a justification that "right now everyone is doing school at home" , they agreed and laughed.
198-205	**G3C4**	**Conflict recognition 4: Weak category** person D asked about balloon in the last swap again for clarification " saying what is balloon as it was part of the activity category	**Conflict recognition 4a: Discard the conflict** They laughed then person A was justifying that it's a kind of activity to play with the balloon (laughing). **Conflict recognition 4b (Relabeling + Discard the conflict))** Person D said everything is roundish suggesting a new label for solution category but he said the car is not roundish he was silent for a bit and said OK whatever (accepting or satisficing) person B said they're all round and the car has wheels.

Story 1 (Experiment E3/ Clipart/ Story / Version 1/ Solving Time = 254 sec)

Line reference	Conflict Code	Observed Conflict	Observed Resolution
234-238	**S1C1**	**Conflict recognition 1 Random item** Person D started offering other members the items that he doesn't need (pumpkin)and started proposing his stories	**Conflict resolution 1: Compromising** They then switch the pumpkin with the pizza from C since pizza fits the theme of the birthday.
240-250	**S1C2**	**Conflict recognition 2: Weak category:** but person C seems not happy with this since he now has too many food items.	**Conflict resolution 2: Affordance-based assumption** Person D proposed a farm story for category 2 that uses the pumpkin with the house and vegetable.
258- 262	**S1C3**	**Conflict recognition 3: Item better fits another category** Person A and C suggested another idea that would fit the coffee in A "cold winter day story" and the tennis in C as a "graduating story"	**Conflict resolution 3: Compromising** They propose switching the tennis from A with the coffee from C.
263-265	**S1C4**	**Conflict recognition 4: Weak category:** Person A recognized a conflict that the beach (waves) won't fit to the theme of winter house.	**Conflict resolution 4: discarding the conflict** Then, person D discarded this as a conflict and justified for that by making the assumption that the waves represent the beach house in a winter day. They all agree on the assumption and laughed.

Story 2 (Experiment F'3/ Clipart/ Story / Version 2/ Solving Time = 373 sec)

Line reference	Conflict Code	Observed Conflict	Observed Resolution
234-244	**S2C1**	**Conflict recognition 1: Item better fits another category** an item fits better in Person B category so he was reluctant to exchange,	**Conflict resolution 1: Compromising** Person B accepted it after discussing the story it seems they dealt with it by compromising which better fits in what category.
257-264	**S2C2**	**Conflict recognition 2: Random items** person B seems to not need the waves (ocean and offers it to other members	**Conflict resolution 2: Compromising** Person D is willing to take the beach.
265-267	**S2C3**	**Conflict recognition 2: Weak category** person A thinks that the haybale might not be a good fit then person B story.	**Conflict resolution 2: Affordance-based assumption** Then, to deal with this conflict, person A suggested that "You could say that the snowy house is like, on a corn farm" they laughed as it seems not to work but then he suggested the house to be a seaside house.
273-275	**S2C4**	**Conflict recognition 1: Random item** person C wasn't sure what to do with the bird	**Conflict resolution 4: Affordance-based assumption:** He came up with a solution that includes the bird.

Story 3 (Experiment F4/ Clipart/ Story/ Version 1 / Solving Time = 276 sec)

Line reference	Conflict Code	Observed Conflict	Observed Resolution
198-201	**S3C1**	**Conflict recognition 1** **Random item** person B asked others if they need the house and the backpack as it seems that he doesn't need these two items	**Conflict resolution 1 : Compromising** Person A switch his fire with the backpack, and they want a house for the regular house.
202-210	**S3C2**	**Conflict recognition 2** **:Random item** Person D indicated two items that he seems to not need which are the Thanksgiving thing (haybale) and the balloon.	**Conflict resolution 2: Compromising** Person C indicated that he needed both and when person D asked him "what do you want to exchange for", person C was flexible to exchange it with any of the three items that include the graduation hat soccer ball or the cup of coffee.
213-216	**S3C3**	**Conflict recognition 3: Item better fits another category** Then person A asked person D if he wants to keep the car, person D asked him what do you want to swap it with person a wanted to swap the car with the bird	**Conflict resolution 3: Compromising** Person D agreed on that and exchanged.
217-223	**S3C4**	**Conflict recognition 4: Item better fits another category:** Person C asked person D if he wants the man. Again, person D asked person C what he wants to exchange the man with and person C replied with the cup of coffee.	**Conflict resolution 4: Compromising** Person D agreed on that and laugh saying that he can probably think of something.